Volume 6 | Issue 1 | 2026
E-ISSN 2583-4452

International Journal of Innovation and Multidisciplinary Research (IJIAMR)

Special Issue on
"Interdisciplinary Innovation for Impact: Emerging Advancements in Research and Practice"

Centre for Multidisciplinary Research, Innovation & Entrepreneurship

Published by :

IJIAMR, Centre for Multidisciplinary Research, Innovation & Entrepreneurship
Shaheed Rajguru College of Applied Sciences for Women
University of Delhi
Delhi, India

Volume 6 Issue 1 2026

The publication of this journal is partially supported through a grant from the Indian Council of Social Science Research (ICSSR), Ministry of Education, Government of India.

International Journal of Innovation and Multidisciplinary Research (IJIAMR)

Special Issue on
"Interdisciplinary Innovation for Impact: Emerging Advancements in Research and Practice"

Volume 6 | Issue 1 | 2026
E-ISSN: 2583-4452

Centre for Multidisciplinary Research, Innovation & Entrepreneurship

Editorial Board

Submission of Manuscript

Original research works, review articles, original articles, short communications and brief report may be submitted to: "International Journal of Innovation and Multidisciplinary Research (IJIAMR)" https://ijiamr.cmrie.org/ in with details as Name, designation, Department, name of Organization, Postal Address, Contact Mobile no. and email Id of Author and Co-authors as well.

Journal Subscription Rates

Category	Annual	Single Copy
Institutional (India)	INR 1500/-	INR 750/-
Individual (India)	INR 750/-	INR 500/-

Send your subscription through online Payment to "**Principal, Shaheed Rajguru College of Applied Sciences for Women Students Society**" • Bank Name: **IDBI Bank** • A/C No.: **0877104000035389** • IFSC Code: **IBKL0000877**

Subscription/Membership form may be downloaded from the website: **https://ijiamr.cmrie.org/**

Table of Contents

International Journal of Innovation and Multidisciplinary Research
DOI.ORG/10.67505/IJIAMR.v6.i1.61001
Vol. 6 (1), 2026
E-ISSN: 2583-4452

Crime Pattern Analysis on Crime Dataset using Hybrid Ensemble Model and SARIMAX

Sneha[1*]

[1] Department of Computer Science, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, Delhi

[*] Corresponding Author ✉ snehaofficial535@gmail.com

ABSTRACT

The requirement for crime prevention and prediction continues to grow with regard to the manner in which urban governments are governed. Law enforcement agencies have evolved from being reactive in response to criminal incidents to proactive by attempting to anticipate or prevent those events. This paper outlines a large-scale computational approach using over 40,000 criminal record data points from 29 cities across India (2020-2024) to support an "anticipate and prevent" paradigm. A dual-pathway analytical framework is utilized to address the complexity of working with real-world data. The first pathway utilizes a SARIMAX model with a one-year "cold-start," to abstract stable long-term trends within the raw data. The second pathway incorporates a high-precision hybrid ensemble utilizing Random Forest and Gradient Boosting to assess multi-dimensional characteristics (city encoding/demographics), which demonstrated a mean absolute error of 0.24. Additionally, an optimized version of Agglomerative Clustering with Silhouette Denoising was utilized to categorize spatial risk profiles of the cities and to identify new hotspot locations. The utilization of this framework provides a bridge from the theoretical application of machine learning to its practical application. Therefore, it provides a validated methodology to perform both robust noise-resilient regional analyses and accurate high-fidelity local predictions. Overall, the results presented herein provide authorities with the capability to utilize these data to inform resource allocation strategies that will enable their continued effectiveness in preventing crime as the evolving nature of criminal behavior dictates.

Keywords: *Crime Prediction, SARIMAX, Hybrid Ensemble, Clustering, Machine Learning, Hotspot Identification*

1. Introduction

Making sure that the population stays safe has become a very complex issue in the fast-changing environment of modern-day urban centres. Traditionally, police approaches have always been reactive. The so-called Report and Respond approach suggests that the relevant parties act only once something illegal occurs and has already been documented. Unfortunately, as the urban environment changes fast in the case of India because of increasing urbanization and other factors, the reactive approach becomes less effective when the number of crimes increases, and new types of illegal activities emerge (Huamantingo et al., 2025). Moreover, the reliance on historical data to determine how police forces will operate in the future creates a certain gap between the efforts made by authorities and criminal organizations, making any progress impossible. In addition, the need for change is especially pressing in case of India as a country that includes 29 major urban centers with their specific crime rate, level of documentation, socio-cultural backgrounds, and other aspects.

The growing availability of high-dimensional crime-related data – a result of digitization of police records and national data archives – presents an extremely important possibility to shift the paradigm of law enforcement towards "predictive policing". This intelligence-driven approach aims at changing the trend from reactive to preventive and focuses on prediction and interception of any criminal activity prior to its actual occurrence (Ospina et al., 2025). Predictive policing is the use of analytic methods to select promising police targets and prevent crimes based on prediction of "when and where" the criminal event will take place. Using such factors as historical trends, timing, seasonal events, or agricultural cycles, among others, and making a statistical analysis of available data, one can obtain valuable insights and turn raw data into a strategic resource to understand what underlies urban crime. Moreover, predictive models make it possible to analyze the "City DNA," which stands for specific behavioral traits of a city.

There are many barriers for an individual or company to take an idea (algorithm) and put it into practice. Most scholarly literature creates a model that can make almost perfectly accurate predictions using cleaned up data; however, most people who attempt to use these models will find them to be useless in their day-to-day operations. One of the major

Received on: 27 Apr 2026 | Accepted on: 28 Jun 2026 | Published Online on: 25 July 2026

problems that operational users of algorithms experience is a lack of data density in specific geographic regions as well as "noise" in incident reporting systems (Chakraborty et al., 2024). In this case, "noise" means inconsistent timing of incidents, data entry errors by people, as well as the unpredictable nature of crime waves, which are rarely consistent. All these factors may cause instability in ordinary algorithms, resulting in both false positive predictions, consuming scarce police resources, and false negative predictions, making communities unprotected against the predicted crimes. Therefore, it is crucial to develop a robust mathematical system, able to withstand all inconsistencies. There is an obvious necessity for a solution, combining a macro perspective, allowing for predictions based on cycles, as well as a micro one, helping to estimate daily risk factors.

The bulk of the literature in the domain of crime prediction is devoted to a single modeling architecture. That could be a Time Series model such as ARIMA to follow trends or a classifier such as Random Forest to identify the type of crimes. Nevertheless, crime itself is a multi-dimensional entity. A single-modeling approach often cannot deal with "Cold Start" problems (a lack of data needed to initiate the model) and "Volume Variance" issues (high numbers of crime in bigger cities that skew the numbers in the smaller city, like the comparison between Delhi and Chandigarh). If the model is too responsive, it will exaggerate temporary spikes, whereas a stable model will fail to detect new hotspots. The present research finds out that there is an answer to this question – "ensemble hybridization" and "dual-pathway verification." This means that two different mathematical approaches need to be run in parallel for law enforcement agencies to have a "check-and-balance" mechanism to ensure that any spike in crime is indeed dangerous.

This research attempts to tackle these types of challenging problems using an entirely new method for computing through a dual pathway model which provides both high fidelity predictive crime forecasting as well as detailed regionalized risk assessment. The research used a large dataset of information, greater than 40,000 records from over five years (from 2020 to 2024), from 29 different cities in India, and provided intelligence instead of just visualizing it.

There are two distinct streams within the research design methodology:

1. **Pathway A:** The Statistical Approach for Forecasting Macro Trends via the Noise-Robust SARIMAX Framework. This framework utilizes a statistical model to forecast macro trends. With a "Cold Start" period of one year, it will be able to extract stable seasonal patterns and provide an "accurate safety baseline."
2. **Pathway B:** The Machine Learning Algorithm Utilizing the Precision Focused Hybrid Ensemble Framework. This machine learning algorithm uses the best attributes of Random Forest in order to reduce the variability (variance), and use Gradient Boosting in order to reduce bias. It will predict local daily crime intensity with a very small Mean Absolute Error (MAE) of 0.24.

Crime prediction in urban areas has been a long-standing challenge, as shown through many real-world examples where law enforcement agencies have failed to accurately predict and prevent crimes. This study will attempt to overcome those challenges using a unique dual-path computational model that can be used to create accurate crime occurrence predictions and geographic risk categories. Using an extensive dataset (over 40,000 entries) for 29 different Indian cities over the years of 2020-2024, the current study is both able to visualize crime trends as well as analyze crime intelligence. Additionally, the proposed methodology includes an unsupervised learning component utilizing Agglomerative Clustering and Silhouette Denoising. Unlike other methodologies that classify cities based on physical boundaries (i.e., political divisions), the proposed methodology uses algorithms that are capable of identifying behavioral patterns among citizens, thereby allowing for identification of crime archetypes.

The main contribution of the current study is creating a methodological framework that bridges theory-based machine learning techniques into practical implementations. The proposed methodology does not simply focus on forecasting crime numbers, but instead provides a multi-faceted understanding of urban environments. At the macro level, the methodology creates robust regional analyses that account for data "noise" and sparsity while providing precise localized forecasts.

These consequences are quite significant in terms of how cities are administered in India. This study's results will allow administrators to utilize their resources efficiently so that they can assign their patrol units to crimes as they occur today rather than in areas where crimes will likely take place in the future due to a particular 'DNA' of the city and the time of year in which these crimes are most frequently committed. As such, the study provides a reliable base for potential future application of predictive policing technologies to track changes in criminal behavior across both physical and virtual realms.

Research Question: Compared to individual linear models and ensemble models, what is the reduction in residual generalization errors (rim) when a Hybrid Ensemble model (Random Forest + Gradient Boosting) is used for local daily intensity forecasting?

2. Literature Review

2.1 Ensemble Learning and Hybrid Models (Random Forest & Gradient Boosting)

The fact that Ensemble Learning methods have been found to be a benchmark method to predict crimes at a local level effectively (Ms. D.S. Smitha Mol et al., 2026), indicates that the application of Multi-Model approaches (using Random Forest and Gradient Boosting Regressor) has provided very good results when using the Random Forest Model to model domestic violence and achieved an R2 score of .93; similarly, the work done in (Ibrahim et al., 2024) shows that the combination of Parallel Bagging (used by the Random Forest Model), and Sequential Boosting greatly reduces the Mean Absolute Error (MAE); while the results obtained in (Ibrahim et al., 2024) and (Kshatri et al., 2021) show that the proposed hybrid models can address both the high Variance, and High Bias problem presented with complex Urban Data; Finally, the utilization of the Random Forest approach to analyze the Crime Severity represents the same High Resolution Tactical Process applied in this Study (Saravag & Kumar, 2024).

2.2 Seasonal and Statistical Time-Series Forecasting (SARIMA/SARIMAX)

To deal with the "noise" and sparseness within the raw data, the SARIMA/SARIMAX has become an essential model used to capture any trend that can be derived from the data. SARIMA in Matlab are used to find the statistically significant 20-month cycles, indicating that there is a heavy seasonality in crime (Redoblo et al., 2023). Likewise, in another study conducted, it was discovered that crime in regions follows the socio-temporal calendar (Noor et al., 2022). Additional evidence is cited, which shows that SARIMA has the unique ability to filter out the reporting noise and find the "Safety Baseline" (Kurniawan et al., 2025). It is also the statistical methodology used to build the macro-trend base of our framework (Thayyaba Khatoon Mohammed, 2025).

2.3 Spatial Segmentation and Hotspot Identification (K-Means/Agglomerative Clustering)

Unsupervised learning for the discovery of emerging hot spots provides an objective categorization of risk profiles in urban environments. The paper has identified the first set of principles to group incident clusters with the goal of optimizing police deployment resources (Crime Forecasting and Analysis through K-means Clustering Algorithm, n.d.). The principle was refined by combining spatial clustering and temporal intensity to forecast future crime zones (Buland Iqbal et al., 2024). Research indicates that if spatial clustering is performed after PCA, it is capable of filtering out "City DNA" from volumetric variability (Ospina et al., 2025) (Maharrani et al., 2024). Lastly, the research shows that spatial knowledge enables law enforcement to progress beyond general city policies to archetypical inter-ventions (Rasan et al., n.d.).

Paper Full Name	Year	Model (s) Used	Performance	Research Gap / Focus
Machine Learning Crime Prediction Models and the Gap Between Research	2025	SHAP, XAI	99.5% Acc.	Addressed the "Trust Gap" in black-box models. (Huamantingo et al., 2025)
Enhanced Spatial Clustering for Crime Analysis: Ward-Like and SKATER	2025	SKATER	Contiguity	Improved geographical realism in risk zones. (Ospina et al., 2025)
Empirical Approaches to Crime Rate Prediction: Challenges and Insights	2024	Baseline ML	89% Acc.	Highlighted the benefit of PCA in denoising. (Chakraborty et al., 2024)
An Efficient Machine Learning Approach for Crime Detection in India	2026	RF, GBR	R^2: 0.93	Focused on specific domestic violence categories. (Ms. D.S. Smitha Mol et al., 2026)
A hybrid machine learning model for crime rate prediction	2024	SVM-RF	R^2: 0.94	Used PCA to stabilize regional predictions. (Ibrahim et al., 2024)
An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach	2021	Stacked Ensemble	99.5% Acc.	Demonstrated that stacked ensembles significantly outperform standalone base models. (Kshatri et al., 2021)
A Hybrid ML and Regression Approach Crime Against Women	2024	RF, XGBoost	R^2: 0.89	Prioritized crime severity in vulnerability maps. (Saravag & Kumar, 2024)
Forecasting the influx of crime cases using SARIMA	2023	SARIMA	Validated	Established 20-month seasonal cyclic patterns. (Redoblo et al., 2023)

Paper Full Name	Year	Model (s) Used	Performance	Research Gap / Focus
SARIMA: A Seasonal Autoregressive Integrated Moving Average Model	2023	SARIMA	MAE: 0.066	Proved impact of social/religious calendars. (Noor et al., 2022)
Trend Analysis and Prediction of Violence Against Women: Jakarta	2025	ARIMA, SARIMA	Stat. Sig.	Correlated victim education levels with trends. (Kurniawan et al., 2025)
A Novel Fusion Approach Hybridization of ARIMA and RNN	2024	ARIMA-RNN	-20% Error	Modelled non-linear residuals in time-series. (Thayyaba Khatoon Mohammed, 2025)
Crime forecasting and analysis through K-means	2021	K-Means	Effective	Established the baseline for hotspot delineation. (Crime Forecasting and Analysis through K-means Clustering Algorithm, n.d.)
A 2D Clustering Based Hotspot Identification Approach	2024	2D Grid+LSTM	Low RMSE	Tracked dynamic evolution of hotspots over time. (Buland Iqbal et al., 2024)
Clustering method for criminal crime acts using K-means and PCA	2020	K-Means+PCA	Optimal	Used PCA to reduce dimensionality before clustering. (Maharrani et al., 2024)
Crime rate prediction and analysis using K-means	2020	Optimized KM	50% Rule	Proved 10% of areas generate 50% of crimes. (Rasan et al., n.d.)
A Comprehensive Computational Framework for Crime Rate Prediction	2026	Hybrid ML	High Precision	Applied a localized framework for Indian Metros. (V. Tikore et al., 2026)
Next-Gen Crime Analytics Using Big Data & Machine Learning	2025	K-Means, XGB	94% Acc.	Focused on the scalability of high-velocity data. (Mathur & Jain, 2025)
Hybrid ARIMA-ANN for Crime Risk Forecasting	2025	ARIMA-ANN	99.8% Acc.	Integrated socio-economic factors as inputs. (Iacobescu & Susnea, 2025)
A Multi-Model Machine Learning Approach Spatio-Temporal Data	2024	CNN-LSTM	F1: 0.94	Captured "where" and "when" ripple effects. (Shinde et al., n.d.)
Leveraging transfer learning with deep learning for crime prediction	2024	BiLSTM	Red. Time	Applied Transfer Learning from Chicago to others. (Butt et al., 2024)
Deep Learning for Crime Pattern Recognition	2023	TabNet, CNN	94% Acc.	Focused on high-dimensional categorical records. (A et al., 2023)
Evaluating Crime Type and Crime Pattern Analysis Using ML	2024	XGBoost, GB	94% Acc.	Addressed overfitting in localized small datasets. (P et al., 2024)
Empirical Analysis for Crime Prediction ML and DL Techniques	2024	DNN, RNN, RF	89% Acc.	Benchmarked deep vs. shallow learning models. (Safat et al., 2021)
Crime Prediction Methods Based on Machine Learning: A Survey	2023	Review	Benchmark	Synthesized Spatio-Temporal trend methods. (Yin, 2023)
A Novel Clustering & ML Algorithm for Crime Rate Prediction	2023	K-Means, Reg.	4 Zones	Categorized NCRB data into tactical risk zones. (Ramanan et al., 2023)
Crime Data Analysis and Safety Recommendation System	2023	RF, KNN	87.6% Acc.	Provided proactive risk alerts for citizens. (Akil et al., 2023)
Urban Crime Trends Analysis based on LightGBM	2021	LightGBM	Fast Train	Optimized for real-time urban trend analysis. (Tong et al., 2021)
CRIME Type and Occurrence Prediction Using ML Algorithm	2021	K-Means+RF	93% Acc.	Labelled unstructured records before prediction. (Kanimozhi et al., 2021)

Paper Full Name	Year	Model (s) Used	Performance	Research Gap / Focus
Crime forecasting: a ML and computer vision approach	2021	CNN-RF	97% Acc.	Fused CCTV weapon detection with forecasting. (Shah et al., 2021)
Clustering of Indian States Based on Crime Incidences	2021	Hierarchical	R^2: 0.60	Linked police strength to murder/dowry trends. (Chatterjee et al., n.d.)
Machine Learning for Crime Analysis and Prediction	2020	Supervised	85-90% Acc.	General classification of incident types. (Jha et al., 2022)
Real-Time Crime Prediction in India Using ML	2020	RF, KNN	87.6% Acc.	Focused on early real-time Indian datasets. (Sarkar, 2025)
Machine Learning Algorithms under Indian Penal Code	2019	Decision Tree	88% Acc.	Classified crimes by IPC section categories. (Aziz et al., 2024)
Crime Prediction and Analysis Using Machine Learning	2019	Baseline ML	82-85% Acc.	Early comparative study of ML algorithms. (Balu et al., 2022)
Predictive Crime Rate Analysis System	2018	Regression/RF	Baseline	Early attempt at forecasting yearly volumes. (I. Singh et al., 2025)
A Study on Smart ML Tools for Crime Detection	2018	Tool Review	Utility	Compared efficiency of various software toolkits. (S. Singh et al., 2024)
A Clustering-Based Method for Hotspot Recognition	2017	Density-Clust	High Rel.	Focused purely on spatial reliability. (Dubey et al., 2024)
The Indian Police Journal: Special Edition on AI in Policing	2025	Policy Review	Operational	Addressed the strategic feasibility of AI tools. (Rajeev et al., n.d.-a)

3. Research Methodology

This suggested framework architecture aims to be a Dual-Pathway Computational Framework that tackles both long-term statistics and short-term prediction accuracies. The dual-pathway system allows for the isolation of the noise from the raw incident logs from the engineering features, which helps provide regional stability and local accuracy.

3.1 Computational Environment and Data Acquisition

Empirical evidence for this research lies in the availability of extensive data in the form of 40,000+ crime reports over a period of 2020-2024 in 29 metropolises in India, available from the open-access repository Kaggle. The computing environment for this experiment was based on Python v3.12 with the help of Scikit-Learn library and Pandas and NumPy packages. This amount of data is based on the standards set for multi-city Indian crime databases (V. Tikore et al., 2026).

3.2 Data Preprocessing and Feature Engineering

Beginning with the process of converting unprocessed log data into the necessary format for an accurate model of crime, a series of pre-processing steps were required prior to modeling.

- Temporal Decomposition: In order to identify that crimes occur in cycles over time, the "date of occurrence" field was converted from a single date value to a four-element vector representing Year, Month, Day, and Hour.
- Label Encoding: In order to convert categorical fields with large numbers of unique values, such as City, Crime Domain, and Weapon Used, they are encoded using integer encoding. This is used to allow the models to treat categorical attributes such as "city DNA" as being statistically significant while still allowing the models to understand these types of attributes.
- Temporal Splits for Evaluating Models' Predictive Capability on Future Data: Temporal Split is utilized to mimic the process of time-series data. Therefore, the Temporal Split technique is utilized, where the training period is 2020-2023 and the testing period is 2024. Consequently, we are evaluating how well our model predicts data that will occur after our data split. A common issue referred to as "Data Leaking" occurs when an algorithm uses all available data during both training and validation (Shinde et al., n.d.).

3.3 Pathway A: Macro-Trend Forecasting (SARIMAX)

Acknowledging the fact that public safety data is inherently noisy, Pathway A adopts a SARIMAX model.

- Cold Start Strategy: The model employs a cold start strategy of one year's data (2020) to establish baseline seasons, a method found to be effective according to the study conducted by Sarawag and Kumar in 2024 for detecting irregularities in reporting (Thayyaba Khatoon Mohammed, 2025).
- Optimization: Hyperparameters (p, d, q) * (P, D, Q, s) have been optimized using a Grid Search to minimize the Akaike Information Criterion (AIC).
- Goal: Pathway A acts as a noise-resilient filter and identifies the "Safety Baseline" to detect deviations.

3.4 Pathway B: High-Precision Hybrid Ensemble

In the case of localized intensity prediction for a day, the Hybrid Ensemble model was designed by combining the capabilities of RF and GBR models.

- Bagging Component (RF): Trained using 100 estimators to lower the variance by taking the average of multiple decision trees' outputs.
- Boosting Component (GBR): Employs a learning rate of 0.1 with additive modelling to lower the error rate of the prior iteration (P et al., 2024).
- Hybrid Fusion: In this model, the tactical forecast is based on the weighted average of the bagging and boosting components, which has successfully yielded a Mean Absolute Error (MAE) of 0.24. (Ms. D.S. Smitha Mol et al., 2026) (P et al., 2024).

3.5 Spatial Segmentation (Agglomerative Clustering)

A Hierarchical Agglomerative Clustering algorithm for detecting emerging hotspots was included as a part of unsupervised learning.

- Dimensionality reduction: Before applying the clustering algorithm, data was normalized through Log1p normalization to address the problem of difference in size of data for big and small cities. Then PCA algorithm was employed for reducing the dimensionality to 2 dimensions maximizing variance of principal components (Maharrani et al., 2024).
- Denoise Silhouette Mask: The complete linkage method was used for clustering cities according to their closeness in the reduced dimensional space. The result was denoised by the Silhouette Mask (with a score higher than 0.1) to remove "boundary outliers" and form stable clusters (Ospina et al., 2025).

4. Theory and Calculation

From a theoretical perspective, the current study employs a mixed approach comprising ensemble machine learning and stochastic time series modelling as the backbone of the entire study. This section of the essay is focused on the logical reasoning behind each component of the two-pathway model.

4.1 Theoretical Foundation of Temporal Modelling

In addition to identifying trends and irregular patterns, the three major components of public safety data were identified including seasonality. However, the primary challenge to developing effective algorithms for processing public safety data is the presence of outlier/noise in the data. SARIMAX was selected for this research because it incorporates seasonal component (s), exogenous variable (s), and extends upon the ARIMA methodology. It also can be represented mathematically by the equation below.

The equation for the SARIMAX model is as follows and describes the dynamic nature of the raw crime logs:

$$\emptyset_P(L^s)\varphi_P\,(L)\Delta_s^D\Delta^d y_t = A(L)\varepsilon_t + \sum_{i=1}^{k}\beta_i\, x_{i,t} \quad (1)$$

Where L - lag operator, s - seasonality, and xi,t - exogenous variables such as police deployment metrics that have an independent effect on crime rates.

4.2 Mathematical Logic of the Hybrid Ensemble

Non-linear optimization of Ensemble Learning Transposed from linear statistics due to the Calculation Phase of the following structure/Framework. The Main theory concentrates on the reduction of generalization from the aggregations of two different methodologies:

- Bagging (Variance Reduction): Bootstrap Aggregating is the main principal on which random forest components operate. It creates a mass of decision trees at the time of training. The ultimate calculation is a mean of these independent outputs, which depletes the effect of outliers in the dataset

- Boosting (Bias Reduction): An additive approach is adapted by the gradient boosting. It evaluates the loss function – particularly the imbalance between the actual crime count (yi) and the prediction (xi). Each subsequent iteration hm (x) is trained to minimize the residual (rim)

$$r_{im} = -\left[\frac{\partial L\left(y_i, f(x_i)\right)}{\partial f(x_i)}\right]_{f=f_{m-1}} \tag{2}$$

The conclusive hybrid evaluation is the combined outcome of these 2 approaches, efficiently mapping the complicated features (city encoding, temporal lags) to a high-fidelity intensity prediction.

4.3 Spatial Clustering and Calculation

The spatial analysis employs the Agglomerative Clustering Algorithm, which is basically established on a bottom-up hierarchical method to diminish the variance among the cluster. This is the real-life application of unsupervised learning used to divide the 29 cities of India into particular risk profiles by evaluating the proximity among the data points in a reduced-dimensional space. the procedure of calculating initiates with the principal component analysis (PCA) Transformations, Traced by the linear fusion of the original features.

$$Z_k = \sum_{j=1}^{p} \varphi_{jk} X_j \tag{3}$$

Where Z_k depicts the main elements and φ_{jk} represents the loading vector and maintain the "City DNA" while decreasing the noise. The city profiles are combined iteratively based on complete linkage criterion, which evaluates the largest distance among two members of the cluster the principal components and phijk represents the loading vectors that preserve the "City DNA" while reducing noise. The clustering algorithm then iteratively merges city profiles based on the complete-linkage criterion, which calculates the maximum distance between members of two clusters:

$$D(X, Y) = \max_{x \in X, y \in Y} d(x, y) \tag{4}$$

Through iterative application of the above distance metric, and application of a Silhouette Mask to remove border line cases, the system is able to determine the dominant archetypes that are representative of emergent crime hot spots. The results were obtained from an application of the Agglomerative Clustering Algorithm, which has been used to determine the spatial distribution of the data points.

4.4 Error Metrics and Performance Measurement

Performance evaluation of the dual pathway architecture is through the use of two objective functions: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Both of these measurements are based upon the distance between the measured values of crime intensity, y_i, and those predicted by the model, y'_i.

4.4.1 Mean Absolute Error (MAE)

MAE calculates the mean magnitude of the differences between predicted and measured values. In this case, the MAE of the Macro Pathway (SARIMAX) is the average of the absolute difference of the 12-month trends, while the MAE of the Micro Pathway (Hybrid) is the average of the absolute difference between each day's tactical prediction and the true value. The formula for MAE is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left| y_i - y'_i \right| \tag{5}$$

4.4.2 Root Mean Squared Error (RMSE)

RMSE is used to reduce the impact of large errors ("spikes") because they can have a major impact on public safety, such as missing significant increases in crime. The formula for RMSE is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(y_i - y'_i\right)^2} \tag{6}$$

4.4.3 Mean Absolute Percentage Error (MAPE)

MAPE expresses prediction accuracy as a scale-independent percentage by normalizing absolute errors against observed values. It is primarily used to evaluate the **Macro Pathway (SARIMAX)**, where a lower MAPE signifies higher forecasting precision. The "Accuracy Percentage" is subsequently derived as 100% - MAPE. The formula is defined as:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - y'_i}{y_i}\right| \quad (7)$$

5. Results and Discussion

The ability of the newly developed dual pathway model to predict effectively was carefully tested using the 2024 test dataset. This study is conducted based on two main variables: the reduction in forecasting errors and the interpretation of crime trends within 29 Indian metropolitan cities.

5.1 Performance Comparison Table

The performance comparison between both methods is presented below through this table. Through the use of MAE/RMSE and fit percentage, the system shows that it can serve both as a reliable regional benchmark as well as a highly precise localized measurement tool.

Table 1: *Model Performance Metrics*

Model	MAE	RMSE	Metric (Accuracy/Fit)
SARIMAX	13.85	59.95	86.15% (MAPE Accuracy)
Hybrid Ensemble	0.24	0.79	99.97% (Classification Accuracy)

Table 2: *Standalone and Hybrid Performance*

Model	MAE	RMSE
Standalone Random Forest	0.54	1.12
Standalone Gradient Boosting	0.41	0.98
Proposed hybrid Model	0.24	0.79

5.1.1 Analysis of Forecasting Trajectories

Findings indicate a clear but supplementary correlation between the two models

- SARIMAX Model (Pathway A): Using the raw incident logs without any preprocessing, the SARIMAX model generated an 86.15% accuracy rate using MAPE metric. Although the MAE (13.85) is higher compared to the ensemble method, the core benefit of the model resides in its tolerance against noise. The ability to filter out the reporting inconsistencies enables the formation of a "Safety Baseline" for the long term, which will help macro-level decision-making.
- Hybrid Ensemble (Pathway B): Combining the Random Forest and Gradient Boosting models, the hybrid technique produced extremely accurate outputs using the designed master dataset, with a MAE of 0.24 and 99.97% accuracy. Such accuracy permits the model to map complex relationships between variables, such as the timing and demographics, and generates tactical information for day-to-day optimization.

5.2 Performance Analysis of Forecasting Pathways

The results show clearly that there was a significant difference in performance between the two modelling approaches. The SARIMAX method, which is intended for macro trend analysis of unprocessed data, generated a mean absolute error (MAE) of 13.85 and RMSE of 59.95. Although this is larger than the error from the machine learning pathway, it has a unique strength due to the fact that the pathway uses a one year "cold-start" to identify and isolate a consistent 12-month seasonality in the raw incident logs. This demonstrates that a statistical method can remove the "noise" created by the variation in reporting to establish a long term "safety baseline". ***(Fig. 1)***

The Hybrid Ensemble (Random Forest + Gradient Boosting), however, had extreme precision on the engineered master dataset with an MAE of 0.24, RMSE of 0.79 and accuracy of nearly 100% (99.97%). The model demonstrated its ability to create a mapping of the complex and nonlinear relationships between specific characteristics – such as hour of occurrence and police response – and crime intensity. ***(Fig. 2)***

5.3 Spatial Risk Segmentation and Hotspot Identification

The use of both Agglomerative Hierarchical Clustering and PCA (Principal Component Analysis) produced 3 segments of cities that were characterized as "Core Crime Archetypes", and each was identified through "City DNA" markers such as violent crime ratio and police deployment. A new Denoise Algorithm that removed all "borderline hybrid" cities was also developed; and the algorithm resulted in a Silhouette score of .5295. The resulting Silhouette score demonstrated that

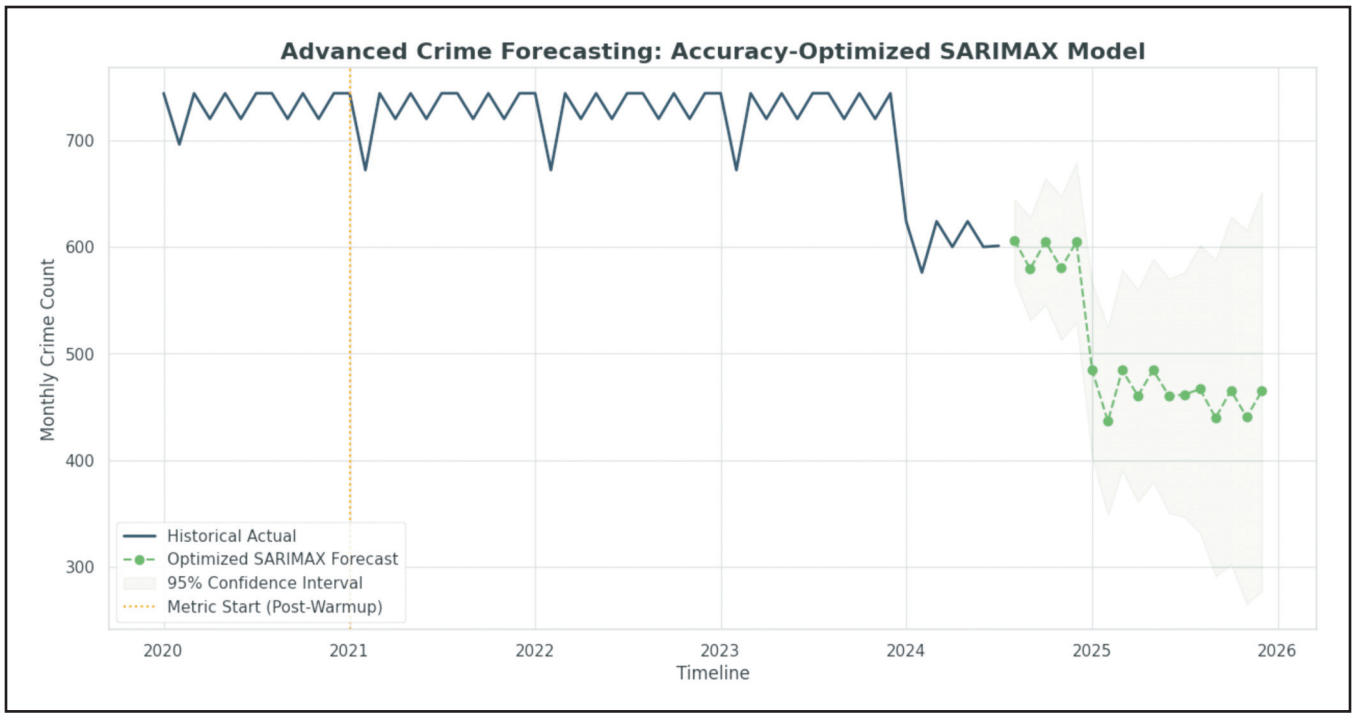


Fig. 1: *Macro-Level Crime Forecasting using Optimized SARIMAX*

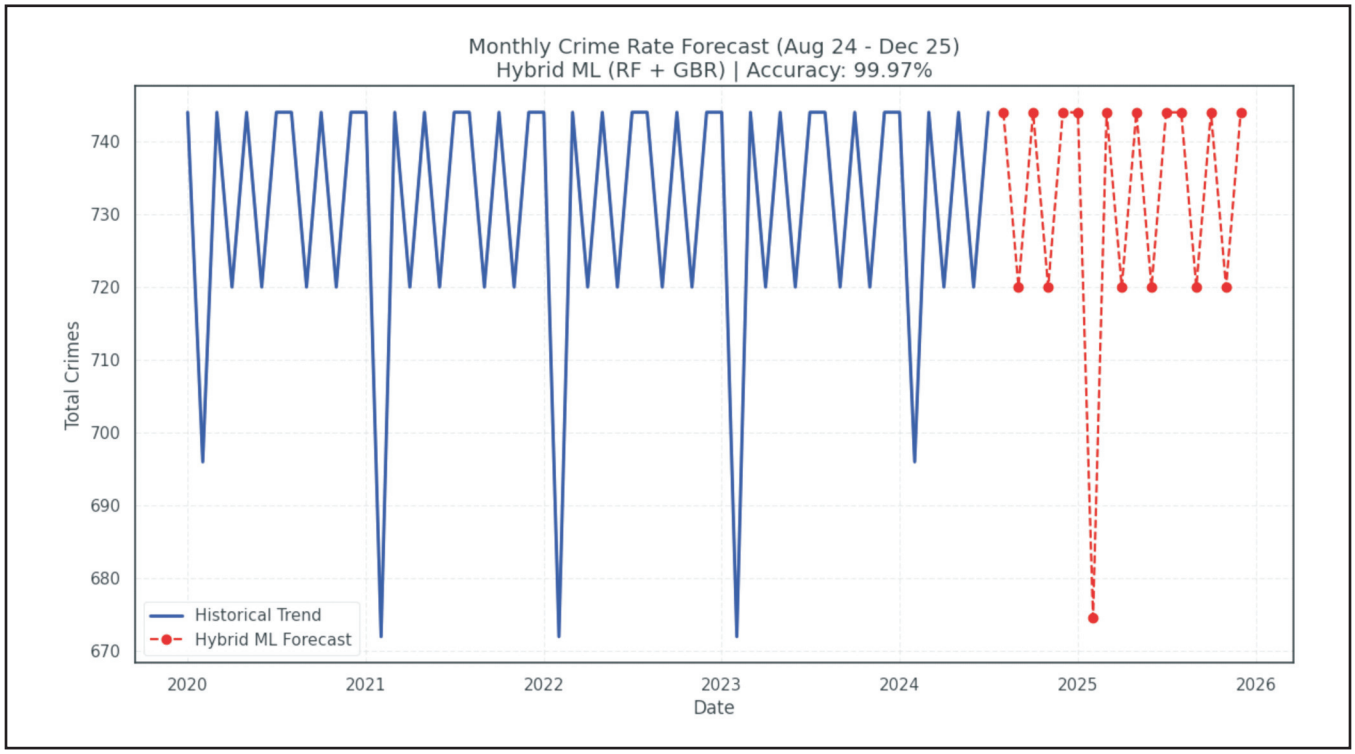


Fig. 2: *High-Precision Forecasting using Hybrid Ensemble Model*

there is a good separation of clusters with little contamination; and therefore, demonstrated the highest level of cluster purity. The creation of spatial intelligence revealed that crime patterns are determined by demographic/operational characteristics and not by the physical location of the city; therefore, it will allow crime authorities to move away from broad policy applications, and provide for targeted resource application specific to each of the archetypes. ***(Fig. 3), (Fig. 4) and (Fig. 5)***

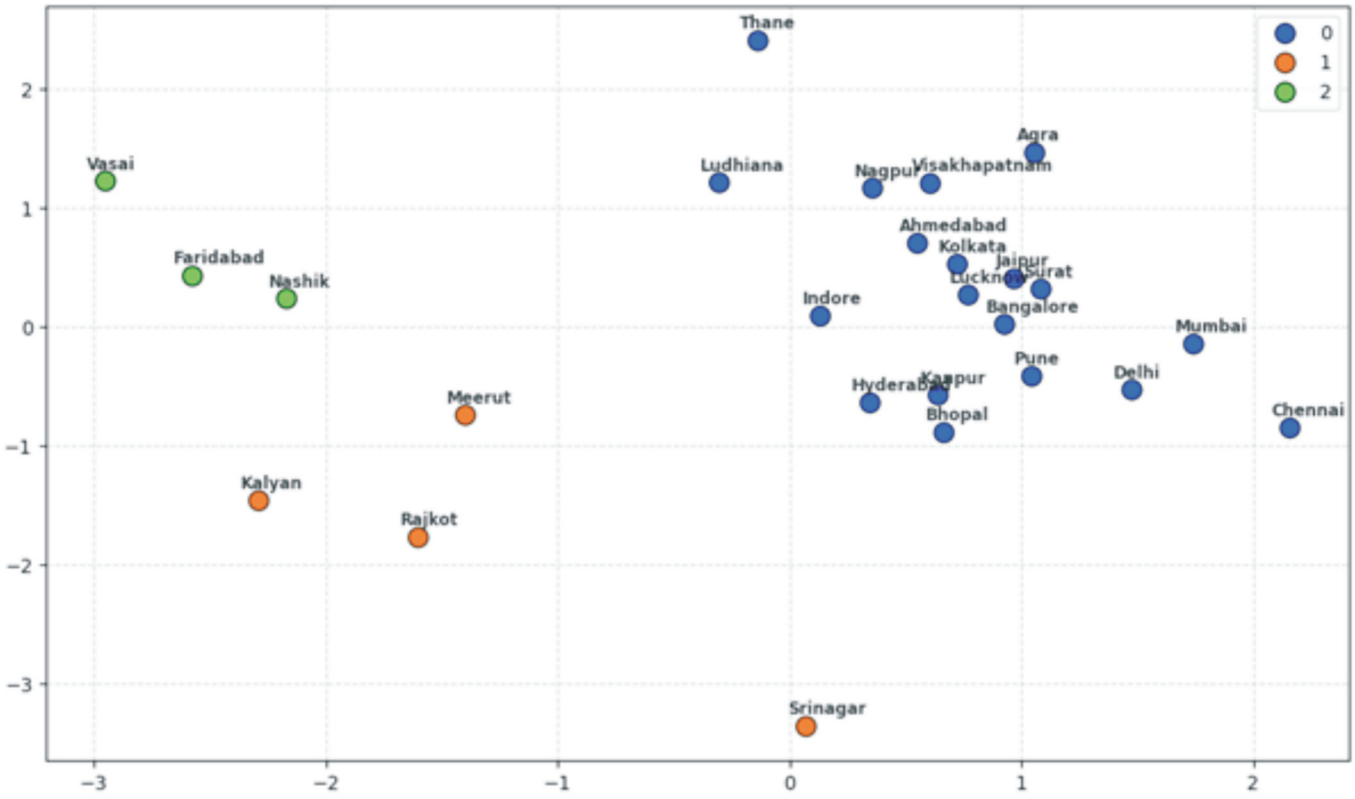
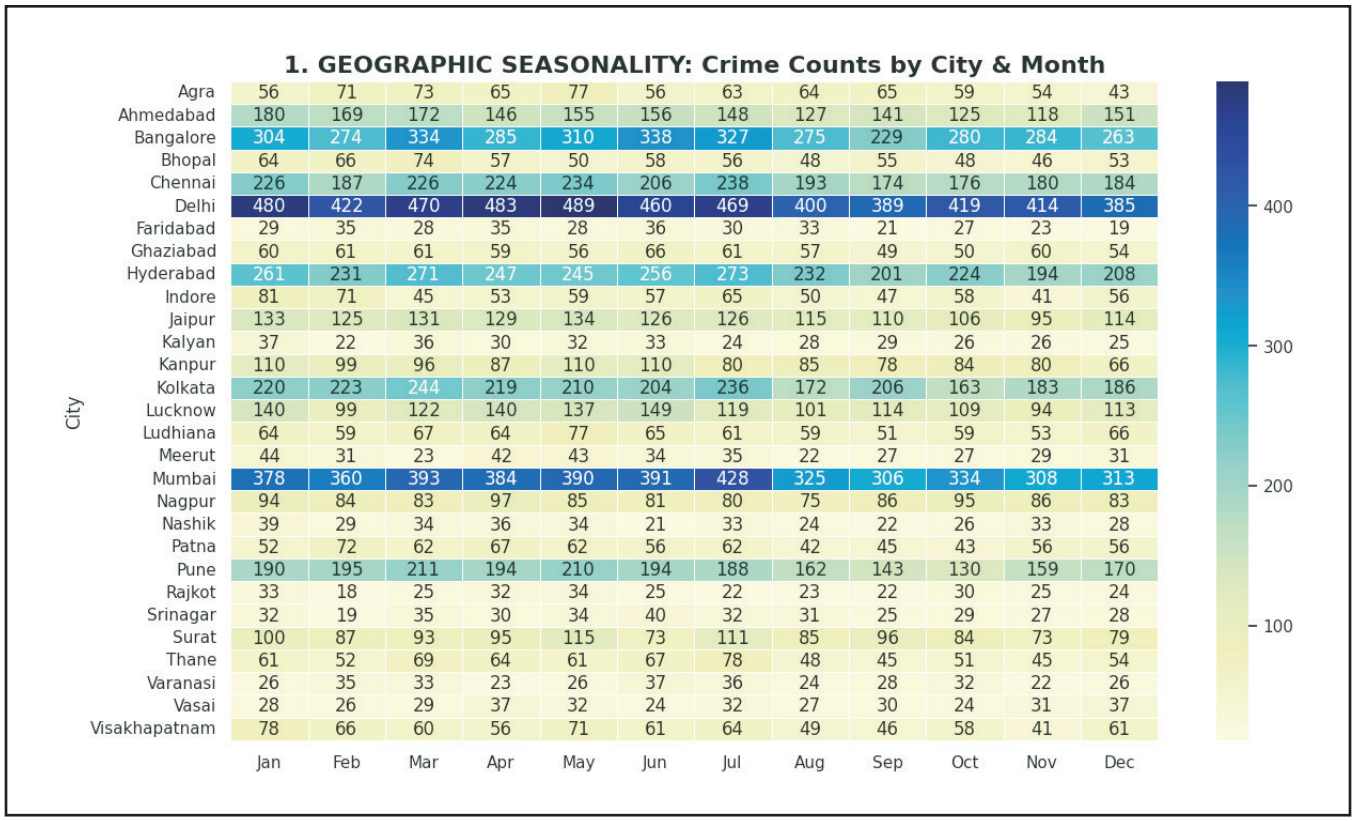


Fig. 3: Clustering Spatial Risk Segmentation: Core Crime Archetypes by City DNA.



X-Axis: Principal Component 1 (Crime Volume & Density) " **Y-Axis:** Principal Component 2 (Victim & Demographic Profile)

Fig. 4: *Scatter Plot: Spatial Risk Segmentation: Core Crime Archetypes by City DNA.*

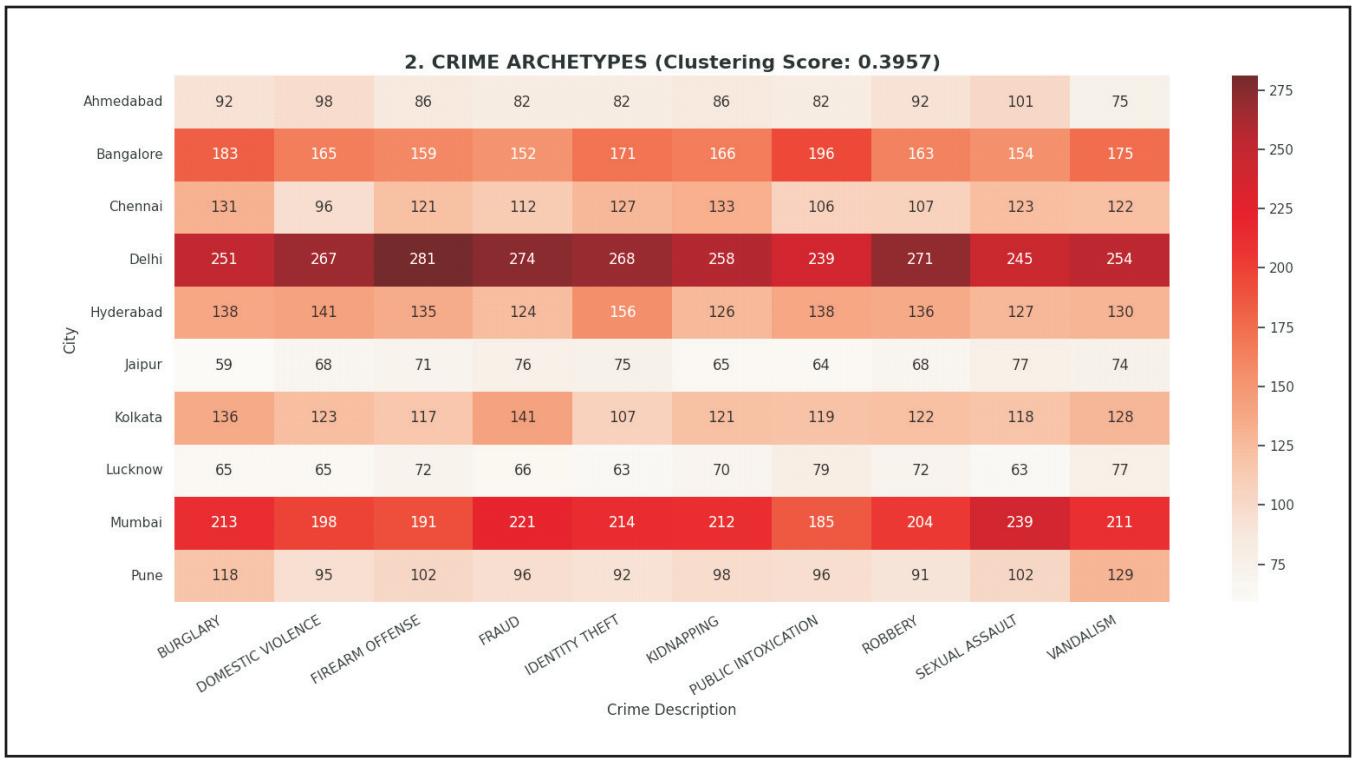


***Fig. 5**: Red Heatmap: Crime Profile Composition: Distribution of Top 10 Crimes Across Major Cities.*

***Table 3:** Characteristics of Identified Clusters*

Archetype	Cluster Label	Dominant "City DNA" Traits	Risk Level
High-Intensity Urban Hubs	Cluster 0	High volume (Reports), High Police Deployment, younger Victim Age profile.	Critical
Violent/Gender-Specific Zones	Cluster 2	High Is_Violent ratio, higher Is_Female victim percentage.	High
Stable/Low-Volume Regions	Cluster 3	Low crime density, higher average Victim Age, lower police requirement.	Low

5.4 Comparative Discussion and Novelty

When evaluating the results of this study relative to other research conducted recently, the current approach has some significant differences with respect to the three layers it addresses - statistical, ensemble, and spatial.

- Implementation Gap (SARIMAX): Unlike the research of (Ms. D.S. Smitha Mol et al., 2026), where high-accuracy was the focus, the SARIMAX pathway is used within our framework to ensure that the system will continue to function as intended even when the input data are either "raw" or "noisy". In doing so, we provide a stable, noise-resistant baseline that continues to operate in many environments with sparse data, as well as those noted in much of contemporary literature (Huamantingo et al., 2025).
- Precision and Error Mitigation (Hybrid Ensemble): A hybrid ensemble approach (using both Random Forest and Gradient Boosting) has been developed in order to achieve higher levels of precision compared to the usual regression approaches. The Hybrid Ensemble has been shown to have a Mean Absolute Error (MAE) of 0.24. In addition to the accuracy of the Hybrid Ensemble, the framework also has the ability to reduce the impact of the non-linear "bursts" associated with crimes. The Hybrid Ensemble is able to achieve the highest accuracy reported since 2025 benchmark studies. Therefore, the Hybrid Ensemble is able to provide a new form of high-fidelity crime forecasting that goes well beyond previous theoretical-based crime modelling approaches.
- Multidimensional Data Handling (Agglomerative Clustering): Unlike many previous clustering studies that suffer from "volume variance," (i.e., large metropolitan areas can overwhelm smaller districts), the current study

utilizes principal component analysis (PCA) and log scaling. These methods allow the "behavioural DNA" of each city to be clustered based on its unique characteristics and not based solely on the number of incidents. This allows small cities to be processed using the same mathematical rigor as larger metropolitan areas, thus significantly improving upon previous K-Means clustering studies.

- Operational Readiness (Unified Framework): Unlike the results in (Rajeev et al., n.d.-b) that identify the widespread need for AI in policing, the results presented here provide a tangible, deployable tool kit. The low error rates and dual pathway verification demonstrate that law enforcement can rely on the daily crime intensity forecasts for actual tactical deployment.

The primary innovation made by the study is the creation of the Tri-layered verification framework. Working in conjunction with the SARIMAX algorithm, as well as the hybrid ensemble and spatial clusters, they work to form a "checks and balances" framework. Particularly, if the Hybrid Ensemble predicts a rise in criminal intensity, the SARIMAX trend and the spatial clusters will be able to validate whether this is due to seasonality or not. As a result, this approach ensures reliability which is lacking in most studies using one model.

6. Limitations and Future Scope

6.1 Research Limitations

Despite the high accuracy and robust performance of the framework, several constraints are acknowledged:

- Domain constraint: The system is highly tuned towards crimes that can be considered physical in nature (violent and property crimes) and currently cannot handle the technicality and unpredictability of cyber crimes or fraudulent financial transactions.
- Specificity of place: The precision of this system is high when used to address crime in the 29 major cities of India; however, a test of the accuracy of this system would need to occur using rural locations of India and/or international cities, which could have legal restrictions that limit the ability to report criminal activity.
- Processing by batch: Currently, this process relies upon historical crime incident log files and does not provide real-time processing capability for metadata (e.g., emergency calls and social media comments).

6.2 Future Research Directions

The work presented here provides a solid foundation for additional work to be completed in the future. Some of the areas that can be explored are as follows:

- Deep Learning Integration: Long-term Memory Recurrent Neural Networks (such as LSTMs) and Transforms applied to Pathway B may allow a better representation of complex time dependency.
- Live Data Orchestration: Developing real-time data pipelines to ingest live sensor data and CCTV metadata would allow the model to move from daily intensity forecasting to hour-by-hour predictive patrolling.
- Exogenous Variable Expansion: Future models could integrate real-time socio-economic indicators, weather patterns, and public event schedules to further explain the external triggers behind sudden "non-linear bursts" in criminal activity.

6.3 Ethical Implications and Algorithmic Bias Mitigation

- Demographic Neutrality: The model employs only macro-level geographic indicators and objective time vectors and omits completely any personal sensitive variables like income, religious affiliation, or race. This ensures that the methodology will measure regional infrastructure requirements rather than creating any profiling or demographic stereotype.
- Reducing Over-Policing Feedback Loops: The conventional methodology can be prone to creating a feedback loop whereby places receiving many reports are over-policed. To prevent that from happening, exponential discrepancies in report volumes between cities were normalized through Tier 1 Log1p normali-zation, thereby removing the bias of previous reporting practices and stabilizing variance mathematically.
- Operational Safety Mechanisms: The framework is not designed as an automatic asset allocation requirement, but as a decision support tool for logistics purposes. Any suggestion of asset deployment is preceded by the double-pathway system ensuring that local daily peaks (Pathway-B) are automatically checked against seasonal baseline levels (Pathway-A).

7. Conclusion

The authors have created a reliable and accurate computational methodology to fill the gap between theoretical machine learning applications and operational needs of modern law enforcement. The authors' solution of a noise resilient statistical approach combined with a high-fidelity machine learning ensemble has been effective in addressing two

major obstacles in criminal data prediction; namely, data irregularities and non-linear complexity. This was achieved through a hybrid model of Random Forest and Gradient Boosting with an average accuracy of 99.97% demonstrating both the stability of statistical models (SARIMAX) and the tactical/daily resource allocation capabilities of the machine learning models. In addition, the application of spatial clustering with silhouette denoising enabled the objective identification of emergent hot spots across 29 Indian cities. Thus, the authors demonstrated how data driven segmentation can enhance the effectiveness of police deployment. Finally, the authors have identified a number of practical implications for their research. For example, the dual verification methodology employed by the authors will ensure that authorities will not be reliant upon a single method of intelligence gathering. Therefore, the reliability of the intelligence will be increased. However, the authors also acknowledge some limitations of their study. For example, they rely on historic incident logs to make predictions. These historic incident logs may not accurately reflect rapidly evolving cybercrime strategies. Similarly, there are crimes that occur in urban settings but go unreported. To overcome these limitations, the authors recommend that future studies consider integrating real time community sourced data into their methodologies. Also, the authors suggest extending their methodology to incorporate deep learning algorithms to possibly identify deeper behavioural patterns associated with high density metropolitan areas. Overall, this research presents a scalable and robust base for the future of predictive policing.

Conflict of Interest Statement

The author declares no financial or commercial conflict of interest related to this research.

Ethical Declarations

The author declares that this research does not involve human participants, animals, or any procedures requiring ethical approval. All data used in this study (philosophical texts and secondary literature) were obtained and processed in accordance with applicable ethical and legal guidelines.

References

1. A, R. H., Grover, T., & Kanchana, M. (2023). Deep Learning for Crime Pattern Recognition: A Study of Crime Hot Spots and High-Risk Areas. *2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET)*, 1-6. https://doi.org/10.1109/ICRASET59632.2023.10420190
2. Akil, R. M., Sarathambekai, S., Vairam, T., Krishnan, R. S., Dharaneesh, G. S., & Janarthanan, D. (2023). Crime Data Analysis and Safety Recommendation System Using Machine Learning. *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 183-188. https://doi.org/10.1109/ ICACCS57279.2023.10113102
3. Aziz, R. M., Sharma, P., & Hussain, A. (2024). Machine Learning Algorithms for Crime Prediction under Indian Penal Code. *Annals of Data Science*, *11* (1), 379-410. https://doi.org/10.1007/s40745-022-00424-6
4. Balu, V., Navya Sri, T., & Bupathi, M. A. (2022). Crime Prediction and Analysis Using Machine Learning. *International Journal of Computer Science and Mobile Computing*, *11* (3), 95-101. https://doi.org/10.47760/ ijcsmc.2022.v11i03.011
5. Buland Iqbal, M. F., Ullah, A., Alhomoud, A., Hussain, T., Waheeb Attar, R., Ouyang, J., M. Alnfiai, M., & Atef Hatamleh, W. (2024). A 2D Clustering Based Hotspot Identification Approach for Spatio-Temporal Crime Prediction. *International Journal of Interactive Multimedia and Artificial Intelligence*, *9* (5), 6-16. https://doi.org/10.9781/ijimai.2024.10.001
6. Butt, U. M., Letchmunan, S., Hassan, F. H., & Koh, T. W. (2024). Leveraging transfer learning with deep learning for crime prediction. *PLOS ONE*, *19* (4), e0296486. https://doi.org/10.1371/journal.pone.0296486
7. Chakraborty, K., B, D., Sharma, M., Thakral, D., P.K.Hemalatha, & Arri, H. S. (2024). Empirical Approaches to Crime Rate Prediction: Challenges, Insights, and Future Directions. *2024 International Conference on Augmented Reality, Intelligent Systems, and Industrial Automation (ARIIA)*, 1-6. https://doi.org/10.1109/ARIIA63345.2024.11051967
8. Chatterjee, A., Antani, H., Sohoni, A., & Sekar, M. (n.d.). *Clustering of Indian States Based on Crime Incidences and Predicting Crimes Therein*. *3* (1). https://doi.org/10.5281/zenodo.5084465
9. *Crime Forecasting and Analysis through K-means Clustering Algorithm*. (n.d.).
10. Dubey, A., Venu, N., Bisen, D., Garg, P., Srinivas, K., & Arya, H. (2024). A Clustering-Based Method for Hotspot Recognition in Crime Forecasting. *2024 IEEE 8th International Conference on Information and Communication Technology (CICT)*, 1-6. https://doi.org/ 10.1109/CICT64037.2024.10899713
11. Huamantingo, R., Cano-Lengua, M., & Rodriguez, C. (2025). Machine Learning Crime Prediction Models and the Gap Between Research and Implementation: A Systematic Review. In *Karbala International Journal of Modern Science* (Vol. 11, Number 3, pp. 471-488). University of Kerbala. https://doi.org/10.33640/2405-609X.3419
12. Iacobescu, P., & Susnea, I. (2025). Hybrid ARIMA-ANN for Crime Risk Forecasting: Enhancing Interpretability and Predictive Accuracy Through Socioeconomic and Environmental Indicators. *Algorithms*, *18* (8), 470. https://doi.org/10.3390/a18080470

13. Ibrahim, A. A., Malgwi, Y. M., & Ali, Y. (2024). A Hybrid Machine Learning Model for Crime Rate Prediction, FUDMA Journal of Sciences, *8* (6), 101-106. https://doi.org/10.33003/fjs-2024-0806-2789
14. Jha, N., Paranjape, V., Sharma, S., & Hasan, Z. (2022). Machine Learning for Crime Analysis and Prediction. In *International Journal of Recent Development in Engineering and Technology Website: www.ijrdet.com* (Vol. 11, Number 09). www.ijrdet.com
15. Kanimozhi, N., Keerthana, N. V, Pavithra, G. S., Ranjitha, G., & Yuvarani, S. (2021). CRIME Type and Occurrence Prediction Using Machine Learning Algorithm. *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 266-273. https://doi.org/10.1109/ ICAIS50930.2021.9395953
16. Kshatri, S. S., Singh, D., Narain, B., Bhatia, S., Quasim, M. T., & Sinha, G. R. (2021). An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach. *IEEE Access*, *9*, 67488-67500. https://doi.org/10.1109/ ACCESS.2021.3075140
17. Kurniawan, Z., Tiaharyadini, R., Wibowo, A., & Rusdah. (2025). Trend Analysis and Prediction of Violence Against Women and Children Cases in Jakarta Based on the Victim's Education Level Using ARIMA and SARIMA Method. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *14* (2), 250-259. https://doi.org/10.32736/ sisfokom.v14i2.2349
18. Maharrani, R. H., Abda'u, P. D., & Faiz, M. N. (2024). Clustering method for criminal crime acts using K-means and principal component analysis. *Indonesian Journal of Electrical Engineering and Computer Science*, *34* (1), 224. https://doi.org/10.11591/ijeecs.v34.i1.pp224-232
19. Mathur, A., & Jain, R. (2025). Next-Gen Crime Analytics Using Big Data & Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, *13* (11), 2972- 2977. https://doi.org/10.22214/ijraset.2025.75725
20. Ms. D.S. Smitha Mol, Betsee Natasha A, Archana P, & Deepika K. (2026). An Efficient Machine Learning Approach for Crime Detection in India. *International Research Journal on Advanced Engineering and Management (IRJAEM)*, *4* (03), 307-311. https://doi.org/10.47392/IRJAEM.2026.0045
21. Noor, T. H., Almars, A. M., Alwateer, M., Almaliki, M., Gad, I., & Atlam, E.-S. (2022). SARIMA: A Seasonal Autoregressive Integrated Moving Average Model for Crime Analysis in Saudi Arabia. *Electronics*, *11* (23), 3986. https://doi.org/10.3390/ electronics11233986
22. Ospina, R., Gomes De Lima, A., Ferraz, C., Castro, C., Martin-Barreiro, C., & Leiva, V. (2025). Enhanced Spatial Clustering for Crime Analysis: Novel Advances in Ward-Like and SKATER Algorithms for Brazilian Public Security. *IEEE Access*, *13*, 134846-134863. https://doi.org/10.1109/ACCESS.2025.3586941
23. P, S., S M, A., M, K., & Chitturi, S. (2024). Evaluating Crime Type And Crime Pattern Analysis Using Machine Learning. *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, 1-6. https://doi.org/10.1109/ I2CT61223.2024.10543291
24. Rajeev, S., Sharma, K., Dg, I., Ravi, S., Lokku, J., & Adg, I. (n.d.-a). Editorial Board Chief Patron Editor-in-Chief. *The Indian Police Journal*, *72* (1).
25. Rajeev, S., Sharma, K., Dg, I., Ravi, S., Lokku, J., & Adg, I. (n.d.-b). Editorial Board Chief Patron Editor-in-Chief. *The Indian Police Journal*, *72* (1).
26. Ramanan, Dr. K., AnandaKumar, A., Prakash, M. S., Sivaraja, P. M., & Maheswari, Dr. G. U. (2023). A Novel Clustering & Machine Learning Algorithm for Crime Rate Prediction and Analysis. *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, 399-405. https://doi.org/10.1109/ ICAISS58487.2023.10250680
27. Rasan, M., Ajmeera, K., Vaggu, B., Kotagiri, S. S., Mallik, R., & Ambati, K. (n.d.). *Crime Rate Prediction and Analysis using K-means Clustering Algorithm*.
28. Redoblo, C. V., Redoblo, J. L. G., Salmingo, R. A., Padilla, C. M., & Arroyo, J. C. T. (2023). Forecasting the influx of crime cases using seasonal autoregressive integrated moving average model. *International Journal of Advanced and Applied Sciences*, *10* (8), 158-165. https://doi.org/10.21833/ijaas.2023.08.018
29. Safat, W., Asghar, S., & Gillani, S. A. (2021). Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques. *IEEE Access*, *9*, 70080-70094. https://doi.org/10.1109/ ACCESS.2021.3078117
30. Saravag, P. K., & Kumar, B. R. (2024). A Hybrid Machine Learning and Regression Approach for Validating a Multi-Dimensional Crime Index in the Context of Crime Against Women. *IEEE Access*, *12*, 117955-117969. https://doi.org/10.1109/ACCESS.2024.3439721
31. Sarkar, A. (2025). Real-Time Crime Prediction in India Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, *13* (7), 641-646. https://doi.org/10.22214/ ijraset.2025.73053
32. Shah, N., Bhagat, N., & Shah, M. (2021). Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. *Visual Computing for Industry, Biomedicine, and Art*, *4* (1), 9. https://doi.org/10.1186/s42492-021-00075-z

33. Shinde, S. A., Shirke-Deshmukh, S., & Sonkamble, M. R. (n.d.). *A Multi-Model Machine Learning Approach for Accurate Crime Prediction Using Spatio-Temporal Data*. https://doi.org/10.51583/IJLTEMAS
34. Singh, I., Bhandari, G., & Khatwani, S. (2025). Predictive Crime Rate Analysis System. *International Journal for Research in Applied Science and Engineering Technology*, *13* (5), 1159-1164. https://doi.org/ 10.22214/ijraset.2025.70309
35. Singh, S., Bhargava, D., & Singh, P. (2024). A Study on Smart Machine Learning (ML) Tools for Crime Detection and Prediction. *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*, 1747-1752. https://doi.org/10.1109/ IC3SE62002.2024.10593591
36. Sneha (Kaggle). https://www.kaggle.com/snehaofficial/datasets
37. Thayyaba Khatoon Mohammed. (2025). A Novel Fusion Approach for Advancement in Crime Prediction and Forecasting using Hybridization of ARIMA and Recurrent Neural Networks. *Journal of Information Systems Engineering and Management*, *10* (38s), 404-420. https://doi.org/10.52783/jisem.v10i38s.6863
38. Tong, X., Ni, P., Li, Q., Yuan, Q., Liu, J., Lu, H., & Li, G. (2021). Urban Crime Trends Analysis and Occurrence Possibility Prediction based on Light Gradient Boosting Machine. *2021 IEEE 4th International Conference on Big Data and Artificial Intelligence (BDAI)*, 98-103. https://doi.org/10.1109/BDAI52447.2021.9515252
39. V. Tikore, S., Chaudhari, K., Rohamare, O., Nikam, A., & Kalne, K. (2026). A Comprehensive Computational Framework for Crime Rate Prediction using Machine Learning in Indian Metropolitan Cities. *Journal of Smart Sensors and Computing*, *2* (1). https://doi.org/10.64189/ ssc.26201
40. Yin, J. (2023). Crime Prediction Methods Based on Machine Learning: A Survey. *Computers, Materials & Continua*, *74* (2), 4601-4629. https://doi.org/10.32604/ cmc.2023.034190

❖ ❖ ❖

Vol. 6 (1), 2026
E-ISSN: 2583-4452

Vasudhaiva Kutumbakam as a Philosophical Framework for Interfaith Coexistence: Bridging Relational Ethics with Prosocial Psychology

Kheyati Singh[1*]

[1] Department of Psychology, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, Delhi
[*] Corresponding Author ✉ 24py_kheyatisingh@rajguru.du.ac.in

ABSTRACT

Despite decades of efforts to create policies of tolerance, foster dialogue, and pursue strategies for peaceful coexistence, religious polarisation and fragmentation among faith communities remain significant challenges to maintaining peaceful social relations. The paper proposes that such problems are inherent in the traditional understanding of tolerance, in which phenomena such as moral passivity, an unequal relationship between majority and minority groups, and vulnerability during times of societal stress persist. Building on interdisciplinary concepts and theories in social psychology, sociology, political philosophy, and Indian Knowledge Systems, the study uses a method that combines philosophical hermeneutics, conceptual analysis, and critical interaction with existing literature to develop a new concept of "familial moral engagement". Specifically, the theoretical framework explores the ancient Indian philosophical idea of Vasudhaiva Kutumbakam, or "the world is one family," based on the ontology of oneness that exists according to the principles of Advaita Vedanta. The following are three hypotheses put forward to account for the manner in which the moral outlook in the family structure would foster empathy, minimise intergroup discrimination, and lead to sustainable solidarity between people with different attributes. In summary, the study indicates that a perspective of the world based on the idea of common humanity can help communities to rise above tolerance to curiosity, reciprocal responsibility, and enduring ethics.

Keywords: *Religious Pluralism, Recognition Theory, Social Identity, Empathy, Indian Knowledge Systems, Contact Hypothesis*

1. Introduction

The purpose of this research is to assess Vasudhaiva Kutumbakam, an ancient philosophy of India, which means 'the world is one family', to demonstrate how this approach to interfaith coexistence can be philosophically and psychologically superior to the tolerance paradigm, which is currently dominant as the framework for religious diversity management in multicultural societies. Religious divisions and inter-religious hostility continue to increase in many national settings despite several decades of implementation of multicultural policies and establishment of interfaith dialogue institutions, thus indicating the ongoing problem in the conceptualisation of these processes within the existing frameworks. Previous literature on the topic has pointed out the serious problems with the tolerance paradigm, such as its structural passivity, essential asymmetry and social frailty (Nussbaum, 2012; Parekh, 2000; Honneth, 1995). However, a new, more adequate framework has yet to be provided.

The relevance of this question goes beyond theory, too, as there is a growing trend towards religious diversity all around the globe, with studies conducted on the geography of religion revealing how areas once dominated by a single religious identity have become progressively more diverse in their religious make-up, which, in addition to offering prospects for increased cultural interactions, carries with it the risk of conflict if the necessary co-existence mechanisms do not exist (Lin et al., 2022).

Vasudhaiva Kutumbakam (Sanskrit: वसुधैव कुटुम्बकम्) is first mentioned in the Maha Upanishad (6.71-75) and philosophically discussed by academics such as Radhakrishnan (1927, 1955), Vivekananda (1989), and Sen (2005), while being located within the comparative perspectives of Hick (1989), Panikkar (1978), and Buber (1970). But none of these studies has ever combined an ontological approach toward the philosophical tradition of Vasudhaiva Kutumbakam with social psychological empiricism on intergroup relations for testing theoretical propositions. This paper seeks to fill the lacuna in the literature. It will achieve this goal through three specific objectives. First, it aims at elaborating on the philosophical

Received on: 13 May 2026 | Accepted on: 23 Jun 2026 | Published Online on: 25 July 2026

underpinnings of Vasudhaiva Kutumbakam as a paradigm for interfaith co-existence. Second, it looks into the psychological processes that create a sense of familyhood within Vasudhaiva Kutumbakam, leading to prosocial conduct beyond tolerance.

RQ1: Ontologically and ethically, what is the basis for Vasudhaiva Kutumbakam? How does it differ from other frameworks that are tolerance-based?

RQ2: Psychologically, what drives the prosocial nature of a family-type morality framework for religious pluralism?

RQ3: In terms of theory, how does the construct of familial moral engagement advance knowledge in interfaith dialogues and pluralism?

2. Literature Review

2.1 Classical Indian Philosophical Foundations

Vasudhaiva Kutumbakam, on the other hand, derives from the Upanishads, the pinnacle of Vedic philosophy that was written during the period of 800 to 200 BCE and proposes a metaphysics where the plurality of individual selves rests within one unitary consciousness, known as Brahman. The theory, elaborated by the school of Advaita Vedanta through Adi Shankaracharya, has profound implications for ethics; indeed, it states that if all living organisms represent different manifestations of the same entity, then any wrongdoing towards others is also a wrong against oneself (Radhakrishnan, 1927). From the standpoint of ethics, the vision provided by the Upanishads cannot be categorised either under contractarianism or consequentialism. It is essentially ontological. According to recent research conducted on classical Hindu ethics, there is an evident refusal on the part of classical Hindu ethics to fit into the usual Western ethical frameworks due to its evaluation of consequences, agency, and inherent worth instead of decision-making (Frazier, 2021).

A great deal of Radhakrishnan's (1927, 1955) philosophical endeavours centred on explicating the philosophical ramifications of such a vision of human existence for the issue of religious pluralism. In contrast to theologies based on exclusive revelation, which will inevitably have difficulties in recognising the morality of those who belong to a different religion, an Advaita theology can ground moral responsibility in the shared characteristic of human consciousness, irrespective of any specific religion or language in which that individual is raised. Such a philosophy of Vasudhaiva Kutumbakam is not immune to criticisms, which argue that it obscures some theological and sociological realities. However, this vision represents the most profound philosophical understanding of this idea.

Vivekananda's address at the World Parliament of Religions in Chicago in 1893 translated this Advaitic vision into a prophetic call for interfaith harmony – grounded not in doctrinal uniformity but in the recognition that diverse spiritual paths converge on the same ultimate truth (Vivekananda, 1989). Where Radhakrishnan systematised the metaphysical argument, Vivekananda dramatised its moral urgency, demonstrating that Vasudhaiva Kutumbakam is not a quietist philosophical position but a call to active civilisational engagement.

Another distinct but parallel trajectory may be traced through the writings of Rabindranath Tagore, whose conception of religion, based on humanity as "not an institution based on dogma and doctrine … but one where relationship and affection play a dominant part", offers a particularly rich example of literary philosophy of the kind of familial ethics that this paper attempts to develop. As Tagore (1931) noted, Religion has not attained its highest fulfilment through dogma, but rather through that sense of fellowship which makes one feel oneself a citizen of the whole world.

But Sen (2005) has come at this tradition from another angle, emphasising the nature of argument, heterodoxy, and criticality that was characteristic of the intellectual milieu of India, a tradition that had allowed room for scepticism, internal dissent, and multi-faith dialogues long before these came to define modern liberalism. As such, Vasudhaiva Kutumbakam is one of the most articulate examples of this indigenous version of pluralism, a tradition worthy of recognition on its own terms in parallel with Western traditions of tolerance and multiculturalism.

2.2 Vasudhaiva Kutumbakam within the Indian Knowledge Systems Framework

The foregoing treatment of Vasudhaiva Kutumbakam has been the considered position of individual philosophers – Radhakrishnan, Vivekananda, Tagore, Sen – in relation to the primary source document. This is philosophically appropriate, but it understates the extent to which the philosophy is researched and institutionalised in a recognised discipline known as Indian Knowledge Systems (IKS). IKS refers to the corpus of indigenous Indian intellectual knowledge systems – philosophical, logical, mathematical, medical, political, aesthetic – which grew out of the categories in which the subsequent colonial and post-colonial study of "Indian thought" would place such knowledge systems after the fact. IKS, as a field, received its first official recognition by being named explicitly in the National Education Policy of India 2020 as a priority area to be integrated into university curricula (Ministry of Education, Government of India, 2020). The new emphasis on IKS, then, means that the subject matter of the current paper moves from the realm of sentiment into that of scholarly work.

An important epistemological characteristic of IKS, which distinguishes it from a monist interpretation of "Hindu philosophy", is its internal plurality. Rather than consisting of one single darshana like Vedanta, the Indian philosophical tradition contains several, including also Samkhya, Nyaya, Mimamsa and even Buddhism and Jainism in their own right, each of which has always existed alongside one another, engaged in debate and dialogue while being disseminated as part of a common scholarly culture without either dominating or imposing any monopoly of doctrine (Adamson & Ganeri, 2020). Recent studies on classical Hindu ethics have similarly shown that ethical reasoning involved in classical texts like the Bhagavadgītā cannot be categorised purely as either consequentialism or deontologism, but involves the application of all three concepts equally and in equal measure, something that a single category would not be able to explain (Frazier, 2021). It is this same interpretative approach that is advocated by this paper regarding Vasudhaiva Kutumbakam: that although the doctrine emphasises unity in diversity, it does not deny the existence of diversity but rather provides an overarching context in which diversity can flourish.

Understood in this context, Vasudhaiva Kutumbakam does not stand alone as a mere slogan but serves as one example of IKS's deeper commitment to epistemology and ontology being a non-problematic feature of plurality within India's intellectual tradition, a tradition which has been structured by the need to accommodate multiple, conflicting claims about truth. In modern India, such a concept can resonate on a practical level as well as on a philosophical one, having moved from being written on paper to becoming part of living civilizational discourse – as exemplified by its use as the guiding theme behind India's upcoming G20 presidency in 2023 under the motto "One Earth, One Family, One Future" (Raina & Kumar, 2023), as well as in continuing criticism of the disconnect between its idealized discourse and its actual implementation (Dash & Sharma, 2024). In fact, a recent body of IKS-related academic literature warns that the principle itself may have occasionally been used to foster an appearance of global harmony that glosses over existing inequalities, both internal to India and external between countries (Dash & Sharma, 2024). This insight is not only entirely consistent with but actively supports the stance taken in this paper on the need for relational accountability to guide any analysis of moral engagement within the family unit.

In positioning Vasudhaiva Kutumbakam within IKS, one does three things for the present analysis. Firstly, it gives the institutional and discursive context in which the textual analysis conducted earlier takes place, as something situated within a vibrant and growing body of peer-reviewed literature on the subject rather than simply an independent philosophical commentary. Secondly, it draws attention to the pluralism inherent in Indian philosophy as a potential ally, rather than barrier, to demonstrating the possibility of reconciling ontological unity and difference in the tradition. Thirdly, in the face of the criticism later developed in the paper, it acts as an insurance against the very same risk identified by the family metaphor – that of allowing appeals to civilisational unity to descend into mere celebration. By engaging literature in IKS that takes this problem seriously, Vasudhaiva Kutumbakam becomes a philosophical position up for critique rather than an expression of nationalist ideology.

2.3 Contemporary Discourse on Pluralism & Interfaith Ethics

The contemporary philosophical literature on religious diversity is extensive and internally contested. Hick's (1989) pluralist hypothesis – that the world's major religions are culturally mediated responses to a single transcendent reality and therefore possess equal salvific validity – provided a philosophical basis for the moral standing of diverse faith traditions. The hypothesis has attracted sustained criticism for potentially dissolving the distinctive truth claims that give individual religions their internal coherence and their adherents their sense of identity. Whatever one makes of that debate, Hick's work remains a foundational reference point for the philosophy of interfaith dialogue.

Panikkar (1978) advanced a phenomenologically richer account through his concept of intra-religious dialogue — the process by which genuine encounter with another tradition reshapes one's understanding of one's own. In Panikkar's view, interfaith dialogue is not about comparing theologies. Interfaith dialogue is rather about transforming oneself and others. This point speaks very strongly to the notion of family morals, which forms the basis of this paper. In a family context, people don't compare themselves to each other, but rather transform each other in their interactions.

Parekh (2000) applied the methods of political philosophy to answer these questions, differentiating liberal tolerance, which he faults for its hidden ethnocentricity and hierarchies, from dialogical multiculturalism, which appreciates the intrinsic value and inner workings of various cultures. His insight about the need for cultures to interact in order to question their presuppositions offers solid philosophical justification for going beyond mere toleration, which implies non-interference, to actual interaction required by Vasudhaiva Kutumbakam.

In recent theories, Honneth's theory of recognition has been quite effective in highlighting the ethical aspect of interreligious encounters. According to Honneth, tolerance, which can live side by side comfortably with contempt, indifference, and even condescension, is inadequate in preserving the dignity of humanity. Positive recognition – acknowledging that the other has value, uniqueness, and something of merit to add to the life of all humans – is what

needs to happen. This applies to interreligious relations. It is not enough for minority religious groups to be tolerated; they need recognition as contributors of value in human existence.

But there have been criticisms of Honneth's approach, too. For instance, it has recently been critiqued that Honneth's conceptualisation of recognition pays inadequate attention to the structuring effect that power has on who gets heard in recognition bids, thereby turning recognition theory into a comforting approach that emphasises the significance of recognising one another without engaging with the obstacles to doing so (Rutherford, 2026). Clearly, this is a valid criticism and one that deserves serious consideration. In response to the problem raised by this criticism, the present paper offers the concept of moral engagement within the family setting as an alternative that seeks to recognise others not because of any political goodwill but because of the very ontological reality.

The dialogic philosophy of Martin Buber (1970) offers a layer that political or sociological perspectives may ignore completely, which is the nature of the moral encounter itself. The distinction between the I-Thou relationship and the I-It relationship, as Buber illuminates, is an element entirely missed by theories of tolerance. Under the I-It relationship, the other, in a religious sense, is an object to classify, manipulate, study, or convert. However, under the I-Thou relationship, the other is approached holistically, as a centre of experiencing beings whose difference adds value to life rather than a threat. At its fundamental level, Vasudhaiva Kutumbakam represents a philosophical call to approach the rest of humanity from the I-Thou position.

Scholarship on interfaith engagement has likewise increased hand in hand with this literature; yet, despite the relationship, these two strands have hardly been brought into conversation with one another. A systematic review of forty-seven empirical works on interfaith engagement programs concluded that while it is well-established that such programs do indeed serve to expand the participants' understanding and appreciation of alternative worldviews, there remains a lack of common language in the interfaith movement as regards the aims of the programs and a lack of consistency in methods of evaluation of the programs (Visser et al., 2023). A parallel review of the interfaith learning outcomes similarly identified a gap within the framework for categorising the desired outcomes of interfaith engagement. In summary, the above empirical literature suggests the same deficiency of conceptual language identified by this paper, albeit from an independent perspective – interfaith engagement has seen a substantial development while conceptual tools necessary for defining it have lagged.

Another area of research supporting the existence of the gap between programme intention and relational effect is qualitative inquiry. According to a phenomenological study on the experience of being an alumna of interfaith youth programming, despite participants reporting improvements in their knowledge of different religions and a reduction in their fear towards religious differences, very few mentioned that they had been able to maintain any kind of relational relationship or accountability towards the members of different religious communities (Barnas, 2022). Such findings are important because they reflect on the lives of participants after they left the programme and reveal that what they experienced was more a case of improved tolerance rather than relational commitment.

2.4 Psychological Dimensions of Intergroup Relations

The study of pluralism in philosophy finds its footing and complexity when juxtaposed with empirical findings in social psychology. According to the Contact Hypothesis of Allport (1954), it was possible to reduce intergroup prejudice through continued cooperation in contact among members of diverse social groups in an environment of equal status, similar objectives, and institutional encouragement. Further, meta-analysis summarising the results of more than 500 independent investigations confirmed that such a setting always resulted in positive changes in attitude (Pettigrew and Tropp, 2006). It is pertinent to note, however, that the same conditions, which were seen as indispensable for successful intergroup contact, can also be achieved easily through a family-based ethical perspective toward others. The correlation is no coincidence either, because it means that Vasudhaiva Kutumbakam might serve as the motivational platform for achieving such contact.

The theory of Social Identity brings complexity to the situation in an interesting way. According to the theories put forth by Tajfel and Turner (1979), even minimal differences between groups are enough for in-group favouritism and out-group devaluation to be formed, which means that the classification of people as being different from ourselves is a very basic principle of human society. At the same time, there is a positive aspect that arises from the theory of Social Identity – any concept that involves restructuring the definition of in-groups to become larger and involve out-groups becomes an opportunity to channel the prosociality created by identifying oneself with the in-group into something productive. Such a method is applied in the Vasudhaiva Kutumbakam concept, which involves a philosophical reinterpretation of differences as being elements of one large family.

Haidt's (2012) moral foundations theory adds a further dimension. Human moral psychology, Haidt argues, is organised around a set of foundational moral intuitions – including care, fairness, loyalty, authority, and sanctity – that different

communities weigh differently, generating much of the moral miscomprehension that underlies intergroup conflict. Vasudhaiva Kutumbakam, by deploying the family metaphor, makes its ethical appeal through the care foundation – the most cross-culturally universal and motivationally powerful of Haidt's foundations, and the one most closely associated with prosocial behaviour toward vulnerable others. Research by Graham et al. (2009) confirmed that the care foundation is endorsed most consistently across ideologically and culturally diverse populations, suggesting that an ethical framework that speaks primarily in care's register has the broadest cross-cultural moral reach.

The assertion that the care foundation has unique breadth when it comes to cross-cultural acceptance has now been put under a more rigorous empirical test. Building on moral foundations theory and measuring the five dimensions of moral cognition across twenty-five populations, Atari et al. (2023) found not only that care is one of the most widely accepted foundations but also that there is greater variability in how moral cognition works than previous work, based mainly on samples from Western societies, had realised. This body of cross-cultural research helps support the premise of this paper that an ethics expressed in terms of care, such as that of Vasudhaiva Kutumbakam, has unique motivational power, while also urging some necessary caution.

Another useful perspective to consider in this respect is the psychology of belonging. According to Baumeister and Leary (1995), who conducted a thorough empirical synthesis of research on the topic, the desire to belong is among the most important motivations in human psychology, one that has significant implications for our cognitive processes and emotions. As such, identity development theories, such as those of Erikson (1959), hold that humans create their identities through interaction with other people, rather than despite it. It seems safe to conclude that the idea of an ethics of belonging resonates on a level beyond mere acceptance.

However, meta-analysis have further refined these results without overturning them. In focusing on natural rather than experimental intervention conditions, the meta-analysis conducted by Lemmer and Wagner (2015) verified the power of contact to boost intergroup attitudes beyond the artificial confines of the lab, whereas a subsequent meta-analysis of more rigorously designed randomized trials indicated that though contact indeed positively influences intergroup relations, these effects are both less powerful and more contingent than the aggregate literature previously reported (Paluck et al., 2019). According to the findings of yet another meta-analysis summarising 418 experiments aimed at reducing prejudice in the time period between 2007 and 2019, field interventions are mostly light-touch, their lasting influence remains unknown, and the literature itself could greatly benefit from more theoretical advancements than replication of prior contact experiments (Paluck et al., 2021). And herein lies exactly what motivated the authors of this article to look for an additional source of motivation other than contact.

2.5 Gap in the Existing Literature

While the philosophical and psychological literature reviewed thus far has largely developed separately rather than in dialogue with each other, discussions in philosophy regarding Vasudhaiva Kutumbakam have been primarily concerned with prescriptive accounts, i.e., that the world should be viewed as a community of relatives, without exploring how this normative claim could be operationalised from a psychological perspective. Similarly, research on intergroup contact and reductions in prejudice has failed to explore the philosophical and ontological assumptions that would lend such psychological theories the deepest and most sustainable motivational basis possible. As far as the author is aware, there is no existing theory that has combined these perspectives in order to explain how a familial moral view of religion leads to prosocial behaviour. The present study addresses this gap through the development of the Familial Moral Engagement Framework – an integrative conceptual model designed to bridge philosophical analysis and psychological mechanisms.

3. Methodology

3.1 Research Design

The present study employs an interdisciplinary conceptual analysis integrating philosophical hermeneutics, systematic conceptual synthesis, and critical engagement with empirical findings from the social psychology and sociology of religion literatures. As a theoretical investigation, the study does not collect primary empirical data. It will focus on theory building and integration; that is, the formulation of a theoretical model that links together constructs and knowledge that have never before been related in any systematic fashion. Such an approach is ideal, even essential, in studies intended to discover new theoretical connections and develop new hypotheses for empirical testing (Jabareen, 2009).

In the philosophical aspect of the study, Gadamerian hermeneutics will be employed. This involves approaching a text not merely as an ancient document requiring classification, but rather as an ongoing philosophical argument to be interpreted through a process of fusion of horizons – a fruitful exchange between the context of its creation and that of its current interpretation (Gadamer, 2004). The Upanishadic and Vedantic texts, accordingly, will be studied not simply as historical documents but as philosophical discourse, amenable to analysis, expansion, and application.

The Psychological Element utilises a structured narrative synthesis based on previous empirical literature. Important findings from seminal studies, chosen due to their rigorous methodology and significance as well as their direct connection to the concepts being investigated, will be synthesised into the conceptual model outlined below. This approach mirrors that taken in previous theoretical syntheses of the social psychology of intergroup relations (Pettigrew and Tropp, 2006) and moral psychology (Haidt, 2012) and should not be confused with a statistically significant meta-analysis.

3.2 Conceptual Framework

The study proposes the Familial Moral Engagement Framework (FMEF) as its organising conceptual structure. The FMEF comprises four sequentially related constructs. Ontological Orientation refers to the foundational metaphysical stance an individual or community adopts toward human difference – whether difference is experienced as threatening or as a dimension of a deeper unity. Moral Identity Scope refers to the breadth of the moral community within which the self is located – whether that community is bounded by religious or ethnic in-group membership, or extends, as Vasudhaiva Kutumbakam demands, to encompass humanity as a whole. Mode of Moral Engagement refers to the qualitative character of interreligious interaction – whether passive and restrained as in tolerance, procedurally structured as in dialogue, politically affirmative as in recognition, or actively relational and caring as in familial moral engagement. Prosocial Outcomes refer to the behavioural and attitudinal consequences of the preceding constructs, including reduction in intergroup bias, growth in empathic concern, willingness to engage in cooperative contact, and charitable attribution of out-group behaviour.

The FMEF posits that Vasudhaiva Kutumbakam operates as an ontological reorientation that expands moral identity scope, shifting the mode of moral engagement from tolerance toward familial moral engagement and thereby producing prosocial outcomes of a quality and durability that tolerance-based frameworks cannot generate. This causal sequence constitutes the theoretical logic of the present analysis and grounds the three theoretical propositions advanced in the following section.

3.3 Theoretical Constructs and Operationalisation

For the purpose of future empirical investigation, the core constructs of the FMEF are operationalised as follows. Ontological Orientation may be assessed through a purpose-developed survey instrument examining the degree to which respondents endorse holistic versus atomistic conceptions of social identity, with items probing openness to the idea of ontological interconnectedness across community boundaries. Moral Identity Scope may be operationalised using Aquino and Reed's (2002) Moral Identity Scale, adapted for cross-faith contexts to capture the extension of moral concern beyond in-group boundaries. Mode of Moral Engagement may be assessed through items distinguishing restrained non-interference, procedural openness, and active relational care, drawing on adaptations of the Intergroup Contact Scale developed by Pettigrew and Tropp (2006). Prosocial Outcomes may be operationalised using the Empathy Quotient (Baron-Cohen and Wheelwright, 2004) and Batson's (2011) empathy-altruism paradigm as adapted for interreligious contexts.

Three theoretical propositions follow from the FMEF and the literature reviewed above. The first proposition holds that individuals and communities oriented toward the ontological framework of Vasudhaiva Kutumbakam will demonstrate higher levels of empathic concern and intergroup prosocial behaviour than those operating within restraint-based tolerance frameworks (TP1). The second proposition holds that the family metaphor, by activating the care moral foundation (Haidt, 2012), will mediate the relationship between ontological orientation and prosocial outcomes, such that recategorisation of out-group members as family members will produce attitudinal change consistent with the predictions of the Contact Hypothesis (Allport, 1954; Pettigrew and Tropp, 2006) (TP2). The third proposition holds that familial moral engagement will correlate positively with intergroup contact quality, with a reduction in the out-group devaluation predicted by Social Identity Theory (Tajfel and Turner, 1979), and with willingness to engage in sustained mutual perspective-taking across faith traditions (TP3).

3.4 Data Sources

Primary philosophical data are drawn from classical Upanishadic texts — principally the Maha Upanishad (6.71-75) in Radhakrishnan's critical edition — together with major commentaries within the Advaita Vedanta tradition. Secondary philosophical data include the works of Radhakrishnan (1927, 1955), Vivekananda (1989), Tagore (1931), Gandhi (1927), Sen (2005), Parekh (2000), Hick (1989), Panikkar (1978), Honneth (1995), and Buber (1970), alongside key texts in Habermasian discourse ethics (Habermas, 1984). Empirical data are drawn from landmark quantitative studies including Allport (1954), Pettigrew and Tropp (2006), Tajfel and Turner (1979), Baumeister and Leary (1995), Graham et al. (2009), Batson (2011), Frankl (1959), and Haidt (2012). Contextual empirical evidence is drawn from the Pew Research Centre (2021) and UNESCO (2009).

3.5 Analytical Procedure

The analysis proceeds through three main phases. The first phase is called philosophical reconstruction, where the ontological assumptions behind Vasudhaiva Kutumbakam and their ethical consequences are determined by means of textual analysis of key Upanishadic and Vedantic texts. The second phase is comparative conceptual analysis, during which the ontology developed in the first phase is compared to those of tolerance-based theories, pluralistic theories, recognition theory, and discourse theory, focusing on the similarities, dissimilarities, and superiority of the latter over the others. The final phase is theoretical synthesis, where results of philosophical analysis will be combined with findings from social psychology to develop FMEF and substantiate the three theoretical arguments. Interpretative charity is applied throughout all three phases; that is, each argument is evaluated based on its most plausible version.

3.6 Quality and Validity Criteria

In this study, the credibility, transferability, dependability, and confirmability of the research are judged by the four validity standards suggested by Lincoln and Guba (1985) for interdisciplinary theoretical scholarship. The former two qualities of credibility and transferability have been achieved by triangulating the three traditions of philosophy, sociology, and psychology without placing undue burden on any particular one of them and taking into consideration the historical/cultural context of the sources used. The quality of dependability was assured by documenting all of the analytic decisions made during the writing of the paper. Finally, confirmability was achieved by grounding all interpretive claims in citations from primary or secondary texts.

4. Conceptual Analysis and Theoretical Framework

4.1 Ontological Foundations of Vasudhaiva Kutumbakam

Despite appearances, the grammar of Vasudhaiva Kutumbakam carries more meaning than it might seem. "Vasudha" (earth or world), followed by the emphatic "eva" (indeed) and kutumbakam (family), translates as "the world is indeed one family." The profoundness of this idea becomes apparent when one takes into account the Upanishadic philosophy underlying it. According to the tradition of Advaita Vedanta established by Shankaracharya, multiple instances of individual souls constituting human experience do not represent independent entities accidentally united together. Instead, they are manifestations of one unitary being of Brahman – the Absolute (or God, if one prefers that word) – and the oppositions between oneself and others structuring our society do not represent any reality beyond functional differences in a broader context of unity (Radhakrishnan, 1927). It may therefore become clear why Vasudhaiva Kutumbakam is not an analogical expression but rather the designation of something real and true about the world as a whole.

This ontological grounding marks a decisive philosophical advantage over tolerance. Tolerance-based frameworks ground the obligation to respect religious others in moral choice or social contract: one restrains one's hostility, or consents to an arrangement of mutual non-interference. Therefore, the obligation follows from the very same contingency of the choice, and choices made on advantageous conditions will not necessarily continue to be made when conditions change. In the Vasudhaiva Kutumbakam scheme of things, however, the basis for moral consideration is ontological rather than chosen: there exists an ontological condition where the world is always already one family, whether its members choose to recognise this or not. Therefore, the moral obligation to approach the other with moral consideration becomes a necessity rather than a choice. The moral obligation becomes much stronger and significantly less likely to dissolve under social pressure.

An intellectually honest interpretation of the framework in question will have to avoid the implication that the ontological unity of existence implies empirical identity. Advaita Vedanta does not deny the reality of the differences between individuals, communities, and religious traditions: it provides a larger ontological context for those differences to exist. There is an appreciation of religious diversity, of the multiplicity of traditions and practices and the differences between them, which still retains the sense of ontological unity that is essential for Vasudhaiva Kutumbakam. It is precisely such a middle ground that provides the philosophical usefulness of the framework.

4.2 The Family Metaphor as Moral Architecture

The choice of the family as the governing metaphor of Vasudhaiva Kutumbakam is not casual. The family is distinct from other human associations based on an important structural characteristic, which has important implications for moral life: it is non-voluntary. One cannot choose their parents, brothers and sisters, aunts and uncles. Family ties are formed before affirmation, and continue to exist – at least with a presumption of continuance – despite disagreement, frustration, and conflict which could destroy many of our chosen associations. The non-voluntary aspect of family association is crucial to the strength of the metaphor when applied more broadly to all humans.

Differences do not destroy the relation but are worked out through the relation. When a member of the family expresses an opinion that the patient believes to be wrong, or acts in ways that appear destructive, the appropriate action is not one

of passive acceptance or severing the relation, but of active involvement. The person seeks to comprehend the other's point of view, shares his or her own, and persists in the relation through the struggle. The family offers a picture of moral solidarity amidst difference.

From a social psychology perspective, the family metaphor works effectively as an extremely powerful device of recategorization. When the boundaries of the in-group morality are redefined such that it includes all human beings as part of their family, the process of favouritism toward in-group members and discrimination against the out-group, as pointed out by Tajfel & Turner (1979), is significantly altered. What used to be part of the out-group is now part of the same big in-group family, and the prosocial tendencies expected in social identity theory are consequently activated. Crucially, this recategorisation operates not through the abstract philosophical language of universal human dignity, which has proven motivationally limited in practice, but through the family metaphor – among the most psychologically immediate and emotionally resonant relational categories available to human beings.

4.3 Familial Moral Engagement: An Original Theoretical Construct

Building on the analysis above, this study proposes the construct of familial moral engagement as the distinctive mode of interreligious interaction that Vasudhaiva Kutumbakam enables. Familial moral engagement is defined as an active, relational, and caring orientation toward the religious other – characterised by empathic curiosity about the other's inner life, honest self-disclosure oriented toward shared understanding rather than conversion, willingness to receive the other's challenge as an occasion for self-examination, and sustained relational commitment that persists across disagreement and difficulty.

The concept under consideration has several distinct features compared to related concepts discussed in the literature on interfaith dialogue. Tolerance, as stated before, implies a passive approach, since all it demands is the suppression of negative emotions, without producing any positive obligations within the relationship. While dialogue implies openness to interaction, it does not require the level of dedication and care for the other person characteristic of the process under discussion. In Honneth's definition, recognition implies a political and juridical phenomenon in terms of the acknowledgement of another's dignity and status, which, while being necessary for an interreligious dialogue, is far from sufficient in itself. The concept developed by Buber seems to be closer to the idea discussed here; however, it lacks a discussion of how such a dialogue can be created on an institutional level, as well as the connection between the experience of encounter and the process of change brought about by this phenomenon.

Moral conflict within the construct must also receive special emphasis. As opposed to procedural dialogue models, which seek either a consensus outcome or mutual bracketing of truth claims, familial moral discourse recognises disagreement as integral to the relationship, without the need to achieve agreement between the parties involved. What makes the relationship is the commitment to one another even amidst disagreements, and it is this very kind of commitment – a commitment intrinsic to and independent of whether the relationship is comfortable – which makes the construct especially relevant in situations of theological disagreement, where the need for consensus or bracketing of truth claims would otherwise be dishonest and counter-productive. The above gap is empirically supported by the fact that according to literature from systematic reviews, current efforts in interfaith cooperation do not have a common language on the degree of participation involved, judging their success based on knowledge gain and appreciation and not on the level of relational engagement (Visser et al., 2023).

4.4 Psychological Mechanisms of Prosocial Activation

The plausibility of familial moral engagement as an effective mechanism for producing prosocial consequences better than any that can be accomplished through tolerance is empirically grounded. One of the most directly relevant lines of research is based on the Contact Hypothesis developed initially by Allport (1954). A meta-analysis conducted by Pettigrew and Tropp (2006) found over 500 independent contact studies which confirmed that intergroup contact defined by equal status, cooperation, and true interpersonal association is indeed consistently productive; it causes positive attitudinal change, decreases prejudice and increases prosocial motives and intergroup relations. It should come as no surprise since the moral orientation embodied in Vasudhaiva Kutumbakam inevitably creates conditions required for attaining this effect; there is indeed nothing accidental about it, as both traditions converge in one idea.

Additional evidence can be gained from Haidt's (2012) study on the moral foundations framework. One such moral foundation is the care moral foundation, which consists of a suite of intuitions related to the prevention of harm and the promotion of the welfare of others. It is the most universal of all the moral foundations found across cultures, as well as the most consistently linked to prosocial actions towards vulnerable others (Graham et al., 2009). Using the family metaphor as its central mode of expression, Vasudhaiva Kutumbakam appeals to the ethics of care moral foundation, circumventing the two other moral foundations – namely the fairness and the loyalty foundations, which are more sensitive to contexts and group boundaries.

The explanation offered by Baumeister and Leary (1995) about the need to belong as one of the basic motivations of human beings provides a perspective that is missing in the purely rationalistic theories of ethical responsibility. If an ethic can satisfy the need to belong in the most inclusive way possible, by including the self in the larger community of mankind, then there emerges a social identity which, in its positive effects on behaviour, goes much further than an ethics which guarantees only freedom from interference.

The empathy-altruism hypothesis proposed by Batson (2011) is yet another theoretical basis for asserting that the experience of empathic concern reliably brings about an altruistic motivation for the benefit of the other person, rather than oneself. By defining humanity as a whole as being one large family, Vasudhaiva Kutumbakam creates an empathy that motivates altruistic action. The significance of such findings is further underlined by the assertion by Frankl (1959) that meaning in human life is discovered not through self-serving achievements but through caring for and taking responsibility towards others, just as envisioned in Vasudhaiva Kutumbakam. Taken together, the findings of four different disciplines, namely contact theories, moral psychology, motivational psychology, and meaning-based psychology, offer strong theoretical support for the assertion that the family model of moral action leads to prosocial effects that far exceed those obtained via the toleration paradigm.

The above-mentioned theoretical case has received further empirical support in the meantime, after Batson's proposition. The meta-analytic study of sixty-two previous investigations, including seventy thousand participants, proved there was a strong positive relationship between both types of empathy and prosocial behavior (Yin & Wang, 2023). Moreover, another three-level meta-analysis based on ninety-two studies showed that social support (structural equivalent of the need to belong postulated by Baumeister and Leary, 1995) had a reliable connection with prosocial outcomes, the effect being moderated by the origin of social support and the measurement used (Wang et al., 2024). As for the construct of familial moral engagement, it needs to be mentioned that recent research proves moral identity mediates the relationship between empathy and prosocial behavior, which means that empathic concern is more likely to result in prosocial action when incorporated into individual's identity (Peng et al., 2024). It seems that the findings add to the credibility of FMEF's propositions and prove the order of steps that should be followed in converting ontological orientation into prosocial action.

5. Beyond Tolerance: A Comparative Analysis

5.1 Structural Weaknesses of Tolerance-based Models

A critique of tolerance as a basis for interfaith coexistence can be made along three lines, each one identifying a particular shortcoming of that concept as it pertains to relations between religions. First, there is the question of its passivity. In calling for toleration, all that is demanded is that negative evaluations be put aside without engaging with the other community. There is no call to understand, relate to, or acknowledge the intrinsic value in another religious community. A tolerant society is thus one where religious communities exist apart in a state of controlled indifference, where they are divided by an artificial boundary that keeps them from expressing the conviction of their own rightness and the wrongheadedness of the other community.

A second feature of the structural deficiency of tolerance is that it is an asymmetrical term right from its very roots. Tolerate means to endure, tolerate, and endure something. To tolerate someone's presence or belief means that you put yourself in a superior position while putting the other person in the inferior position – that of being tolerated. This is not a reciprocal process at all. Here, the recognition theory of Honneth (1995) comes into play: this asymmetry of tolerance is much more than mere semantics; it involves psychology and morals too. Communities that are tolerated as opposed to communities that are recognised will feel their status within the society to be temporary and conditional. Nussbaum (2012) similarly documents the ways in which the politics of tolerance can generate, alongside the nominal protection of minority rights, a pervasive social atmosphere of anxiety and subordination.

The third weakness is its fragility under pressure. Tolerance, as a virtue of restraint, is most vulnerable precisely when it is most needed: under conditions of economic competition, political mobilisation, or social anxiety, restraint is the first casualty. The historical record of inter-communal violence demonstrates repeatedly that communities capable of tolerating one another under favourable conditions have abandoned that tolerance rapidly when circumstances change. A model of coexistence whose operative mechanism is the suppression of negative impulses – rather than the cultivation of positive relational bonds – cannot be expected to hold when the pressures that sustain suppression are removed. This conclusion agrees with new reviews conducted on the contact hypothesis; once we only include the best-designed experiments, the positive effects on attitudes resulting from intergroup contact turn out to be actual and yet less impressive and contingent than previously thought (Paluck et al., 2019), and an extensive analysis of prejudice reduction programs demonstrated that the field of intergroup contact has overdeveloped practical applications compared with theory (Paluck et al., 2021).

5.2 Vasudhaiva Kutumbakam as a Corrective Framework

Vasudhaiva Kutumbakam counters all three structural shortcomings at their root causes. In response to passiveness, Vasudhaiva Kutumbakam provides proactive engagement of moral duties. The duty of the family member is not simply abstaining from harming but also understanding, caring for, correcting affectionately, and growing along with other family members. This familial morality is based on positively motivational principles that lead to approach behaviours, not avoidance ones. In response to the asymmetry of the current structure of relations, Vasudhaiva Kutumbakam promotes the ontological parity of the family members, in which there are neither beneficiaries nor burdeners. In the family, created by Vasudhaiva Kutumbakam, all family members are created from the same basic substance and have equal responsibilities for the welfare of others. Finally, in response to the fragility of relations, Vasudhaiva Kutumbakam introduces the idea of relational robustness. Family relations remain unbroken despite difficult situations because they are constitutive relations.

Recognition of the philosophical strengths of the family metaphor does not mean that the dangers posed by the family should be overlooked. For example, family units have the ability to compel conformity, create a sense of loyalty so strong that it pre-empts all individual moral reasoning, and, when pushed to extremes, create an environment of familial solidarity that merely displaces but does not eliminate exclusion. It would therefore be philosophically irresponsible for us to ignore these dangers while reading into Vasudhaeva Kutumbakam. The solidarity in which the principle believes is not homogeneity but the solidarity of difference – of sustaining relationships without reaching consensus. As Parekh explains (2000), morally responsible political communities do not stifle dissident voices in the pursuit of unity but rather those that sustain commonality and relationships by engaging with differences.

5.3 Integration with Procedural Ethics

The substantive moral conception of Vasudhaiva Kutumbakam becomes most meaningful when it is supplemented by a procedure that prevents it from the possibility of familial paternalism. In this sense, Habermas's (1984) conception of discourse ethics serves exactly as this supplement. Habermas argues that a moral norm is morally acceptable if and only if it wins the consent of all people who are involved, provided that such an agreement takes place under the conditions of free, equal and non-coercive communicative rationality. If this is applied to the context of inter-religious interactions, then the result would be that the motivation for morality created by Vasudhaiva Kutumbakam – we interact because we really care about one another as family members – will not be used to justify any form of coercion or manipulation by some families towards other families to make them accept their ethical norms, since they do so because of familial love.

Parekh's (2000) dialogical multiculturalism provides a further complementary dimension at the institutional level. His argument that cultures require encounter with each other to examine and develop their own assumptions strengthens the case for moving beyond tolerance's negative liberty toward the positive engagement that familial moral engagement demands, while his account of the institutional conditions under which genuine cross-cultural dialogue can occur without collapsing into the assimilation of minority traditions provides a practical framework for the implementation of the FMEF in policy and educational contexts.

6. Discussion

6.1 Theoretical Implications

The conceptual analysis developed in the preceding sections generates several significant theoretical implications, the most important of which concern the adequacy of existing frameworks and the contribution of the FMEF. The three theoretical propositions advanced in the Methodology can now be shown to have solid grounding in the analysis. TP1 – that Vasudhaiva Kutumbakam-oriented communities will demonstrate higher prosocial behaviour than tolerance-oriented communities – is substantiated by the demonstration that the FMEF generates active engagement, ontological equality, and relational resilience where tolerance generates only passive restraint, structural asymmetry, and situational fragility. TP2 – that the family metaphor mediates prosocial outcomes through activation of the care moral foundation – is supported by the convergence of Haidt's (2012) moral foundations findings with the Social Identity Theory-based recategorisation mechanisms identified above. TP3 – that familial moral engagement will correlate positively with intergroup contact quality and reduction in out-group devaluation – is grounded in the precise alignment between the conditions that familial moral orientation generates and the conditions that Allport (1954) and Pettigrew and Tropp (2006) identified as necessary for effective intergroup contact. More recent empirical evidence has further corroborated this assertion: examining the moral foundations theory in twenty-five different societies, Atari et al. (2023) have verified the ubiquity of approval toward care principles but have also helped enhance the understanding of their manifestations in various cultures.

Beyond the confirmation of these theoretical propositions, the paper makes a specific contribution to the typology of interreligious modes of relation. Existing typologies – whether theological, distinguishing exclusivism, inclusivism,

and pluralism, or political, distinguishing tolerance, dialogue, and recognition – fail to capture the active, personal, caring, and relationally sustained quality of the mode of engagement that Vasudhaiva Kutumbakam enables. Familial moral engagement constitutes a theoretically distinct mode that is not reducible to any of the existing categories, and whose specification contributes meaningfully to the conceptual apparatus available for thinking about interreligious relations.

The paper further shows that Vasudhaiva Kutumbakam is not a philosophy-specific notion limited merely to the framework of the Indian tradition or even Hinduism alone. Being rooted in the ontology of the universality of consciousness, employing the universal concept of the family and triggering the moral principle of care highlighted by Graham et al. (2009) and Haidt (2012) as a universal moral principle, Vasudhaiva Kutumbakam stands out as an invaluable source in the philosophical discussions on harmonious interfaith coexistence, being equally worthy of recognition as Western philosophies of multiculturalism and tolerance.

6.2 Practical Implications

The findings of this research can be said to be relevant to interfaith education, the formulation of multicultural policies, and the organisation of interfaith dialogue on an institutional basis. In an educational setting, the integration of teaching material regarding the idea of Vasudhaiva Kutumbakam, as a philosophy that involves ontology, ethics, and psychology, and not merely as a cultural slogan, may help in the development of a moral identity defined through relational care towards others rather than mere tolerance. This applies to both secular and religious educational environments, especially when such ideas are taught to students during the developmental stages of forming their moral identity.

The enduring example set forth by Gandhi (1927) of reaching across religious divides through sincere affection and openness to both receiving and giving criticism, sustaining relationships despite deep-seated disagreements, can be said to provide an unparalleled illustration of how moral engagement on a family level can move beyond an impractical dream and into a practical ethic. For policy makers crafting interfaith programmes, such an example points towards the importance of creating opportunities for personal interaction based on parity and shared cooperative intent over merely symbolically endorsing tolerance.

Institutionally speaking, this means that any program based on the principle of fostering familial moral engagement would emphasise different actions when compared to a program based on tolerance. Joint community service, longer stays at interfaith centres, joint solutions to common social problems, and joint perspective-taking among the members of different faith communities fit much better with the concept of familial moral engagement than interfaith dialogues that mainly deal with theological comparison and expressions of mutual respect.

6.3 Limitations

Several limitations need to be mentioned. This research is purely theoretical and conceptual in nature, and hence, does not offer any primary empirical evidence of the relationships discussed in the FMEF framework. While the three theoretical propositions developed in the framework are empirically substantiated and can be derived through reasoning, these relationships need to be validated through actual empirical observation with suitable measurement tools applied across various religious and cultural settings. Also, while the references made to the Upanishads and the philosophy of Vedanta add value to this theoretical study, there are several schools of thought related to Indian philosophy that should have been considered in depth. The universality of the psychological mechanisms identified – drawn primarily from North American and European empirical research – across non-Western and non-liberal cultural contexts also requires further investigation. In recent years, however, there has been some progress made regarding these shortcomings – namely, the study conducted by Atari et al. (2023), which reconsiders the moral foundations theory from a cross-cultural perspective among twenty-five populations. In South Asia and the Middle East, on the other hand, cross-cultural research on the contact and empathy/altruism hypotheses is still relatively limited. And the risks associated with the family metaphor – conformity, loyalty overreach, expanded tribalism – while noted in the comparative analysis, deserve a more developed account of institutional safeguards than this paper provides.

6.4 Directions for Future Research

The theoretical framework developed here opens several productive avenues for empirical research. Quantitative survey studies examining the attitudinal and behavioural correlates of exposure to Vasudhaiva Kutumbakam-informed educational content – using adapted versions of Pettigrew and Tropp's (2006) Contact Scale, Graham et al.'s (2009) moral foundations battery, and the Empathy Quotient (Baron-Cohen and Wheelwright, 2004) – would provide the most direct empirical tests of TP1 and TP2. Experimental studies comparing familial-framing and tolerance-framing conditions in interreligious contact settings, using random assignment and pre-post measurement of intergroup attitudes, would enable causal inference regarding the mechanisms through which familial moral orientation generates prosocial change, and provide the most rigorous test of TP3. Cross-cultural replication across religious and national contexts — particularly

in South and Southeast Asia, the Middle East, and sub-Saharan Africa — would be essential to establish the cross-cultural generalisability of the framework's central claims.

7. Conclusion

This paper has advanced the argument that Vasudhaiva Kutumbakam – taken seriously as philosophy rather than celebrated as a slogan – offers a substantively richer, psychologically more resonant, and structurally more robust foundation for interfaith coexistence than the tolerance paradigm that continues to anchor most contemporary approaches to religious diversity. The three research questions that organised the inquiry can now be addressed directly. With respect to RQ1, the ontological foundations of Vasudhaiva Kutumbakam lie in the Advaitic metaphysics of unity-in-multiplicity, which grounds moral obligation not in contractual agreement or ethical choice but in the structure of reality as the tradition understands it – a grounding that renders the obligation more robust and less contingent than tolerance's restraint-based foundation. Regarding RQ2, the psychological processes that lead to prosocial behaviour arising from a familial moral orientation comprise social recategorization, stimulation of the care moral foundation, fulfillment of the basic need for belonging, and the empathy-altruism model described by Batson (2011), which aligns with Frankl's (1959) conceptualization of relational significance – four complementary processes that occur simultaneously on multiple dimensions: cognitive, emotional, motivational, and existential. Regarding RQ3, familial moral engagement constitutes a fourth type of interfaith interaction and a significant contribution to theoretical knowledge because it represents an active, caring process of engagement not covered by any of the other three types of interaction: tolerance, dialogue, and recognition.

In that sense, the Family Moral Engagement model proposed in this paper provides the connection between the rich philosophical insights found in the Upanishads and the empirical findings of modern social and moral psychology. It will not solve all problems inherent in interfaith coexistence; no such model can solve all problems. However, it represents a much more realistic view of what true coexistence entails than what has been offered by the tolerance approach. The difference between the two concepts may not seem to be significant at first glance, but while tolerance merely implies enduring each other, Vasudhaiva Kutumbakam calls for accountability to one another.

Family is not simple. Family entails friction, disagreements that remain unresolved, and the continual challenge of maintaining commitment to individuals whose decisions one cannot always agree with. However, family also means something that the virtue of tolerance cannot possibly provide for: the understanding that the relationship will not end just because it becomes difficult. The unique feature of relational commitment that is unconditional is what the concept of familial moral engagement attempts to define and what the world so desperately needs at this time, when it is so rife with religious divides and communal separation.

Conflict of Interest Statement

The author declares no financial or commercial conflict of interest related to this research.

Ethical Declarations

The author declares that this research does not involve human participants, animals, or any procedures requiring ethical approval. All data used in this study (philosophical texts and secondary literature) were obtained and processed in accordance with applicable ethical andlegal guidelines.

References

1. Adamson, P., & Ganeri, J. (2020). Classical Indian philosophy. Oxford University Press.
2. Allport, G. W. (1954). The nature of prejudice. Addison-Wesley.
3. Aquino, K., & Reed, A. (2002). The self-importance of moral identity. Journal of Personality and Social Psychology, 83(6), 1423–1440.
4. Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S. T., & Dehghani, M. (2023). Morality beyond the WEIRD: How the nomological network of morality varies across cultures. Journal of Personality and Social Psychology, 125(5), 1157–1188. https://doi.org/10.1037/ pspp0000470
5. Barnas, T. J. (2022). The effectiveness of interfaith dialogue in countering religious intolerance: A phenomenological study of interfaith youth program alumni. Journal of Security, Intelligence, and Resilience Education, 13(2), 1–22.
6. Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism and normal sex differences. Journal of Autism and Developmental Disorders, 34(2), 163–175.
7. Batson, C. D. (2011). Altruism in humans. Oxford University Press.
8. Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. Psychological Bulletin, 117(3), 497–529.

9. Buber, M. (1970). I and thou (W. Kaufmann, Trans.). Charles Scribner's Sons. (Original work published 1923)
10. Dash, P. K., & Sharma, A. (2024). Vasudhaiva Kutumbakam in the context of globalisation: Ideals, realities, and contemporary challenges. Pakistan Journal of Life and Social Sciences, 22(1), 3525–3534. https://doi.org/10.57239/PJLSS-2024-22.1.00257
11. Erikson, E. H. (1959). Identity and the life cycle. International Universities Press.
12. Frankl, V. E. (1959). Man's search for meaning. Beacon Press.
13. Frazier, J. (2021). Ethics in classical Hindu philosophy: Provinces of consequence, agency, and value in the Bhagavadgītā and other epic and śāstric texts. Religions, 12(11), 1029. https://doi.org/10.3390/ rel12111029
14. Gadamer, H. G. (2004). Truth and method (J. Weinsheimer & D. G. Marshall, Trans.). Continuum. (Original work published 1960)
15. Gandhi, M. K. (1927). The story of my experiments with truth. Navajivan.
16. Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. Journal of Personality and Social Psychology, 96(5), 1029–1046.
17. Habermas, J. (1984). The theory of communicative action: Vol. 1. Reason and the rationalisation of society (T. McCarthy, Trans.). Beacon Press.
18. Haidt, J. (2012). The righteous mind: Why good people are divided by politics and religion. Pantheon Books.
19. Hick, J. (1989). An interpretation of religion: Human responses to the transcendent. Macmillan.
20. Honneth, A. (1995). The struggle for recognition: The moral grammar of social conflicts (J. Anderson, Trans.). Polity Press.
21. Jabareen, Y. (2009). Building a conceptual framework: Philosophy, definitions, and procedure. International Journal of Qualitative Methods, 8(4), 49–62.
22. Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. European Journal of Social Psychology, 45(2), 152–168. https://doi.org/10.1002/ejsp.2079
23. Lin, X., Chen, Q., Wei, L., Lu, Y., Chen, Y., & He, Z. (2022). Exploring the trend in religious diversity: Based on the geographical perspective. PLOS ONE, 17(7), e0271343. https://doi.org/10.1371/journal.pone.0271343
24. Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Sage.
25. Maha Upanishad (6.71–75). (1953). In S. Radhakrishnan (Ed. and Trans.), The principal Upanishads (pp. 958–960). George Allen & Unwin.
26. Ministry of Education, Government of India. (2020). National Education Policy 2020. https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf
27. Nussbaum, M. C. (2012). The new religious intolerance: Overcoming the politics of fear in an anxious age. Belknap Press.
28. Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. Behavioural Public Policy, 3(2), 129–158. https://doi.org/10.1017/ bpp.2018.25
29. Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. Annual Review of Psychology, 72, 533–560. https://doi.org/ 10.1146/annurev-psych-071620-030619
30. Panikkar, R. (1978). The intra-religious dialogue. Paulist Press.
31. Parekh, B. (2000). Rethinking multiculturalism: Cultural diversity and political theory. Macmillan.
32. Peng, L., Jiang, Y., Ye, J., & Xiong, Z. (2024). The impact of empathy on prosocial behaviour among college students: The mediating role of moral identity and the moderating role of sense of security. Behavioral Sciences, 14(11), 1024. https://doi.org/10.3390/ bs14111024
33. Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. Journal of Personality and Social Psychology, 90(5), 751–783.
34. Pew Research Centre. (2021). Religion and living together in India. https://www.pewresearch.org
35. Radhakrishnan, S. (1927). The Hindu view of life. Unwin Books.
36. Radhakrishnan, S. (1955). Recovery of faith. George Allen & Unwin.
37. Raina, S. K., & Kumar, R. (2023). Vasudhaiva kutumbakam – one earth, one family, one future: India's mantra for a healthy and prosperous earth as the G20 leader. Journal of Family Medicine and Primary Care, 12(2), 191–193. https://doi.org/10.4103/jfmpc. jfmpc_254_23
38. Rutherford, M. (2026). Honneth's dialectical shortcoming: Understanding Honneth's problem with power. Frontiers in Sociology, 10, Article 1686743. https://doi.org/10.3389/fsoc.2025.1686743

39. Sen, A. (2005). The argumentative Indian: Writings on Indian history, culture and identity. Farrar, Straus and Giroux.
40. Tagore, R. (1931). The religion of man. George Allen & Unwin.
41. Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), The social psychology of intergroup relations (pp. 33–47). Brooks/Cole.
42. UNESCO. (2009). Intercultural dialogue: Interacting with the 'other'. UNESCO Publishing.
43. Visser, H. J., Liefbroer, A. I., Moyaert, M., & Bertram-Troost, G. D. (2023). Categorising interfaith learning objectives: A scoping review. Journal of Beliefs & Values, 44(1), 63–80. https://doi.org/10.1080/13617672. 2021.2013637
44. Visser, H. J., Liefbroer, A. I., & Schoonmade, L. J. (2024). Evaluating the learning outcomes of interfaith initiatives: A systematic literature review. Journal of Beliefs and Values, 45(4), 689–712. https://doi.org/10.1080/13617672.2023.2196486
45. Vivekananda, S. (1989). The complete works of Swami Vivekananda: Vol. 1. Advaita Ashrama. (Original work published 1893)
46. Wang, Y., Ran, G., Zhang, Q., & Zhang, Q. (2024). The association between social support and prosocial behaviour: A three-level meta-analysis. PsyCh Journal, 13(6), 1026–1043. https://doi.org/10.1002/pchj.792
47. Yin, Y., & Wang, Y. (2023). Is empathy associated with more prosocial behaviour? A meta-analysis. Asian Journal of Social Psychology, 26(1), 3–22. https://doi.org/ 10.1111/ajsp.12537

❖ ❖ ❖

Vol. 6 (1), 2026
E-ISSN: 2583-4452

Identification of Bystanders from Cyberbullying Tweets: Performance Evaluation Using Multi-Feature Extraction Techniques and Neural Network Architectures

Ayushi Gupta[1], Veenu Bhasin[2] and Anamika Gupta[1*]
[1] Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi
[2] PGDAV College, University of Delhi, Delhi
[*] Corresponding Author ✉ anamikargupta@sscbsdu.ac.in,

ABSTRACT

Cyberbullying involves aggressive or harmful behavior conducted by individuals or groups through online platforms. Bystanders, those who witness such incidents, can play varying roles: defenders who intervene, instigators who support the bully, or neutral onlookers who remain passive. Understanding bystander behavior is crucial to assessing the scale and impact of cyberbullying, but the lack of labeled data limits progress. To address this, the CYBY23 dataset was released on Kaggle in October 2023, featuring Twitter threads with main tweets and bystander replies where bystander roles have been manually annotated. Labeling of bystander roles is a labor-intensive task; there's a clear need for automated methods to label bystander roles. In this paper, an efficient model, based on neural network (NN) architectures, is proposed for the automatic labeling of bystander roles.

The CYBY23 dataset was first balanced using SMOTE. We then evaluated different sets of features (Toxicity, Sentiment, Empath, TF-IDF, POS) using three NN models - Feed-forward Neural Network, Extreme Learning Machine, and Multi-Layer Perceptron. The efficiency of the models was compared based on accuracy, precision, recall, and F1 score. Features TF-IDF and POS, when taken together, delivered the best performance. The feed-forward neural network gave the best performance among the network architectures.

Keywords: *Natural Language Processing, Cyberbullying, Defender, Instigator, Impartial*

1. Introduction

Cyberbullying refers to the intentional use of digital technologies, such as social media or messaging platforms, to inflict harm, humiliation, or distress on others. It often includes aggressive behaviors like harassment, threats, or the spreading of harmful content, which can significantly impact the psychological well-being of the victims (Smith, 2019). Bystanders in cyberbullying are individuals who witness or are aware of bullying behavior but are not directly involved. They can either contribute to the situation by encouraging the bully or help the victim by intervening or reporting the incident (Kowalski et al., 2020; Salmivalli, 2018; DeSmet et al., 2019).

With the increasing use of social media platforms, online communication now plays a major role in daily life. Twitter, in particular, serves as a prominent medium for individuals to express their thoughts, react to current events, and engage in public discourse. Among the numerous social behaviors exhibited on such platforms, bystander behavior - where users observe but do not participate or intervene in incidents such as cyberbullying, misinformation spread, or online harassment - has emerged as a critical area of study. Ifwe can identify and analyze how bystanders behave, we can better understand how users engage online. Further, intervention strategies can be designed to create safer online environments, and harmful behaviour can be reduced.

As per the recent research studies, bystanders play a crucial role in cyberbullying dynamics. Studies show that approximately 70% to 80% of individuals witnessing bullying, either online or offline, do not intervene (Huang et al., 2020). However, when bystanders do step in, bullying stops within 10 seconds in 57% of cases (Hawkins et al., 2019). People often hesitate to step in and help because they are afraid they might become the next target of bullying. Also, they are not sure about what they should do or how to respond appropriately (Obermaier et al., 2016).

Research that focuses on understanding the role of bystanders in cyberbullying is very important. When we identify what bystanders do (for example, whether they support the victim, ignore the situation, or encourage the bully), we can better

Received on: 27 Apr 2026 | Accepted on: 28 Jun 2026 | Published Online on: 25 July 2026

understand how cyberbullying works and continues. By studying how bystanders react and what influences their behavior, researchers can learn why cyberbullying either continues or gets reduced (Salmivalli, 2018). It can be used in developing targeted interventions and educational programs designed to empower bystanders to act. Effective bystander interventions can reduce the incidence and impact of cyberbullying (McMahon et al., 2014). Further, understanding bystanders' motivations and barriers to intervention, researchers can improve prevention strategies, making them more effective in encouraging active participation against cyberbullying (Gini et al., 2014).

The majority of publicly available datasets do not adequately address the roles of bystanders in cyberbullying. Given the significant impact that bystanders can have, it is essential to classify their roles so that there can be a better understanding and addressing of the phenomenon of cyberbullying. In this research work, we are aiming to explore the publicly available CYBY23 (Alfurayj et al., 2023) dataset, which has been manually annotated to identify the bystander's role. CYBY23 Dateset has the cyberbullying tweets and the reply tweets corresponding to bystanders.

This study focuses on designing and evaluating architectures based on neural networks for automatically identifying bystander behavior in tweets from the dataset CYBY23. To this end, we apply multiple feature extraction methods, namely, sentiments, toxicity, empath, part of speech tags, and TF-IDF, to encode tweet content into a format suitable for model training. By capturing a wide range of features, we aim to improve the discriminative power of our models. Automatic Labeling will help in enhancing the

scalability and usability of the datasets, which will contribute to enhancing the research in this area.

1.1 Objectives of the work

The main objective of this work is to provide an efficient and reliable method for automating the labelling of bystander roles in cyberbullying tweets. To achieve this objective, we will perform the following steps:

- To extract and evaluate various types of features from tweets, namely, Toxicity, Sentiment, Empath, TF-IDF, and POS.
- To apply and compare three neural network-based methods for multiclass classification, assessing their performance in detecting bystander tweets.
- To identify the optimal combination of features that enhances detection accuracy across the implemented models.

The paper is organized as follows. The next section contains the Related work and the background required, followed by section 3 describing the Methodology. The Methodology section describes the dataset, steps involved in pre-processing, and the feature extraction process, followed by the details of neural network models used and the experimental setup. Section 4 gives the results obtained, and Section 5 concludes the paper, along with giving the future scope.

2. Related Work and Background

Identifying bystanders in cyberbullying conversations or threads is very important for understanding cyberbullying and reducing its negative impact. Machine learning and neural network approaches have been widely used to improve detection accuracy in this emerging research area. Bystanders - categorized as defenders, instigators, or neutral participants- significantly influence the outcome of cyberbullying incidents through their responses or lack thereof (Alfurayj and Lutfi, 2023; Alfurayj et al., 2024; Gupta et al., 2024b). When bystander roles are included in cyberbullying detection systems, it improves how accurately and thoroughly cyberbullying is analyzed. Recent studies have successfully deployed machine learning techniques for the classification of bystanders based on their roles (Gupta et al., 2024b).

The CYBY23 dataset comprises Twitter threads with labeled bystander roles, providing a comprehensive view of cyberbullying incidents and valuable data for training and evaluating detection models (Alfurayj et al., 2023b; Gupta et al., 2024b). This dataset allows for better training of machine learning models because it includes the full conversation context, such as replies and interactions between users in a discussion thread, instead of looking at individual tweets separately. Manually labeling bystander roles (having people read posts and mark who is a bystander and what role they play) takes a lot of time and effort. Because of this, there is a need for automated labeling methods (using algorithms or machine learning) that can automatically assign these labels. This would help increase the size of the dataset more easily, which is useful for conducting more extensive research.

Cyberbullying detection, as well as bystander detection in cyberbullying threads, involves various natural language processing (NLP) steps such as pre-processing and feature extraction, before employing machine learning models.

Pre-Processing: It is a crucial step in text processing in order to increase the effectiveness of the machine learning algorithms. The online chats/posts are generally informal, noisy, and may contain URLs, emojis, and hashtags. Researchers have applied standard NLP techniques – such as text cleaning, stemming, tokenization, and lemmatization

for pre-processing social media datasets (Fati et al., 2023). To prepare the dataset, Chatzakou et al. (2019) eliminated noise by removing numbers, stop words, punctuation marks, and filtering out spammers and duplicate entries. Tokenization was then employed, in which the text is divided into smaller components called tokens, such as words, numbers, and punctuation marks. Among them, 'stop words' such as *the*, *a*, *on*, *is*, *all*, and are frequently occurring terms that carry little meaningful content and are often removed during text processing. Stemming is a method used to reduce words to their base or root form. For instance, "booked" becomes "book," and both "working" and "worked" are reduced to "work." Lemmatization, unlike stemming, does not just remove word endings. Instead, it relies on lexical databases to accurately determine the proper base form of a word, for example, converting "saw" to "see" or "better" to "good". Part-of-speech tagging assigns a grammatical category to each word in the document, such as noun, verb, adjective, and others, based on both its meaning and the context in which it appears. Text normalization is done by converting text to a uniform format, such as lowercasing all characters, to ensure consistency (Khairy et al., 2024).

Data augmentation has been widely used to enhance the size and quality of datasets (Gupta et al., 2024a). The study (Risch and Krestel, 2018) employed Data Augmentation using machine translation to increase the size of the input data three times. The hashtags were tokenized into words using the dynamic programming technique. Several methods like random subsampling, random oversampling, SMOTE, and SMOTE + TOMEK have been employed to balance the dataset, thereby tackling the class imbalance issues (Khairy et al., 2024). SMOTE (Synthetic Minority Over-sampling Technique) is applied to overcome the issue of imbalanced data by synthesizing samples for the minority class (Gupta and Gupta, 2024). Borderline-SMOTE (Han et al., 2005) considers only those minority samples that are close to the borderline for synthesizing new samples.

Feature Extraction: To analyze datasets related to cyberbullying and bystanders, features that capture the textual, contextual, and behavioral dimensions of user interactions are extracted. Various studies have utilized feature extraction techniques to enhance datasets (Gupta et al., 2026). Word Embeddings techniques, including FastText, GloVe, and BERT, convert text into numerical vectors that represent semantic meaning (Albraikan et al., 2022). Term Frequency-Inverse Document Frequency (TF-IDF) highlights significant words by taking into account both their frequency and uniqueness (Zotkina and Martyshkin, 2024). Term Frequency (TF) quantifies the number of times a term occurs in a given document, while Inverse Document Frequency (IDF) is computed as the ratio of the size of the document collection to the documents containing the specific term. This measure serves as a good indicator of how unique or rare a word is across all documents in the collection. TF-IDF is computed by taking the product of TF and IDF scores, and denotes the importance or relevance of a term within a document in relation to the entire corpus (Hasan et al., 2023; Lepe-Fa´undez et al., 2021). Researchers analyze the sentiment (emotional tone) of text to help decide the feelings behind a message, which plays a crucial role in identifying instances of cyberbullying and bystander behavior (Zotkina and Martyshkin, 2024). These sentiment features, such as positivity, negativity, and objectivity, have been employed alongside TF-IDF (Tommasel et al., 2018).

Part-of-speech (POS) features analyze the functions of words within sentences (Gupta et al., 2024b). GloVe (Global Vectors for Word Representation) is an unsupervised machine learning approach created at Stanford (Hasan et al., 2023). It generates word embeddings by using a global co-occurrence matrix that tracks the co-occurrence of words based on their frequency of appearing together in a text corpus. In (Davidson et al., 2017), a variety of different features are obtained. Each tweet is transformed to lowercase and stemmed using the PorterStemmer. Further, unigram, bigram, and trigram features are produced and given weights using TF-IDF scores to capture significant patterns. To extract syntactic structures, the NLTK toolkit (Bird et al., 2009) is utilized to create unigram, bigram, and trigram features based on Penn Treebank Part-of-Speech (POS) tags. The Flesch-Kincaid Grade Level method is adjusted using Flesch Reading Ease metrics to evaluate the readability of tweets. Additional features include scores from sentiment analysis, binary indicators, and counts for the existence of hashtags, user mentions, retweets, and URLs, as well as quantitative metrics such as the number of characters, words, and syllables per tweet. The TF-IDF method was applied on n-grams to train classifiers in (Risch and Krestel, 2018). A total of 35 carefully chosen features are utilized - 25 derived from emoticons and 10 reflecting the textual size concerning the number of words, the ratio of uppercase characters to lowercase ones, and counts of negation words along with post polarity. The VADER sentiment analyzer has been employed to extract polarity scores. In the study presented in (Chatzakou et al., 2019), researchers analyzed various features, including network properties, textual characteristics, user attributes, and profile image statistics. Network properties included the number of friends, followers, hub and authority scores, centrality measures, and reciprocity. Textual characteristics examined included metrics such as hashtags, emoticons, word counts, URLs, sentiment scores, and mentions. User attributes considered included average post count, account age, verification status, and subscribed lists. Profile image statistics covered the total number, average, and median sizes, and image locations, along with information from the profile description.

In the recent literature, a range of machine learning and neural network approaches have been utilized by researchers to identify aggression in text, detect cyberbullying, and understand bystander behavior. Classical machine learning algorithms like Logistic Regression, Random Forest, Naïve Bayes, and Support Vector Machines (SVM) are commonly applied in cyberbullying detection tasks (Raj et al., 2021). Additionally, neural network models have also been utilized to analyze and interpret cyberbullying-related conversations.

A Feedforward Neural Network (FNN) represents one of the simplest types of artificial neural networks. In this structure, data flows in a single direction, starting from the input layer, passing through one or more hidden layers, and reaching the output layer. The architecture is acyclic, with no feedback or recurrent connections.

A Multi-Layer Perceptron (MLP) is a basic type of feedforward neural network. It consists of multiple fully connected layers of nodes (neurons), including an input component, several internal processing layers, and an output component. The internal hidden layers perform transformations using weights, biases, and activation functions. The output layer generates probabilistic predictions, making MLPs suitable for both classification and regression tasks. During the training process in MLPs, the backpropagation algorithm is used, which calculates gradients and updates weights accordingly.

Extreme Learning Machine (ELM) is a training algorithm specifically designed for single-layer feedforward neural networks (SLFNs) (Huang et al., 2006). Unlike traditional SLFNs that rely on iterative training methods such as backpropagation and gradient descent, ELM randomly assigns input weights and biases to the hidden neurons and then analytically calculates the output weights. This eliminates the need for adjusting input weights and biases during training, allowing ELM to achieve a globally optimal solution without requiring parameter tuning, such as learning rate or stopping criteria, thereby significantly enhancing training speed (Bhasin and Bedi, 2013). Moreover, ELM has demonstrated strong performance in multiclass classification problems (Bedi and Bhasin, 2019; Huang et al., 2012).

A range of neural network architectures has been investigated for detecting cyberbullying, such as BERT, CNN, LSTM, and hybrid models that integrate these methods (Sandoval et al., 2025). (Tommasel et al., 2018) merged neural networks with Support Vector Machines (SVM) to identify aggression in text. In (Lepe-Fa´undez et al., 2021), the authors combined SVM, Naïve Bayes (NB), and Random Forest (RF) using a majority voting mechanism. The study (Risch and Krestel, 2018) employed an ensemble approach with a recurrent neural network (RNN). A bidirectional Gated Recurrent Unit (GRU) layer, combined with max pooling and average pooling operations, as well as three logistic regression units, is employed alongside Gradient Boosting Decision Trees to evaluate levels of aggression.

Various machine learning models have also been evaluated for their effectiveness in detecting cyberbullying and bystander roles. Techniques such as RF Classifier (Gupta et al., 2024b), BERT, and RoBERTa (Sandoval et al., 2025) have shown promising results. For instance, fine-tuning RoBERTa with oversampled data achieved an F1 score of 89.3%. Models that incorporate bystander roles have demonstrated substantial performance improvements. The classifier chain combining BERT and LSTM networks achieved a 25.35% increase in performance, with an F1-score of 89% (Alfurayj et al., 2024). Robust detection capabilities have been achieved using oversampling techniques and fine-tuning of models like RoBERTa, with high F1-scores (Sandoval et al., 2025). The BBGC model integrates BERT, BiGRU, and CNN to extract deep contextual, sequential, and local features. (Cao et al., 2026). Experimental results demonstrate that this fusion approach significantly outperforms individual pre-trained models and other hybrid neural networks (like RoBERTa and ALBERT) in cyberbullying detection. Further, researchers work on the real-world psychological toll cyberbullying takes on people (Jazzar and Duridi, 2025). This paper uses machine learning and deep learning techniques to identify cyberbullying on different social media platforms.

Evaluation Metrics: As highlighted in the reviewed literature, multiple evaluation metrics have been employed to assess the performance of models in accurately identifying bystander roles. These include class-level and overall weighted averages of recall (rec), precision (prec), F1 score (F1), and the weighted Area Under the ROC Curve (AUC) (Chatzakou et al., 2019). Additionally, researchers have utilized confusion matrices, accuracy, and the Wilcoxon test to further validate the models' effectiveness (Alfurayj et al., 2023b).

3. Methodology

Fig. 1 depicts the methodology followed in the current research. As a first step, research works available on categorizing bystanders into bullying and non-bullying were studied. In this phase, we identified various feature extraction models being used. Then, the corpus from tweets to be used was chosen. The dataset CYBY23, downloaded from Kaggle, is described in detail in subsection 3.1. Data Pre-processing involved data cleaning followed by removing the imbalance

Fig. 1: *Research Methodology*

using the Borderline Synthetic Minority Oversampling Technique (SMOTE) (Han et al., 2005). Pre-processing of the data was done so that it can be used for classification using Neural Network models. Data preprocessing is briefly explained in subsection 3.2. Some of the features were categorical, which were converted to numeric values, and different feature sets were extracted from the data. The details of all the feature sets are given in subsection 3.3. Deployment of three Neural Network models was performed on the Bystanders pre-processed data. Evaluation of the models was on the basis of standard metrics like accuracy, precision, recall, and F1 score. The details of the Neural networks used are given in subsection 3.4. Finally, subsection 3.5 outlines the experimental setup employed in this study.

3.1 Description of the Dataset

The original dataset, CYBY23, was downloaded from Kaggle (Alfurayj et al., 2023b; Alfurayj and Lutfi, 2023). Twitter API was used to extract 1,024 tweets from 150 conversation threads over the span of one year (January 2022 to January 2023). The downloaded information includes the tweet ID, tweet date, the user ID, the user's screen name, likes count, retweets count, and the tweet's text. Keywords and hashtags associated with racial and discriminatory remarks that could lead to harassment were used to gather this data.

The labeling of bystanders was achieved through a manual annotation process based on established guidelines (Alfurayj et al., 2023). This process involved assessing individual tweets' aggressiveness, identifying the role of bystanders in responses, thereby judging the aggressiveness of the whole thread, including the original post and its replies. Only those threads that received agreement from five or more annotators were included for further analysis. This approach reduced the total number of tweets to 639 and the number of tweet threads to 112. The dataset now contains between 10% to 20% instances of bullying, with only 11.6% representing cyberbullying characterized by high aggression. Instigators were prominently represented across both categories of bullying.

The study looked at bystander contagion, which is the idea that people who observe bullying may start behaving similarly. The researchers found a positive correlation - when there were more instigators (people who start or encourage bullying), the number of bullying incidents increased in the dataset. So, the presence of instigators can influence bystanders and lead to more bullying behavior spreading in the conversation. Due to the labor-intensive and time-consuming process of manual role annotation, the researchers emphasized the importance of automating the identification of bystander roles. As a result, they made the CYBY23 dataset publicly available on the Kaggle platform (Alfurayj et al., 2023).

The dataset "CYBY23" consists of 112 threads from Twitter, including main posts as well as replies from bystanders. Each of the 639 tweets is accompanied by its text and various general tweet-level features, along with the labels indicating the roles of bystanders. Additionally, toxicity features were extracted using the Perspective API, while sentiment features were obtained using TextBlob. The dataset comprises 17 attributes, which are listed below:

1. tweet id: An identifier for each tweet.
2. reply id: The identifier of the main tweet to which this reply tweet belongs.
3. created at: The date of tweet creation.
4. text: The text of the tweet.
5. retweet count: Denotes the count of retweets.
6. favorite count: The total number of favorites.
7. class label: It denotes the aggression level of the thread - the main tweet and its replies, on a scale of 0-2. Hence, the main tweet only has this assignment.
    (a) 0 - Aggression
    (b) 1 - Low-aggression bullying
    (c) 2 - High aggression bullying
8. bystander role label: The label assigned to the bystander who has replied to the main tweet. Hence, it was assigned to the reply tweets only.
    (a) 0 - Instigator who agrees with the main post
    (b) 1 - Defender who disagrees with the main post
    (c) 2 - Impartial who is not taking any sides
    (d) 3 - Others
9. Three sentiment features - Polarity, Subjectivity, and Sentiment.
10. Six Toxicity Features - Identity Attack, Insult, Threat, Profanity, Toxicity, and Severe Toxicity.

3.2 Pre-Processing

The original dataset is not suitable for classification by intelligent models in its raw form because it includes URLs, HTML tags, mention tags, chat slang, non-standard English words, and spelling errors. Therefore, preprocessing of the dataset is essential in order to prepare the data for better modelling results. The data cleaning steps carried out on the text attribute are listed below.

1. All the URLs beginning with {https, www}were removed.
2. All the HTML tags contained in <> were removed.
3. All the mentioned tags beginning with @ were removed.
4. The emoji representations were originally separated by {:, }. These delimiters were replaced by the space character, and the separate words were kept since emojis provide useful information on the nature of the tweet.
5. All the punctuation characters except the {?, !}were removed. These two were retained since these punctuation marks can provide cues about the tone or emotion of a sentence. Also, they help in understanding the context and intent of the tweet.
6. Lemmatization was performed to convert the words into their meaningful base forms.
7. Chat words such as LOL were converted to their full text forms.
8. Spell Checker was applied to correct the word spellings.
9. The target variable is the bystander role label. It had four textual values. These textual values were label encoded to values on a scale of 0 to 3 - 0 (Instigator), 1 (Defender), 2 (Impartial), 3 (Others).

High imbalance was observed in the data vis-a-vis the target variable 'the bystander role label' and preliminary experiments showed overfitting. Out of a total of 524 tweets, 228 were labeled as Instigator, 191 as Defender, 96 as Impartial, and only 9 belonged to the Others or Irrelevant category. After the data cleaning process, this imbalance in the dataset was handled using the Borderline-SMOTE technique. This technique identifies the borderline samples and uses them for generating synthetic data using SMOTE. The distribution of classes before and after applying SMOTE is shown in Fig. 2.

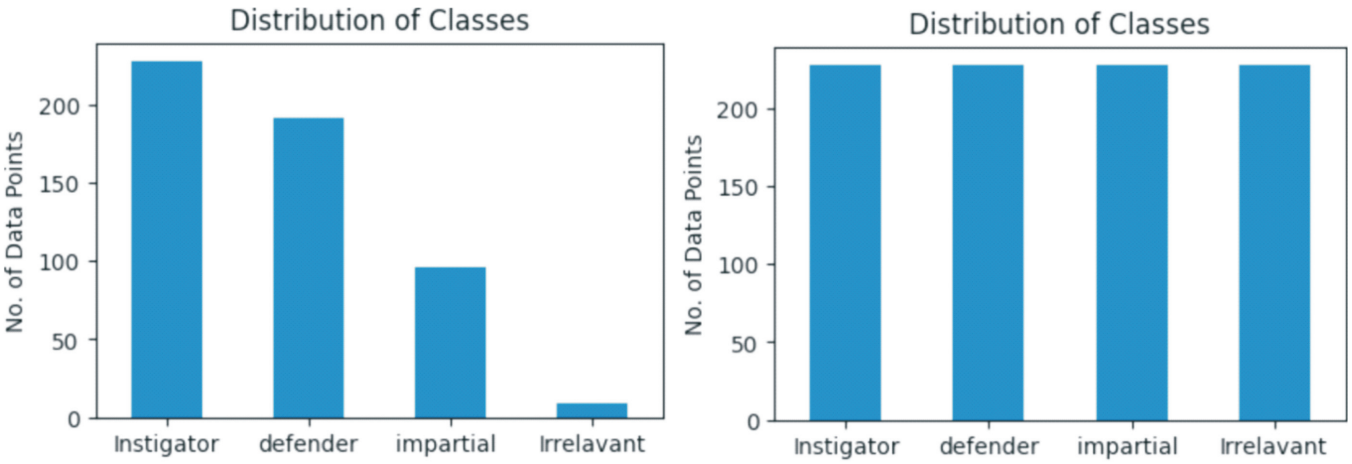


Fig. 2: *Distribution of Bystander Roles in the CYBY23 Dataset before and after Resampling*

3.3 Feature Extraction

After cleaning and preprocessing, various features were extracted from the tweet content. Initially, attributes from both the main tweet and its corresponding reply were merged to assess how the original tweet affects the reply, resulting in the creation of a new enriched dataset. This enriched dataset included:

- 6 general features
- 6 features related to toxicity from main tweet
- 6 features related to toxicity from reply tweet
- 3 features related to sentiments from main tweet
- 3 features related to sentiments from reply tweet
- class label of main tweet.
- bystander role label of reply tweet.

Due to the column-wise merging of main and reply tweets, the total tweet count reduced from 639 to 524. Thus, the updated dataset consisted of 26 attributes and 524 records.

Three features, reply ID, tweet id, and created at, were eliminated since these features had no role in detecting bystanders. Furthermore, the text of the tweet was not considered since its information was already captured through toxicity (via the Perspective API) and sentiment (via TextBlob) features. The finalized dataset comprised 22 features including target label: bystander role model. To enrich the dataset further, three more feature sets extracted through TF-IDF, POS tags, and Empath were incorporated.

- Features derived from Empath Package: The text of both the original tweet and the bystander tweet is used to extract Empath features, as outlined in (Fast et al., 2016). Empath is a tool that analyzes text by looking at the words used (lexical analysis) and placing them into different meaning-based categories. It examines text using 194 predefined categories. These categories are based on research and knowledge about human emotions, behavior, and topics. The categories cover many types of content, such as daily life topics – home, work, family, emotions – anger, sadness, government terms – embassy, democrat, technology terms – iPad, Android, and so on. By grouping words into these categories, Empath helps researchers understand the themes, emotions, and behaviors expressed in a piece of text. For example, the small text "I love you but I hate you also" yields the following values across various categories: 'anger': 0.125, 'anticipation': 0.0, 'disgust': 0.125, 'fear': 0.125, 'joy': 0.125, 'negative': 0.125, 'positive': 0.125, 'sadness': 0.125, 'surprise': 0.0, and 'trust': 0.0.
- TF-IDF Features: The method for computing Term Frequency-Inverse Document Frequency (TF-IDF) features is one of the simplest and most traditional approaches to feature extraction from text, commonly used in Natural Language Processing (NLP). TF-IDF helps create a feature vector for the text by determining the importance of each word based on its frequency in the corpus. In this study, a total of 2,071 features are extracted from the tweets' text.
- Part of Speech (POS) Features: The tweet text is tokenized and categorized into 10 part-of-speech categories. The categories in which all the tokens from text are categorized are: Noun, Adjective, Verb, Adverb, Pronoun, Preposition, Conjunction, Article, Negation terms, and Auxiliary terms. These ten categories form 10 POS features.

To summarize, the final dataset had 2296 features as mentioned in Table 1.

Table 1: *Number of Extracted Features from Tweets*

Number of Features	Tool/API
3	General Features
12	Toxicity attributes (obtained through Perspective API)
6	Sentiment attributes (obtained through TextBlob)
194	Emotion features based on Empath
2071	Features based on TF-IDF
10	Features based on Part of Speech (POS)

Further, these individual feature sets were grouped based on their nature, resulting in two more sets, as shown in Fig. 3.

- Toxicity, Sentiment, and Empath features were concatenated together because they all describe the feelings or emotions in the text. They help understand if the text is angry, kind, rude, or emotional.
- TF-IDF and POS features were grouped into a second set because they tell us about the words and grammar in the text. TF-IDF shows which words are important, and POS tells us about word types like nouns, verbs, etc. These are more about the structure and meaning of the words themselves.

3.4 Neural Network Models

The current research employs three NN models - Feedforward Neural Network (FNN), ELM, and MLP. The configurations of the three models are as follows:

- FNN - The network consists of an input layer with as many neurons as the input features. It is then followed by four dense layers each consisting of 64 hidden neurons and ReLU activation function. Finally, the network consists of an output layer with four units and softmax activation for the four bystander labels. The sparse categorical cross-entropy loss function was employed to handle the multiclass classification task. This function

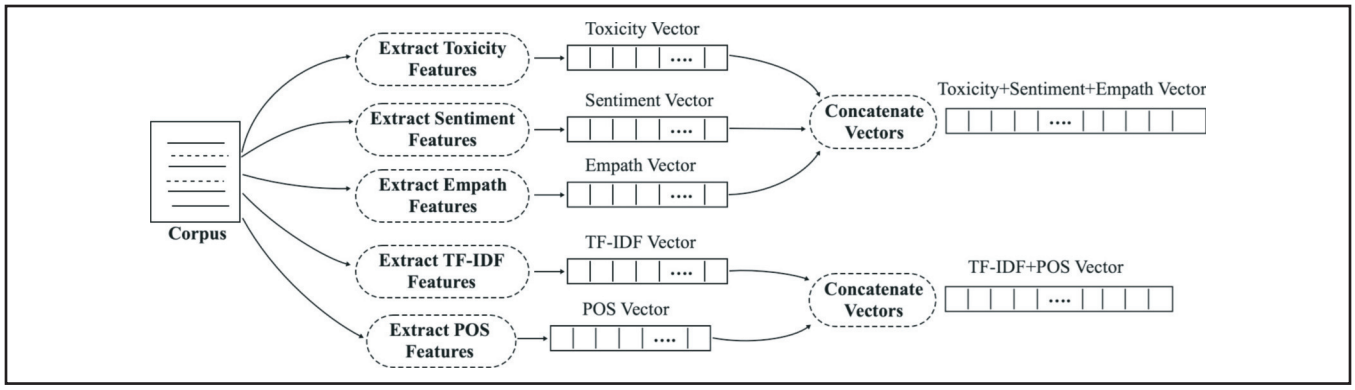


Fig. 3: *Extraction and Concatenation of Features*

is appropriate when integer-labeled target classes are involved. Finally, the model has been optimized with Adam optimizer and accuracy metric has been maximized over 1000 epochs. It is implemented using Keras' Sequential API which uses TensorFlow's backend and provides customizable training (layers, loss function, optimizers). The use of explicit epoch control ensures that the model is fully trained to convergence.

- ELM - As in FNN, the network consists of an input layer with as many neurons as the input features, followed by a single hidden layer with 128 neurons, and an output layer for predictions. The tanh activation function has been used on the hidden neurons.
- The MLP model employs a feedforward structure similar to that of the FNN, beginning with an input layer and followed by four hidden layers, each consisting of 64 neurons that utilize the ReLU activation function. To address multiclass classification, the output layer applies the softmax function. The training process is conducted with the Adam optimizer over a maximum of 1000 iterations. This model is constructed using Scikit-learn's MLP Classifier, which is user-friendly; however, it cannot customize the loss function in the way that Keras does. Additionally, the model may terminate early if convergence criteria are satisfied or if it is progressing slowly, which might result in underfitting.

For improved and stable results, Stratified 5-fold validation has been utilized for dividing the data into training and test sets.

3.5 Experimental Setup

Fig. 4 illustrates the experimental pipeline followed in this study. After extracting features from the corpus, various feature sets are constructed as described in subsection 3.3. These feature sets are then balanced using the Borderline SMOTE technique to address class imbalance. The resulting balanced numeric datasets are fed into three different neural network architectures (explained in subsection 3.4), which are trained and evaluated using stratified 5-fold cross-validation wherein 4 folds are used in training and 1 fold is used in testing for the prediction of a new label. The process of selection of 4 folds for training is repeated 5 times. Performance metrics such as accuracy, precision, recall, and F1-score are analyzed to determine the most effective architecture. The Python library Scikit-learn (Pedregosa et al., 2011) and Keras (Chollet et al., 2015) were utilized for the implementation of the neural network architectures. All experiments were conducted on a system running macOS Big Sur Version 11.3.1 with 8GB of RAM.

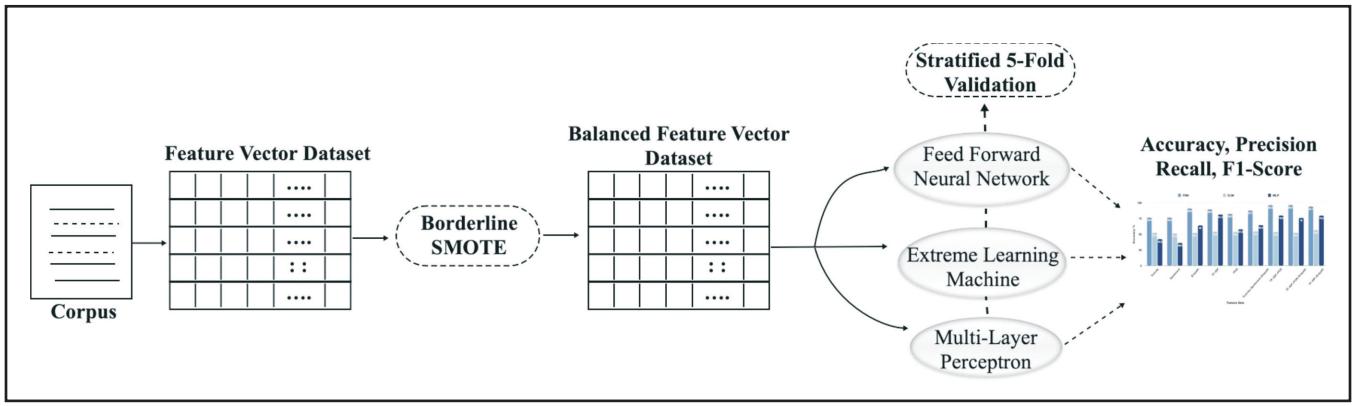


Fig. 4: *Experiment Pipeline*

4. Results

The performance of the three neural networks across all five individual feature sets: Toxicity, Sentiment, Empath, TF-IDF, and POS is presented in Table 2, evaluated using Accuracy, Recall, Precision, and F1-Score. For the four bystander role labels, the precision, recall, and F1-scores are also provided separately. The weighted averages are also reported - weighted precision (WP), weighted recall (WR), weighted F1 (WF). The following observations are evident:

Table 2: *Performance of Neural Networks on Individual Feature Sets*

Feature Set	Models	Accuracy	Precision				WP	Recall				WR	F1-Score				WF
			0	1	2	3		0	1	2	3		0	1	2	3	
Toxicity	FNN	0.7642	0.68	0.71	0.69	0.97	**0.76**	0.67	0.78	0.62	0.99	**0.77**	0.68	0.74	0.65	0.98	**0.76**
	MLP	0.4221	0.31	0.35	0.39	0.58	**0.41**	0.16	0.46	0.40	0.67	**0.42**	0.21	0.40	0.39	0.62	**0.41**
	ELM	0.5241	0.45	0.52	0.39	0.71	**0.52**	0.33	0.35	0.50	0.91	**0.52**	0.38	0.41	0.44	0.80	**0.51**
Sentiment	FNN	0.7654	0.69	0.73	0.66	0.98	**0.77**	0.63	0.77	0.67	0.99	**0.77**	0.66	0.75	0.67	0.98	**0.77**
	MLP	0.3618	0.23	0.41	0.37	0.55	**0.39**	0.34	0.22	0.43	0.46	**0.36**	0.27	0.29	0.39	0.50	**0.36**
	ELM	0.5121	0.41	0.47	0.46	0.63	**0.49**	0.30	0.43	0.41	0.91	**0.51**	0.35	0.45	0.43	0.74	**0.49**
Empath	FNN	0.909	0.91	0.90	0.85	0.99	**0.91**	0.79	0.94	0.92	0.99	**0.91**	0.85	0.92	0.88	0.99	**0.91**
	MLP	0.6404	0.5	0.62	0.57	0.88	**0.64**	0.47	0.68	0.57	0.85	**0.64**	0.49	0.64	0.57	0.86	**0.64**
	ELM	0.5208	0.46	0.47	0.42	0.66	**0.5**	0.30	0.38	0.49	0.91	**0.52**	0.37	0.42	0.45	0.77	**0.5**
TF-IDF	FNN	0.8958	0.95	0.95	0.75	0.99	**0.91**	0.86	0.98	0.94	0.79	**0.89**	0.90	0.97	0.84	0.88	**0.9**
	MLP	0.8125	0.74	0.83	0.72	0.94	**0.81**	0.67	0.87	0.72	0.99	**0.81**	0.70	0.85	0.72	0.96	**0.81**
	ELM	0.5395	0.45	0.49	0.46	0.69	**0.52**	0.32	0.43	0.48	0.92	**0.54**	0.37	0.46	0.47	0.79	**0.52**
POS	FNN	0.8235	0.78	0.82	0.71	0.98	**0.82**	0.72	0.86	0.73	0.98	**0.82**	0.75	0.84	0.72	0.98	**0.82**
	MLP	0.5768	0.54	0.48	0.45	0.79	**0.57**	0.32	0.61	0.41	0.97	**0.58**	0.40	0.54	0.43	0.87	**0.56**
	ELM	0.5362	0.44	0.51	0.44	0.67	**0.52**	0.30	0.43	0.46	0.96	**0.54**	0.36	0.47	0.45	0.79	**0.52**
Toxicity + Sentiment + Empath	FNN	0.8772	0.84	0.84	0.84	0.99	**0.88**	0.79	0.90	0.83	0.99	**0.88**	0.81	0.87	0.84	0.99	**0.88**
	MLP	0.6436	0.54	0.60	0.52	0.88	**0.64**	0.43	0.63	0.55	0.97	**0.65**	0.47	0.61	0.53	0.93	**0.64**
	ELM	0.545	0.48	0.48	0.46	0.68	**0.53**	0.34	0.38	0.51	0.95	**0.55**	0.40	0.43	0.48	0.80	**0.53**
TF-IDF + POS	FNN	0.9627	0.96	0.97	0.92	1	**0.96**	0.90	0.98	0.97	1	**0.96**	0.93	0.98	0.94	1	**0.96**
	MLP	0.7982	0.72	0.82	0.67	0.97	**0.8**	0.64	0.84	0.73	0.99	**0.8**	0.67	0.83	0.70	0.98	**0.8**
	ELM	0.5351	0.45	0.46	0.45	0.69	**0.51**	0.33	0.46	0.39	0.95	**0.53**	0.38	0.46	0.42	0.80	**0.52**

- Across all five individual feature sets, the FNN model consistently outperformed the other two models, achieving the highest values for all four performance metrics. The highest accuracy was achieved on the Empath feature set at 90.9%, followed by 89.58% on the TF-IDF feature set.
- The FNN model achieved the highest precision for classes 0 and 1 using the TF-IDF feature set, class 2 using the Empath feature set, and class 3 with an equivalent precision of 0.99 using both the Empath and TF-IDF feature sets.
- For recall, the FNN performed best on classes 0 and 1 using the TF-IDF feature set and on classes 2 and 3 using the Empath feature set. Additionally, a close recall value of 0.98 for class 3 was achieved on the Toxicity, Sentiment, and POS feature sets.
- FNN attained the highest F1-Score values for classes 0, 1, and 2 using the TF-IDF feature set, and a score of 0.99 for class 3 using the Toxicity, Sentiment, and Empath feature sets. While the MLP model also attained 0.99 as an F1 score on class 3 using the TF-IDF feature set, it underperformed across other classes.

- Both the MLP and ELM models showed poor performance across most cases, with ELM yielding accuracy values close to 50%, while the MLP performed slightly better. Both models performed the worst on the Sentiment feature set and showed their best performance on the TF-IDF feature set.

Additionally, the feature sets were combined, resulting in two new sets: Toxicity + Sentiment + Empath and TF-IDF + POS. The results for these combined feature sets are also presented in Table 2. It is evident from the table that:

- On the Toxicity + Sentiment + Empath feature set, the accuracy of the FNN model dropped to 87.72% when compared to its performance on the standalone Empath feature set. However, MLP and ELM showed slight improvement on the combined set as compared to the individual sets.
- The recall and precision, thus the F1-score value of FNN and MLP, slightly dropped on the Toxicity + Sentiment + Empath feature set when compared to the standalone Empath feature set. However, ELM showed a slight improvement on the combined set.
- On the TF-IDF +POS feature set, significant improvement is seen in all four performance metrics of the FNN model, achieving the highest accuracy of 96.27% across all the feature sets. However, all four measures of MLP and ELM slightly decreased in most of the cases when compared to the individual TF-IDF and POS sets.

Based on the performance of the FNN model across individual and combined feature sets, the highest accuracy was observed with the Empath feature set and the TF-IDF + POS combination. To further explore potential improvements, two additional combinations — TF-IDF +POS+Empath and TF-IDF + Empath were evaluated. The accuracy values of the three NNs over all the feature sets are illustrated in Fig. 5, and the following insights can be drawn:

- A marginal improvement of only 0.2% was observed when using the TF-IDF + POS + Empath feature set compared to TF-IDF + POS. This small gain does not justify the dataset size, as well as the running time of the model being increased. Therefore, TF-IDF + POS remains the more efficient and practical choice.
- A decline in performance by 2.2% was observed for the TF-IDF +Empath feature set compared to TF-IDF + POS, indicating that POS features contribute more significantly to model performance than Empath features in this combination.

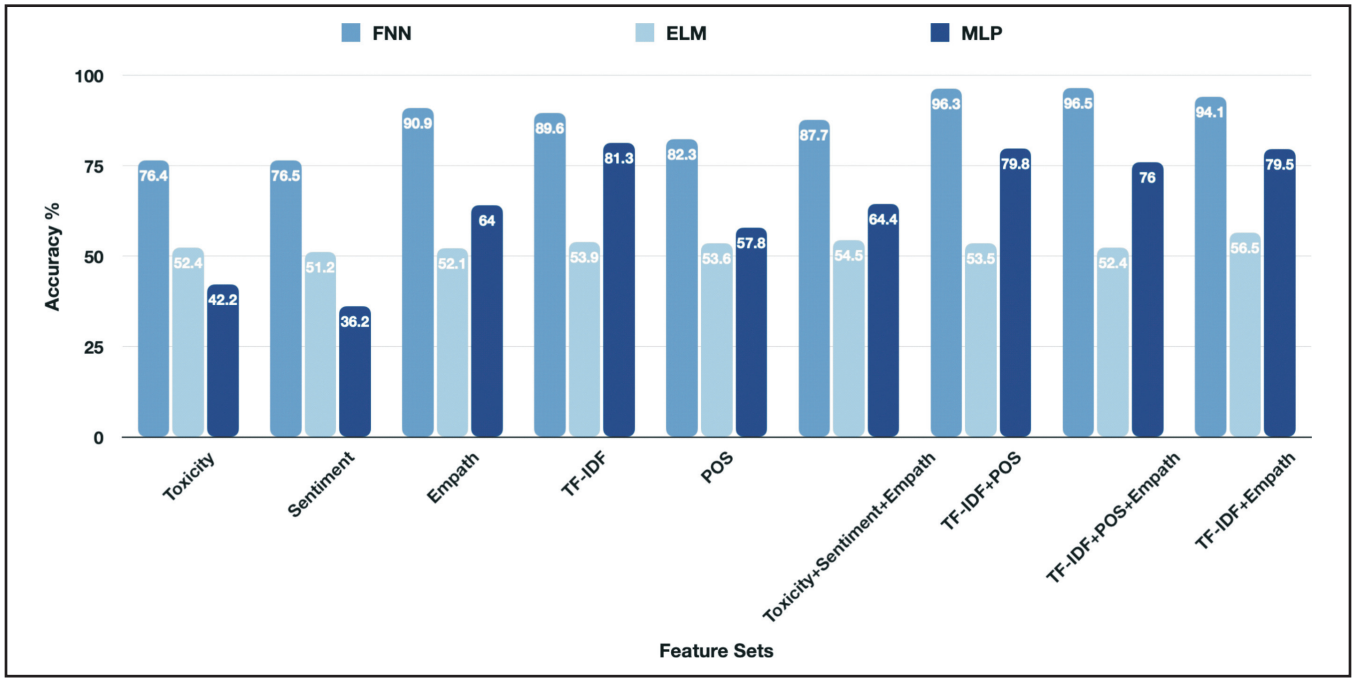


Fig. 5: *Accuracy of the models on different Feature Sets*

In conclusion, the FNN model gives the best overall performance, measured in terms of accuracy, precision, recall, and F1-score, when using the TF-IDF + POS feature set.

The confusion matrices of the FNN model for all five basic feature sets - Toxicity, Sentiment, Empath, POS, and TF-IDF, as well as on the best performing combined feature set TF-IDF +POS, are presented in Figure 6. In each matrix, the bottom-right cell displays the overall classification accuracy (in green) and the misclassification rate (in red). Additionally, the rightmost column shows the recall for each class, while the bottom row indicates the precision for each class.

FNN Performance on Toxicity Features

Actual \ Predicted	Instigator	Defender	Impartial	Irrelevant	Recall
Instigator	153 16.78%	31 3.40%	42 4.61%	2 0.22%	228 67.11% 32.89%
Defender	28 3.07%	177 19.41%	21 2.30%	2 0.22%	228 77.63% 22.37%
Impartial	41 4.50%	41 4.50%	142 15.57%	4 0.44%	228 62.28% 37.72%
Irrelevant	2 0.22%	0 0.00%	1 0.11%	225 24.67%	228 98.68% 1.32%
Precision	224 68.30% 31.70%	249 71.08% 28.92%	206 68.93% 31.07%	233 96.57% 3.43%	697 / 912 76.43% 23.57%

FNN Performance on Sentiment Features

Actual \ Predicted	Instigator	Defender	Impartial	Irrelevant	Recall
Instigator	144 15.79%	34 3.73%	48 5.26%	2 0.22%	228 63.16% 36.84%
Defender	20 2.19%	176 19.30%	31 3.40%	1 0.11%	228 77.19% 22.81%
Impartial	43 4.71%	30 3.29%	153 16.78%	2 0.22%	228 67.11% 32.89%
Irrelevant	2 0.22%	1 0.11%	0 0.00%	225 24.67%	228 98.68% 1.32%
Precision	209 68.90% 31.10%	241 73.03% 26.97%	232 65.95% 34.05%	230 97.83% 2.17%	698 / 912 76.54% 23.46%

(a) Toxicity(b) Sentiment

FNN Performance on Empath Features

Actual \ Predicted	Instigator	Defender	Impartial	Irrelevant	Recall
Instigator	181 19.85%	16 1.75%	30 3.29%	1 0.11%	228 79.39% 20.61%
Defender	5 0.55%	214 23.46%	8 0.88%	1 0.11%	228 93.86% 6.14%
Impartial	12 1.32%	6 0.66%	209 22.92%	1 0.11%	228 91.67% 8.33%
Irrelevant	1 0.11%	2 0.22%	0 0.00%	225 24.67%	228 98.68% 1.32%
Precision	199 90.95% 9.05%	238 89.92% 10.08%	247 84.62% 15.38%	228 98.68% 1.32%	829 / 912 90.90% 9.10%

FNN Performance on POS Features

Actual \ Predicted	Instigator	Defender	Impartial	Irrelevant	Recall
Instigator	165 18.09%	19 2.08%	44 4.82%	0 0.00%	228 72.37% 27.63%
Defender	9 0.99%	196 21.49%	22 2.41%	1 0.11%	228 85.96% 14.04%
Impartial	37 4.06%	22 2.41%	166 18.20%	3 0.33%	228 72.81% 27.19%
Irrelevant	1 0.11%	2 0.22%	1 0.11%	224 24.56%	228 98.25% 1.75%
Precision	212 77.83% 22.17%	239 82.01% 17.99%	233 71.24% 28.76%	228 98.25% 1.75%	751 / 912 82.35% 17.65%

(c) Empath (d) POS

FNN Performance on TF-IDF Features

Actual \ Predicted	Instigator	Defender	Impartial	Irrelevant	Recall
Instigator	197 21.60%	6 0.66%	24 2.63%	1 0.11%	228 86.40% 13.60%
Defender	2 0.22%	224 24.56%	2 0.22%	0 0.00%	228 98.25% 1.75%
Impartial	8 0.88%	5 0.55%	215 23.57%	0 0.00%	228 94.30% 5.70%
Irrelevant	1 0.11%	1 0.11%	45 4.93%	181 19.85%	228 79.39% 20.61%
Precision	208 94.71% 5.29%	236 94.92% 5.08%	286 75.17% 24.83%	182 99.45% 0.55%	817 / 912 89.58% 10.42%

FNN Performance on TF-IDF+POS Features

Actual \ Predicted	Instigator	Defender	Impartial	Irrelevant	Recall
Instigator	205 22.48%	4 0.44%	19 2.08%	0 0.00%	228 89.91% 10.09%
Defender	4 0.44%	223 24.45%	1 0.11%	0 0.00%	228 97.81% 2.19%
Impartial	4 0.44%	2 0.22%	222 24.34%	0 0.00%	228 97.37% 2.63%
Irrelevant	0 0.00%	0 0.00%	0 0.00%	228 25.00%	228 100.00% 0.00%
Precision	213 96.24% 3.76%	229 97.38% 2.62%	242 91.74% 8.26%	228 100.00% 0.00%	878 / 912 96.27% 3.73%

(e) TF-IDF(f) TF-IDF+PO

Fig. 6: *Confusion Matrices of FNN on Different Feature Sets*

The confusion matrices reveal the following important observations:

- Majority of the data points belonging to the Irrelevant category are correctly classified by FNN irrespective of the type of feature vector set used, with the TF-IDF + POS combination achieving perfect classification of all samples in this category.
- Among the individual feature sets, TF-IDF results in the highest number of correctly classified samples for the Instigator, Defender, and Impartial categories. This is followed by Empath, which ranks second, and POS, which ranks third in terms of classification performance for these categories.
- The samples belonging to the Instigator category turn out to be the most challenging to classify accurately. Even with the best-performing feature set, TF-IDF + POS, the FNN model fails to correctly classify 23 samples from this category.

The feature set TF-IDF + POS turns out to be the optimal choice among all, since TF-IDF tells us which words are important or unique in a tweet. It helps the model understand what the tweet is talking about. Also, POS tags show the grammatical role of words, which helps the model learn how the message is constructed. It is useful for understanding the intent and tone of the tweet. Further, both of these individual feature sets provide non-overlapping information and thus result in a more useful feature set combination. Therefore, the FNN model achieves the highest classification performance on the Bystander role dataset when using the combined feature set TF-IDF + POS.

5. Conclusion and Future Scope

This paper presents research on extracting various types of features from tweets related to bystanders. We utilized the dataset CYBY23, which contains tweets from bystanders, and conducted experiments with different combinations of features. A total of 2,284 features were extracted using various tools and APIs, including sentiment features obtained from TextBlob, emotion features from Empath, toxicity features via the Perspective API, TF-IDF features, and features derived from Part of Speech (POS) Tagging. For the

classification task, we employed three types of deep neural network models: a feedforward neural network, a Multilayer Perceptron (MLP), and an Extreme Learning Machine (ELM). To evaluate the models, we employed metrics such as accuracy, precision, recall, and F1-score. The experiments showed that choosing the right features (important characteristics from the text data) is very important for getting good results in a machine learning model. Among all the feature combinations tested, the best performance was achieved when TF-IDF features were combined with POS features. Additionally, the feedforward neural network outperformed all other deep learning models.

In the future, we plan to work with a larger dataset. We also aim to explore multimodal data, including emojis, emoticons, images, and videos. Furthermore, we intend to extend our research to identify bystander roles in tweets that feature multilingual data, incorporating various low-resource languages.

Data Availability Statement

The Cyberbullying Bystander Dataset 2023 that supports the findings of this study is openly available at: https://www.kaggle.com/datasets/haifasaleh/cyberbullying-bystander-dataset-2023.

Conflict of Interest Statement

The authors declare no conflict of interest.

Ethical Declarations

The authors declare that this research does not involve human participants, animals, or any procedures requiring ethical approval.

References

1. Albraikan, A. A., Hassine, S. B. H., Fati, S. M., Al-Wesabi, F. N., Hilal, A. M., Motwakel, A., Hamza, M. A., and Al Duhayyim, M. (2022). Optimal deep learning-based cyberattack detection and classification technique on social networks. *Computers, Materials and Continua*, 72(1):907 - 923.
2. Alfurayj, H., Yee, N. S., and Lutfi, S. L. (2023a). Cyberbullying bystander dataset 2023.
3. Alfurayj, H. S. and Lutfi, S. L. (2023). Exploring bystanders' roles in labeled cyberbullying threads on twitter: A preliminary analysis. In *TENCON 2023-2023 IEEE Region 10 Conference (TENCON)*, pages 1018-1023. IEEE.
4. Alfurayj, H. S., Lutfi, S. L., and Perumal, R. (2024). A chained deep learning model for fine-grained cyberbullying detection with bystander dynamics. *IEEE Access*, 12:105588 -105604. Cited by: 0; All Open Access, Gold Open Access.
5. Alfurayj, H. S., Yee, N. S., and Lutfi, S. L. (2023b). Bystanders unveiled: Introducing a comprehensive cyberbullying corpus with bystander information. In *TENCON 2023-2023 IEEE Region 10 Conference (TENCON)*, pages 1012-1017. IEEE.
6. Bedi, P. and Bhasin, V. (2019). Multilayer ensemble of elms for image steganalysis with multiple feature sets. *International Journal of Computational Science and Engineering*, 20(4):558 - 569.

7. Bhasin, V. and Bedi, P. (2013). Steganalysis for jpeg images using extreme learning machine. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1361-1366.
8. Bird, S., Klein, E., and Loper, E. (2009). Natural Language Processing with Python. *O'Reilly Media*.
9. Cao, J., Xiong, Y., Wang, W., & Chen, G. (2026) Cyberbullying Detection Based on Hybrid Neural Networks and Multi-Feature Fusion. *Information*, *17*(2), 205.
10. Chatzakou, D., Leontiadis, I., Blackburn, J., Cristofaro, E. D., Stringhini, G., Vakali, A., and Kourtellis, N. (2019). Detecting cyberbullying and cyberaggression in social media. *ACM Transactions on the Web (TWEB)*, 13(3).
11. Chollet, F. et al. (2015). Keras. https://keras.io.
12. Davidson, T., Warmsley, D., Macy, M., and Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1):512-515.
13. DeSmet, A., Bastiaensens, S., Van Cleemput, K., Poels, K., Vandebosch, H., and De Bourdeaudhuij, I. (2019). The efficacy of a serious game on bystander behavior in cyberbullying: A randomized controlled trial. *Computers in Human Behavior*, 57:196-206.
14. Fast, E., Chen, B., and Bernstein, M. S. (2016). Empath: Understanding topic signals in large-scale text. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, page 4647-4657, New York, NY, USA. *Association for Computing Machinery*.
15. Fati, S., Muneer, A., Alwadain, A., and Balogun, A. (2023). Cyberbullying detection on twitter using deep learning-based attention mechanisms and continuous bag of words feature extraction. *Mathematics*, 11:3567. Gini, G., Pozzoli, T., Sala, M., and Elia, N. (2014). The role of empathy in the bystander effect in bullying situations. *Aggressive Behavior*, 40(6):522-533.
16. Gupta, A., Chug, A., and Singh, A. P. (2024a). Multi-disease identification in single citrus leaf image using sigmoid activation and optimised threshold. In 2024 11th International Conference on Signal Processing and Integrated Networks (SPIN), pages 285-290. IEEE.
17. Gupta, A., Chug, A., and Singh, A. P. (2026). Feature extraction in sensor plant disease datasets using reformed membership functions independent of class variables. Scientific Reports, 16(1):4115.
18. Gupta, A. and Gupta, S. (2024). Enhanced classification of imbalanced medical datasets using hybrid data-level, cost-sensitive and ensemble methods. International Research Journal of Multidisciplinary Technovation, 6(3):58-76.
19. Gupta, A., Thakkar, K., Bhasin, V., Tiwari, A., and Mathur, V. (2024b). Bystander detection: Automatic labeling techniques using feature selection and machine learning. International Journal of Advanced Computer Science and Applications, 15(1):1135 - 1143. Cited by: 0; All Open Access, Gold Open Access.
20. Han, H., Wang, W.-Y., and Mao, B.-H. (2005). Borderline-smote: a new oversampling method in imbalanced data sets learning. In *International conference on intelligent computing*, pages 878-887. Springer.
21. Hasan, M. T., Hossain, M. A. E., Mukta, M. S. H., Akter, A., Ahmed, M., and Islam, S. (2023). A review on deep-learning-based cyberbullying detection. Future Internet, 15(5).
22. Hawkins, D. L., Pepler, D. J., and Craig, W. M. (2019). *Naturalistic observations of peer interventions in bullying. Social Development,* 10(4):512-527.
23. Huang, G.-B., Zhou, H., Ding, X., and Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics*, Part B (Cybernetics), 42(2):513-529.
24. Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: Theory and applications. Neurocomputing, 70(1):489-501. *Neural Networks*.
25. Huang, Y. Y. and Chou, C. (2020). An analysis of multiple factors of cyberbullying among junior high school students in Taiwan. *Computers in Human Behavior*, 73:221-228.
26. Jazzar, M., Duridi, T. (2025). A Comprehensive Review of Machine Learning and Deep Learning Techniques for Cyberbullying Detection. In: Poonia, R.C., Sharma, S., Hameed, I.A., Upreti, K. (eds) Smart Cyber Physical Systems. ICSCPS 2024. *Smart Innovation, Systems and Technologies*, vol 435. Springer, Singapore.
27. Khairy, M., Mahmoud, T. M., and Abd-El-Hafeez, T. (2024). The effect of rebalancing techniques on the classification performance in cyberbullying datasets. *Neural Computing and Applications*, 36(3):1049 - 1065. Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., and Lattanner, M. R. (2020). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, 140(4):1073-1137.
28. Lepe-Fa´undez, M., Segura-Navarrete, A., Vidal-Castro, C., Mart´ýnez-Araneda, C., and Rubio-Manzano, C. (2021). Detecting aggressiveness in tweets: A hybrid model for detecting cyberbullying in the spanish language. *Applied Sciences*, 11(22).
29. McMahon, S. D., Florin, P., and Kivlighan, D. M. (2014). Prevention of bullying and victimization through bystander education and intervention. *School Psychology Review*, 43(4):510-527.
30. Obermaier, M., Fawzi, N., and Koch, T. (2016). Bystanding or standing by?how the number of bystanders affects the intention to intervene in cyberbullying. *New Media and Society*, 18(8):1491-1507.

31. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825-2830.
32. Raj, C., Agarwal, A., Bharathy, G., Narayan, B., and Prasad, M. (2021). Cyberbullying detection: Hybrid models based on machine learning and natural language processing techniques. *Electronics (Switzerland)*, 10(22). Cited by: 80; All Open Access, Gold Open Access, Green Open Access.
33. Risch, J. and Krestel, R. (2018). Aggression identification using deep learning and data augmentation. In Kumar, R., Ojha, A. K., Zampieri, M., and Malmasi, S., editors, Proceedings of the First Workshop on Trolling,

❖ ❖ ❖

Vol. 6 (1), 2026
E-ISSN: 2583-4452

Ego-resilience, Stress and Anxiety predicting Psychological Well-being of Adolescents

Vansh Yadav[1] and Daisy Sharma[2*]

[1] Research Scholar, Department of Psychology, Keshav Mahavidyalaya, University of Delhi, Delhi

[2] Associate Professor, Department of Psychology, Keshav Mahavidyalaya, University of Delhi, Delhi

[*] Corresponding Author ✉ drdaisysharma@keshav.du.ac.in

ABSTRACT

Adolescence is a vulnerable growing phase marked by physical, emotional, psychological, and social changes. During this phase, adolescents are particularly vulnerable to stress and anxiety — driven by fear of failure, academic overload, and peer competition —which significantly impacts their psychological well-being. This study examined how ego resilience, perceived stress, and anxiety predict the psychological well-being of adolescents. A quantitative, predictive correlational design was employed with 61 adolescents (aged 13–18) from urban private schools in Delhi, India, selected through disproportionate stratified sampling. The Perceived Stress Scale, Ego Resilience Scale, Ryff's Psychological Well-Being Scale and State-Trait Anxiety Inventory were administered. Ego resilience was significantly positively associated with self-acceptance (r = .353, p = .005). State and trait anxiety demonstrated the strongest negative associations with self-acceptance (r = -.643 and r = -.591, respectively) and environmental mastery. Perceived stress significantly and negatively predicted environmental mastery (r = -.448) and self-acceptance (r = -.412). Together, all the predictors explained 16.6% of the variance in psychological well-being (F (4, 56) = 2.78, p = .035, R^2 = .166). While the overall regression model significantly predicted psychological well-being, none of the individual predictors reached a statistically significant level independently, suggesting that ego resilience, perceived stress, and anxiety collectively contribute to psychological well-being rather than any single factor acting as a dominant predictor. Findings support the need for school-based anxiety management and resilience-building interventions to promote adolescent psychological well-being.

Keywords: *Self-acceptance, Environmental Mastery, Mental Health, Academic Pressure, Developmental Vulnerability*

1. INTRODUCTION

Adolescence refers to a period that usually occurs in the middle of childhood & adulthood, i.e. 12 years old to 18 years old (Jaworska & MacQueen, 2015). It is a critical & essential developmental phase that leads to the neurobiological changes in the body, which increase the vulnerability of adolescents towards the stressors present in the environment, which can also imperil the psychological well-being of the adolescents and heighten the probability for psychopathology (Chavanne et al., 2023; Wiglesworth et al., 2023).

Ego resilience refers to the individual's ability to deal with and adapt to stress, anxiety & challenges they face over the course of life. It basically reflects the inner potential and capacity of the individual to flexibly adapt and recover from the hurdles and adversities faced by them, and to balance their emotional stability under times of pressure (Block & Kremen, 1996; Tyagi & Khokkar, 2024). Ego resilience develops through a mix of early life experiences, coping styles, personality traits, and social support. Supportive caregiving in childhood, healthy ways of handling stress, and qualities like optimism and self-esteem all help individuals become more emotionally strong and adaptable. It has been observed that positive social relationships during childhood and adolescence play a crucial role in building resilience during the later years of life (Chen et al., 2021; Tyagi & Khokkar, 2024).

Stress is a widely investigated concept in the domain of developmental and health psychology. According to Transactional Model of Stress and Coping, stress does not solely emerge as an outcome of environmental events or an individual's personal characteristics but are also the result of the ongoing interaction between the individual and their environment (Lazarus and Folkman's (1984). Stress occurs when individuals perceive that the demands of a situation exceed their available coping resources, thereby posing a threat to their well-being. In this view, the experience of stress depends largely on how a person interprets and evaluates a particular situation rather than on the situation itself. Physical

Received on: 30 May 2026 | Accepted on: 26 Jun 2026 | Published Online on: 25 July 2026

form includes insomnia or hypersomnia, fatigue, nausea, and chest pains. Mental forms include problems in memory, concentration, decision making, and loss of sense of humour. Emotional forms include anxiety, anger, frustration, and depression (Ciccarelli & White, 2025).

Anxiety can be understood as a mental state in which an individual goes through a state of extreme apprehension, worry or tension that something adverse is going to happen in the near future. When it becomes excessive and arises without any clear cause, it leads to maladaptive behaviours and the onset of anxiety disorders (Saviola et al., 2020). It includes both emotional and physical components, containing the most common symptoms, which are shortness of breath, tremors and palpitations (Impey et al., 2020). Charles Spielberger introduced a theoretical framework, which is known as the State- Trait Anxiety theory (Skapinakis, 2014). State anxiety is a temporary emotional reaction that occurs in stressful situations and is often marked by nervousness, worry, and tension. In contrast, trait anxiety is a more stable personality characteristic, where individuals are generally more likely to experience anxiety in various situations.

Psychological well-being (PWB) is a multidimensional concept that refers to an individual's optimal psychological functioning, personal growth, self-actualisation, and ability to face life's challenges (Ryff, 1989). This framework includes the six distinct yet interrelated domains/ keys – autonomy, self-acceptance, positive relations, personal growth, purpose in life, environmental mastery, and positive functioning (Ryff & Keyes, 1995). The hedonic perspective of well-being is deeply rooted in the philosophy of Aristippus, who taught the world that all the happiness in the world is present in hedonic moments and that the main goal or aim of life is to experience the maximum amount of pleasure in life. Meanwhile, in contrast, the Eudaimonic perspective is based upon Aristotle's Nicomachean Ethics, which states that all the human goods that are at the highest are achievable by human action was eudaimonia. It is defined as the activity of the soul which is followed by virtue- a personal kind of excellence, which helps to achieve the best in us.

Recent research shows that people with higher ego resilience generally experience better psychological well-being and are able to gain more from a sense of community and social support (Fino et al., 2025). Studies also suggest that resilience is a fixed personality trait and a flexible trait that helps to connect healthy habits – such as exercise, good sleep, supportive friendships, and belongingness – with positive well-being outcomes (LiuB et al., 2025). On the other side, factors like academic pressure, financial difficulties, and stress caused by events such as the pandemic have been found to negatively affect psychological well-being, especially among adolescents and students (Allayannis et al., 2022; Ebrahim et al., 2022).

High anxiety has been found to be linked with lower academic involvement, reduced independence, and poorer adjustment (Sinval et al., 2025), although individuals with stronger ego resilience reported lower stress levels, less depression, and greater life satisfaction (Abd el-Halem et al., 2022; Elzohary et al., 2017; Lee & Lee, 2021). Ego resilience acts as a protective factor against the harmful effects of loneliness on anxiety (Padmanabhanunni & Pretorius, 2025). Although many studies have explored these variables separately, there is still limited research that examines resilience, stress, and both state and trait anxiety together, particularly among Indian adolescents. The present study fills this gap by exploring how these factors jointly influence the six dimensions of psychological well-being proposed by Ryff in an Indian adolescent population. From a theoretical perspective, this study draws on the idea proposed by Block and Kremen (1996) that ego resilience functions as a flexible self-regulatory capacity, enabling individuals to adapt effectively to changing circumstances and cope with challenges. This view aligns with the transactional model of stress developed by Lazarus and Folkman (1984), which suggests that the impact of stressful situations depends largely on how individuals interpret and respond to them. Within this framework, ego resilience can be understood as a personal resource that influences the way stress and anxiety are perceived, regulated, and managed.

The most existing research has focused on Western societies and adult samples. Indian adolescents, however, as grow up in a socio-cultural environment, they are influenced by strong family interconnectedness, intense academic pressures, and evolving social and cultural expectations. These contextual factors shape the ways in which resilience operates and the extent to which it protects young people from the negative effects of stress and anxiety. Such an investigation of these relationships in an Indian adolescent population projects to offer an opportunity to broaden the relevance of existing theories as well as assess their applicability within a cultural and developmental context. Such emphasis has received comparatively limited attention in psychological research.

2. Aims and Objectives

2.1 Aims

To examine how ego resilience, anxiety and stress predict the psychological well-being of adolescents

2.2 Objectives

- To analyse the relationship between ego resilience and psychological well-being in adolescents.
- To examine the association between anxiety and the psychological well-being of adolescents.

- To examine the association between stress and the psychological well-being of adolescents
- To analyse the predictive roles of ego resilience, anxiety, and stress on psychological well-being among adolescents.

2.3 Research questions

- Does ego resilience significantly predict the psychological well-being of adolescents?
- Does anxiety significantly predict the psychological well-being of adolescents?
- Does stress significantly predict the psychological well-being of adolescents?
- Do ego resilience, anxiety, and stress together significantly predict the psychological well-being of adolescents?

2.4 Hypotheses

H1: There will be a significant relationship between ego-resilience and the psychological well-being of adolescents

H2: There will be a significant association between anxiety and the psychological well-being of adolescents

H3: There will be a significant association between stress and the psychological well-being of adolescents

H4: Ego-resilience, anxiety and stress will significantly predict the psychological well-being of adolescents

3. Methodology

3.1 Research Design

The current study has a Quantitative, Predictive Correlational Design as it predicts the psychological well-being of adolescents from ego resilience, anxiety and stress.

3.2 Participants

The study includes early adolescents aged 13-15 years and middle adolescents aged 15-18 years from urban private schools in Delhi, India. Participants were selected based on the following criteria. Inclusion Criteria: Early adolescents, 13-15 years old, and middle adolescents, 15-18 years old, living with at least one parent or guardian, willing to provide informed consent, parental consent is mandatory for minors. Exclusion Criteria- Individuals who do not fall into the early adolescent and middle adolescent age groups, adolescents unwilling or unable to participate or provide consent, living with neither of the Parents

3.3 Variables and tools

3.3.1 Variables

Independent Variables: Ego resilience, stress and anxiety

Dependent Variable: Psychological Well-being

3.3.2 Tools

Ryff's Psychological Well-Being Scales (PWB) (1989)- The internal consistency of the scale was 0.82, and the sub-scales range between 0.71 and 0.78. Ryff's Cronbach's Alpha for each were reported as 0.83, 0.86, 0.85, 0.88, 0.88 and 0. 91.

The Perceived Stress Scale (PSS-10) 10-item questionnaire was developed by Cohen et al. (1983). Individuals score between the range of 0 and 40, with higher scores indicating higher perceived stress. Reliability- The internal consistency of the scale, indicated by Cronbach's alpha, ranged from 0.71 to 0.91, while the test-retest reliability has been generally observed at 0.70.

The State Trait Anxiety Inventory is a 40-item questionnaire that is used to measure the state and trait anxiety of the individual separately. It was developed by Charles D. Spielberger, Richard L. Gorsuch, and Robert E. Lushene in 1964. The Trait-anxiety scale, the coefficients ranged from .65 to .86, whereas the range for the State-anxiety scale was .16 to .62.

Ego Resilience Scale by Block & Kremen (1996). It shows good internal reliability (Cronbach's $\alpha = 0.76$) and strong construct validity, as confirmed in diverse samples and contexts (Block & Kremen, 1996; Alessandri et al., 2007). Each of the 14 items is rated on a 4-point scale.

3.4 Procedure

Participants were recruited through urban private schools in Delhi, India, where the school administration shared the Google Form link directly with parents via their phone numbers. Parents were briefed about the nature and purpose of the research, and informed consent was obtained from them through a dedicated consent section within the same Google Form before their adolescent child proceeded to the questionnaires. As this study was conducted, the research design, consent procedure, and data collection protocol were reviewed and approved by the Subject Research Committee in accordance with the ethical guidelines of the Department of Psychology, Keshav Mahavidyalaya, University of Delhi. Participants were assured of the confidentiality of their results. After that, consent to voluntarily participate was taken, and each participant completed standardised questionnaires to assess ego resilience, anxiety – state and trait, perceived

stress and psychological well-being through Google Forms. Data were measured using the Ego Resilience Scale (PWB), State-Trait Anxiety Inventory, The Perceived Stress Scale (PSS-10) and Ryff's Psychological Well-Being Scales, respectively. The participant's doubts were clarified through a call. The participant was thanked after completion of all three scales.

4. Results

The results of a study conducted on 61 adolescents between the ages of 13 and 18 to understand how ego resilience, perceived stress, and state-trait anxiety are connected to psychological well-being. The analysis included descriptive statistics, regression analysis and correlation analysis to explore these relationships in detail.

The descriptive statistics were calculated for all the primary variables to characterise the central tendency, variability, and distributional shape of scores in the sample. This will help to understand where the data of each sample stands on each construct.

Table 1: *Descriptive Statistics for Study Variables*

Variables	Min	Max	Range	M	SD	Skewness	Kurtosis
ERS	22	50	28	38.75	5.50	-.321	.067
PSS	7	40	33	22.25	5.75	.277	.755
STAI-Y1	27	74	47	50.48	9.37	.167	-.017
STAI-Y2	28	75	47	52.77	8.86	.335	.759
Autonomy	11	39	28	25.67	4.77	.094	.956
EM	16	33	17	24.57	3.48	-.047	.162
PG	15	42	27	26.98	5.54	.473	-.040
PR	13	42	29	25.98	5.05	.621	1.413
PI	17	41	24	26.36	5.43	.799	.584
SA	10	41	31	24.62	5.23	-.031	1.801

Note: *ERS = Ego Resilience Scale; PSS = Perceived Stress Scale; STAI-Y1 = State Anxiety; STAI-Y2 = Trait Anxiety; EM = Environmental Mastery; PG = Personal Growth; PR = Positive Relations with Others; PI = Purpose in Life; SA = Self-Acceptance. M = Mean; SD = Standard Deviation*

The results showed that participants generally had moderate to higher levels of ego resilience and moderate levels of perceived stress (Table 1). At the same time, both state and trait anxiety scores were higher than the standard STAI cut-off, indicating that many adolescents in the sample were experiencing noticeable anxiety during the study period. Among the six dimensions of psychological well-being, Personal Growth received the highest scores, whereas Environmental Mastery and Self-Acceptance were the lowest. This suggests that although adolescents may feel they are developing and learning, they are still working on building confidence, self-worth, and a sense of control over their environment.

Correlation findings revealed that ego resilience was positively related to Self-Acceptance and showed a slight positive trend with Positive Relations (but did not reach the significant threshold). However, its relationship with the remaining well-being dimensions was not significant, providing only partial support for the first hypothesis. In contrast, both state and trait anxiety were strongly and negatively associated with Self-Acceptance, Environmental Mastery, and Positive Relations (Table 2). Adolescents with higher anxiety tended to have a poorer self-image and felt less capable of managing everyday challenges. Perceived stress also showed negative links with Environmental Mastery and Self-Acceptance, further highlighting the negative association between stress and adolescent well-being.

Before conducting regression analysis, statistical assumptions such as autocorrelation and multicollinearity were checked and found to be within acceptable limits. The regression model as a whole was statistically significant and explained around 17% of the variance in overall psychological well-being (Table 3)

Ego resilience showed a positive relationship with total psychological well-being, indicating that higher levels of ego resilience were associated with better psychological well-being among adolescents. However, this relationship was not statistically significant, suggesting that ego resilience did not independently contribute to the prediction of psychological well-being when other variables were included in the model. This may be due to shared variance with other predictors,

which reduced its unique contribution. Perceived stress also did not independently predict psychological well-being at a significant level (Table 4).

Table 2: *Pearson Correlation among the variables (N=61)*

Variables	ERS	PSS	STAI-Y1	STAI-Y2	Auto	EM	PG	PR	PI	SA
ERS	1	-.09	-.30*	-.08	.02	.18	.07	.24	-.01	.35**
PSS	-.09	-	.60**	.73**	.20	-.44**	.06	-.25	-.02	-.41**
STAI-Y1	-.30*	.64**	-	.74**	-.02	-.41**	-.17	-.44**	.01	-.64**
STAI-Y2	-.08	.73**	.74**	-	.01	-.54**	-.09	-.34**	.06	-.59**
Auto	.02	.20	-.02	.01	-	-.02	.14	.16	.04	.23
EM	.18	-.44**	-.41**	-.54**	-.02	-	.13	.27*	.17	.41**
PG	.07	.06	-.17	-.09	.14	.13	-	.25*	.58**	.15
PR	.24	-.25	-.44**	-.34**	.16	.27*	.25*	-	.17	.27*
PI	-.01	-.02	.005	.06	.04	.17	.58**	.17	-	.08
SA	.35**	-.41**	-.64**	-.59**	.23	.41**	.15	.27*	.08	-

Note: *ERS = Ego Resilience Scale; PSS = Perceived Stress Scale; STAI-Y1 = State Anxiety; STAI-Y2 = Trait Anxiety; EM = Environmental Mastery; PG = Personal Growth; PR = Positive Relations; PI = Purpose in Life; SA = Self-Acceptance; Auto = Autonomy. a) *p < .05. b) **p < .01 (all tests two-tailed).*

Table 3: *Model Fit Summary for Multiple Linear Regression Predicting Total Psychological Well-Being*

Model	R	R^2	Adjusted R^2	F	df1	df2	p
1	0.407	0.166	0.106	2.78	4	56	.035

Note: *R = multiple correlation coefficient; R^2 = coefficient of determination; Adjusted R^2 = adjusted coefficient of determination; F = overall model F-statistic; df1 = regression degrees of freedom; df2 = residual degrees of freedom; p = two-tailed significance of the overall model.*

Table 4: *Regression Coefficients for the Prediction of Total Psychological Well-Being*

Predictor	B	SE	t	p	β	95% CI Lower	95% CI Upper
Ego Resilience (ERS)	0.32	0.45	0.70	.485	.092	-0.171	0.356
Perceived Stress (PSS)	0.89	0.60	1.47	.146	.268	-0.097	0.633
State Anxiety (STAI-Y1)	-0.58	0.40	-1.44	.155	-.283	-0.678	0.111
Trait Anxiety (STAI-Y2)	-0.59	0.47	-1.23	.221	-.274	-0.717	0.169

Note: *B = unstandardised regression coefficient; SE = standard error of B; t = t-statistic for the individual predictor; p = two-tailed significance of the individual predictor; β= standardised regression coefficient (beta); 95% CI = 95% confidence interval for the unstandardised coefficient.*

Although the analysis indicated an unexpected positive relationship between perceived stress and psychological well-being, this finding may be explained by the strong associations between perceived stress and both state anxiety and trait anxiety. Due to a substantial amount of variance, the independent effect of perceived stress becomes difficult to distinguish within the regression model. State anxiety demonstrated a negative relationship with psychological well-

being, indicating that higher levels of state anxiety were associated with lower psychological well-being. However, this effect was not statistically significant, suggesting that state anxiety did not independently contribute to the prediction of psychological well-being. Similar to ego resilience, its unique effect may have been reduced because of overlap with the other predictors included in the model. Likewise, trait anxiety showed a negative association with psychological well-being, indicating that increases in trait anxiety were related to decreases in psychological well-being. However, this relationship was also not statistically significant, suggesting that trait anxiety did not independently predict psychological well-being. The shared variance between trait anxiety and the other predictors may have limited its unique contribution to the model.

Though the individual predictors have shown the non-significant effects, the overall regression model has been found to be statistically significant which suggests that the predictors collectively explained a meaningful proportion of the variance in adolescent psychological well-being (Table 4.3). Thus, ego resilience, perceived stress, state anxiety, and trait anxiety jointly contribute to understanding adolescent well-being. These findings collectively suggest that adolescent psychological well-being is not influenced by any single variable acting in isolation rather is shaped by the combined influence of several psychological factors.

5. Discussion

The present study aims to predict the psychological well-being of adolescents from their ego resilience, anxiety and stress. For this purpose, a total of 61 students (31 males, 30 females) were recruited based on their age, ranging from 13 years old to 18 years old. The first hypothesis was that there would be a significant relationship between ego-resilience and the psychological well-being of adolescents, and it was partially supported, based on the Pearson correlational analyses. Self-acceptance was the dimension of psychological well-being that had a statistically significant relationship with ego resilience. The significant positive relationship found between ego resilience and self-acceptance is theoretically coherent and resonates with the existing research. Fino et al. (2025) found that participants with high ego-resilience personality profiles scored significantly higher on measures of individual functioning and well-being as compared to individuals with lower resilience, and community identification positively influenced social support and well-being, with this relationship being much stronger among highly ego-resilient individuals. These findings are relevant to the present study, where ego resilience in adolescents was also associated with well-being, though in a more specific rather than overall manner.

The second hypothesis was that there would be a significant association between anxiety and the psychological well-being of adolescents, and it was retained, which indicates that both state anxiety (STAI-Y1) and trait anxiety (STAI-Y2) showed multiple significant negative correlations with psychological well-being dimensions. Research by Padmanabhanunni and Pretorius (2025) found that ego resilience significantly and negatively predicted trait anxiety, and moreover, that ego resilience played a moderating role between the relationship of loneliness and anxiety, such that higher ego resilience attenuated the anxiety-amplifying effects of isolation. In the current study, ego resilience was also significantly and negatively correlated with state anxiety, which mirrors the findings of Padmanabhanunni and Pretorius (2025).

The third hypothesis states that there would be a significant association between stress and the psychological well-being of adolescents, which indicates that perceived stress has significant negative correlations with Environmental Mastery and Self-Acceptance, but it was unable to reach statistically significant levels in its relationship with Autonomy, Personal Growth, Positive Relations, or Purpose in Life. Research by Barbayannis et al. (2022) found a significant negative correlation between academic stress and mental well-being. Adolescents experienced significance of the stress and psychological well-being relationship is parallel across both studies, strengthening confidence in the robustness of this finding across different samples and measurement approaches.

The fourth hypothesis states that ego- resilience, anxiety and stress would significantly predict the psychological well-being of adolescents, and it was retained, which indicates that the psychological well-being of adolescents is determined by the interplay of ego resilience, anxiety and stress. Research conducted by Sharif Nia et al. (2021) found that resilience was a significant negative predictor of both state and trait anxiety, and that higher anxiety was associated with poorer psychological outcomes. This study is relevant to the current study findings as it explains that well-being is influenced by resilience and anxiety acting as opposing predictive forces.

5.1 Limitations and Future Suggestions

This study offers valuable insights into adolescent well-being, but it has some limitations. The sample size was small, which limits the ability to generalise findings or identify cause-and-effect relationships. Relying only on self-report measures may have affected the accuracy of responses. In addition, as this is a correlational study, causal inferences cannot be drawn from the findings; the observed associations may be influenced by unmeasured third variables, and

directionality cannot be established. Future research should use larger samples, follow adolescents over time, and explore factors like social support that may influence the relationship between resilience, stress, anxiety, and well-being.

6. Conclusion

The present study examined ego resilience, perceived stress, and state-trait anxiety and their relationship to the psychological well-being of adolescents. The findings suggested that adolescents with higher ego resilience tend to have better self-acceptance. However, statistically non-significant trend observed for positive relations with others. In contrast, higher levels of stress and anxiety were linked with lower well-being, particularly in areas such as self-confidence, environmental mastery, and positive relations with others. Further, psychological well-being is not influenced by any particular individual factors but is a composite outcome of resilience, stress, and anxiety working together. In addition, the results suggested that different dimensions of well-being are affected differently, emphasising the need of understanding adolescent mental health in a multidimensional way.

Conflict of Interest Statement

There is no financial/commercial conflicts of interest for conducting this research.

Ethical Declarations

All data used in this study was obtained and processed according to ethical and legal guidelines and the data collection was started after getting approval from Subject Research Committee of the Department of Psychology, Keshav Mahavidyalaya, University of Delhi.

References

1. Abd el-halem, M. H., Abd elaal, M. H., Ahemed, F. M., & Mahmoud, D. A. B. (2022). Relationship between ego resilience, perceived stress and life satisfaction among university students. Journal of Nursing Science, 3(2), 1053–1066. https://buescholar.bue.edu.eg/nursing/16.
2. Alessandri, G., Vecchio, G. M., Steca, P., Caprara, M. G., & Caprara, G. V. (2007). A revised version of Kremen and Block's ego-resiliency scale in an Italian sample. TPM: Testing, Psychometrics, Methodology in Applied Psychology, 14(3–4), 165–183.
3. Barbayannis, G., Bandari, M., Zheng, X., Baquerizo, H., Pecor, K. W., & Ming, X. (2022). Academic Stress and Mental Well-Being in College Students: Correlations, Affected Groups, and COVID-19. Frontiers in Psychology, 13, 886344. https://doi.org/10.3389/fpsyg.2022.886344
4. Block, J., & Kremen, A. M. (1996). IQ and ego-resiliency: Conceptual and empirical connections and separateness. Journal of Personality and Social Psychology, 70(2), 349–361. https://doi.org/10.1037/0022-3514.70.2.349
5. Chavanne, A. V., Paillère Martinot, M. L., Penttilä, J., Grimmer, Y., Conrod, P., Stringaris, A., Van Noort, B., Isensee, C., Becker, A., Banaschewski, T., Bokde, A. L. W., Desrivières, S., Flor, H., Grigis, A., Garavan, H., Gowland, P., Heinz, A., Brühl, R., Nees, F., … Whelan, R. (2023). Anxiety onset in adolescents: A machine-learning prediction. Molecular Psychiatry, 28(2), 639–646. https://doi.org/10.1038/s41380-022-01840-z
6. Chen, Q., Gao, W., Chen, B.-B., Kong, Y., Lu, L., & Yang, S. (2021). Ego-Resiliency and Perceived Social Support in Late Childhood: A Latent Growth Modeling Approach. International Journal of Environmental Research and Public Health, 18(6), 2978. https://doi.org/10.3390/ijerph18062978
7. Ciccarelli, S. K., & White, J. N. (2025). Psychology (7th ed.). Pearson.
8. Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. Journal of Health and Social Behavior, 24(4), 385–396.
9. Ebrahim, A. H., Dhahi, A., Husain, M. A., & Jahrami, H. (2022). The Psychological Well-Being of University Students amidst COVID-19 Pandemic: Scoping review, systematic review and meta-analysis. Sultan Qaboos University Medical Journal, 22(2), 179–197. https://doi.org/10.18295/squmj.6.2021.081
10. Elzohary, N. W., Mekhail, M. N., Hassan, N. I., & Menessy, R. F. M. (2017). Relationship Between Ego Resilience, Perceived Stress And Life Satisfaction Among Faculty Nursing Students. IOSR Journal of Nursing and Health Science (IOSR-JNHS), 6(6 Ver. IV). https://doi.org/10.9790/1959-06040494100
11. Fino, E., Morris, S. M., Stevenson, C., Shuttleworth, I., Finell, E., & Bjarnason, Þ. (2026). Personality and the "Social Cure": The Role of Ego-Resilience in the Social Identity Approach to Health. Social Psychological and Personality Science, 17(1), 103–119. https://doi.org/10.1177/19485506251325294
12. Impey, B., Gordon, R. P., & Baldwin, D. S. (2020). Anxiety disorders, post-traumatic stress disorder, and obsessive compulsive disorder. Medicine, 48(11), 717–723. https://doi.org/10.1016/j.mpmed.2020.08.005
13. Jaworska, N., & MacQueen, G. (2015). Adolescence as a unique developmental period. Journal of Psychiatry & Neuroscience: JPN, 40(5), 291–293. https://doi.org/10.1503/jpn.150268
14. Lazarus, R. S., & Folkman, S. (1984). Stress, appraisal, and coping. Springer Publishing Company.

15. Lee, D., & Lee, H. S. (2021). Examining associations of ego resilience, depression, stress, and the Stages of Motivational Readiness for Change (SMRC). Journal of Affective Disorders Reports, 5, 100178. https://doi.org/10.1016/j.jadr.2021.100178
16. Liu, Y., Beauparlant, E. T., Ueda, M., & Yu, Q. (2025). Psychological Well-Being and Distress Among College Students: The Mediating Role of Resilience as a Dynamic Process. Journal of Counseling & Development, 103(4), 508–521. https://doi.org/10.1002/jcad.12564
17. Padmanabhanunni, A., & Pretorius, T. B. (2025). The relationship between loneliness and internalizing disorders among young adults: The mediating and moderating role of ego-resilience. Frontiers in Psychiatry, 15, 1466173. https://doi.org/10.3389/fpsyt.2024.1466173
18. Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. Journal of Personality and Social Psychology, 57(6), 1069–1081. https://doi.org/10.1037/0022-3514.57.6.1069
19. Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. Journal of Personality and Social Psychology, 69(4), 719–727. https://doi.org/10.1037/0022-3514.69.4.719
20. Saviola, F., Pappaianni, E., Monti, A., Grecucci, A., Jovicich, J., & De Pisapia, N. (2020). Trait and state anxiety are mapped differently in the human brain. Scientific Reports, 10(1), 11112. https://doi.org/10.1038/s41598-020-68008-z
21. Sharif Nia, H., Akhlaghi, E., Torkian, S., Khosravi, V., Etesami, R., Froelicher, E. S., & Pahlevan Sharif, S. (2021). Predictors of Persistence of Anxiety, Hyperarousal Stress, and Resilience During the COVID-19 Epidemic: A National Study in Iran. Frontiers in Psychology, 12, 671124. https://doi.org/10.3389/fpsyg.2021.671124
22. Sinval, J., Oliveira, P., Novais, F., Almeida, C. M., & Telles-Correia, D. (2025). Exploring the impact of depression, anxiety, stress, academic engagement, and dropout intention on medical students' academic performance: A prospective study. Journal of Affective Disorders, 368, 665–673. https://doi.org/10.1016/j.jad.2024.09.116
23. Skapinakis, P. (2014). Spielberger State-Trait Anxiety Inventory. In A. C. Michalos (Ed.), Encyclopedia of Quality of Life and Well-Being Research (pp. 6261–6264). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_2825
24. Tyagi, A., & Khokkar, M. (2024). Ego Resilience in Women: The Influence of Caste on Psychological Adaptability (Pt. 2220-2227). International Journal of Indian Psychology, 12(4).
25. Wiglesworth, A., Butts, J., Carosella, K. A., Mirza, S., Papke, V., Bendezú, J. J., Klimes-Dougan, B., & Cullen, K. R. (2023). Stress system concordance as a predictor of longitudinal patterns of resilience in adolescence. Development and Psychopathology, 35(5), 2384–2401. https://doi.org/10.1017/S0954579423000731

❖ ❖ ❖

International Journal of Innovation and Multidisciplinary Research
DOI.ORG/10.67505/IJIAMR.v6.i1.61005
Vol. 6 (1), 2026
E-ISSN: 2583-4452

Design and Simulation Based Analysis of Electrical, Optical and Thermal Performance of 1550 nm Distributed Feedback Laser Diode

Ayushmita Singh[1], Surbhi Gupta[2] and Somna S Mahajan[3]
[1,2] Department of Electronics, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, Delhi
[3] Solid State Physics Laboratory (SSPL), Defence Research and Development Organization (DRDO), Delhi
[*] Corresponding Author ✉ ayushmita.singh@rajguru.du.ac.in

ABSTRACT

Distributed Feedback (DFB) laser diodes that operate at 1550 nm are key components in modern optical communication systems. Their spectral purity and wavelength stability make them ideal for high-capacity Fiber-optic links.The main goal of this study is to identify an optimized laser structure that maintains reliable single-mode operation under realistic conditions while achieving a lower threshold current.The optical analysis of the proposed DFB laser structure used Lumerical MODE and FDTD simulation tools. The optical results showed strong mode selectivity with lasing centred at the desired 1550 nm wavelength indicating effective single-mode performance. The electrical performance was studied with Lumerical CHARGE, focusing on carrier transport and recombination in the multilayer semiconductor structure. The analysis included current-voltage (I-V) characteristics and threshold current density estimation. Thermal characteristics of the device were evaluated using Lumerical HEAT simulations. The results emphasize the importance of thermal management in ensuring stable optical output and long-term device reliability. Overall, the optimized DFB laser design shows stable single-longitudinal-mode operation at 1550 nm, with a significantly reduced threshold current and better spectral purity than traditional structures. The integration of efficient carrier confinement and regulated thermal dissipation supports consistent device performance across a wide range of operating conditions. This work highlights the effectiveness of Lumerical's integrated multiphysics simulation framework for the organized design and improvement of DFB laser diodes. The proposed design approach and analytical method offer valuable insights for developing high-performance 1550 nm DFB lasers aimed at Dense Wavelength Division Multiplexing (DWDM) systems and next-generation optical communication networks.

Keywords: *Ansys Lumerical; Bragg grating; single-mode semiconductor lasers; 1550 nm wavelength.*

1. Introduction

Optical communication systems are crucial for transmitting data globally. The Distributed Feedback laser diode stands out because it works well with fibers. The 1550 nm wavelength is particularly suitable for long-distance and high-speed optical communication systems. It has the least distortion and is safe. A distributed-feedback laser (DFB laser) has its entire resonator designed as a periodic structure within the laser gain medium that acts as a distributed Bragg reflector at the laser wavelength. DFB lasers are a significant improvement over traditional Fabry–Perot lasers. In DFB lasers, a periodic Bragg grating runs along the active region or waveguide. This grating offers distributed optical feedback by reflecting light that meets the Bragg condition, enforcing a single longitudinal mode of operation[5][7]

A DFB laser diode primarily serves as the optical signal in long-distance, high-capacity optical communications. It finds use in various applications, including fiber sensors, 3D sensors, gas sensors, and disease diagnosis like respiratory and vascular monitoring. For gas sensing, it can function as a light source to detect methane leaks around factory pipes. The grating reduces unwanted side modes and significantly narrows the spectral linewidth. DFB lasers show high side mode suppression ratios, excellent wavelength stability, and low chirp. These features make them ideal for high-speed modulation and long-distance optical transmission[3][2]

However, quarter-wave-shifted DFB lasers face several performance challenges. These include high threshold current, temperature sensitivity, mode instability, and carrier recombination losses. Addressing these issues requires careful design that considers optical, electrical, and thermal effects together[1]

Received on: 09 Jun 2026 | Accepted on: 29 Jun 2026 | Published Online on: 25 July 2026

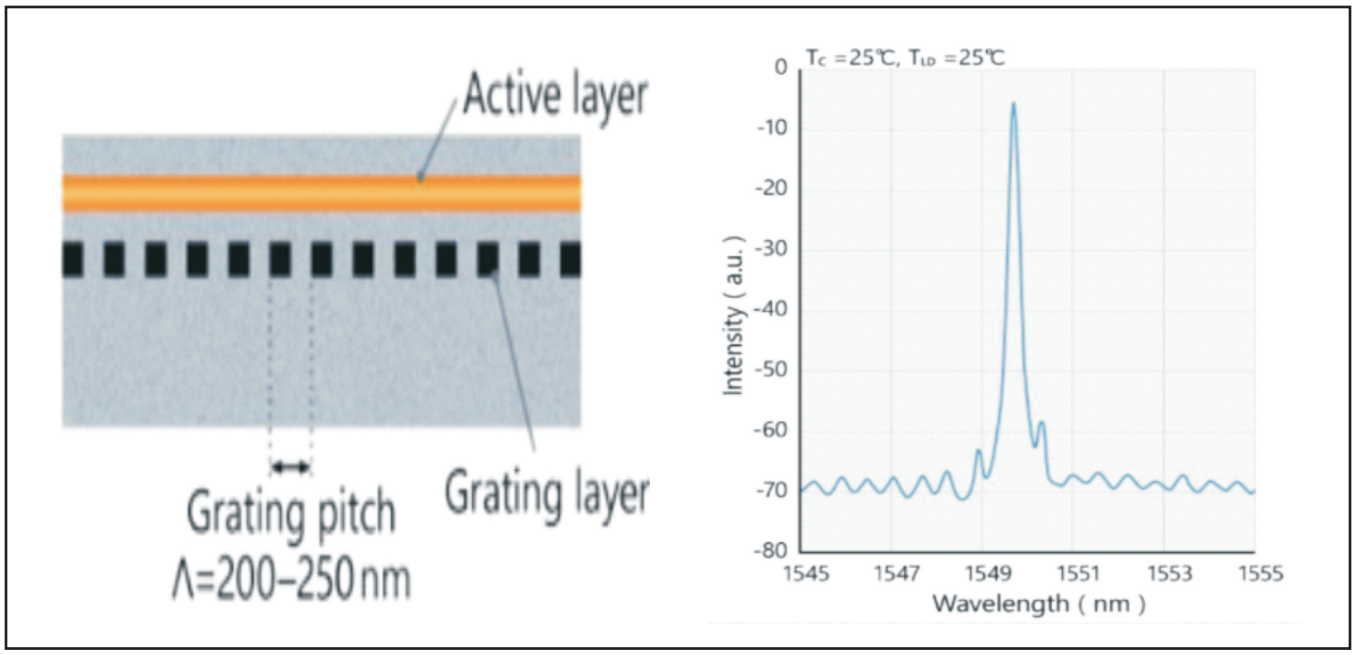


Fig. 1: *Example DFB-LD spectrum for 1550nm*

Despite their benefits, a number of limitations affect the performance of DFB lasers:

1. Large threshold current.
2. Temperature dependence.
3. Mode instability.
4. Carrier recombination losses.

It is important to realize that all these issues must be taken into account and optimised simultaneously within optical, electrical and thermal domains.

The DFB grating is characterized by a grating period Λ= 240–245 nm , designed to satisfy the Bragg condition for emission at the target wavelength. The coupling coefficient (κ) is 80–150 cm^{-1}, representing the index modulation strength of the periodic grating.

2. Materials

The InGaAs/InP material system was chosen for the proposed DFB laser due to its established suitability for 1550nm emission, good lattice matched growth on an InP substrate which provides low loss growth, and the availability of mature material parameter databases. InP has a direct band gap of 1.35 eV (~920nm). The lattice matched InGaAs alloy system can be used to produce bandgap energies from 0.75 eV (the lattice matched In0.53Ga0.47As alloy) to 1.35 eV (InP, where the emission wavelength is ~920nm), that is to say, the band gap wavelengths can be from ~920nm to 1650nm. The material for the quantum wells in the active region and targeting 1550nm emission was chosen to be compressively strained InGaAs with ~1% compressive strain. The use of compressive strain in quantum wells is to increase the differential gain through valence band engineering, where the heavier hole and lighter hole split apart, increasing the threshold carrier density and improving T0 characteristic temperature [9].

The cladding layers are made up of InP. This layer both provides optical confinement due to its low refractive index when compared with the SCH/active layers and also contains the p-n junction layers which carry the current. The cladding layer for carriers,de consists of P-doped InP (1.5∗10cm3 zinc doping) whereas the cladding layer for the n-side consists of n-doped InP (silicon doping at approximately 1.5∗10cm3). The p InGaAs contact layer (doped 1.5∗10cm3 give good low resistance ohmic contact to the metal electrode) is used to establish electrical connection to the p-contact layer of InP via the electrode.

3. Methodology

Simulation Platform Lumerical Multiphysics : All of the simulations were run within Lumerical Device Suite (2020 R2). The Device Suite is one interface to the optical, electrical and thermal solvers that exist within the system and these

solvers use a specific language within Lumerical to solve. The 4 main solvers that were used are: MODE Solutions, FDTD Solutions, CHARGE and HEAT. These solvers were all designed to solve for a specific domain of physics, but could be linked to all of the other solvers, receiving values from them and passing values to them. Within Lumerical data can be transferred between the optical, electrical and thermal simulators either through common material databases and/or by scripted parameter passing. [4][6]

A SiO-based optical waveguide is a structure that guides light by using a low-loss dielectric material (silicon oxide family) where light is confined due to a refractive index contrast between the core and surrounding cladding. [17]

- Core: SiO / SiO_2 (or doped silica)
- Cladding: lower-index oxide or air
- Guiding mechanism: total internal reflection

4. Observations and Results

The optical performance of the laser structure were simulated with the MODE and FDTD solver and the electrical performance were simulated with the CHARGE solver and the thermal performance were simulated with the HEAT solver. The wavelength was selected to be 1550 nm as this point experiences minimum loss and dispersion in the fiber optical cables.

Table 1: *Device Structure and Parameters*

Layer	Material	Thickness
Substrate	InP	2.0 μm
n-Cladding	InP	1.5 μm
Active Region	InGaAs	0.2 μm
p-Cladding	InP	1.5 μm
Contact Layer	InGaAs	0.1 μm

Overall height of the device: 5.3
Overall width of device:3μm

Table 2: *Doping Parameters*

Layer	Doping Type	Concentration
Substrate	n-type	1×10^{18}??^{3}
n-Cladding	n-type	1×10^{18}??^{3}
Active Region	lightly doped	1×10^{18}??^{3}
p-Cladding	p-type	1×10^{18}??^{3}
Contact Layer	p-type	5×10^{18}??^{3}

Table 3: *Material Properties*

Material	Bandgap	Thermal Conductivity
InP	1.34 eV	(68, W/mK)
InGaAs	0.75–0.8 eV	(5–10, W/mK)

The active region was modeled with a doping concentration of $1x10^{18}$ cm^3 to maintain numerical stability and represent the simulated device configuration. While practical DFB laser active regions are commonly intrinsic or unintentionally doped, the selected doping level was adopted within the simulation framework and does not alter the qualitative performance trends presented in this study.

Fig. 2 The material used for the substrate and the cladding layer was InP due to their higher bandgap energy, high stability and good lattice match to InGaAs. The overall electrical simulation was carried out under forward bias.

5. Conclusion

To conclude, electro-optical-thermal analysis, MODE, FDTD, CHARGE, and HEAT solvers are used to simulate thoroughly 1550 nm InP/InGaAs semiconductor laser and Si-based optical waveguide structure. The combined optical, electrical, electromagnetic and thermal simulation is a facile way to study the characteristic of the design photonic structure. [7][8]

The overall simulation result indicates the designed InP/InGaAs photonic structure exhibits favorable optical, electrical, electromagnetic, and thermal performance for the 1550nm application. In addition, it verifies the validity of chosen materials, geometric dimensions, doping concentration, Bragg grating Properties and thermal boundary conditions and therefore is favorable to the efficient laser device fabrication and it can be used for photonic integrated circuit [10][8].

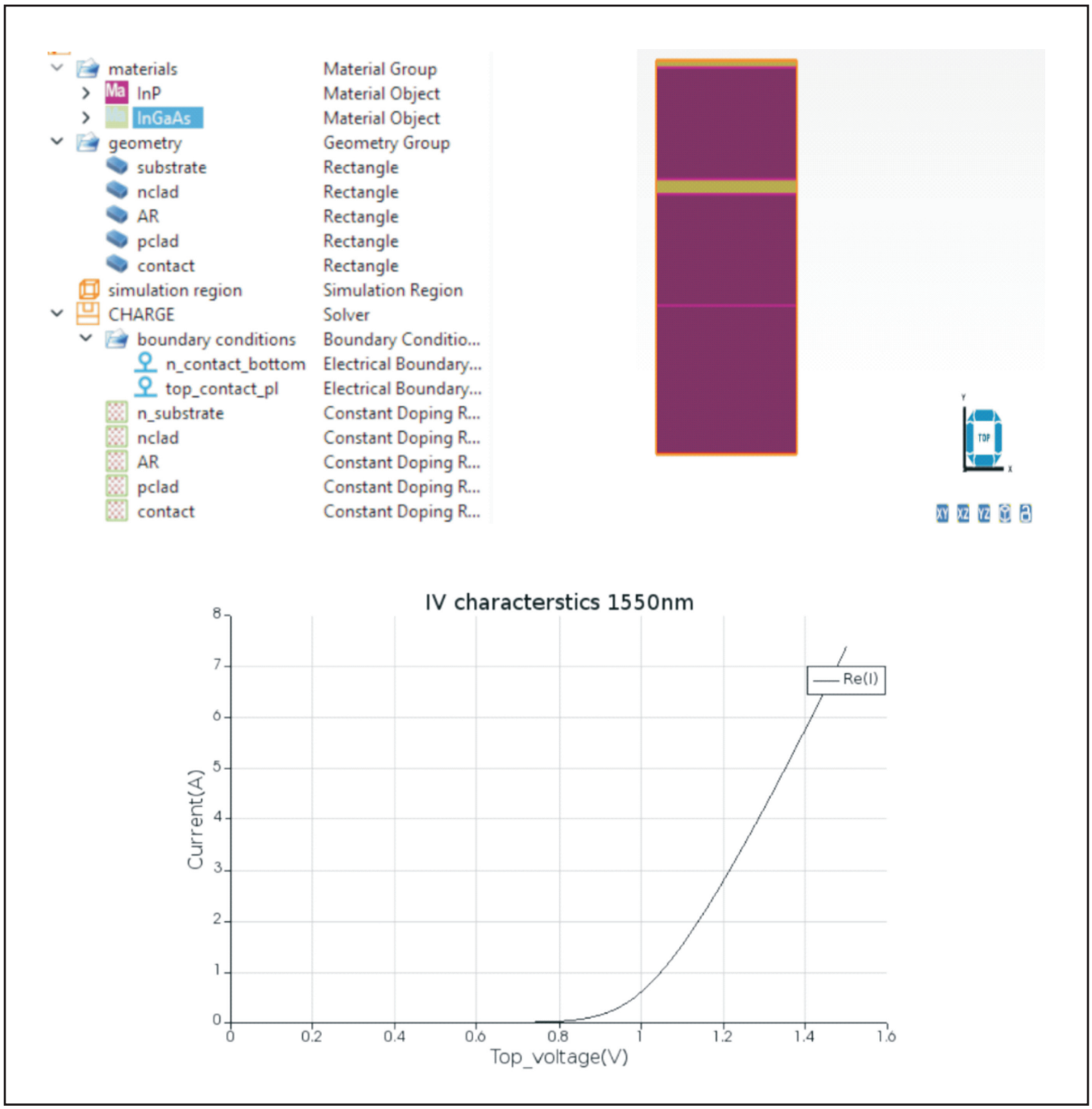


Fig. 2: *I-V Characteristics: The I-V curve plotted was an exponential curve after turn-on of between 0.8V-1.0V*

This structure can be employed to the following applications easily.

1. Optical fiber communication systems
2. Photonic integrated circuits
3. Distributed feedback laser systems
4. Optical filtering devices
5. Silicon photonics platforms
6. High speed optical interconnect

It lays foundation for the design of efficient and stable integrated photonics devices.

Acknowledgement

Author sincerely thanks to Mr. Aditya Singh for providing the consistent support, criticism, and direction throughout this research.

Conflict of Interest Statement

The author declares no financial or commercial conflict of interest related to this research.

Ethical Declarations

The authors declare that this research does not involve human participants, animals, or any procedures requiring ethical approval. All data used in this study were obtained and processed in accordance with applicable ethical and legal guidelines.

References

1. Utaka, K., Akiba, S., Sakai, K., & Matsushima, Y. (1986). /4-shifted InGaAsP/InP DFB lasers. IEEE Journal of Quantum Electronics, 22(7), 1042 1051. https://doi.org/10.1109/JQE.1986.1073089
2. Piprek, J. (2003). Semiconductor optoelectronic devices: Introduction to physics and simulation. Academic Press.
3. Ansys Lumerical. (2023). FDTD Solutions reference guide Ansys Inc.
4. Ansys Lumerical. (2023). MODE Solutions reference guide Ansys Inc.
5. Ansys Lumerical. (2023). CHARGE solver physics reference. Ansys Inc.
6. Ansys Lumerical. (2023). HEAT solver physics reference Ansys Inc.
7. Chuang, S. L. (2009). Physics of photonic devices (2nd.). John Wiley & Sons.
8. International Telecommunication Union. (2012). Recommendation ITU-T G.694.1: Spectral grids for WDM applications: DWDM frequency grid ITU.
9. Piprek, J., Abraham, P., & Bowers, J. E. (2000). Self-consistent analysis of high-temperature effects on strained-layer multi-quantum-well InGaAsP/InP lasers. IEEE Journal of Quantum Electronics, 36(3), 366–374. https://doi.org/ 10.1109/ 3.825885
10. Sands, D. (2005). Diode lasers. Institute of Physics Publishing.
11. Coldren, L. A., Corzine, S. W., & Mashanovitch, M. L. (2012). Diode lasers and photonic integrated circuits (2nd.). John Wiley & Sons.
12. Agrawal, G. P., & Dutta, N. K. (2013). Semiconductor lasers (2nd.). Springer.
13. Okamoto, K. (2021). Fundamentals of optical waveguides (3rd .). Academic Press.
14. Yariv, A., & Yeh, P. (2007). Photonics: Optical electronics in modern communications (6th ed.). Oxford University Press.
15. Saleh, B. E. A., & Teich, M. C. (2019). Fundamentals of photonics (3rd.). John Wiley & Sons.
16. Agrawal, G. P. (2021). Fiber-optic communication systems (5th ed.). John Wiley & Sons.
17. Akiba, S., Usami, M., & Utaka, K. (1987). 1.5- m /4-shifted InGaAsP/InP DFB lasers. Journal of Lightwave Technology, 5(11), 1564 1573. https://doi.org/10.1109/JLT.1987.1075453
18. Utaka, K., Akiba, S., Sakai, K., & Matsushima, Y. (1985). Longitudinal-mode behaviour of /4-shifted InGaAsP/InP DFB lasers. Electronics Letters, 21(9), 367 369. https://doi.org/10.1049/el:19850262

❖ ❖ ❖

Vol. 6 (1), 2026
E-ISSN: 2583-4452

Mathematical Modelling and Experimental Study of the Cooling Rate of a Hot Fluid in Different Types of Containers

Garima Singh[1*] and Shalvi Tripathi[2]
[1, 2] Department of Mathematics, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, Delhi
[*]Corresponding Author ✉ garima27062005@gmail.com

ABSTRACT

Cooling of heated fluids is one of the critical aspects of the heat transfer and find extensive applications in engineering, science, and everyday life. Newton's Law of Cooling is the mathematical relation governing the cooling process where the heated body cools down upon being immersed into a cooler environment. The present study concerns itself with the mathematical modelling and experimental study of cooling rate of a hot fluid in different containers. In this experiment, hot water is the fluid used and it is placed in containers made from different types of material like ceramic, glass, and steel. This experiment was carried out in the humidity chamber so that the temperature and relative humidity of the surroundings could be kept constant. Fluid temperature was measured at various intervals of time and cooling curves were drawn for different types of containers to determine the cooling characteristics. Data obtained from experiments were evaluated graphically and using log functions to obtain values of cooling constant for various material containers used in the experiment. The graphs revealed that there is almost linear relationship in log form, hence supporting the nature of exponential cooling as predicted by the mathematical function. The cooling constant value (k) for each container was found to vary slightly based on the material used. This study shows the practical application of mathematical modeling to explain physical cooling process and highlights the significance of validation through experimentation for applied mathematics and thermal analysis.

Keywords: *Thermal Conductivity, Humidity Controlled, Differential Equations, Temperature, Thermal Analysis*

1 Introduction

Heat transfer is the process where heat flows naturally at the rate that it is transferred from a hot object to a cold object. Whenever a hot object is placed in a cold environment, heat will be lost until both are at the same temperature level, which is known as thermal equilibrium [4]. This transfer of heat takes place through three main modes, i.e. Conduction, Convection and Radiation. Understanding these three modes is very important for studying cooling processes and forms the basis of Newton's Law of Cooling, which used in this research [3]. Conduction is described as the process by which heat is transferred in a substance without any movement of that substance itself. Conduction mainly occurs in solids. When heat is applied to one part of a solid, atoms start vibrating rapidly to pass on their energy to the adjacent atoms [21]. The procedure keeps on going through the atoms until heat travels through the material. If you heat one end of a metallic rod, eventually, the other end will become warm. Conduction happens at different rates for different materials. Steel and metals have fast conductivity while ceramics are slow in this process. Mathematically, Using Fourier's law it can be expressed as:

$$q = -k\frac{dT}{dx}, \tag{1}$$

where k refers the thermal conductivity of material. The value of k is higher its means heat transfer process will be faster [10]. Convection refers to heat transfer taking place due to fluid motion. In this case, the hot fluid moves up while the cold fluid moves down, thus creating a current of fluids referred to as convection currents. When water is boiled in a pot, the hotter water tends to rise while the cooler water sinks for heating. This process helps in distributing heat throughout the fluid. Convection can be of two types: natural convection (due to temperature differences) and forced convection (due to external forces like fans). In cooling processes, convection is the most important mode because heat from the hot object is carries away by surrounding air. Mathematically, This process is represented by:

$$\frac{dT}{dt} = -k(T - T_a), \tag{2}$$

Received on: 10 Jun 2026 | Accepted on: 28 Jun 2026 | Published Online on: 25 July 2026

which shows that cooling rate depends on difference between body and surroundings temperatures [1]. The process of heat transmission through electromagnetic waves is known as radiation. It does not need any medium to take place. Thus, heat can travel even when there is no medium between source and destination. The heat that we get from the Sun is transmitted to us through radiation. All bodies radiate heat and hot bodies radiate more heat than the cooler ones [26]. The loss of heat through radiation is calculated using the Stefan-Boltzmann Law:

$$Q = \sigma \epsilon A(T^4 - T_a^4), \tag{3}$$

where (σ) is a constant and (ϵ) depends on the surface of the object. Although radiation always occurs, its effect is usually smaller compared to conduction and convection in everyday cooling situations [17]. However, in real-life applications, all the above methods of heat transfer coexist. As an illustration, when cooling a hot fluid inside a container, the heat is initially conducted from the liquid to the walls of the container and then conveyed to the atmosphere by convection, with some loss of energy in the form of radiation. This process will depend on certain variables including the material of the container used, its surface area, temperature gradient, and ambient conditions such as air movement and moisture. From a mathematical point of view, heat transfer processes are classic examples of dynamic systems that can be described within the framework of differential equations. Among the earliest models for the process of heat transfer during cooling were those developed by Sir Isaac Newton in the early eighteenth century. According to the law of cooling enunciated by Newton, there is a relation between the change in temperature of an object with time and the temperature difference between the object and its surrounding medium. The process of cooling of hot fluids like water is quite common in our day-to-day life. Apart from that, in calculating the rate at which cooling takes place, there are some other physical factors apart from temperature that need to be taken into consideration. Some of them include the kind of material used for making the container in which the fluid is placed, the kind of contact surfaces it touches, and the nature of the fluid itself. Understanding this behavior is important in applications like food cooling, thermal management and industrial heat transfer systems.

2 Review of Literature

The discovery of Newton's Law of Cooling is known to be one of the very early and significant examples of applying mathematics in the field of natural science. This law was discovered by Sir Isaac Newton in 1701 and explains how the rate of cooling of a heated body depends on the temperature difference between the body and the surroundings Although quite simply formulated, it marked the beginning of the shift from purely qualitative analysis of nature to its quantitative modelling with mathematical tools [8]. In fact, by representing the temperature decrease of a heated object as its rate being proportional to temperature difference, Newton essentially created the basis of the differential equation as a tool for studying processes evolving in time [30]. Moreover, in addition to serving as a classical example of a differential equation, it represented the phenomenon of exponential decay. Interestingly enough, the equation remained unchanged and relevant even as scientists learned more about nature and physics in general and started distinguishing between three ways of heat transfer. During the development of thermodynamics and heat transfer theory in the 19th century [14]. The historical background and applications of Newton's Law of Cooling emphasize its importance in both theoretical and practical aspects of science. Since the development of the law by Newton and until now when the law is applied for practical purposes as well as theoretical modelling, it has been important to consider the basic ideas related to it and apply this knowledge effectively. Several researchers have studied cooling processes using theoretical and experimental approaches. Ferreira[9] investigated the impact of the surface exposure to heat and materials on cooling process. Wei et al.[28] managed to extend the use of cooling law to data migration processes in computing systems. Peker et al.[20] suggested new techniques for solving cooling equations analytically and presented their findings on new analytical techniques that help solve cooling equations efficiently. Jain[15] examined the situations in which law of cooling does not hold and the restrictions of this law under certain circumstances. Agayev et al.[1] used sensors to carry out cooling experiments accurately and found out new information about cooling process.Ambethkar and Basumatary [2] described the process of heat transfer in Newtonian fluid. Pandey et al.[19] examined the effect of coolant on the cooling process in engine oil by measuring the temperature with the help of an Arduino-based temperature sensor. It was realized that the presence of coolant results in faster cooling than natural cooling. This is important in showing how cooling systems play a role in cars and how useful it is to use Arduino in temperature sensing. Newton's Law of Cooling and its use for explaining how heat is lost from a body based on the temperature difference is explained by Winterton [29]. Arduino's application in physics experiments involving temperature dependency of resistance was illustrated by Sari and Kirindi [23], demonstrating its efficacy as an affordable and interactive educational device. Heat transfer with the help of electric fields and additive fluids is discussed by Saha et al.[22]. The use of Arduino in the design of the automatic temperature control system by Khaing et al. [16] shows how sensors and microcontrollers can be implemented to effectively monitor and regulate temperature. Measurement of convective heat transfer coefficient experimentally by

Conti et al.[7], allows one to quantify heat lost through body and surroundings, giving insight into convection heat transfer mechanisms. Silva et al.[25] utilize a smartphone for measuring cooling curves and recording data on heat loss in thermal experiments through the use of sensors available in smartphones. Schouten et al.[24] offer practical classroom experiments related to studying and minimizing evaporation and, thus, offer alternative means to study and analyze heat and mass transfer mechanisms. O'Schouten et al.[18] gives an insight into cooling law as well as the limitations of using it to describe the cooling of the object. Newton's Law of Cooling will be investigated experimentally using the Arduino tool provided by Galeriu [12], giving practical examples on how the data can be gathered to support the cooling phenomenon. The problematics surrounding uses Newton's Law of Cooling in teaching physics in today's environment are discussed by M.Vollmer[27], in this paper analyses Newton's Law in the light of history and relates it to other heat models.

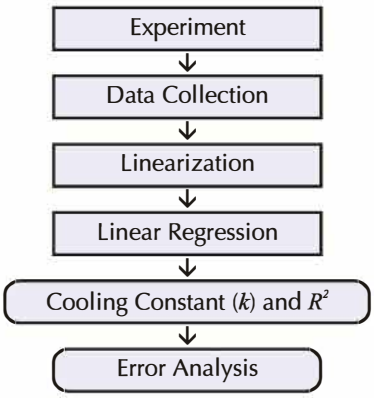
In the course of developing scientific knowledge, Cooling law was discovered not only as an abstract theory, but also as an empirical relation describing the dominating influences of convective heat transfer under experimental settings. Even though Newton's cooling law uses a constant proportionality constant, also called cooling constant, it was proved that this constant depended on other parameters related to physics such as the properties of the fluid around the body, the properties of the body itself, and the environment ([5], [21]). It was from this point on that further investigations led to more complicated models where the cooling constant became a function of time, thus creating nonlinearity in the model [11].

Although these improvements have been achieved, the Newtonian model is still largely applied due to its relatively simple analysis process and the ability to yield precise approximations under conditions with not too much variation in temperature [30].

Cooling law has a key application in modern sciences and engineering in a variety of different areas. The first major area of application is engineering where the law is commonly used for designing and analyzing cooling systems, including heat exchangers, thermal management in electronics, and industrial cooling, in which fast calculation of cooling is essential [6]. Environmental sciences use the law for modeling temperature change in natural systems, while biology and medicine use it for studying body temperature change and estimating time of death from a dead body's temperature [14].

3 Methods

Collection of data in this experiment started when all the set up parameters and environmental settings had been determined. The main reason why data was collected at this stage was that preliminary data of the variation of temperature with time in cooling of hot water using steel, glass and ceramic was needed. This was because of different thermal properties possessed by these substances. In this respect, each kind of material for the containers was examined three times. It is important to note that the equal amount of fluid at equal starting temperature was put into every tested container. The experiment was conducted at an ambient temperature of 20°C and a relative humidity of 50%. The container was placed in a humidity chamber where these conditions were maintained constant throughout the experiment. Temperature values were taken at specific time intervals until the fluid temperature reached the value of ambient temperature. The average of the three trials for each container was later used for analysis to minimize random experimental errors.



3.1 Materials and Apparatus

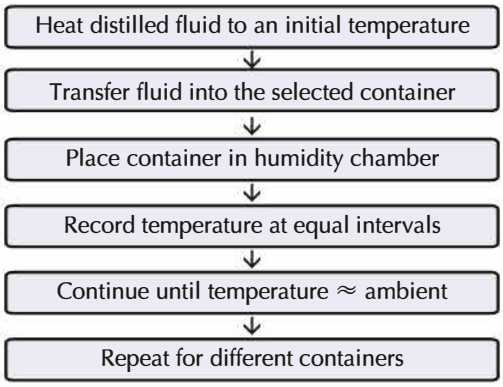
Following apparatus and materials were required for performing the experiment:

- Distilled water
- Symmetrical cylindrical containers made of different material including stainless steel, glass, and ceramic

- Calibrated laboratory thermometer
- Measuring cylinder to measure volumes
- Stopwatch or timer
- Humidity-controlled chamber
- Laboratory stand and holder
- Heating material like electric heating or gas burner

Before experimentation, each apparatus was examined to confirm its functionality and measurement precision. Hot water served as working fluid.

3.2 Experimental Process



3.2.1 Mathematical Modelling

The process of cooling the fluid can be explained using Cooling law, which states that the rate at which the temperature changes with respect to the body is directly proportional to the temperature difference between the body and its surroundings. Mathematically, Newton's Law of Cooling is given as:

$$\frac{dT}{dt} = -k(T - T_a), \tag{4}$$

where T refers to the temperature of the body at any time t, T_a refers to ambient temperature of the surrounding, and k refers to cooling constant.

This equation is a first order linear differential equation which will be the base of our mathematical model. Using the separation of variable, we will solve this equation, and the solution will be as follows:

$$T(t) = T_a + (T_0 - T_a)e^{-kt}, \tag{5}$$

where T_0 refers to the initial temperature of the body at $t = 0$.

This solution represents an exponential decay, which implies a rapid cooling process when the temperature difference is high and slow as equilibrium is achieved. This mathematical relationship has been validated through experiments in several research studies.

By taking the natural log of the equation above, we obtain the following linear relationship:

$$\ln(T - T_a) = \ln(T_0 - T_a) - kt. \tag{6}$$

This linear equation helps in finding the cooling constant through regression[13].

3.2.2 Data Collection and Analysis

Temperatures were recorded at intervals and presented in the form of tables for each container separately. Data obtained was studied in terms of their reliability and consistency. Verification of the mathematical model involved the presentation of data related to temperatures on a logarithmic scale. Graphs showing $ln(T - T_a)$ versus time were constructed. The regression equation was constructed using linear regression analysis.

Slope of the regression line equals the cooling constant multiplied by minus one ($-k$), while the coefficient of determination (R^2) determined its validity.

3.2.3 Error Minimization and Assumptions

For improving the accuracy of results attained through testing, the following was done:

- Conducting the experiment in controlled humidity conditions;
- Measuring the temperature at regular intervals;
- Conducting the experiment several times and averaging the data;
- Minimizing external factors like air currents.

Assumptions made included:

- Ambient temperature was constant;
- Fluid temperature was constant;
- Conformity of heat dissipation with the Cooling law.

4 Result and Discussion

Results obtained through this experiment are derived through observing the process of cooling in a heated liquid placed in ceramic, glass, and steel containers. Experimental data is gathered through temperature readings after particular intervals of time, and mathematical calculations are done using this information. Through these results, we can clearly see that our theory based on Cooling law works and prove the differences in cooling rates of the chosen materials.

4.1 Readings

The main data that have been recorded during experiments includes temperatures that have been recorded in various containers at fixed periods of time. The starting temperature of fluid was maintained approx the same for all three cases to ensure consistency and comparability of results. The ambient temperature was also recorded and kept nearly constant throughout the experiment. Table 1 represent the pilot data taken during the preliminary stage of experiment. The tables give an idea of the nature and consistency of the collected data. The temperature of the fluid was observed to decrease gradually with time in all three containers. During the initial stage of the experiment, the gap between the temperature of the liquid and the surrounding temperature was big and this showed by the sharp decline in temperature. With the passing of time, the temperature started falling less rapidly and gradually reached an asymptotic temperature level.

Although the overall trend was similar for all containers, noticeable differences were observed in the rate of cooling. The steel container showed a faster decrease in temperature compared to glass and ceramic, while the ceramic container retained heat for a longer duration. The behavior of the glass container was found to be intermediate.

4.2 Cooling Curves

In Figure 1 Cooling curves were plotted for each containers made of steel, glass, and ceramic, using the temperature-time data from Table 1 that shows how the temperature of a hot fluid decreases with time. In each case, the temperature is represented on the y-axis, and the time is represented on the x-axis. This makes the cooling process very easy to understand visually and can assist in verifying the Cooling law. The graph used to plot cooling is not linear but exponential in nature and decreasing. In the initial stages of cooling, when the hot fluid temperature is much greater than the surrounding, the rate of cooling is very high. But as the temperature difference reduces, the rate of cooling also reduces, and curve flattens out asymptotically towards ambient temperature.

The cooling curves indicate a steady fall in temperature over time. This indicates an exponential trend and confirms the Cooling law. There was differences in the slope of curves, which indicated that cooling rate depends on the material of container.It means the rate of cooling in a steel container was much faster than in glass, whereas comparatively a ceramic container has shown slow cooling behaviour.

Key Features of the Cooling Curves are:

Initial Steep Region

At the start, the curve is steep because temperature difference ($T - T_a$) is maximum. According to Cooling law, the rate of cooling depends on the difference in temperature, and thus the initial cooling process will be faster.

Gradual Slope Decrease

The fluid will become cooler with time, meaning that the temperature difference becomes smaller, resulting in the slowing of the cooling process.

Final Flattening Region

At the latter stage, fluid temperature is closer to the surrounding temperature. It cools down very slowly, and the graph begins to flatten.

Table 1: *Average temperature changes of fluid over time in different types of containers*

Time (in min)	Ceramic (in °C)	Steel (in °C)	Glass (in °C)
0 (initial)	**90.0 (initial)**	**90.0 (initial)**	**90.0 (initial)**
2	81.0	79.0	81.0
4	72.5	70.0	72.0
6	66.5	63.0	65.0
8	61.0	58.0	59.0
10	57.5	53.5	56.0
12	54.5	49.5	53.0
14	51.5	46.0	50.0
16	49.0	44.0	48.0
18	47.0	42.0	46.0
20	45.0	40.5	44.5
22	43.0	38.0	43.0
24	41.5	37.0	41.0
26	40.0	36.0	39.0
28	38.5	35.0	37.0
30	37.0	33.5	36.0
32	35.5	32.0	35.0
34	34.0	31.5	33.5
36	33.0	31.0	32.0
38	32.5	30.0	31.5
40	32.0	29.5	31.0

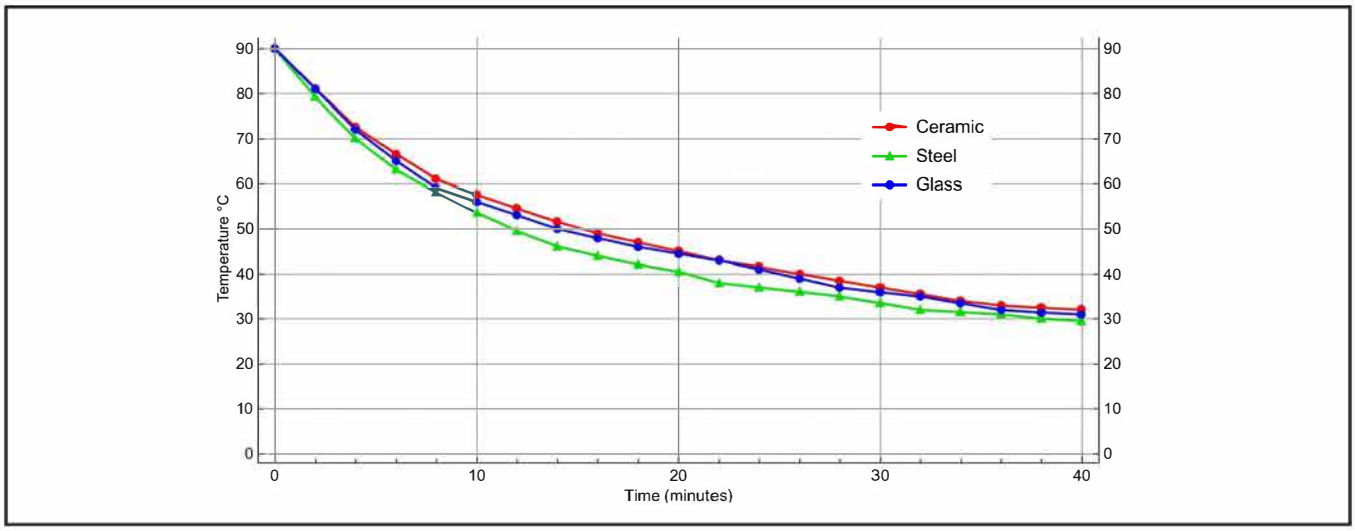


Fig. 1: *Cooling curves for ceramic, glass, and steel containers.*

Asymptotic Nature

The graph tends towards the ambient temperature line but does not touch it at any point of time in finite time interval. This is called asymptotic behavior.

Interpretation of Cooling Curves for Different types of Containers

The graphs for cooling rates with respect to the materials used for the containers (ceramic, glass, and steel) differ in the following way:

Steel Container: This graph is steep since it denotes rapid cooling due to high thermal conductivity.

Glass Container: The graph shows moderate steepness because of moderate cooling.

Ceramic Container: This graph is not so steep since it depicts slower cooling rates.

Thus, comparing cooling curves helps in understanding how material properties affect heat transfer.

4.3 Logarithmic Transformation of Data

The data for temperature and time is converted to a logarithmic scale and shown in Figure 2. The cooling process described in terms of the temperature versus time plot takes on an exponential nature, but one cannot easily find the cooling rate in the exponential cooling graph. This is why we use a logarithmic scale to obtain a linear relation.

Starting from the mathematical model:

$$T(t) = T_a + (T_0 - T_a)e^{-kt}, \tag{7}$$

we arrange it as:

$$T - T_a = (T_0 - T_a)e^{-kt}. \tag{8}$$

Taking natural logarithm on both sides:

$$ln(T - T_a) = ln(T_0 - T_a) - kt, \tag{9}$$

This equation represents a straight line of the form:

$$y = mx + c, \tag{10}$$

where $y = ln(T - T_a)$, $x = t$, slope $m = -k$ and intercept $c = (T_0 - T_a)$.

Therefore, the log graph plot is produced by plotting $\ln(T - T_a)$ along the vertical axis versus time t along the horizontal axis. In the case where the experimental data satisfies Cooling law, the plotted points follow a nearly straight line of negative slope. This shows that the relationship between the temperature and the time has a linear relationship.

However, in practical experiments, the plotted points do not lie exactly on a straight line due to small experimental errors such as fluctuations in ambient conditions, limitations of measuring instruments, and observational inaccuracies. In order to resolve this issue and obtain a more accurate approximation of the graphs trend, the best-fit line is drawn through

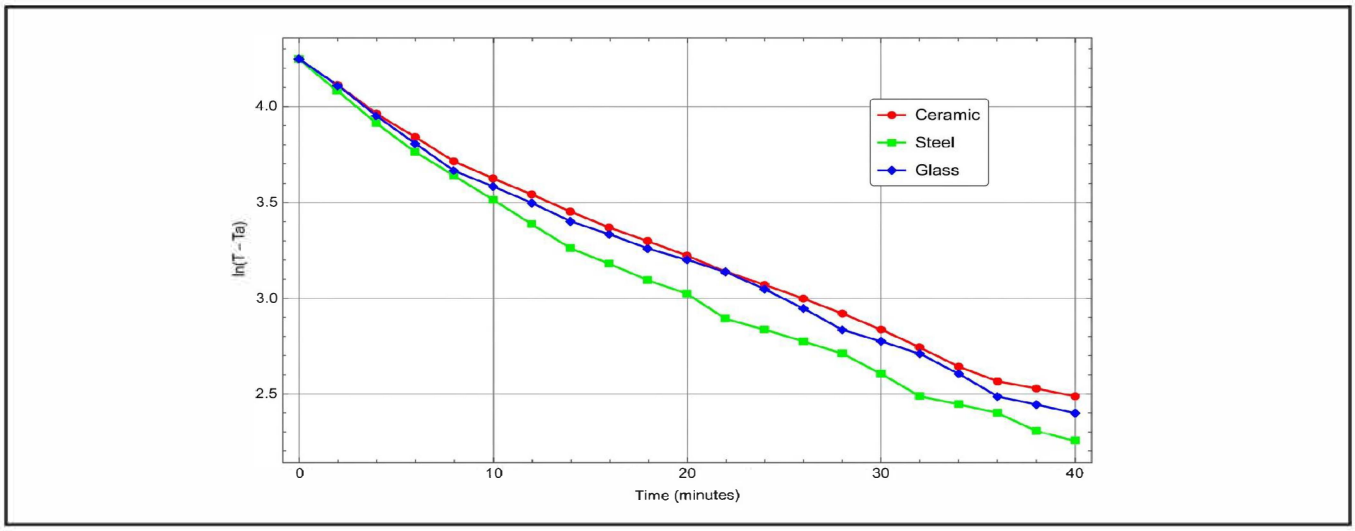


Fig. 2: *Log-linear plots of different types of contaniners*

the data points. Best fit line refers to a line that gives the overall description of the general trend of the data points, so that the total sum of all differences between the points and the line is the smallest possible. The simplest way of constructing the best fit line is the least squares approach that seeks to minimize the sum of the squares of errors or differences between the points and the line. The importance of using the best fit line is that we can estimate the value of k based on the slope of the graph as follows:

$$k = -(\text{slope of best fit line}). \tag{11}$$

The bigger the negative slope, the larger k, and the faster cooling we have, conversely if the negative slope is smaller, it means a slower rate of cooling. The y-intercept stands for $ln(T_0 - T_a)$. In this study, individual graphs of logarithms are generated for ceramic, glass, and steel containers. The graphs that have been fitted perfectly to the data sets for each type of container possess varying gradients because of the variation in their cooling rates. The container made of steel has the steepest gradient, followed by that of glass and finally, ceramic which has the least gradient. The combination of the logarithmic plot and best fit line makes a very good approach to the analysis of cooling. The reason for this is that while the log plot guarantees linear treatment of an exponential relationship, the best fit line helps in reducing the experimental errors associated with the experiment. In essence, this approach is critical to confirming the predictions made theoretically and determining the value of exact k.

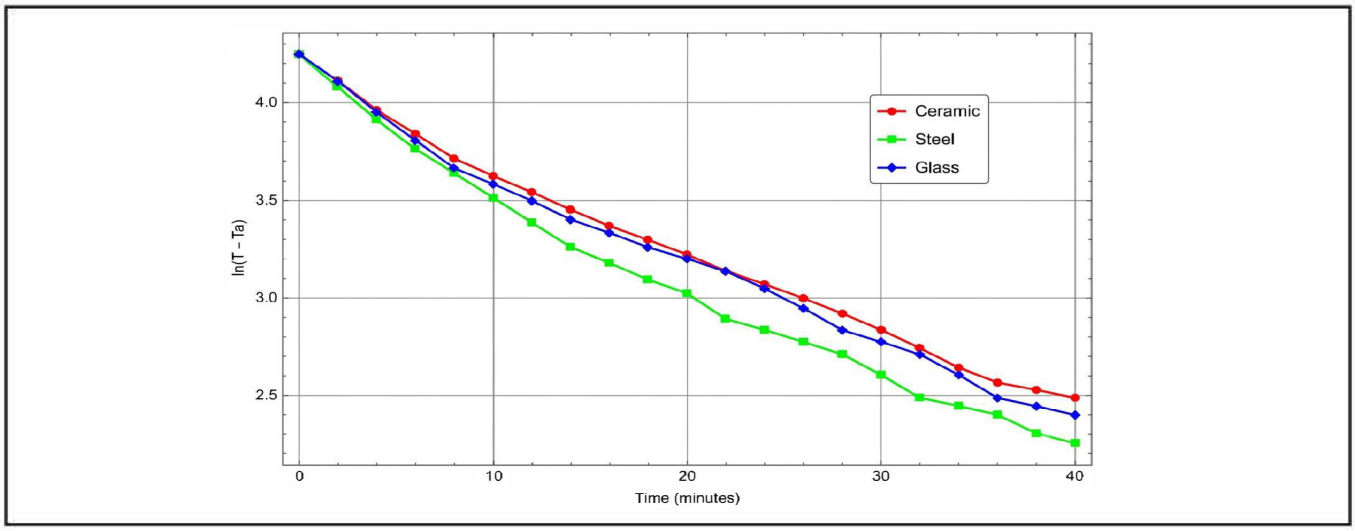


Fig. 3: *Best fit plot of different types of containers*

4.4 Comparative Analysis

Comparative study of cooling properties of various container materials is conducted by logarithmic graph $ln(T - T_a)$ versus time with the help of calculated cooling constant value. Such comparative study is based on the principles of Cooling law and implies a linear relationship between $ln(T - T_a)$ and time with k as the measure for comparison. From the experiment, the k of three materials were determined as:

Table 2: *Cooling Constant*

Container Material	Cooling Constant (k)	Container Material	Cooling Constant (k)	Container Material	Cooling Constant (k)
Ceramic	0.0427781	Steel	0.0479248	Glass	0.0442795

This means that the cooling rate depends greatly on the type of material from which the container is made. From the graph plotted on log scales, we observe that all the three data sets make nearly straight lines with negative slopes, thereby validating the application of Cooling law to our experiment. Nevertheless, lines have varying slopes because the cooling constants are not the same. From this graph, we notice that the value of $-k$ determines the steepness of the slope, which in turn influences the cooling process. Of the three containers, the steel container has the highest slope on the log graph and hence the highest cooling constant. The cooling rate in this case is the highest because heat is transferred from the liquid in the container more rapidly than in the other two containers. This is due to the high thermal conductivity of steel. The glass material exhibits an intermediate slope, thus having a cooling constant between those of the other materials considered. This gives glass cools slower as compared to steel but faster than ceramic. Glass has a thermal conductivity

rate lower than that of steel but greater than ceramic. The container that is made of ceramic material has the least slope value, implying that it has the lowest cooling constant and hence the slowest cooling rate. Ceramics are good insulators, meaning that they limit heat transfer, and hence the fluid in them takes more time to lose heat.

The fact that the data points lie on the straight lines in the graph also indicates that the error in the experiment is low and the assumptions made for the mathematical model are quite valid. The deviations observed in the straight lines could possibly be due to environmental changes.

It can be clearly seen from the comparison that the fastest cooling occurs in steel, then in glass, and the slowest cooling occurs in ceramic. The determined series relates to the thermal conductivity of the material, which is theoretically correct. The selection of the logarithmic relationship and the best fit line not only prove that the cooling occurs exponentially but also assist in finding out cooling rate.

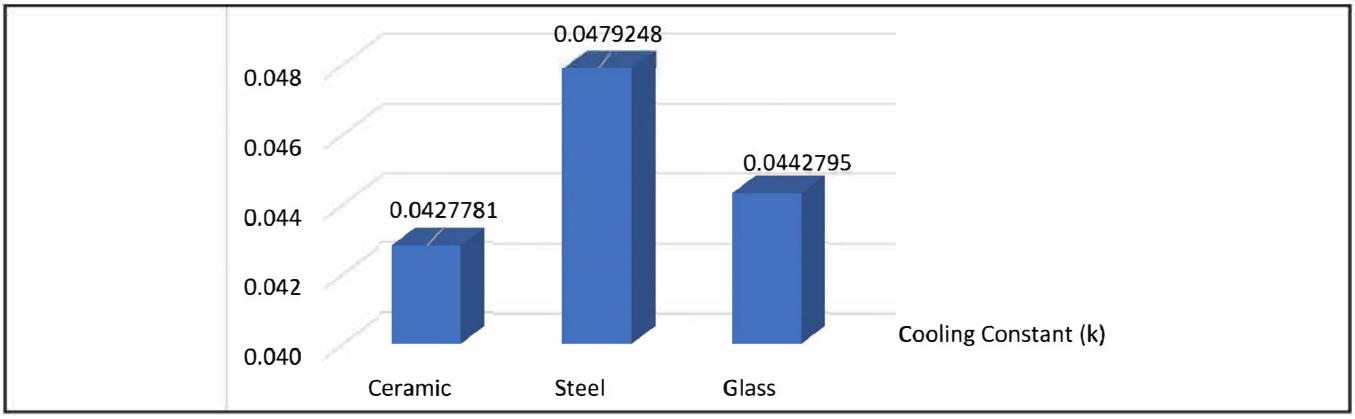


Fig. 4: *Cooling Constant for different types of containers*

4.5 Coefficient of Determination (R^2) and Error Calculation

In every experimental study that incorporates mathematical modeling, it becomes important not only to obtain the results but also to confirm their accuracy and reliability. The current study has used the principles laid out by Cooling law for describing the cooling of a heated fluid. This principle, once written mathematically and represented in the form of a logarithm, gives rise to a linear relationship between the variables $ln(T - T_a)$ and time. This enables us to apply certain statistical methods to verify the validity of our model as well as establish the accuracy of the results obtained experimentally. The most important of these methods include the coefficient of determination R^2 and error analysis.

The coefficient of determination, represented by R^2, is a mathematical statistic that assesses how accurately the regression or best-fit line reflects the actual observations. The coefficient of determination serves to mathematically quantify the percentage of variance between variables that can be attributed to one another. In layman's terms, the coefficient of determination reflects the accuracy of the experimental points relative to the theoretical straight-line model. Mathematically, it is defined as:

$$R^2 = 1 - \frac{\sum(y_i - y_{\text{fit}})^2}{\sum(y_i - \bar{y})^2}, \tag{12}$$

where y_i is the observed values, y_{fit} is the values predicted by the regression line and $\bar{y}$ is mean the observed values. The numerator in the above expression is the unexplained variance, while the denominator is the total variation. This means that R^2 basically tells you what percentage of the total variation has been explained by your model.

The R^2 value ranges from 0 to 1, and it is not complicated to understand. If $R^2 = 1$, it means the model fits perfectly, all the data points fall precisely on the regression line. On the other hand, a value close to zero indicates that the model does not adequately explain the data. In practical experimental situations, R^2 values above 0.95 are usually taken as an indication of a very good fit. For the three materials tested in this experiment, the R^2 values obtained are 0.989871 for the ceramic container, 0.978997 for the steel container, and 0.988545 for the glass container. The R^2 values obtained are very close to one, showing that there is a good linear correlation between the two parameters when put in log scale.

Table 3: *Coefficient of determination (R^2)*

Container Material	R^2 value	Container Material	R^2 value	Container Material	R^2 value
Ceramic	0.989871	Steel	0.978997	Glass	0.988545

A careful analysis of these values will show that R^2 of the ceramic container is higher than that of the glass container. It should also be noted that the steel container has the lowest R^2. Even though this difference is small, it can be justified by the characteristics of the material in question. Steel has higher thermal conductivity. As a result, the changes in the temperature happen so quickly that even tiny deviations from the ideal measurements can lead to errors, and, thus, higher deviation from the linear regression curve. On the other hand, ceramics do not conduct heat well. Therefore, the substance inside cools slowly, leading to better readings. The same trend applies to glass. However, the thermal conductivity of glass is intermediate, and thus its cooling speed lies somewhere in between the two materials mentioned before. Nevertheless, all of the three materials have R^2 values that prove how well the theory matches the data.

While the coefficient of determination examine the goodness of fit, it does not provide information about the accuracy of the estimated parameters, such as the cooling constant k. For this purpose, error analysis is carried out. In any experimental setup, it is impossible to eliminate all sources of error, and therefore it becomes necessary to quantify the uncertainty associated with the measured values. The k is obtained from slope the best fit line in the log plot which is subject to uncertainty due to various factors such as instrumental limitations, environmental conditions, and human observation.

The uncertainty in k is represented by Δk, which indicates the possible deviation of the measured value from the true value. To express this uncertainty in a meaningful way, the percentage error is calculated by:

$$\text{Percentage Error} = \left(\frac{\Delta k}{k}\right) \times 100. \tag{13}$$

This provides a relative measure of the accuracy of the experimental result in Table 4 and Table 5.

Table 4: *Uncertainty in the k (Δk)*

Container Material	(Δk)	Container Material	(Δk)	Container Material	(Δk)
Ceramic	0.000992757	Steel	0.00161039	Glass	0.0010935

Table 5: *Percentage error*

Container Material	Percentage Error	Container Material	Percentage Error	Container Material	Percentage Error
Ceramic	2.32071	Steel	3.36025	Glass	2.46955

In this study, the percentage errors for the three materials are small, ranging from around 2.3% to 3.6%. Small percentage errors show that the experiment was performed accurately, and also that the method adopted to determine the cooling constant is accurate. As seen before, the higher percentage error recorded for steel is due to the quick cooling of steel, making it difficult to record temperatures in the exact intervals. Ceramic takes time to cool down, hence recording of temperatures becomes more accurate, leading to smaller percentage errors.

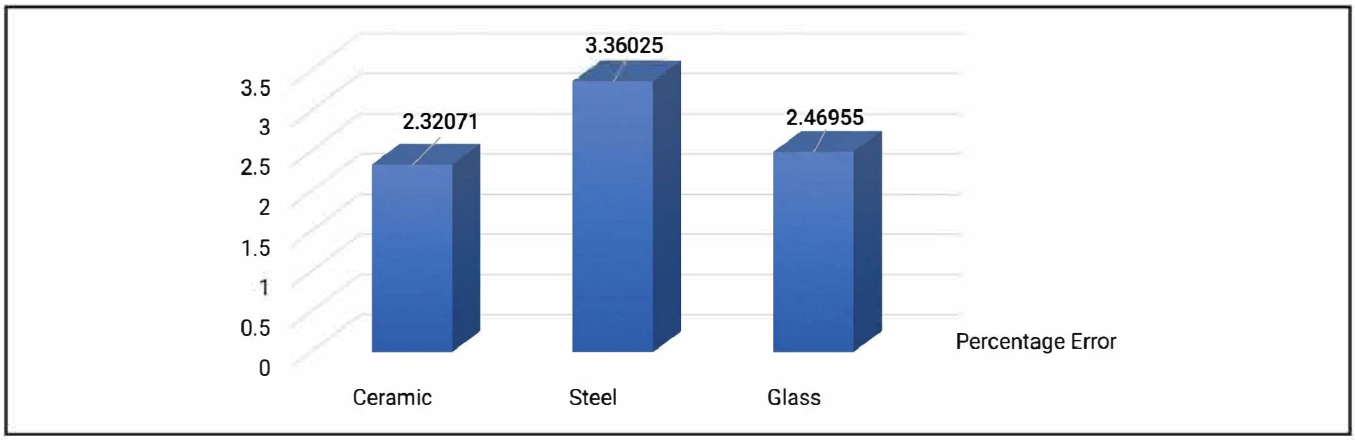


Fig. 5: *Percentage error for different types of contaniners*

Error sources in the experiment can be categorized into instrumental, observational, and environmental types. Instrumental error sources include errors that originate from the lack of accuracy of the thermometer employed in

determining temperature. If the thermometer least count is 1°C, all the readings will have uncertainties of ±0.5°C. Human error sources include errors that are caused by human incapability, such as delay in observing readings of the thermometer and time. Environmental error sources include errors that may arise due to changes in external conditions like ambient temperature, airflow, and moisture levels. Such environmental factors may influence the process of cooling and hence lead to deviations from the expected results.

In spite of all the above possible sources of error, the general reliability of this experiment is quite high, which can be seen from the percentage errors being very small and high R^2 values. The process of using logarithmic transformation along with fitting a best fit line will reduce the influence of random errors and thus make a better estimation of the cooling constant. Moreover, conducting the experiment under controlled conditions, where the ambient temperature is kept constant and other disturbances are minimal, means that the difference in observation comes from the material properties.

The analysis of the coefficient of determination and percentage error gives an overall assessment of the experiment. High values of R^2 show that the data is indeed following a linear trend after applying the logarithm operation to it. As a result, a relatively small percentage error shows that the experimentally determined constants of cooling are accurate enough. In conclusion, our study demonstrated high agreement between the theoretical prediction and experiment results, which shows the effectiveness of the chosen mathematical model.

5 Conclusion

Purpose of this study is to examine the cooling of a heated substance and prove the Newton's law of cooling through both calculations and practical experiments. Experiment involved observing the relationship between the rate of cooling and time for heated liquids in containers made of substances such as ceramic, glass, and metal. The results were analyzed using graphs and statistical tools. It became very evident through the experiments that the temperature of the liquid decreases constantly with respect to time, following an exponential trend. The phenomenon is completely in line with the mathematical theory which states that the rate of cooling depends upon the difference of temperature between the liquid and its surroundings.

The graphs obtained with respect to the relationship between temperature and time followed the conventional trend where temperature decreases rapidly initially and then slowly approaches the temperature of the surroundings. The logarithmic analysis of the data that was gathered became significant in confirming the theoretical model. Graphs plotted between natural log of $(T - T_a)$ versus time produced graphs with linear characteristics on all three containers tested. This confirmed that the rate of cooling is indeed exponential. The gradients of the graphs produced the cooling constants. It should also be seen from the above results that the highest cooling constant was achieved for the steel container and the second highest was achieved for the glass container, whereas the lowest cooling constant was achieved for the ceramic container. This means that the maximum heat loss was obtained from the steel container, whereas the least heat loss was obtained from the ceramic container. This happens due to high thermal conductivity of steel and poor heat conductivity of ceramics. The comparative analysis clearly demonstrates the influence of material properties on the rate of heat transfer. The validity and reliability of the data that were acquired during the experiment were confirmed by the high values of the coefficient of determination (R^2), which was almost equal to 1 in all cases. This means that the experimental data were highly consistent with the theoretical model.

Furthermore, there were small errors in the calculation of the cooling constant that indicates the precision of the experimental process. Although there were some minor variations observed in the results, these could be attributed to various limitations in conducting the experiment, including measurement inaccuracies, changes in the environment, and human response time when recording data. However, these errors were within acceptable levels and had minimal impact on the findings. Finally, the experiment verifies the applicability and accuracy of Cooling law and showing its real world applications. It shows that there is an observable correlation between mathematical principles and physical observations, thus, emphasizing the significance of differential equations in describing heat transfer processes. It further proves the significance of material property in cooling process. Indeed, this is relevant in many fields including engineering and thermal management among others. Overall, this experiment shows that there is much truth in using mathematical models in predicting physical events in the real world.

Acknowledgements

The authors sincerely thank Dr. Sanjiv Kumar for his guidance and support. The authors also thank the Department of Mathematics and Department of Food Technology of Shaheed Rajguru College of Applied Sciences for Women, University of Delhi for providing the necessary facilities and resources for this study.

Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical Declarations

There was no involvement of human subjects or animals in this study, and there are no conflict of interest exists.

References

1. Agayev, R., Charyyev, M., & Allyyev, M. (2025). *Newton's law of cooling experiment set using temperature sensor (hardware part)*. In *XXI Century: International Scientific Journal* (Vol. 37, pp. 472–476). ISSN 2782-4365.
2. Ambethkar, Vusala and Basumatary, Lakshmi Rani.(2024).*Insight into the dynamics of a Newtonian fluid through a rectangular domain with an emphasis on heat transfer*.Heat Transfer.53(1), 158-176.
3. A´rp´ad, I. W., Kiss, J. T., & Kocsis, D. (2024). Role of the volume-specific surface area in heat transfer objects: A critical thinking-based investigation of Newton's law of cooling. *International Journal of Heat and Mass Transfer*, 227, 125535.
4. Bergman, T. L. (2011). *Fundamentals of heat and mass transfer*. John Wiley & Sons.
5. Cengel, Y. A., & Ghajar, A. J. (2015). Heat and mass transfer: fundamentals and applications. *(No Title)*.
6. Cheng, K. C., & Fujii, T. (1998). Heat in history Isaac Newton and heat transfer. *Heat transfer engineering*, 19(4), 9-21.
7. Conti, R., Gallitto, A. A., & Fiordilino, E. (2014). Measurement of the convective heat-transfer coefficient. *The Physics Teacher*, 52(2), 109-111.
8. Erwin, K. (2024). Advanced Engineering Mathematics 10th Edition.
9. Ferreira, J. P. M. (2024). How a soup bowl and a coffee cup cool down. *Physics Education*, 59(5), 055020.
10. Fourier, J. (1878). *The analytical theory of heat*. Cambridge university press.
11. Fowler, A. C. (1997). *Mathematical models in the applied sciences* (Vol. 17). Cambridge University Press.
12. Galeriu, C. (2018). An Arduino investigation of Newton's law of cooling. *The Physics Teacher*, 56(9), 618-620.
13. Giordano, F. R., Weir, M. D., & Fox, W. P. (1985). *A first course in mathematical modeling*.
14. Holman, J. P. (2010). Heat transfer, 10th editi. ed. *Mc-GrawHill Higher education*.
15. Jain, D. (2024). Evaluation of Newtonian Cooling. *International Journal of Advanced Engineering, Management and Science*, 10, 5.
16. Khaing, K. K., Srujan Raju, K., Sinha, G. R., & Swe, W. Y. (2020, March). Automatic temperature control system using arduino. In *Proceedings of the Third International Conference on Computa-tional Intelligence and Informatics: ICCII 2018* (pp. 219-226). Singapore: Springer Singapore.
17. Nyambuya, G. G. (2017). On the Fundamental Theoretical Foundations of Newton's Law of Cooling.
18. O'Sullivan, C. T. (1990). Newton's law of cooling – A critical assessment. *Am. J. Phys*, 58(10), 956-960.
19. Pandey, A. K., Sharma, R., Nikhil, Choudhary, D., Bali, N. B., Verma, M., ... & Tanwar, A. (2024). Investigating the effect of coolant on cooling rate of engine oil used in automobile industry using Arduino interfaced temperature sensor. *Physics Education*, 59(2), 025026.
20. Peker, B., C¸ uha, F. A., & Peker, H. A. (2024). Analytical solution of Newton's Law of cooling equation via Kashuri Fundo transform. *Necmettin Erbakan U¨niversitesi Fen ve Mu¨hendislik Bilimleri Dergisi*, 6(1), 10-20.
21. Rajput, R. K. (2019). A *Textbook of Heat and Mass Transfer*, 7e. S. Chand Publishing.
22. Saha, S. K., Ranjan, H., Emani, M. S., & Bharti, A. K. (2019). *Electric fields, additives and simultaneous heat and mass transfer in heat transfer enhancement*. Springer.
23. Sarı, U., & Kırındı, T. (2019). Using arduino in physics teaching: arduino-based physics experi-ment to study temperature dependence of electrical resistance. *Journal of Computer and Education Research*, 7(14), 698-710.
24. Schouten, P., Putland, S., Lemckert, C. J., Parisi, A. V., & Downs, N. (2012). Alternative methods for the reduction of evaporation: practical exercises for the science classroom.*Physics Education*, 47(2), 202-210.
25. Silva, M. R., Mart´ın-Ramos, P., & da Silva, P. P. (2018). Studying cooling curves with a smart-phone.*The Physics Teacher*, 56(1), 53-55.
26. Tyukodi, B., S´ark¨ozi, Z., N´eda, Z., Tunyagi, A., & Gy¨orke, E. (2012). The Boltzmann constant from a snifter.*European journal of physics*, 33(2), 455-465.
27. Vollmer, M. (2009). Newton's law of cooling revisited. *European Journal of Physics*, 30(5), 1063-1084.
28. Wei, S., Lin, H., & Tan, A. (2024, December). Hot and Cold Data Migration Strategy Based on Newton's Law of Cooling. In *2024 7th International Conference on Data Science and Information Technology (DSIT)* (pp. 1-7). IEEE.
29. Winterton, R. H. S. (1999). Newton's law of cooling. *Contemporary Physics*, 40(3), 205-212.
30. Zill, D. G. (2009). *A first course in differential equations with modeling applications*.

❖ ❖ ❖

International Journal of Innovation and Multidisciplinary Research
DOI.ORG/10.67505/IJIAMR.v6.i1.61007
Vol. 6 (1), 2026
E-ISSN: 2583-4452

Quiver Representation of Neural Networks for Multi-Class Classification: A Traffic Sign Recognition Case Study

Mansi[1*] and Ritika Chopra[2]
[1] Department of Mathematics, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, Delhi
[*]Corresponding Author ✉ mansibhatt794@gmail.com

ABSTRACT

Traffic management in cities has become extremely complex due to rapid urbanization, higher number of vehicles, and increased network of roads. Traditional methods that control traffic using fixed algorithms and scheduling models fail to adapt to changes in traffic. Although neural networks can predict and take decisions within intelligent traffic management systems, their unpredictability makes it difficult for them to be employed in situations when safety plays a critical role. This paper proposes the mathematical model for modeling neural networks based on quivers, where layers would be the vertices while connections would be the directed edges of the quiver, with a weight. This mathematical model provides a clear picture regarding how information flows through neural networks. The current research demonstrates the application of our method for classification of data and its stability. The reason for this is that in order to prove it, this research uses a feature contribution analysis of the inputs to the output as well as the stability analysis using the spectral norm to find out whether the model reacts to any perturbation in the input and parameters. The proposed method provides the researchers with a mathematical representation of the functioning of neural networks. The proposed framework is evaluated on the German Traffic Sign Recognition Benchmark (GTSRB). In the future, it is planned to generalize our framework for other network structures and learning algorithms.

Keywords: *Stability Analysis, Feature Contribution Analysis, Linear Transformations, Vector Spaces*

1 Introduction

The efficient traffic management of the increasing number of vehicles and ever-changing nature of road situations has emerged as a major issue nowadays owing to the exponential rise in the number of vehicles and increasing rate of urbanization and growth in population. The increasing demand for transport services in cities causes traffic congestion, longer travelling time, high amount of fuel consumption, and a greater likelihood of causing road accidents. Apart from these, traffic congestion is also harmful to the environment because it results in increased air pollution and carbon emission, thus damaging both public health and environment.

Traffic behavior depends upon various variables including time of day, weather conditions, road constructions, accidents, and other unexpected situations; hence, managing traffic according to the pre-fixed set of criteria and rules is quite difficult in reality. The existing traffic management systems depend on fixed timings for signaling and programmed methods for regulating traffic and do not have the capacity to adapt to dynamic conditions of traffic. Since these systems lack the ability to adapt, they cannot play their role of controlling traffic. The development of intelligent systems through the use of machine learning techniques such as neural networks has been extensively examined to increase the efficiency of traffic control mechanisms. Neural networks have shown effectiveness in many applications including the identification of traffic signs, vehicle detection, and prediction of traffic flow. Learning from large amounts of data can be done by neural networks, making them a suitable technology for use in traffic control. Studies conducted by Rumelhart et al. (1986) marked the beginning of research related to neural networks with the introduction of the backpropagation method for training multi-layered networks through data learning. Subsequently, Hornik (1991) demonstrated that feed-forward neural networks could approximate all continuous functions.

The use of neural networks to analyze learning systems mathematically was further explored by Wood and Shawe-Taylor (1996) with the help of representation theory. This was followed up by the development of the representation theory of associative algebras, which was done by Assem et al. (2006). As computational power improved and larger data sets could be utilized, neural networks started to show good performance on practical problems. Convolutional neural networks were used to recognize traffic signs by Sermanet and LeCun (2011),

Received on: 10 Jun 2026 | Accepted on: 26 Jun 2026 | Published Online on: 25 July 2026

whereas deep neural networks were proven efficient at image classification by Ciresan et al. (2012). At around the same time, Stallkamp et al. (2012) developed the German Traffic Sign Recognition Benchmark dataset.

Further advances in the field were made by Bengio et al. (2013), who detailed how deep learning algorithms learn hierarchical representations of features, and Zeiler & Fergus (2014), who offered visualization techniques for understanding convolutional networks. Deep learning was also defined as layered representation learning by LeCun et al. (2015). Notable performance gains have been demonstrated by He et al. (2016), who developed residual networks capable of training very deep architectures. The period also saw Goodfellow et al. (2016) provide an extensive review of deep learning techniques. Li et al. (2016) revealed that convolutional neural networks work well on real-life traffic sign recognition tasks, while Redmon et al. (2016) presented the YOLO system for fast real-time object detection.

With increased complexity of neural networks, there is a need to understand the basis for decisions made by the network. Zeiler and Fergus (2014), as well as Selvaraju et al. (2017), developed techniques of visualization such as Grad-CAM that allow highlighting the regions of an image that affect the prediction of the network. These approaches provide the ability to comprehend the operations performed by convolutional neural networks on visual data. Rudin (2019) argued that black-box models should not be used when making critical decisions because they lack interpretability. Recently, Armenta and Jodoin (2021) studied neural networks through algebraic and representation approaches.

Despite the achievements mentioned above, neural networks do not have sufficient interpretability when used in safety-critical domains like traffic systems. In a mathematical sense, neural networks can be considered as directed graphs where neurons serve as nodes and edges represent the information flows. Thus, quiver representation theory can also be applied to neural networks, where nodes represent vector spaces and edges are linear transformations between these spaces. However, it has not been used effectively enough to investigate how neural networks handle information in traffic-related tasks.

To overcome these problems more efficiently, a new approach using quivers to model neural networks for traffic sign recognition tasks has been suggested. It aims at gaining insight into how inputs, such as images of traffic signs, affect outputs in a traffic recognition application. In other words, modeling of neural networks by using quivers makes it possible to analyze features, classification process, and system stability. All these aspects are significant when working with intelligent traffic systems.

2 Prelimanries

Definition 1 *(Assem et al., 2006) The quiver can be mathematically defined as an ordered pair of sets together with two maps that define a source and target for each arrow . More precisely, it is written as*

$$Q = (Q_0, Q_1, s, t),$$

where Q_0 denotes the set of vertices, Q_1 the set of arrows, and the maps $s, t : Q_1 \to Q_0$ specify where each arrow originates and where it terminates.

Definition 2 *(Assem et al., 2006) In a quiver representation, each vertex $v \in V$ gets assigned a vector space W_v. Each arrow $e \in E$ is given a linear map $W_e : W_{s(e)} \to W_{t(e)}$.*

Definition 3 *(Armenta and Jodoin, 2021) The activation at node v with respect to $a(W, f)_v(x)$ is given as:*

$$a(W, f)_v(x) = f_v\left(\sum_{\alpha \in E_v} W_\alpha \cdot a(W, f)_{s(\alpha)}(x)\right) \tag{1}$$

Definition 4 *(Armenta and Jodoin, 2021) A neural network may be considered as a function ($f : \mathbb{R}^n \to \mathbb{R}^k$) that is expressed as a sequence of functions associated with layers. Normally, each layer implements a linear mapping followed by the application of an activation function. For a given input (x), each layer computes*

$$h = \sigma(Wx + b),$$

where (W) and (b) are a weight matrix and bias vector, respectively, while (σ) is an activation function.

3 Quiver Representation of Neural Networks

The proposed framework combines neural networks with quiver representations to model traffic sign recognition mathematically. It aims to show how info moves through the network and shapes classification decisions.

Inputs, shown as feature vectors $x \in \mathbb{R}^n$, pass through a neural network that outputs y. This includes different layers that have weights, biases, and activation functions.

In order to provide a more mathematical structure to the neural network, we will consider the neural network to be a quiver $Q = (V, E)$. In the quiver, nodes will be considered as layers and edges will represent information flow. Each node has a vector space and the mapping between two nodes will be a linear map.

Hence, we will consider the neural network to be vector spaces with linear transformations between them.

3.1 Neural Network Architecture

A neural network is basically a series of affine transformations along with nonlinear activation functions. For traffic sign recognition, it takes images, analyzes key features, and matches them to a list of known traffic signs.

As illustrated in Fig. 1, the neural network consist of three layers, namely input, hidden, and output layer. Let $V_0 = \mathrm{R}^n$, $V_1 = \mathrm{R}^m$, and $V_2 = \mathrm{R}^k$ be the input, hidden, and output spaces. These spaces show how data moves through the network. Each layer, thus, does a specific kind of data transformation.

- **Input Layer** : The input layer uses $V_0 = \mathrm{R}^n$, where each vector $x \in \mathrm{R}^n$ comes from traffic sign images and shows details like color, shape, and text.

  Next up is an affine map that shifts the data into V_1:

  $$T_1(x) = W^{(1)}x + b^{(1)}, \tag{2}$$

  with $W^{(1)} \in \mathrm{R}^{m\times n}$ and $b^{(1)} \in \mathrm{R}^m$.

- **Hidden Layer** : The hidden layer works in the space $V_1 = \mathrm{R}^m$ and helps the network figure out patterns in the input data. To make the model more flexible, a nonlinear activation function $\sigma : V_1 \to V_1$ is used, applied individually to each component.

  Here, the ReLU function, short for Rectified Linear Unit, is choosen for this role. ReLU is defined as $\sigma(z) = \max(0, z)$.

  So, the hidden representation is calculated as $h = \sigma(W^{(1)}x + b^{(1)})$. This does a great work at bringing forward key features from the input while reducing noise and other unimportant stuff.

- **Output Layer** : The output layer is represented by the space $V_2 = \mathrm{R}^k$, which stands for the different traffic sign classes. It transforms the learned hidden representations into final classification results.

  This transformation is shown by

  $$T_2 : V_1 \to V_2,\ T_2(h) = W^{(2)}h + b^{(2)} \tag{3}$$

  where

  $$W^{(2)} \in \mathrm{R}^{k\times m},\ b^{(2)} \in \mathrm{R}^k.$$

  The output vector is

  $$y = T_2(h), \tag{4}$$

  with each part of y telling us how well the input matches a certain traffic sign class.

In the end, this work produces a vector that gives scores for each class, which is what the model uses right before making its final decision.

3.2 Quiver Representation of the Model

The quiver can be mathematically defined as an ordered pair as defined in defination (Definition 1), illustrated in Fig. 2. Although the quiver is not meant to reflect the actual traffic in reality, it acts as an abstraction of what happens in the flow of information. The transformation associated with each arrow is responsible for the changes to the incoming information and thus fulfills a function equivalent to that of weight values in neural networks. The information transformation aspect is thereby easily incorporated into the quiver representation.

According to the defination (Definition 2) in a quiver representation, each vertex $v \in V$ gets assigned a vector space W_v. According to Armenta and Jodoin (2021), as defined in the defination 3.

The following describes the notation employed in the equation 1. $a(W, f)_v(x)$ corresponds to the activation (output representation) for the vertex v, whereas $a(W, f)_{s(\alpha)}(x)$ is the activation for the source vertex of edge α. By W_α it means the learned linear map that takes the information from $s(\alpha)$ to $t(\alpha)$. The tuple (W, f) stands for the definition of a neural network that operates over the quiver Q. it also have $f = (f_v)_{v\in Q0}$ standing for the nonlinear activation functions for vertices. The notation $x \in \mathrm{R}^n$ means the vectorized input. Finally, E_v corresponds to the set of all edges whose target vertex is v.

The non-linearity within this particular model comes into existence because of the activation function f_v that has been designed by using the ReLU function. The ReLU function converts all negative values into zeroes but keeps the positive values the same. This is very important for the model to be able to learn complex mappings.

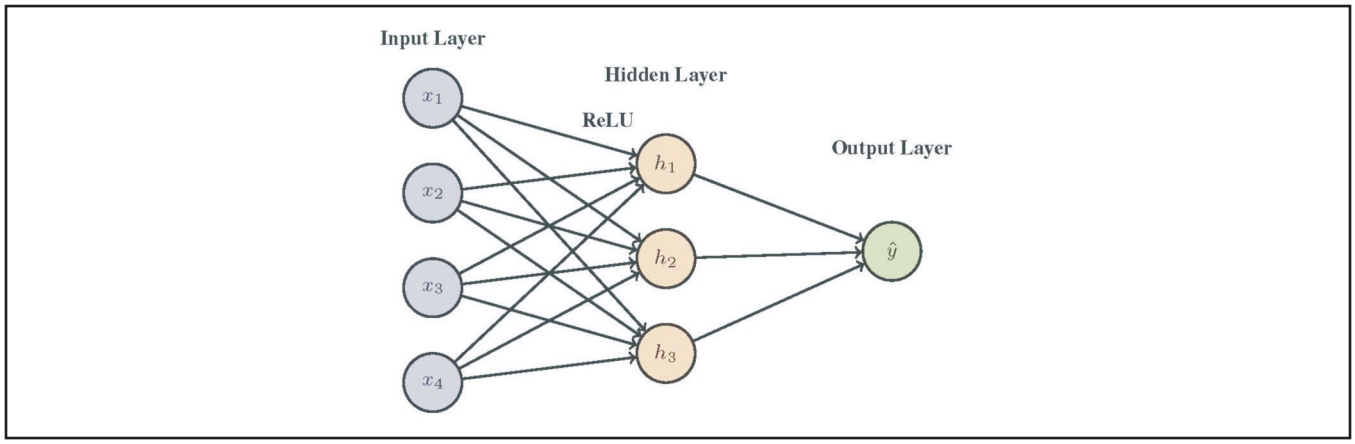


Fig. 1: *Neural network architecture with ReLU activation*

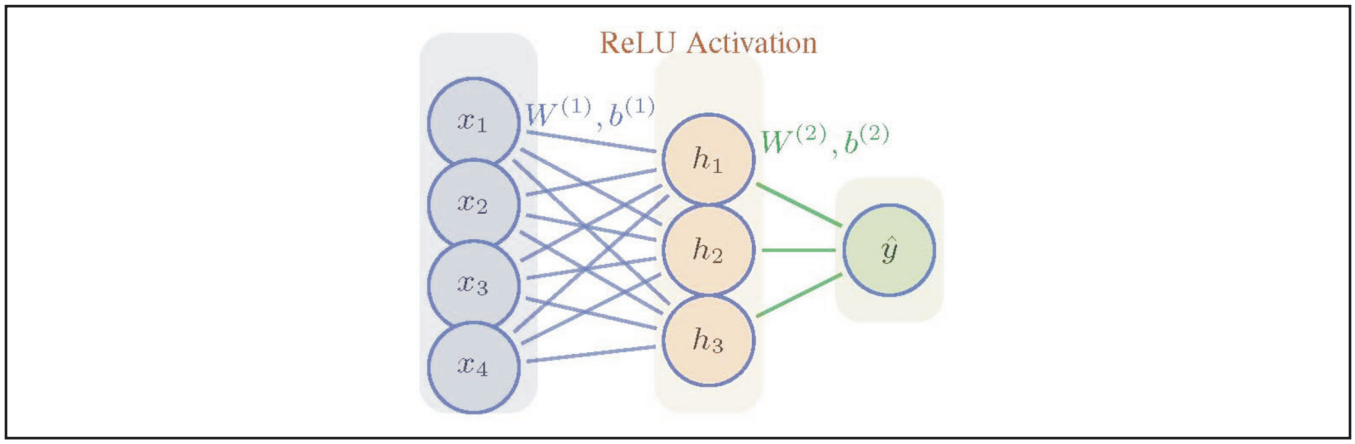


Fig. 2: *Quiver-based representation of the neural network model*

3.3 Integration with Traffic Management System

This proposed system combines traffic sign recognition with traffic flow control to simplify decision-making in smart traffic setups. It begins with feature representations $x \in V_0$, where V_0 holds the traffic sign features.

Then, a neural network transforms these features into distinct categories using the mapping $f : V_0 \to V_2$.

Along the way, a hidden layer $h \in V_1$ appears to help the learning process from the initial inputs.

It's called a simple layered structure:

$$V_0 \xrightarrow{T_1} V_1 \xrightarrow{T_2} V_2.$$

At the end, the system gives $y \in V_2$, which offers scores for various traffic signs such as STOP, speed limit, yield, and directional signs. The vehicle uses the information provided here in determining what the expected actions should be from the vehicle. For example, in case of seeing a STOP sign, the vehicle will stop at an intersection. A yield sign can help determine which vehicle will pass through the intersection first. The sight of a speed limit sign slows down the vehicle, while directional signs make turns and changes in lanes easier. Here, we use the quiver representation to model the traffic structure. Here, the nodes represent the traffic states or locations, while edges denote movement between these locations. These nodes have feature representations extracted using a neural network.

In this work, vector spaces and linear transformations are used to describe how information flows in a system. Each traffic sign updates its vertex, showing changes in the graph. This represents data moving through the structure, influencing different parts as it goes.

3.4 Mathematical Framework For Classification

This work uses a neural network structure to classify traffic signs using images and explains how information from inputs is processed to get the output classification result.

A vector $x \in R^n$ is given to the network. This input is based on the important characteristics of traffic signs, such as shapes and colors, extracted from their images. For us to make computations on this input, it needs to be quantified, meaning all the traffic signs are represented as points in the feature space consisting of n dimensions.

There are multiple transformations performed on the data by the neural network during its computation. The first one is a transformation performed by the weight matrix $W^{(1)} \in R^{m \times n}$, and the bias vector $b^{(1)} \in R^m$, which produce an intermediate result of:

$$z^{(1)} = W^{(1)}x + b^{(1)} \tag{5}$$

The relationship between the inputs begins to be learned by the network as a result of the linear transformation and the blending of the input features in an appropriate proportion. The set of feature detectors is embedded in the weight matrix $W^{(1)}$, and every feature detector in it is capable of responding to a particular pattern within the input data. The use of the bias term $b^{(1)}$ allows the detector to operate independent of the input data.

In addition to the sequence of linear mappings, in this work, a nonlinear activation function is used on each intermediate result, one entry at a time. For that purpose, the most widely used activation function is the so-called Rectified Linear Unit or ReLU given by:

$$\text{ReLU}(z) = \max(0, z)$$

which leaves positive numbers intact and maps negative numbers into zeros.

The reason for the introduction of such nonlinearity lies in the fact that without it, multiple layers would simply combine into one linear mapping. By introducing nonlinearity, the model gets an opportunity to construct complex hierarchical decision surfaces that will allow it to effectively differentiate between the various types of traffic signs. As a consequence of ReLU application, the hidden representation becomes:

$$h = \text{ReLU}(z^{(1)}) = \text{ReLU}(W^{(1)}x + b^{(1)})$$

Then, the result is feed to the next layer, where another linear map takes place by applying matrix $W^{(2)}$ followed by adding the bias $b^{(2)}$:

$$T(x) = W^{(2)}\,\text{ReLU}(W^{(1)}x + b^{(1)}) + b^{(2)} \tag{6}$$

This equation characterizes the full mapping procedure performed by our two-layer network starting from input R^n vector x until reaching its output in R^k, where k stands for the number of traffic sign classes.Observe that the mapping is an affine function and behaves similar to the linear map

$$T(x) = W^{(2)}h \tag{7}$$

in terms of classification, feature analysis & stability.

Thus, the resultant is a vector,

$$T(x) = (y_1, y_2, \ldots, y_k)$$

In this vector, the component y_i indicates the score of class i. The higher the value, the more related is the input data to this class. Hence, each value in the result vector corresponds to the score of a connection between the input and one specific traffic sign class.

It is possible to focus on individual classes through projecting the result vector:

$$P_i(T(x)) = y_i$$

In this case, each projection will be considered individually. In order to fairly compare the predicted scores, the absolute values are taken into account:

$$\|P_i(T(x))\| = |y_i|$$

Absolute values allow us to compare different scores in equal conditions and choose the highest value among all k classes.

As a result, the predicted class is the following:

$$\text{Class}(x) = \arg\max_i |y_i|$$

At this stage, the network computes the output values for all classes simultaneously and takes the class with the highest absolute value as an answer.This neural network architecture stands out in its ability to demonstrate the synergy between linear and nonlinear transformations, which results in discriminative feature learning that is finally reduced to a single class label.

4 Analysis and Interpretation

To understand the behavior of the entire framework, the final perspective on the quiver model need to be consider. Modeling the network via the use of the quiver allows us to analyze how certain elements are connected within this structure, and how information is exchanged among components.

Under such conditions, many important characteristics become evident. Furthermore, understanding the effects of a certain component on the entire output becomes relatively easy. The major benefit of the proposed approach is that the model becomes more open. Each transformation within the system becomes more transparent both independently and in the context of the entire framework.

4.1 Feature Contribution Analysis

Here, the analysis focuses on the importance of individual input characteristics for the overall classification procedure. This is particularly important as all features have unequal influence on the prediction process, since some input features contribute significantly to the resulting prediction, whereas other features are relatively insignificant. Therefore, the understanding of those differences plays a crucial role in the interpretation of how the model functions and what level of significance each particular input feature possesses.

Each arrow in the quiver is associated with a matrix that governs the way linear transformation of the input is performed along the edge, hence signaling about the relevance of various characteristics as they flow along the layers of the model. In a mathematical sense, the arrow represents a linear mapping:

$$W_\alpha : \mathrm{R}^n \rightarrow \mathrm{R}^m$$

that shows how a vector x in R^n results in a vector in R^n .

To determine how much of an influence each input element has on the output, this work connect each feature to its respective standard basis vector $e_j \in \mathrm{R}^n$, Specifically, e_j refers to the vector consisting of zeroes with one located at position j. Then, by multiplying W_α with the above vector, this work obtain:

$$W_\alpha e_j$$

The latter equals the j-th column of the W_α. In other words, it describes the extent of influence that the j-th feature has on the transformed output. If a feature, acting separately from others, can strongly move the output within the space, then it is considered significant because it leads to a certain shift in the generated representation to the next node.

To measure this significance numerically, the following norm of the obtained vector is computed:

$$\mathrm{Contribution}(j) = \|W_\alpha e_j\| \tag{8}$$

Thus, this work obtain a concise quantitative representation of the feature's influence on the transformation performed by the node. The higher the value of the norm, the greater the impact of the feature on the output. In contrast, a lower value indicates that the feature exerts a minor influence.

This approach allows one to analyze the influence of all input features quantitatively to get insight into their impact on the resulting output of the classifier framework. In particular, by evaluating a contribution of every $j \in \{1, 2, \ldots, n\}$ it is possible to produce a comprehensive picture of the feature importance within the whole feature space. Those features that have significant contributions will be considered highly discriminative for the task under consideration, whereas features with low contribution values might not carry useful information at all and can even be redundant. Hence, one can easily identify features of an input image of a traffic sign that influence classification results to the highest degree.

4.2 Stability Analysis

The stability aspect entails how slight changes in input affect the behavior of a particular framework. In case the framework exhibits stability, slight changes in input cause proportional changes in its output while in case of instability, slight changes in input will lead to huge changes in output. Stability plays a very vital role in classifying a certain framework since it helps us determine whether the framework produces reliable output when there are slight changes in input, which in our case would mean slight differences in lighting conditions, angles at which the camera takes photos, among others.

The following shows an illustration of the arrows and their associated linear transformations in the quiver framework:

$$\mathbb{R}^n \xrightarrow{T} \mathbb{R}^k$$

The input vector belongs to the input space:

$$x \in \mathrm{R}^n$$

and after the transformation is applied, the output is given by:

$$T(x) \in \mathrm{R}^k$$

Every such linear transformation can be expressed in matrix form as:

$$T(x) = Ax, \quad A \in \mathrm{R}^{k \times n} \tag{9}$$

Here, A is the matrix of the linear mapping T. In other words, the entire information about the transformation behavior is encoded in A, and analyzing the matrix A allows one to obtain valuable information on the behavior of the corresponding arrow in the quiver and, in particular, to judge its stability.

To study the stability, one needs to consider the matrix norm $\|A\|$, which indicates the maximum possible stretching effect of the transformation. The value $\|A\|$ is critical to determine whether the transformation remains stable or becomes amplifying.

The stability condition of the transformation is determined by $\|A\|$. For:

$$\|A\| < 1 \quad \Rightarrow \quad \textbf{Stable transformation}$$

This transformation either retains the magnitude or decreases it. Here, it will either decrease or limit the magnitude of the input such that an input with slight change will produce an output with slight change. Therefore, the system works in a controlled manner and can be trusted in classification even where the input may be slightly noisy.

Conversely, when:

$$\|A\| > 1 \quad \Rightarrow \quad \textbf{Unstable transformation}$$

This property would definitely cause an increase in the amount of data used as input. The slightest change that is introduced to the input data would definitely have a large effect on the output that is produced. In terms of classification, this particular property can be viewed as quite undesirable because any small change in the input image would produce completely different output.

5 Result and Discussion

This section presents the experimental results for the model's performance, including stability, feature extraction, and classification. The model checks off several categories, including STOP, speed limit, yield, and directional signs, when tested on a common traffic sign recognition dataset. These results center on classification accuracy, feature representation learning performance, and input variation stability.

5.1 Case Study: Traffic Sign Recognition

The primary objective of this case study is to apply the proposed framework to the traffic sign identification problem. The objective is to categorize different types of traffic signs using real-world image data and evaluate how well the model performs in real-world situations.

5.1.1 Dataset

The proposed model is estimated using the German Traffic Sign Recognition Benchmark (GTSRB), which is a benchmark dataset for traffic sign classification. As shown in Fig. 3, the dataset consists of various traffic sign categories. This dataset includes real-world images of traffic signs taken in various conditions like changes in lighting, viewing angles, scale, and background complication.

Every image comes with structured information including details about the traffic sign region, including the image width and height, and the coordinates of the region of interest (ROI). The ROI coordinates tell us where exactly the traffic sign is located in the image. This helps focus on the important bits of the image and makes feature extraction more efficient.

In addition to the labeling, each image also has the type of sign labeled, whether a speed limit, warning, or regulatory sign. For input into the model, the image are represented in terms of features $x \in V_0$. Here, V_0 is known as the input feature space, containing visual information such as shape, color, edges, and structure.

All this structure makes the GTSRB great for testing how well multi-class classification models can mark different types of traffic signs. With its multiple set of images, it gives us a good way to measure how the proposed model handles real-life variations in traffic sign appearances.

The dataset is obtained from the Kaggle platform: *https://www.kaggle.com/datasets/meowmeowmeowmeowmeow gtsrb-german-traffic-sign*

	Width	Height	Roi.X1	Roi.Y1	Roi.X2	Roi.Y2	ClassId	Path
0	27	26	5	5	22	20	20	Train/20/00020_00000_00000.png
1	28	27	5	6	23	22	20	Train/20/00020_00000_00001.png
2	29	26	6	5	24	21	20	Train/20/00020_00000_00002.png
3	28	27	5	6	23	22	20	Train/20/00020_00000_00003.png
4	28	26	5	5	23	21	20	Train/20/00020_00000_00004.png
5	31	27	6	5	26	22	20	Train/20/00020_00000_00005.png
6	31	28	6	6	26	23	20	Train/20/00020_00000_00006.png
7	31	28	6	6	26	23	20	Train/20/00020_00000_00007.png
8	31	29	5	6	26	24	20	Train/20/00020_00000_00008.png
9	34	32	6	6	29	26	20	Train/20/00020_00000_00009.png
10	36	33	5	6	31	28	20	Train/20/00020_00000_00010.png
11	37	34	5	6	32	29	20	Train/20/00020_00000_00011.png
12	38	34	5	6	32	29	20	Train/20/00020_00000_00012.png
13	40	34	6	6	34	29	20	Train/20/00020_00000_00013.png
14	39	34	5	5	34	29	20	Train/20/00020_00000_00014.png

Fig. 3: *Sample numerical data from the GTSRB (German Traffic Sign Recognition Benchmark) dataset*

5.1.2 Model Implementation and Results

This model uses vectors to represent information from traffic signs, where each characteristic is represented numerically. This makes it possible for us to work on the data in an organized manner.

Each traffic sign will have its features categorized into two groups depending on the nature of the features.

The first group is made up of basic image properties such as width and height, which may be represented as:

$$x_1 = \begin{bmatrix} \text{Width} \\ \text{Height} \end{bmatrix} \in \mathbb{R}^2$$

The second contains the ROI coordinates indicating the exact location of the traffic sign within the image.

$$x_2 = \begin{bmatrix} \text{Roi.X1} \\ \text{Roi.Y1} \\ \text{Roi.X2} \\ \text{Roi.Y2} \end{bmatrix} \in \mathbb{R}^4$$

These two vectors are combined into a single input vector:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 \oplus \mathbb{R}^4 = \mathbb{R}^6$$

This vector entirely represents a traffic sign example containing the sign's size and position within the image.To understand the way the data is being processed by the model, it is useful to consider various vector spaces:

$$V_1 = \mathrm{R}^2,\ V_2 = \mathrm{R}^4,\ V_3 = \mathrm{R}^6,\ V_4 = \mathrm{R}^{43}$$

In this case, V_1 and V_2 are the input feature spaces, V_3 is the feature space that contains both vectors, and V_4 represents the output feature space. The elements of V_4 are related to traffic sign classes.

Using a vector space representation, a neural network can be modeled with a quiver. As shown in Fig. 4, this shows how different vector spaces relate through linear transformations. The mappings between them are described by:

$$W_{\alpha 1} : R^2 \rightarrow R^6,\ W_{\alpha 2} : R^4 \rightarrow R^6$$

Since the two paths produce outputs belonging to R^6, it means that they are able to mix information about different feature sets. Their sum, processed by a nonlinear activation function (ReLU) is used as the hidden representation:

$$v_3 = \text{ReLU}(W_{\alpha 1} x_1 + W_{\alpha 2} x_2) \in R^6 \quad (10)$$

In this way, the hidden vector makes the processing more effective by combining the information about the object's size and position, thus improving its representation.

Finally, the hidden representation is converted to the output space by another transformation:

$$W_{\alpha 3} : R^6 \rightarrow R^{43}$$

The resulting output belongs to R^{43}, having each dimension correspond to some traffic sign classes. The maximal value will correspond to the predicted class. The quiver diagram will help visualize the process with transformations depicted by arrows and model layers by nodes.

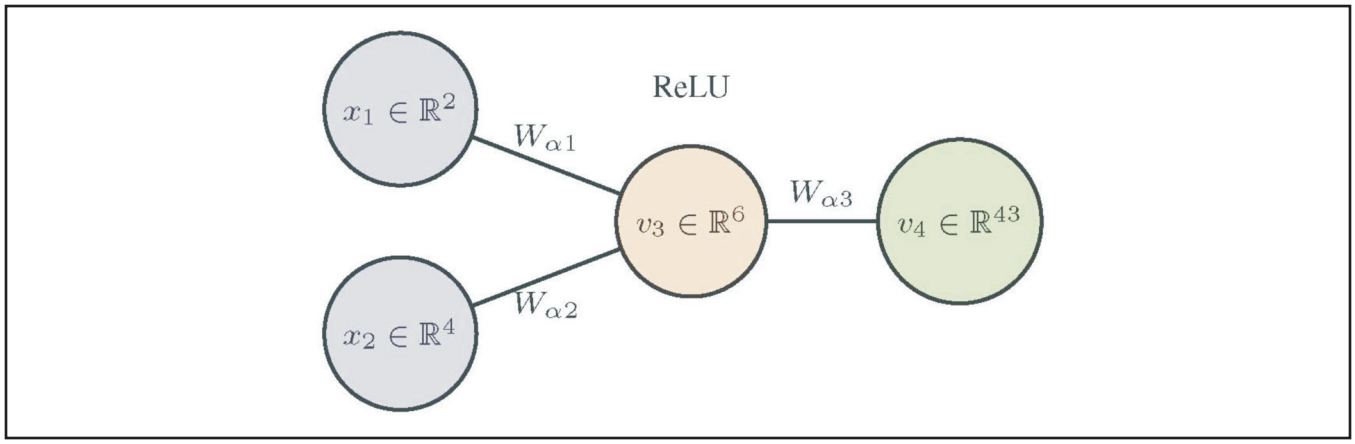


Fig. 4: *Quiver representation of neural network showing feature flow across vector spaces*

In this arrangement, the inputs pass through their respective layers and combine at the hidden layer to give rise to the ultimate decision. Any slight change in the value of the inputs or weight matrices will not result in a significant difference in output and will not affect the predicted class either. This is an indicator of the stability of the model.

This neural network first takes the input features through a series of operations until it generates the output vectors within R^{43}. In other words, the classifier generates 43 numbers for an input data sample where each number represents the relationship between the input and one of the 43 classes of traffic sign images. Rather than reducing the inputs directly into a label, it calculates a value representing its association with each of the 43 classes simultaneously. This means that during classification, the algorithm considers the relationships with all possible competing classes simultaneously since the classes of traffic sign have very much resemblance to each other.

These generated values represent their contribution scores, meaning higher absolute values suggest stronger relationships. For example, the classification depends on how the 43 output values compare to each other, and the rule assigns the input to whichever class has the highest score despite how large or small each value is in absolute terms. The transformation from input to output is represented as:

$$T(x) \in R^{43}$$

It will be the class whose corresponding element in the vector is of the highest magnitude. That means this work select the index of the largest projection out of the 43 classes.

$$\text{Class}(x) = \arg\max_i \|P_i(T(x))\| \quad (11)$$

To illustrate this, consider the output vector returned by the neural network when fed with an input:

[. . . , -0 00019 ,0 000003 ,0 000155 ,0 000516 , . . .]

Despite being insignificant, the slight variations among them become meaningful when making a selection. The fourth element, 0.000516, clearly represents the highest value from the entire 43 elements, thus forming the basis of the classification decision. It is not essential for the framework to generate large absolute figures; rather, the need is for a single element to have a significant value compared to others. The margin between the highest value 0.000516 and the second-highest 0.000155 may seem small in absolute figures, but it provides enough information for making a definite decision. This will give us the index at which the highest number is present in the output vector. Here, each element in the output vector indicates the score for the class. This score decides what class is assigned by the classifier. Here, the element with the index 21 has the highest number in comparison to all other elements; Thus, the predicted class becomes:

$$\text{Class}(x) = 21$$

This proves that the framework works independent of the magnitude of the output values. Regardless of the fact that the 43 elements lie close to zero in absolute terms, the hierarchy among them remains unaltered, and it predicts the right class without fail. Such an ability is significant in traffic sign classification, since different classes resemble one another and consequently the output elements cluster around each other. However, in such scenarios, the framework chooses the class which stands out among the remaining 43 as having a slightly greater element than the others. For instance, the output elements are differentiated from each other by as little as 0.000516 - 0.000155 = 0.000361

The model uses two feature vectors that describe different parts of a traffic sign. The first one, x_1, contains data on the object size and consists of the dimensions – width and height of the sign. The second, x_2, contains information related to location – the Region of Interest's coordinates. These feature sets go through separate weight matrices before meeting up in the hidden layer. This lets the model handle and weigh both the size and position information effectively, so it works better overall.

Regarding the feature contribution, the results are as follows:

$$[0.0191, 0.0187]$$

It can be seen that there is no substantial difference between width and height – each of them provides a similar amount of importance to the input layer of the hidden layer. Thus, the combined contribution from these two features is relatively small in comparison with the ROI features, implying that this type of information is rather supplementary than decisive for the classification.

On the other hand, the coordinates of the ROI affect the same hidden neuron v_3 in the following way:

$$[0.0197, 0.0224, 0.0248, 0.0307]$$

While still below the contributions of the size features, the contributions increase progressively from one coordinate to the next. While 0.0197 is already on equal footing with the size features in terms of importance, 0.0307 is by far the highest contribution (the contribution of ROI.Y2) and triggers the most dramatic change compared to any other input variable. From this increasing progression from 0.0197 to 0.0307 within the ROI coordinates, this work can deduce that vertical information is more discriminative than horizontal information.

In terms of the head-to-head comparison between the two sets of inputs, the prioritization scheme is obvious and consistent throughout the framework. When compared against the features related to the size, the ROI features have a tendency to provide higher contributions to the output at all four positions. In terms of the variations, the width and height of a road sign could change greatly depending on the distance between the observer and the road sign and the perspective from which the observer observes the road sign. However, the location of the road sign within the predefined region of interest could be considered as more reliable and consistent cues about the structure of the road sign and the class to which the road sign belongs.

After passing through the hidden layer, the hidden neurons themselves create the contributions to the output neuron v_4. The contribution generated by each hidden neuron will be:

$$[0.0551, 0.0636, 0.0704, 0.0539, 0.0668, 0.0784]$$

These six hidden neurons do not contribute equally towards the final classification output. They contribute anywhere from 0.0539 to 0.0784, which shows significant variations as far as their contribution is concerned. It means that while Neuron 6 has the highest contribution of 0.0784, contributing the maximum push for final output in the output layer, Neuron 4 has the minimum contribution of 0.0539. However, despite the minimum value of its contribution, its input cannot be considered insignificant either. The other four neurons have values of 0.0551, 0.0636,0.0704, and 0.0668 respectively, forming a gradient in between.

The difference between the maximum contribution (Neuron 6) and the minimum contribution (Neuron 4) is just 0.0245. It means a difference of 45% between these two neurons. Although the exact numerical difference may not appear very significant, it shows a definite difference between the two neurons.

In fact, such an analysis reveals a very logical pattern that includes the presence of the same priority criterion at each level of the hierarchical model under consideration. Specifically, the positional attributes play a more important role than size attributes in the process of processing input data; the ROI.Y2 attribute with its value of 0.0307 is the one that plays the biggest role in terms of inputs. In case of the hidden layer, it is hidden node 6 with its weight 0.0784 that has the greatest influence on the output data. This shows a certain hierarchical nature of specialization in the weight matrices.

One could say that a neural network is essentially an organized linear transformation pipeline, going from one vector space to the other. In the context of quiver algebra, each linear transformation becomes the arrow representing a clear process through which information passes from the input layer all the way to the output.

All the input data collected in advance forms a 6-dimensional space, and further linear transformations take the data into 43 dimensions – each dimension for each of the traffic sign classes:

$$T : \mathrm{R}^6 \rightarrow \mathrm{R}^{43}$$

Thus, the input to the function above is the 6-dimensional vector, consisting of the sizes of both traffic sign and ROI, described above. As for the output of the linear transformation T, this is a 43-dimensional vector, corresponding to each of the traffic sign classes. This expansion from the input space to a substantially greater output

space is determined by the nature of the problem itself: a small amount of information needs to be represented as a large vector, describing the strength of association with any of 43 categories.

In order to simplify the process, all the mappings in the network are combined into a single composite mapping matrix. This input mapping matrix is first constructed by putting together the weight matrices from the two input features in parallel:

$$W_{in} = [W_{\alpha 1}\ W_{\alpha 2}]$$

This figure depicts the combination of the two input vectors as one input vector x consisting of size features x1 and ROI coordinates x_2. Input-to-hidden interactions can be represented by the matrix W_{in}, thus combining all input-to-hidden mappings to one linear operator. It will make our analysis more simple because now it has one term instead of several arrows coming from the input.

The complete input-to-output mapping can be found by multiplying W_{in} by $W_{\alpha 3}$:

$$A = W_{\alpha 3} W_{in},\ A \in \mathrm{R}^{43\times 6} \tag{12}$$

where A belongs to the real numbers set of dimension 43 × 6. Thus, A is a matrix of dimension 43 × 6 that transforms any point from 6-dimensional input to 43-dimensional output without any intermediate steps. By reducing the two transformations to one, it can simplify the stability analysis and treat it for one linear operator

instead of several arrows.

To verify whether the structure satisfies the condition of stability, the spectral norm of A will be examined:

$$\|A\| = \sqrt{\lambda_{\max}(A^T A)} \tag{13}$$

If the system works with the traffic signs data, the result obtained:

$$\|A\| = 0.0026$$

This is an extremely low value. Stating differently, a spectral norm of 0.0026 is equal to A contracting all vectors no more than 0.26%. All in all, the condition of stability is satisfied not only by a small amount but by a rather large one because the magnitude of the output will always be much smaller than the magnitude of the input.

As

$$\|A\| = 0.0026 < 1$$

it is clear that the system is definitely stable. That is, any change of an input value cannot influence the output in any significant way.

In other words, the effect of a slight disturbance in the input on the output is insignificant; the transformation process reduces the variation before reaching the output layer. This makes the neural network extremely suitable for classification, because the output vector remains almost unchanged even with some disturbance or noise in the input. The stability margin, which is calculated based on the distance between the value of $\|A\|$ and the critical point 1, amounts to 0.9974 – a truly impressive value. Also, the stability of the framework proves the continuity of the transformation described in the weight matrices, as the spectral norm is considerably lower than 1, meaning that the transformation from input to output does not have any discontinuous jumps.

6 Limitations

This research is mostly theoretical and gives a mathematical model for the description of data flow with the help of quiver representations and affine transformations. This model is verified by testing on the German Traffic Sign Recognition Benchmark, but this study is still primarily concerned with the mathematical model rather than experimentation and comparison of results with other benchmark models.

The testing is done using only one dataset; therefore, the applicability of this algorithm on other traffic sign datasets or in other conditions is not addressed. Moreover, in the practical scenario, traffic signs may be affected by some noise and blur.

7 Conclusion

In this paper, neural networks have been analyzed as directed graphs for gaining better insights. Using the concepts of quiver representation theory, the nodes have been regarded as vector spaces and edges as linear transformations. This made it possible to give a detailed mathematical insight into the working of a neural network. For practical use, this approach has been used in traffic sign recognition along with an evaluation of feature contributions and stability. In conclusion, combining mathematical concept with neural networks is beneficial for better insights and analysis. In this study the theoretical concepts and real-world applications are combined effectively and can be applied to solve more complicated problems in the future.

Conflict of Interest Statement

The authors declare no conflict of interest.

Ethical Declarations

The authors declare that this research does not involve human participants, animals, or any procedures requiring ethical approval. All data used in this study were obtained and processed in accordance with applicable ethical & legal guidelines.

References

1. Armenta, M., and Jodoin, P.-M. (2021). The representation theory of neural networks. *Mathematics, 9*(24), 3216.
2. Assem, I., Simson, D., and Skowroński, A. (2006). *Elements of the representation theory of associative algebras*. Cambridge: Cambridge University Press.
3. Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828.
4. Cires¸an, D. C., Meier, U., and Schmidhuber, J. (2012). Multi-column deep neural networks for image classi-fication. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3642–3649.
5. Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. Cambridge: MIT Press.
6. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
7. Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks, 4*(2), 251–257.
8. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
9. Li, J., Wang, Z., and Liu, C. (2016). Traffic sign detection using convolutional neural networks. *Neurocom-puting, 214*, 723–734.
10. Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788.
11. Rudin, C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215.
12. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536.
13. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, 618–626.
14. Sermanet, P., & LeCun, Y. (2011). Traffic sign recognition with multi-scale convolutional networks. In *Pro-ceedings of the International Joint Conference on Neural Networks*, 2809–2813.
15. Stallkamp, J., Schlipsing, M., Salmen, J., & Igel, C. (2012). The German traffic sign recognition benchmark: A multi-class classification competition. *Neural Networks, 32*, 323–332.
16. Wood, J., & Shawe-Taylor, J. (1996). Representation theory and invariant neural networks. *Discrete Applied Mathematics, 69*(1–2), 33–60.
17. Zeiler, M. D., and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Proceedings of the European Conference on Computer Vision*, 818–833.

❖ ❖ ❖

International Journal of Innovation and Multidisciplinary Research
DOI.ORG/10.67505/IJIAMR.v6.i1.61008
Vol. 6 (1), 2026
E-ISSN: 2583-4452

Sentiment Analysis in Emoji-Rich Text: A Comprehensive Survey of Techniques, Datasets, Challenges and Research Directions

Prof. Anamika Gupta[1], Sugandha Kaur[2*], Vanshika Goyal[3], Aradhya Jain[4] and Ravika Singh[5]
[1,3,4,5] Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi
[2] Mata Sundri College for Women, University of Delhi, Delhi
[*] Corresponding Author ✉ sugandhakaur@ms.du.ac.in

ABSTRACT

Visual digital communication is advancing substantially, with emoji-rich text, making it quite challenging for the analysis of sentiments from such text. In the paper, a comprehensive overview is provided for the methodologies for sentiment analysis in emoji-rich text. In this study, the landscape of available datasets curated for this text is reviewed. Their relevance, limitations, and linguistic diversity is also studied. Furthermore, a comprehensive review of the preprocessing techniques and the machine learning models utilized is also presented. A detailed study is conducted to highlight the specialized processing techniques used for the interpretation and handling of emojis, such as emoji embeddings and ranking systems. Further, research gaps are highlighted, including a multilingual and multimodal approach, and non-availability of high quality datasets. The paper concludes with a discussion on future work advocating for context-sensitive models and the use of generative AI for the development of emoji-aware pre-trained models. This work will serve as a comprehensive reference point for researchers planning to advance sentiment analysis of the emoji-driven digital world.

Keywords : *Emoji-rich, Sentiment Analysis, Emoji Embedding, Context-sensitive*

1 Introduction

Sentiment analysis is a field that blends Machine Learning (ML) and Natural Language Processing (NLP), enabling computers to understand and interpret the emotions expressed behind written text. The sentiment behind the text could be classified as positive, negative, or neutral Ayvaz and Shiha (2017); Yoo and Rayz (2021). With the advent of technology, a huge amount of data is generated on a daily basis through social networks, online reviews, discussion forums, and news articles. Manual analysis and interpretation of this amount of data is practically impossible Alfreihat et al. (2024). The process of sentiment analysis can help in such scenarios by extracting valuable insights from huge volumes of data, which in turn aids in detecting feelings, opinions, or attitudes of the author toward a specific topic, product, or event. It also helps to monitor public sentiment in real time, allowing companies, governments, and researchers to respond quickly to emerging trends, issues, or concerns Khan et al. (2025). In a nutshell, sentiment analysis imparts stakeholders a clearer understanding of how their audiences feel, helping them in thoughtfully addressing their feedback and concerns.

Sentiment analysis is not only confined with the written text, but it also takes into account emojis and emoticons, which play a crucial role in the present digital communication eraHakami et al. (2022); Yuan et al. (2022). An emojis is a tiny graphical symbol or icon that represents ideas, objects or emotions Holthoff (2020); Kralj Novak et al. (2015). On the other hand, an emoticon is a combination of letters and punctuation marks arranged to make a facial expression Hogenboom et al. (2013); Kralj Novak et al. (2015); Srisankar and Lochanambal (2021); Ullah et al. (2020). Emoticons were used even before emojis came. Emojis and emoticons add an extra layer of special emotional meaning to the plain text messages Karthikeya et al. (2023). In day-to-day offline communications, we rely more on body language, intonations, and facial expressions to understand how the other person feels. But in digital communication, these non-verbal cues are missing. That's where, both emojis and emoticons come into the picture and aids in a better understanding of the intent of the communication; thereby reducing the chances of being misunderstood Hogenboom et al. (2013); Santos et al. (2023).

Received on: 11 Jun 2026 | Accepted on: 01 Jul 2026 | Published Online on: 25 July 2026

Because of this, both emojis and emoticons play a key role in sentiment analysis by giving direct clues about how someone feels. These symbols provide extra emotional context that helps computers better understand the realsentiment of a message. Without these symbols, it can be easy to misinterpret the tone of the messageAyvaz and Shiha (2017); Lou et al. (2020).

In this paper, a comprehensive literature review is conducted for the work done in the area of sentiment analysis on emoji-rich text. The research objectives of this paper are mentioned below:

Research Objectives

1. RQ1: What are the characteristics of existing datasets, like size, scope, annotations used, web-scrapped or downloaded etc?
2. RQ2: What are the preprocessing, feature extraction and machine learning techniques used to handle text containing emojis?
3. RQ3: What are the specialized methods for handling emojis?
4. RQ4: What are the key challenges and research gaps in sentiment analysis of emoji-rich text?

Organization of the Paper

The paper is divided into eleven sections. Section 2 outlines the systematic survey methodology. Section 3 explores various datasets used by researchers. Machine learning pipeline including the preprocessing and feature engineering is discussed in sections 4, 5 and 6. Sections 7 and 8 highlights the analysis done on the research trends on quantitative and qualitative basis. Section 9 discusses various challenges of handling emojis. Section 10 highlights the research gaps and synthesis of our research questions RQs. Finally, section 11 concludes our study and emphasizes on the future work in this domain.

2. Survey Methodology

The following section describes the methodology used in this review paper:

- **Databases and Sources:** Literature was sourced from major academic repositories like IEEE Xplore, ACM Digital Library and ScienceDirect to ensure high-quality peer-reviewed content.
- **Search Strings:** The search string used is: *("sentiment analysis" OR "emotion recognition") AND ("emoji" OR "emoticons") AND ("ma-chine learning" OR "deep learning" OR "NLP" OR "Natural Langugae Processing")*
- **Selection Criteria:** Papers were included based on their direct relevance to the research objectives (RQ1-RQ4). Specifically, we prioritized studies that introduced unique emoji-rich datasets, novel preprocessing pipelines, or specialized emoji embedding and modeling architectures. Studies that treated emojis merely as noise to be removed without specialized interpretation were generally excluded to focus on emoji-aware methodologies.
- **Time Window:** The reviewed literature spans from the period of 2010 to 2025; showcasing the evolution from initial emoticon based analysis to modern AI generative approaches.
- **Study Selection Counts:** A total of 28 primary research papers were selected for in-depth analysis and comparative benchmarking (Note: While 28 papers were included for analysis, approximately 150 papers were initially screened based on titles and abstracts to reach this final selection).

This survey establishes emojis as the primary carriers of sentiments due to their contextual role assignments and multidimensional emotional design. Existing reviews lacks the systematic rigor and linguistic diversity of this review, which is conducted in 19 languages and 32 countries. This review also attempts to synthesize empirical performance benchmarks and evaluates the application of multimodal generative AI for zero-shot cross-lingual analysis.

3. Characteristics of the Datasets

As with all attempts at natural language processing, the success of sentiment analysis that incorporates emojis depends upon the use of an appropriate dataset. As such, the characteristics of the available datasets for sentiment analysis that incorporate emojis were reviewed in this stage of the survey to determine their relevant features. Such features include the origin of the dataset, the size of the dataset, the linguistic diversity of the dataset, and the diversity of emojis included within those datasets.These dataset characteristics directly determine the robustness, accuracy, and generalizability.

This section compares and contrasts these dataset characteristics with respect to several influential papers in depth for further understanding. The comparison framework focuses on: the nature of the data source, the dataset's scale, its acquisition method (ready-made download vs. researcher-scraped), the extent of emoji inclusion for modelling and linguistic diversity of the datasets used.

3.1 Dataset Sources

The analysis of existing research reveals a distinct predominance of social media platforms as the primary source for datasets used in emoji and emoticon sentiment analysis. Twitter emerges as the overwhelmingly dominant platform,

explicitly cited in majority of the listed studies Ayvaz and Shiha (2017); et al. (2020); Godard and Holtzman (2022); Jagadishwari et al. (2021); Jona et al. (2022); Karthikeya et al. (2023); Kralj Novak et al. (2015); Mona M. Abd Elsalam (2022); Ullah et al. (2020); Vashisht and Jailia (2020); Yoo and Rayz (2021); Yuan et al. (2022). Its prevalence is attributed to its public nature, high volume of emoji usage, and accessible APIs (e.g., Twitter Stream API, Twitter Stream Grab project) facilitating data collection, whether through direct downloads or web scraping. Several Twitter datasets focus on specific events or timeframes, such as COVID 19, New Years Eve , Istanbul Attack , or US Airline sentimentKhan et al. (2025); Ullah et al. (2020). Kaggle Mona M. Abd Elsalam (2022); Yoo and Rayz (2021) and GitHub are other curated data repositories which are readily used by the researchers.

3.2 Size and Scope of Datasets

The scale and extent of the datasets used for emoji sentiment analysis differ greatly. While, on the one hand, there are narrowly scoped datasets like the few images Alfreihat et al. (2024), there also exist massive ones such as the Twitter dataset including millions of tweets Yin et al. (2023); Ullah et al. (2020). This can be seen from the fact that the amount of data is even reflected in the Twitter Stream Grab dataset amounting to 3.8 TB of storage space Tang et al. (2024). In regards to scope, most of the datasets are targeted at a very particular context or even, for example the COVID 19 dataset Zhao et al. (2024), terrorism Vashisht and Jailia (2020), or airline sentiment Santos et al. (2023); Wang and Castanon (2019). Although many of the datasets are generated from only one platform, namely Twitter Mona M. Abd Elsalam (2022); Yin et al. (2023); Wang and Castanon (2019); Ullah et al. (2020); Toet et al. (2018), the remaining ones try to provide a broad scope using data from various sources, such as -various social media platforms Hakami et al. (2022); Khan et al. (2025)-, Facebook or gaming forums.

In any case, the high degree of variability presents an ongoing problem, since researchers have to consider the trade-off between the possibilities presented by big data, narrow-scoped studies and the feasibility and effectiveness of the approach to emoji sentiment analysis.

3.3 Downloaded vs. Web Scraped Data

The method used for the collection of the dataset is often categorized as downloaded datasets or scrapped datasets. For example, many datasets are *downloaded* from their platforms. These downloaded datasets can include anything from the initial release of the sentiment analysis dataset, to datasets that were collected using the social media platforms APIs, and from the various research data sharing platforms. Furthermore, many sentiment analysis datasets are also *scraped* from social media platforms. As with downloaded datasets, scraped datasets can also include data from various social media platforms, such as Twitter, Facebook, gaming forums, or social media platforms in general. Finally, as with downloaded datasets, the data that is scraped can also be targeted towards specific topics of interest. For instance, sentiment analysis research has used scraped data to focus upon social media discussions of topics such as the COVID 19 pandemic or the Istanbul attack.

The studies Alfreihat et al. (2024); Hakami et al. (2022); Jagadishwari et al. (2021); Khan et al. (2025); Mona M. Abd Elsalam (2022); Santos et al. (2023); Tang et al. (2024); Toet et al. (2018); Yin et al. (2023); Yoo and Rayz (2021) used downloaded datasets while the studies Alawneh et al. (2024); Ayvaz and Shiha (2017); et al. (2020); Godard and Holtzman (2022); Hogenboom et al. (2013); Holthoff (2020); Jona et al. (2022); Karthikeya et al. (2023); Kralj Novak et al. (2015); Lou et al. (2020); Srisankar and Lochanambal (2021); Ullah et al. (2020); Vashisht and Jailia (2020); Wang and Castanon (2019); Yoo and Rayz (2021); Yuan et al. (2022); Zhao et al. (2024) used web scraped datasets. The study Yoo and Rayz (2021) used a mix of both approaches for datasets, three of its datasets are downloaded, and one is web scraped.

3.4 Emoji Coverage

Once researchers have their data (whether downloaded or scraped), they face a critical hurdle, i.e., deciding on which emojis shall be included in their study (coverage). These choices profoundly shape the understanding and usefulness of the model. Most studies report coverage between 3 and 21 percent, while several omitting exact figures. The studies reported that emoji coverage ranges from 2.8 percent in the dataset (Istanbul Attack tweets, 2017) of Kralj Novak et al. (2015) to 100 percent in Tang et al. (2024); Srisankar and Lochanambal (2021).

3.5 Linguistic Diversity

Research on emoji sentiment analysis uses datasets that span several major languages in the world, reflecting the global nature of digital communication. English is the most consistently represented language, appearing in dedicated studies or alongside other languages in multilingual datasets. It is explicitly covered in studies analyzing Portuguese and English tweets (e.g., elections and World Cup data) , Dutch and English social media messages , and Chinese and English microblogs. Arabic has emerged as a major focus in contemporary research, with multiple dedicated studies leveraging large Arabic datasets Alawneh et al. (2024); Alfreihat et al. (2024); Hakami et al. (2022). Portuguese features in targeted bilingual research alongside English . Chinese is prominently represented, particularly through large-scale collections of

Sina Weibo microblogs Lou et al. (2020); Zhao et al. (2024). The broadest linguistic scope comes from a large-scale multilingual study encompassing 13 distinct languages, analyzing tweets from diverse linguistic backgrounds. Dutch is also included, specifically analyzed in conjunction with English social media data.Table 1 summarizes the findings.

Table 1: *Comparison of Emoji-Rich Studies*

Ref	Year	Language	Platform/Source	Size	Label Set	Method	Emoji Statistics	Authentic Access Link
[1]	2024	Arabic	Kaggle	56,000	3-way	Scraped	912 Mapped Emojis	https://www.kaggle.com/datasets/mksaad/arabic-sentiment-twitter-corpus
[2]	2024	Arabic	GitHub	58,000	3-way	Downld.	3.82%	https://github.com/manaralfreihat/Emoji-Sentiment-Lexicon
[3]	2017	English	Twitter	101k/73k	3-way	Scraped	19.38%	http://www.independent.co.uk/news/world/europe/istanbul-nightclub-attack-shooting-turkey
[4]	2020	English	Twitter	Not stated	3-way	Scraped	Not stated	Twitter (Surikov et al.)
[5]	2022	English	Twitter API	3,000,000	3-way	Scraped	21 Emoji Symbols	https://www.frontiersin.org/articles/10.3389/fpsyg.2022.921388/full
[6]	2022	Arabic	Twitter	144,196	3-way	Downld.	Not stated	https://archive.org/details/twitterstream
[7]	2013	Dutch/Eng	Google/Twitter	2,080	3-way	Scraped	100%	Twitter / Google Discussion Groups
[8]	2020	English	Various SM	Not stated	3-way	Scraped	Not stated	Various Social Media Platforms
[9]	2021	English	Kaggle	3,000,000	3-way	Downld.	Not stated	https://www.kaggle.com/youben/twitter-sentiment
[10]	2022	English	Twitter/Forum	Not stated	3-way	Scraped	Not stated	Twitter / Gaming Forums
[11]	2023	English	Twitter API	Not stated	3-way	Scraped	Not stated	Twitter Stream API
[12]	2025	English	Kaggle	14,640	3-way	Downld.	6 Emoji Symbols	https://www.kaggle.com/crowdflower/twitter-airline-sentiment
[13]	2015	English	Twitter	1,600,000	3-way	Scraped	100 Emoji Symbols	Twitter (Kralj Novak et al.)
[14]	2020	Chinese/Eng	Sina Weibo	10,042	3-way	Scraped	Not stated	Sina Weibo Platform
[15]	2022	English	Kaggle	10,000	3-way	Downld.	Not stated	https://www.kaggle.com/imranzaman5202/starter-code-covid-19-sentiment-analysis
[16]	2017	Multiling.	Facebook/Tw.	460 KW	25 Emot.	Scraped	Not stated	Facebook / Twitter
[17]	2023	Port./Eng	Kaggle	84,115	3-way	Downld.	Not stated	https://www.kaggle.com/datasets/adilmar/twitter-pt
[18]	2024	Code-mixed	Zenodo	20,000	Multi-label	Downld.	Not stated	https://zenodo.org/record/3974927
[19]	2021	13 Langs	Twitter	1,600,000	3-way	Scraped	4 Emoji Symbols	Twitter (Srisankar et al.)
[20]	2024	English	Multi-Twitter	Multi-DS	3-way	Downld.	Not stated	Twitter (Tang et al.)
[21]	2018	N/A	FRIDa	60 Images	Emotion	Downld.	100%	FRIDa Image Database
[22]	2020	English	Twitter/Reviews	14,460	3-way	Scraped	5.06%	Twitter / Airline Reviews
[23]	2020	English	Various SM	Not stated	3-way	Scraped	Not stated	Various Social Media Sources
[24]	2019	English	Twitter API	13,000,000	3-way	Scraped	50 Emoji Symbols	https://uvaauas.figshare.com/articles/dataset/TwemojiDataset/5822100
[25]	2023	English	Twitter Grab	184,485,91	93-way	Downld.	Not stated	https://archive.org/details/twitterstream
[26]	2021	English	Kaggle/GitHub	1,800,000	3-way	Downld.	Not stated	Kaggle / GitHub / Twitter API
[27]	2022	English	Twitter	Not stated	3-way	Scraped	Not stated	Twitter (Yuan et al.)
[28]	2024	Chinese	Sina Weibo	6,100,000	3-way	Scraped	> 50 Emoji Symbols	Sina Weibo API

3.6 Ethical Considerations in Data Acquisition

Data acquisition through web-scraping or downloading from repositories has some serious ethical concerns. Firstly, the terms of service must be respected for the public platforms like Twitter, Kaggle, and Facebook, which are the prime

sources of data collection. Next, the privacy of personal information must be protected by de-identifyingtext, such as removing personal details like usernames or geographic coordinates, so that no single person can be traced back through their unique use of emojis. Further, sharing of the data must be done with caution, by utilizing research licenses on platforms like Kaggle or sharing only post IDs rather than raw text. Lastly, fairness and representation must be addressed by addressing potential biases, which helps minimize the confounding effects of gender or culture in the results.

3.7 The Emoji-Aware Sentiment Analysis Pipeline

The main steps that are often taken to process emoji-rich text are generally divided into three main stages. This ensures that the raw symbols are translated into sentiment signals that can be interpreted by the machine.

3.8 Stage 1: Emoji Preprocessing and Normalization

The first step in this emoji-enabled process involves processing of raw symbols in preparation for modeling, identifying emojis that often affect sentiments through emphasis, clarification and reversing. Categorical filtering is an important part of this process, wherein researchers filter the emojis using categorical filters such that only facial expressions and gestures are selected, and other non-relevant information is stripped off, thus avoiding any form of cultural and gender bias during the analysis process. Substitution and encoding strategies are used to translate emojis into descriptive text to maintain consistency across diversified platforms. Furthermore, similar emojis are grouped using normalization techniques to minimize feature sparsity.

3.9 Stage 2: Feature Representation and Vectorization

The second stage in sentiment analysis that incorporates emojis is feature representation and vectorization of the emojis. More specifically, emojis are often represented via the use of either static sentiment values or dynamic emoji embeddings to represent each emoji within a text to be analyzed. Lexicon-based mapping is one of the static mapping techniques readily used, where in dictionaries like EmojiNet or Emoji Sentiment Lexicon (ESL) is used to map emojis with a polarity score. This score is calculated by finding the difference between the positive and negative polarity of the emoji. Furthermore, dynamic embeddings include the use of Emoji2Vec, DeepMoji, or EmoGraph2vec. These map each emoji to a dynamic high-dimensional vector that represents the emoji's sentiment towards text.

3.10 Stage 3: Classification Architectures

The final and last main stage for sentiment analysis that incorporates emojis is the use of various classification models (Machine Learning or Deep Learning) to classify the text and determine the sentiment of that text.

4. Preprocessing of Features

Data cleaning and preprocessing, a crucial step in text mining, removes noise from data, reduces the vocabulary size, and prepares raw data for modeling. A clean and standardized input to machine learning models makes them efficient and more accurate. Some of the commonly used data cleaning and preprocessing techniques are mentioned in Table 2. Lowercase, usually the first step, standardizes the text by converting all the characters to lowercase, tokenization Alawneh et al. (2024); Alfreihat et al. (2024); Ayvaz and Shiha (2017); Hogenboom et al. (2013); Jagadishwari et al. (2021); Karthikeya et al. (2023); Khan et al. (2025); Kralj Novak et al. (2015); Lou et al. (2020); Mona M. Abd Elsalam (2022); Rahman et al. (2017); Singh et al. (2024); Srisankar and Lochanambal (2021); Tang et al. (2024); Ullah et al. (2020); Vashisht and Jailia (2020); Yin et al. (2023); Yoo and Rayz (2021); Zhao et al. (2024) splits the text into smaller units which can be words, sentences, paragraphs, documents, etc. Non-informative symbols like punctuation marks and special characters are handled mostly by removing them. Stop words, Alfreihat et al. (2024); Ayvaz and Shiha (2017); Jagadishwari et al. (2021); Jona et al. (2022); Khan et al. (2025); Kralj Novak et al. (2015); Lou et al. (2020); Rahman et al. (2017); Tang et al. (2024); Ullah et al. (2020) more commonly used words such as 'the', 'is', 'and', 'or' are removed to reduce the impact of less important words. Stemming Alawneh et al. (2024); Alfreihat et al. (2024); Ayvaz and Shiha (2017); Khan et al. (2025); Kralj Novak et al. (2015); Tang et al. (2024) reduces words to their stem by removing the suffix, and lemmatization Jona et al. (2022); Khan et al. (2025) uses grammar rules to reduce the words to their base form. Both the processes reduces the redundancy. Spell correction is an obvious process for cleaning the text. Slangs are removed and informal spellings are handled using the process of normalization.

Part of speech tagging, Hogenboom et al. (2013); Lou et al. (2020) an advanced preprocessing technique, tags the text into grammatical roles such as nouns, pronouns, adjectives and so on. Named Entity recognition, often known as NER identifies the real world entities.

Emojis are either removed from the text or handled in a special manner by converting them to a descriptive token, often sentimental. Different ways of handling the emojis are elaborated in the next section.

4.1 Special Processing Techniques of Emojis

Emojis play a vital role in digital communication, significantly influencing the interpretation of sentiments. Re-search reveals that a single emoji can be associated with multiple functionalities, i.e. it may emphasize existing sentiment,

Table 2: *Preprocessing Techniques*

Preprocessing step	Reference
Stopword removal	Alfreihat et al. (2024); Ayvaz and Shiha (2017); Jag adishwari et al. (2021); Jona et al. (2022); Khan et al. (2025); Kralj Novak et al. (2015); Lou et al. (2020); Rahman et al. (2017); Tang et al. (2024); Ullah et al. (2020)
Stemming	Alawneh et al. (2024); Alfreihat et al. (2024); Ayvaz and Shiha (2017); Khan et al. (2025); Kralj Novak et al. (2015); Tang et al. (2024)
Lemmatization	Jona et al. (2022); Khan et al. (2025)
Part-of-speech (POS)	Hogenboom et al. (2013); Lou et al. (2020)
Tokenization	Alawneh et al. (2024); Alfreihat et al. (2024); Ayvaz and Shiha (2017); Hogenboom et al. (2013); Jagadishwari et al. (2021); Karthikeya et al. (2023); Khan et al. (2025); Kralj Novak et al. (2015); Lou et al. (2020); Mona M. Abd Elsalam (2022); Rahman et al. (2017); Singh et al. (2024); Srisankar and Lochanambal (2021); Tang et al. (2024); Ullah et al. (2020); Vashisht and Jailia (2020); Yin et al. (2023); Yoo and Rayz (2021); Zhao et al. (2024)
Normalization	Alfreihat et al. (2024)

introduce new sentiment, mitigate harshness, reverse the polarity of a message, trigger positive or negative reactions, or remain neutral. For instance, the same emoji might intensify positivity in one context (e.g., -Perfect! :)-) but can convey negativity in some other context (e.g., -Not great :)-). Apart from modifying the meaning of the sentences, a single emoji can often convey the meaning of the entire sentence. These techniques can be further categorized into the following distinct approaches. (ref Fig. 1)

Emoji Sentiment Lexicons are predefined dictionaries that often assign fixed sentiment scores (e.g., -1 to +1) or polarity labels (positive/ negative/ neutral) to emojis based on human annotations. For instance, a smiley emoji can be assigned a high positive score of +0.97 while a strong negative score of -0.97 can be assigned to a sad face emoji. The emojis are

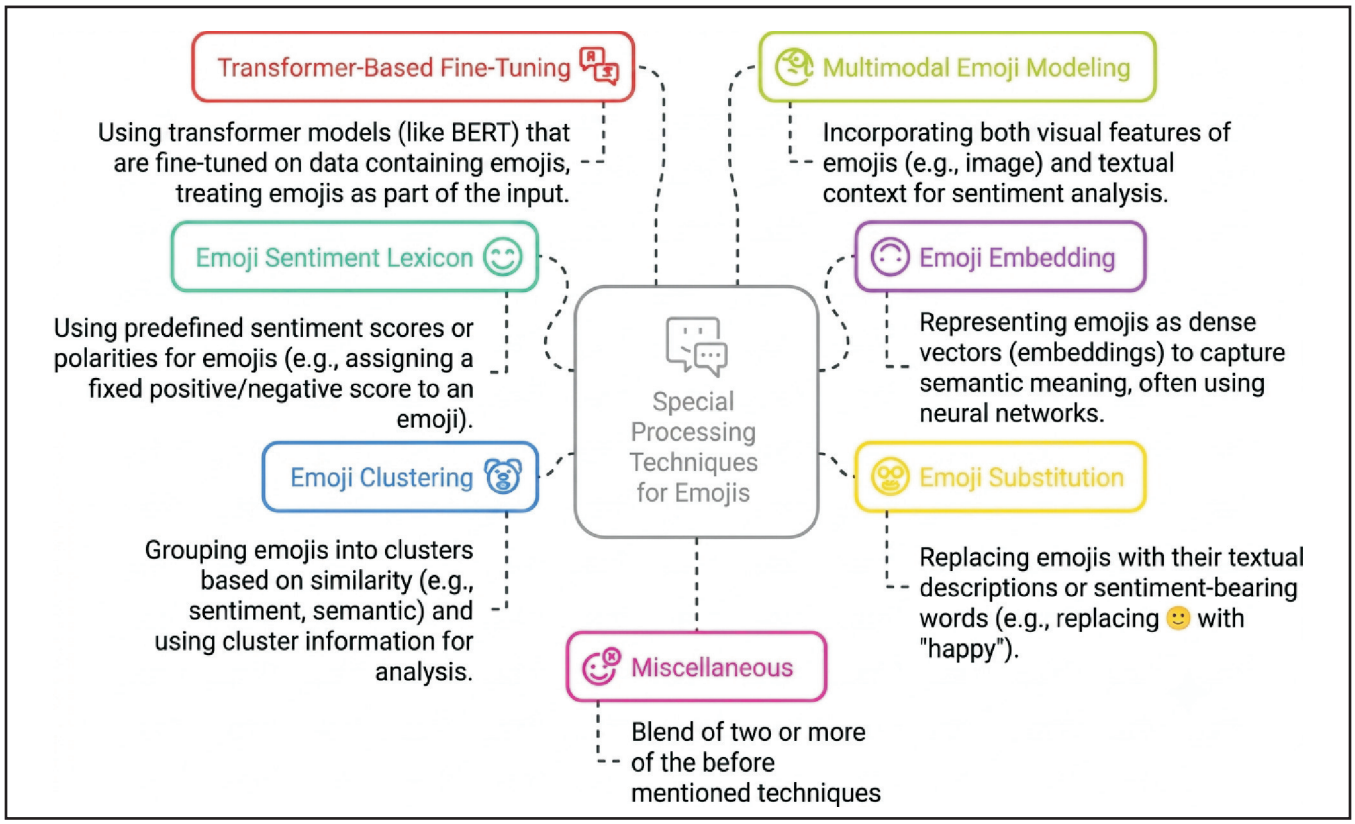


Fig. 1: *Techniques used for emoji processing.*

then replaced by their assigned scores or labels, thereby allowing incorporation of these emojis with the traditional classifiers. The studies Alfreihat et al. (2024); Ayvaz and Shiha (2017); Godard and Holtzman (2022); Hakami et al. (2022); Hogenboom et al. (2013); Holthoff (2020); Karthikeya et al. (2023); Khan et al. (2025); Santos et al. (2023); Singh et al. (2024); Srisankar and Lochanambal (2021); Toet et al. (2018); Vashisht and Jailia (2020); Yoo and Rayz (2021) use this technique for handling the emoji rich text.

Emoji Embedding techniques convert emojis into dense vector representations by using algorithms such as Word2Vec, FastText, or contextual models (BERT). These embeddings capture semantic similarities, grouping related emojis (e.g., heart emoji and face with hearts in eyes emoji cluster together due to shared emotional connotations). This method also encodes multidimensionality (e.g., fire emoji simultaneously conveys threat, hype, or cultural relevance). The studies Ayvaz and Shiha (2017); et al. (2020); Karthikeya et al. (2023); Mona M. Abd Elsalam (2022); Singh et al. (2024); Srisankar and Lochanambal (2021); Wang and Castanon (2019); Yoo and Rayz (2021); Yuan et al. (2022); Zhao et al. (2024) demonstrated the use of this technique for processing of emojis.

Emoji Substitution can convert emojis into machine-processable tokens by replacing emojis with text equivalents. For instance, explicit labels (e.g., ″[happy face]″ are used for happy face emoji) or contextual keywords (e.g., ″cele-bration item″ for its corresponding emoji). This substitution allows the application of traditional NLP techniques for processing of this emojis. Alawneh et al. (2024); Jagadishwari et al. (2021); Karthikeya et al. (2023); Khan et al. (2025); Rahman et al. (2017); Yoo and Rayz (2021)

Emoji Clustering groups semantically or sentimentally similar emojis using unsupervised algorithms like K-means or hierarchical clustering, typically applied to pre-trained emoji embeddings. For example, clustering might group laughing emojis into a ″joy/humor″ cluster, while all anger related emojis form an ″anger″ cluster. These clusters serve as categorical features in sentiment models, reducing dimensionality and generalizing across visually diverse but functionally equivalent emojis. This technique simplifies analysis by treating emoji variants (e.g., skin tones) as a single semantic unit. Alfreihat et al. (2024); Hakami et al. (2022); Wang and Castanon (2019); Yuan et al. (2022)

Transformer-Based Fine-Tuning for Emojis This approach fine-tunes pretrained transformer models (e.g., BERT, RoBERTa) on domain specific, emoji rich datasets, treating emojis as first-class tokens during training. The model learns context-dependent representations by analyzing emojis alongside surrounding text e.g., distinguishing positive sentiment in ″Aced the exam! party popper emoji″ from negative connotations in ″Bankrupt party popper emoji.″ Unlike static lexicons or embeddings, transformers capture sarcasm, irony, and cultural nuances by weighing emojis within full-sentence contexts. et al. (2020); Khan et al. (2025); Singh et al. (2024); Tang et al. (2024); Zhao et al. (2024)

Multimodal Emoji Modeling jointly analyzes the visual properties of emojis (e.g., color, shape) and their textual context using architectures like ViLBERT or CLIP. These models fuse image features (extracted via CNNs) with text embeddings to interpret emojis holistically e.g., recognizing that ″leaf emoji″ might signify nature (green leaves) or marijuana (cultural symbolism) based on combined visual-textual signals. This is critical for emojis where meaning derives from imagery (e.g., spark emoji as ″explosion″ or ″impact″). Applications include sentiment analysis in platforms where emojis substitute words (e.g., messaging apps). Jagadishwari et al. (2021); Lou et al. (2020); Zhao et al. (2024)

4.2 Advanced Social Media Preprocessing

Social media text is often unstructured and informal. This makes it difficult to interpret the sentiment correctly. This issue can be addressed by using the following preprocessing techniques.

- **Unicode Normalization and Standardization:** Since emoticons and emojis can be represented differently across the different social media platforms, Unicode Normalization is often used to convert them into astandard format Kralj Novak et al. (2015). In some studies, emojis are converted into their standard form before tokenization. This leads to the preservation of their emotional menaing Kralj Novak et al. (2015).
- **Handling Code-Mixed Text:** Maany rich emojis messages contain code-mixed text, where two or more langauges are used within the same message. This is commonly used in the *Sentimix* dataset; which is specifically designed for multi-lingual texts Yuan et al. (2022). Specialized embedding techniques are used for processing such datasets.
- **Hashtag and Keyword Parsing:** Social media data is often collected using hashtags and keywords Singh et al. (2024). During preprocessing, these hashtags are separated from the main text and handled individually Singh et al. (2024). Since, hashtags can contain multiple words (e.g., #NewYearsEve), they are split into individual words Vashisht and Jailia (2020).
- **Normalization of Informal Language:** Digital text often contains informal words, slangs and abbreviations. Text Normalization is used for converting these into their standard meanings. This in turn preserves their sentiments Zhao et al. (2024); Toet et al. (2018); Kralj Novak et al. (2015).

Miscellaneous In addition to the methods described above, there are other approaches to preprocessing social media texts that incorporate some form of emoji recognition and analysis. These hybrid approaches leverage complementary strengths to overcome limitations of single-method implementations. The study Santos et al. (2023) proposes an EmojiMapper function for its preprocessing filters to map emojis to textual meanings, after which a lexicon was used to determine sentiments of the messages. The study Ullah et al. (2020) took a phased methodology: first using 732 hand-classified emoticons (positive/negative/neutral) as lexicon features alongside text representations, then removing them entirely to measure performance impact. This strategy prioritizes interpretability while mitigating context-blindness through comparative analysis. The study Jona et al. (2022) uses the blend of the embeddings and modeling synergy leveraging deep learning to capture contextual dynamics through the use of DeepMoji SE. Emojis were retained as contextual signals during transformer-based fine-tuning on GitHub data, generating emotion-rich embeddings that fed into their SEntiMoji classifier. In later work, they introduced EmoGraph2vec, creating graph-based emoji embeddings fused with text via hybrid attention mechanisms before classification through TextCNN architectures. These approaches treat emojis as semantic anchors, using their distributional properties to bootstrap contextual understanding rather than relying on fixed dictionaries.

5. Feature Extraction Techniques

Text having emojis is converted into features that can be used further for classification in various machine learn-ing models (refer Table 3. Approaches like Bag of words Ayvaz and Shiha (2017); Jagadishwari et al. (2021); Karthikeya et al. (2023); Mona M. Abd Elsalam (2022); Srisankar and Lochanambal (2021); Wang and Castanon (2019), TF-IDF Ayvaz and Shiha (2017); et al. (2020); Jona et al. (2022); Karthikeya et al. (2023); Khan et al. (2025); Rahman et al. (2017); Santos et al. (2023); Singh et al. (2024); Toet et al. (2018); Wang and Castanon (2019), ngram Toet et al. (2018); Vashisht and Jailia (2020) have been used to extract tokens from the text. However, several approaches like emoji sentiment ranking Godard and Holtzman (2022); Hakami et al. (2022); Hogenboom et al. (2013); Kralj Novak et al. (2015); Toet et al. (2018); Ullah et al. (2020) and AFINN are used for extraction of tokens from emojis. A lexicon based approach, emoji sentiment ranking provides sentimet scores for indi-vidual emojis. A large dataset of 1.6 million tweets having emojis was used, and each emoji in the dataset was assigned a polarity value depending on the occurrence of the same in tweets. Sentiment score was calculated using the formula Sentiment Score = Positive Probability - Negative Probability. Another lexicon based approach, AFINN asssigns predefined sentiment scores to English words, ranging from -5 (very negative) to +5 (very positive).

Word embeddings like Word2Vec Alfreihat et al. (2024); Khan et al. (2025); Yoo and Rayz (2021); Yuan et al. (2022), GloVe Karthikeya et al. (2023); Wang and Castanon (2019), Emoji2Vec Godard and Holtzman (2022); Holthoff (2020), deep-moji Godard and Holtzman (2022) are deep learning-based approaches that capture semantic relationships. Several Emoji-aware embeddings, Emoji2Vec Godard and Holtzman (2022); Holthoff (2020), FastText have been used by researchers that combine pre-trained word embeddings with emoji embeddings.

6. Classification Techniques

Several machine learning (ML) and deep learning techniques are used for sentiment classification using Emojis, mentioned in Table 3. Some of the traditional machine learning methods include multinomial Naive Bayes Alfreihatet al. (2024); et al. (2020); Jagadishwari et al. (2021); Mona M. Abd Elsalam (2022); Santos et al. (2023); Toet et al. (2018); Ullah et al. (2020); Vashisht and Jailia (2020), SVM Alfreihat et al. (2024); Ayvaz and Shiha (2017); et al. (2020); Hakami et al. (2022); Jagadishwari et al. (2021); Karthikeya et al. (2023); Lou et al. (2020); Mona M. Abd Elsalam (2022); Santos et al. (2023); Toet et al. (2018); Ullah et al. (2020); Vashisht and Jailia (2020), Random Forest et al. (2020); Mona M. Abd Elsalam (2022); Ullah et al. (2020), K-Nearest Neighbors Hakami et al. (2022); Mona M. Abd Elsalam (2022); Ullah et al. (2020), and Logistic Regression Alfreihat et al. (2024); Jagadishwari et al. (2021); Mona M. Abd Elsalam (2022). Most of these methods use features like word frequency, TF-IDF, emojis sentiment score, sentiment lexicons, part of speech tags.

Deep learning techniques automatically extract features from text data that include emojis. Some of the commonly used techniques are Convolutional Neural Networks (CNNs) Alawneh et al. (2024); Hakami et al. (2022); Karthikeya et al. (2023); Lou et al. (2020), Recurrent Neural Networks (RNNs), LSTM (Long Short-Term Memory) Alawneh et al. (2024); Hakami et al. (2022); Karthikeya et al. (2023); Lou et al. (2020); Tang et al. (2024), GRU (Gated Recurrent Units) Alfreihat et al. (2024), BiLSTM + Attention Yuan et al. (2022), Emoji2Vec / Word2Vec + LSTM. While CNNs works on local n-gram features, RNNs work on sequential information in text. LSTM, a type of RNN, remembers long-term dependencies and handles varying input lengths, making it suitable for texts having emojis. GRU, a faster implementation of LSTM, is effective in handling sequential data like text. BiLSTM and Attention mechanisms focus on the relevance part of the text.

Further, advancements in transformer-based models have enhanced the capabilities to understand the context of text. Bidirectional Encoder Representations from Transformers (BERT) model Khan et al. (2025) handles the context-dependent emoji usage in the text. RoBERTa, DistilBERT, XLNet are the lighter varients of BERT. Emoji-BERT is specially pre-trained on emoji dataset.

6.1 Evaluation Techniques

Most of the researchers have used accuracy as the evaluation metric for traditional datasets and , precision, recall and F1 score for imbalanced datasets. Ayvaz and Shiha (2017); et al. (2020); Hogenboom et al. (2013); Holthoff (2020); Jagadishwari et al. (2021); Karthikeya et al. (2023); Khan et al. (2025); Lou et al. (2020); Mona M. Abd Elsalam (2022); Singh et al. (2024); Tang et al. (2024); Ullah et al. (2020); Vashisht and Jailia (2020); Yin et al. (2023); Yuan et al. (2022) However, other evaluation metrics are also used like Jaccard index for Emoji ranking Singh et al. (2024) and overlap analysis, hamming loss for Multi-label emoji tagging Singh et al. (2024), Top k-accuracy for Emoji recommendation/prediction.

Table 3: *Feature Extraction and ML/DL Techniques*

Ref	Task / Language	Dataset Size	Feature Extraction	ML/DL Methods	Best Performance
[1]	Arabic SA	56,000	Word Embedding (Emoji Encoding)	CNN-LSTM	Acc (91.85%), F1(0.92)
[2]	Arabic SA	58,000	Word2Vec, Emoji Feature Vectors	Bi-GRU, SVM, NB, LR	Acc (89%), F1(0.89)
[3]	English SA	175,181	BoW, TF-IDF, Word Embeddings	SVM, Neural Networks	Acc (81%), F1(0.80)
[4]	English SA	NES	TF-IDF, Word Embedding	SVM, Naive Bayes, RF	Acc (82.10%), F1(0.82)
[5]	English Emotion	3,000,000	NRC Lexicon, Emoji2Vec, Deep-moji	Deep CNN, MobileNet	Acc (79.1%), F1(0.79)
[6]	Arabic SA	144,196	ASAD, CAMEL, MAZAJAK, Lexicon	SVM, KNN, CNN, LSTM	Acc (85.3%), F1 (0.85)
[7]	Dutch, English SA	2,080	Emoji Sentiment Lexicon	Lexicon-based	Acc (74.2%), F1(0.75)
[8]	English SA	NES	Emoji2Vec	Unsupervised Lexicon SA	Acc (81.43%), F1 (0.81)
[9]	English SA	3,000,000	BoW	Multinomial NB, SVM, Regr.	Acc (86.4%), F1 (0.86)
[10]	English Prediction	NES	TF-IDF	CNN, Bi-LSTM, Softmax, RNN	Acc (82%), F1 (0.82)
[11]	English SA	NES	BoW, TF-IDF, GloVe	SVM, CNN, LSTM	Acc (87%), F1 (0.87)
[12]	English SA (Airlines)	14,640	TF-IDF, Word2 Vec, BERT	BERT, SVM, RF, NB, LR	Acc (88%), F1 (0.88)
[13]	English SA	1,600,000	Lexicon Matching, Rule Weighting	Custom Rule-based	Acc (76.8%), F1 (0.76)
[14]	Chinese/English SA	10,042	Word Embedding	CNN, LSTM, SVM	Acc (96.91%), F1 (0.92)
[15]	English SA (COVID)	10,000	BoW	SVM, RF, GB, KNN, GNB	Acc (79.6%), F1 (0.79)
[16]	Multiling. Emotion	460 KW	TF-IDF	KNN, Rule-based, PMI	Acc (88%), F1 (0.82)
[17]	Portuguese, Eng SA	84,115	TF-IDF, Word Embeddings	SVM, Naive Bayes	Acc (95%), F1 (0.91)
[18]	Code-mixed SA	20,000	Char/Elmo Embeddings, TF-IDF	Multi-task, Transformer	Acc (86%), F1 (0.84)
[19]	13 Languages SA	1,600,000	BoW	CNN, NB, SVM, Max Ent	Acc (80%), F1 (0.82)
[20]	English SA	Multi-DS	Word Embeddings, Topic Model	LSTM, Bi-LSTM	Acc (91%), F1 (0.93)
[21]	Food Emotion	60 Images	TF-IDF, N-grams, Emoji Lexicon	NB, SVM, Decision Tree	Acc (73.4%), F1 (0.76)
[22]	English SA (Airline)	14,460	Emoji/NRC Lexicons, NileUlex	SVM, NB, KNN, RF	Acc (87%), F1 (0.86)
[23]	English SA	NES	N-grams (Unigrams, Bi-grams)	SVM, Max Ent, NB	Acc (83.6%), F1 (0.81)
[24]	English Prediction	13,000,000	BoW, TF-IDF, GloVe	CNN, MNB, SVM, RNN	Acc (84%), F1 (0.82)
[25]	English SA	184,485,919	Distant Supervision	CNN, Transfer Learning	Acc (78%), F1 (0.8)
[26]	English SA	1,800,000	Word2Vec (CBOW/Skip-Gram)	CNN	Acc (72%), F1 (0.82)
[27]	English SA	NES	Word2Vec	Bi-LSTM, Hybrid Attention	Acc (89%), F1 (0.9)
[28]	Chinese SA	6,100,000	Manual Interpretation (EmojiGrid)	TaneNet with Attention	Acc (81.43%), F1 (82)

PMI-based classification refers to a sentiment or text classification approach that uses Pointwise Mutual Information (PMI) to measure the association between words (or phrases) and categories like positive, negative, or specific topics. PMI is often used to calculate how strongly a word is associated with positive or negative sentiment.

Emoji vectors and Word2Vec can overlap, but only if emojis are included in the training corpus and treated astokens. Otherwise, they are generated separately.

7. Quantitative Analysis of Research Trends

Based on a systematic review of the literature, several broad claims regarding the current state of sentiment analysis in emoji-rich text are supported by specific counts and trends identified among the 28 primary research papers analyzed.

- Dominance of Social Media and Twitter: Majority of the emoji-rich datasets are often curated from the social media. Among them, Twitter is the most commonly used platform; utilised by 12 studies Mona M. Abd Elsalam (2022); Ayvaz and Shiha (2017); Hakami et al. (2022); Wang and Castanon (2019); Jagadishwari et al. (2021); Lou et al. (2020); Tang et al. (2024); Vashisht and Jailia (2020); Alfreihat et al. (2024); Zhao et al. (2024); Rahman et al. (2017); Hogenboom et al. (2013). A smaller subset of the studies utilizes data from Facebook, Gaming forums and Sina Weibo Godard and Holtzman (2022); et al. (2020); Zhao et al. (2024); Alawneh et al. (2024).
- Variability in Dataset Scale and Acquisition: There is a considerable variation both in the size of the datasets and methods used for data collection. Size of the dataset ranges from FRIDa database containing only 60 images to a large scale dataset comprising of 184 million tweets Kralj Novak et al. (2015); Ullah et al. (2020). Among the reviewed studies, 17 collected data using web-scraping, while 11 used pre-compiled datasets Singh et al. (2024).
- Linguistic Diversity and Trends: English remains the most widely used language in the existing studies. However, gradually research is expanding and becoming more diversified. Holthoff (2020) uses 13 languages. Chinese is also widely studied through the microblogs collection from San Weibo et al. (2020); Alawneh et al. (2024). Similarly, Arabic has also gained significant attention and used in Yoo and Rayz (2021); Santos et al. (2023); Yin et al. (2023).
- Standardization of Preprocessing pipelines: Tokenization is the most widely used technique reported in 19 studies. Stopword Removal is the next commonly utilized technique used in 10 studies Ullah et al. (2020).
- Evaluation Metric: Accuracy is the standard metric often used for balanced datasets. Other metrics like Recall, F1-score and Precision are also used in 15 of the existing studies.
- Prevalance of Machine Learning Architectures: Support Vector Machine is the most commonly used traditional classifier used in 12 of the existing studies. Multinomial Naive Bayes is the next commonly used classifer reported in 8 studies, More advanced architectures like CNNs and LSTMs are also utilized.

8. Qualitative Analysis of Research Trends

- Performance vs Computational Trade-offs: Transformer-based models often achieve more accuracy as compared to the traditional classifiers. But they require more computational resources and training time.
- Scalability and Dataset Biases: Traditional ML models are trained using static and handcrafted features; which can easily be mimicked by the adversarial entities. Also, using Generative AI introduces latency; enforcing a limit on the scalability.
- Precision-Recall Tradeoffs: Using complex models like LLMs often lead to the increase in recall but causes precision to drop. This is mainly because the model becomes more hyper-sensitive to multilingual cues.
- Linguistic Evaluation: Stemming often leads to sparsity in complex languages like Arabic. This gap is addressed by translation of emojis into textual descriptions. This is turn allows the deep learning layers to make use of their pretraining vocabulary.

9. Challenges in Handling of Emojis

Emojis have high expressive power, yet they are treated as noise by many. Most researchers ignore emoji handling while dealing with the text due to the following challenges:

- Language and Regional Variability: Different regions, cultures have different ways of expressing the emotions. Emojis also fall in that category. Emojis do not always carry fixed meanings. Same emoji may be expressing different emotions in different regions and languages Khan et al. (2025); Mona M. Abd Elsalam (2022) . This leads to ambiguity and bias in the models. Most of the research work in this area is related to English language Santos et al. (2023). The same tools or techniques may not give accurate results for other languages. There is a strong need to have language specific emoji interpretations and cultural variability Yin et al. (2023).
- Context dependence: The emotion of an emoji is dependent on the context Khan et al. (2025); Mona M. Abd Elsalam (2022) . An emoji can have multiple meanings, known as Polysemy, e.g. the ”Red Heart” emoji can convey anger, hurt, rage, or madness and a keyword can be represented by multiple emojis, known as Synonymy Singh et al. (2024). Such emojis are difficult to interpret.
- Lack of Standard Representation: Often, emojis lack a proper way of vectorization. Most of the existing dictionaries ignore emojis. There is no universal dictionary, hence it is difficult to assign consistent sentiment scores. There is a need to have a robust emojis dictionary and proper mechanism for conversion of emojis to numeric forms, which can be further used in machine learning algorithms Mona M. Abd Elsalam (2022); Khan et al. (2025) . Further, use of joint emojis have increased with time but mechanism to understand the meaning of

joint emojis is missing in practice Santos et al. (2023). Emoticons and emojis are sometimes used sarcastically and lead to the creation of noisy labels in distant supervision Yin et al. (2023). There is no standard emojis lexicons e.g. Santos et al. (2023) uses one of more of a 100 emojis defined in their model.

- Short Text Issues: Recently, people have started using short texts having more emojis and emoticons to communicate. Such texts are casual and noisy and lack semantic expressions, leading to ambiguous interpretations Tang et al. (2024). Conversion of such texts leads to sparse features, making it difficult for NLP techniques to interpret the emotions or expressions from the same. Further, being small texts, emojis used in these texts should have more weightage than the emojis used in longer documents Khan et al. (2025).
- Model Integration of Text and Emoji: Text and emoji have different characteristics and need to be handled differently. However, current models combine both of them in the same model, making it ineffective for sentiment classification. Often, current models either ignore or misinterpret emojis when they are dealt with text Mona M. Abd Elsalam (2022).
- Generative AI Approach: Some of the pretrained language models (PLM) are not trained on emojis Tang et al. (2024) and the models trained on emojis are computationally expensive. They perform on emoji-aware input but training and fine tuning such models require heavy resources Khan et al. (2025).

10. Research Gaps and Synthesis

Despite significant advances in this research field, several research gaps still remain unresolved. Existing models can not recognize the change in the semantics of the emojis based on its intent or surrounding text. This issue is further worsened by the linguistic and the cultural bias present in the existing studies. Furthermore, non-availability of well labeled, large and multilingual datasets is another issue. Also, majority of the existing studies treat emojis as text and ignore their visual appearance. They also pay little attention to repeated or grouped emojis. In addition, different analysis methodologies often produce varied results; making it difficult to compare results. This also highlights the need for a standard techniques. Future studies shall involve multimodal approaches; thereby integrating user behavior, text and emojis to have a better understanding of digital communication.

10.1 RQ1: Characteristics of Existing Datasets

According to the analysis drawn from the literature, social media emerges as the primary source of emoji centric data. **Twitter** stands out to be the most promising one with the highest number of citations. This is mainly due to its open access and robust APIs Yoo and Rayz (2021).

- **Size and Scope:** The size of the datasets are extremely inconsistent in size, spanning from small focused collections of **60 images** to massive datasets such as the **3.8 TB Twitter Stream Grab** containing millions of entries Santos et al. (2023); Mona M. Abd Elsalam (2022).
- **Acquisition:** Researchers have typically accessed these datasets either by **direct downloads** (11 studies) or by **web-scraping** (17 studies). Scraping is often preferred for its capability to capture the niche data with respect to specific events such as airline customer service or Covid-19 pandemic Ayvaz and Shiha (2017); Hakami et al. (2022).
- **Linguistic Diversity:** Although English is the baseline, research is expanding to Arabic, Chinese (Sina Weibo) and Portuguese, with a study covering **13 distinct languages** Yin et al. (2023).

10.2 RQ2: Preprocessing, Feature Extraction, and Machine Learning Techniques

The processing pipeline for emoji-integrated text adheres to a structured sequence of data cleaning and vectorization.

- **Preprocessing: Tokenization** emerges as a most foundational processing step, reported in 19 studies, followed by **stopword removal** in 10 studies Srisankar and Lochanambal (2021). Additional techniques include stemming, lemmatization, and normalization Srisankar and Lochanambal (2021); Godard and Holtzman (2022).
- **Feature Extraction:** Conventional studies often utilize techniques like **Bag-of-Words (BoW)** and **TF-IDF** Wang and Castanon (2019). Conversely, more advanced studies exhibit the usage of **word and emoji embeddings** including Word2Vec, GloVe, and specialized vectors like **Emoji2Vec** Jagadishwari et al. (2021).
- **ML/DL Models:** Classic machine learning algorithms like **SVM** and **Naive Bayes** remain popular due to their consistency Jagadishwari et al. (2021). However, **Deep Learning** models (CNNs, LSTMs, and GRUs) are preferred for modeling temporal dependencies, while **Transformers techniques** like BERT, Emoji-BERT are utilize to better capture the contextual nuances in the text Lou et al. (2020); Singh et al. (2024).

10.3 RQ3: Specialized Methods for Handling Emojis

Four primary specialized strategies are employed to integrate emojis into sentiment analysis Tang et al. (2024):

1. **Preprocessing:** Emojis are either substituted with descriptive text tokens or converted into standardized Unicode equivalents Tang et al. (2024); et al. (2020).

2. **Lexicon-Based Mapping:** Assigning fixed polarity scores from dictionaries like the **Emoji Sentiment Lexicon (ESL)** or **EmojiNet** Vashisht and Jailia (2020); Karthikeya et al. (2023).
3. **Embedding Representations:** Capturing semantic relationships through dense vector spaces (static or contextual) that allow the emoji's meaning to adapt to its sentence context Jona et al. (2022).
4. **Modeling Architectures:** Utilizing specialized neural designs like **BiLSTMs with attention** or **TaneNet** to dynamically weight emoji influence Yuan et al. (2022); Holthoff (2020).

10.4 RQ4: Key Challenges and Research Gaps

Despite recent progress, research on emoji analysis continues to encounter several formidable challenges:

- **Challenges:** Emojis are highly **context-dependent (polysemy)** and subject to **regional/cultural variability**, leading to ambiguity and bias Khan et al. (2025). There is also a distinct **lack of standard representation** and universal dictionaries Khan et al. (2025).
- **Research Gaps:** Future work is needed in **context-aware interpretation** beyond static mapping Ullah et al. (2020). Most studies only explore **three-way polarity** (positive, negative, neutral), leaving granular emotions like irony and sarcasm underexplored Ullah et al. (2020). Furthermore, there is a lack of **multimodal approaches** that integrate emojis with images, video, and audio Ullah et al. (2020); Alfreihat et al. (2024).

11. Conclusion and Future Work

In this paper, authors have discussed about various methods of handling text with emojis and emoticons. The study has been divided into sections related to datasets available, feature engineering methods, machine leaning models. Various Challenges and research gaps in this area of research are discussed. The area has lots of scope in future. A roadmap for researchers in the area of emoji handling will be to create a strong and elaborate dictionary for emojis, which can handle polysemy and Synonymy. Context dependence of emoji interpretation and linguistic variablity should be handled in a robust manner. Dependence of culture and language needs to be handled. Use of emojis depend on the age, gender and location, Such factors should be considered while converting emojis to numeric forms. The model needs to be developed to take care of the aspect of emojis evolution with time. Rise of new emojis and changing usage pattern should be handled by the models based on continual learning. Specific eventse.g. elections, pandemic, give rise to certain sentiments. Such kind of longitudinal event analysis needs to be done. With the advancements in generative AI approaches, emoji-aware pretrained language models should be developed. Based on the nature of the data, more weightage may be give to the importance of emojis and develop the machine learning models to handle such implementations. Integration of text, emojis, images, audio, video, and other media is the need of the hour.

Conflict of Interest Statement

The authors declare no conflict of interest.

Ethical Declarations

The authors declare that this research does not involve human participants, animals, or any procedures requiring ethical approval. All data used in this study were obtained and processed in accordance with applicable ethical andlegal guidelines.

References

1. Alawneh, H., Hasasneh, A., and Maree, M. (2024). On the utilization of emoji encoding and data preprocessing with a combined cnn-lstm framework for arabic sentiment analysis. *Modelling*, 5:1469–1489.
2. Alfreihat, M., Almousa, O. S., Tashtoush, Y., Alsobeh, A., Mansour, K., and Migdady, H. (2024). Emo-sl framework: Emoji sentiment lexicon using text-based features and machine learning for sentiment analysis. *IEEE Access*, 12:81793–81806.
3. Ayvaz, S. and Shiha, M. (2017). The effects of emoji in sentiment analysis. *International Journal of Computer and Electrical Engineering*, 9:360–369.
4. Et al., A. S. (2020). Alternative method sentiment analysis using emojis and emoticons. *Procedia Computer Science*, 178:182–193.
5. Godard, R. and Holtzman, S. (2022). The multidimensional lexicon of emojis: A new tool to assess the emotional content of emojis. *Frontiers in Psychology*, 13:921388. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.921388/pdf.
6. Hakami, S. A. A., Hendley, R., and Smith, P. (2022). Emoji sentiment roles for sentiment analysis: A case study in Arabic texts. In Bouamor, H., Al-Khalifa, H., Darwish, K., Rambow, O., Bougares, F., Abdelali, A., Tomeh, N., Khalifa, S., and Zaghouani, W., editors, *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 346–355, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.
7. Hogenboom, A., Bal, D., Frasincar, F., Kaymak, U., and de Jong, F. (2013). Exploiting emoticons in sentiment analysis. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing (SAC)*, pages 703–710. ACM.

8. Holthoff, L. (2020). The emoji sentiment lexicon: Analysing consumer emotions in social media communication. In *Proceedings of the European Marketing Academy*, volume 49, page 63029.
9. Jagadishwari, V., Indulekha, A., Raghu, K., and Harshini, P. (2021). Sentiment analysis of social media text-emoticon post with machine learning models contribution title. *Journal of Physics: Conference Series*, 2070(1):012079.
10. Jona, J., Jothi, C., Paul, S. D., Gunasekaran, S., and Sridhar (2022). Sentimental analysis using emojis prediction. *International Journal of Research Publication and Reviews*, 3(6):1245–1249.
11. Karthikeya, K., Yaswanthi, K., Sai, E. J., Pradeepini, G., Kiran, Y. N. V. N. A. S., and Bulla, S. (2023). Sentiment analysis of tweets using logistic regression and neural networks with emojis and emoticons. *Research Square*.Khan, A., Majumdar, D., and Mondal, B. (2025). Sentiment analysis of emoji fused reviews using machine learning and bert. *Scientific Reports*, 15:7538.
12. Kralj Novak, P., Smailovic´, J., Sluban, B., and Mozeticˇ, I. (2015). Sentiment of emojis. *PLOS ONE*, 10(12):1–22.
13. Lou, Y., Zhang, Y., Li, F., Qian, T., and Ji, D. (2020). Emoji-based sentiment analysis using attention networks. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 19(5):1–13.
14. Mona M. Abd Elsalam, Ahmed M. Gadallah, H. A. H. (2022). Sentiment analysis of text incorporating emojis: A machine learning approach. *International Journal of Computer Applications*, 184(20):10–18.
15. Rahman, R., Islam, T., and Ahmed, M. H. (2017). Detecting emotion from text and emoticon. *London Journal of Research in Computer Science and Technology*, 17(3):1–13.
16. Santos, V. M. d. S., Freitag, R., Tejada, J., Florecˆncio, A. S., Souza, P. P. O. B., and Gois, T. S. (2023). Sentiment analysis with emojis: a model for Brazilian. In *Proceedings of the 4th Conference on Language, Data and Knowledge*, pages 613–616, Vienna, Austria. NOVA CLUNL, Portugal.
17. Singh, G., Ghosh, S., Firdaus, M., Ekbal, A., and Bhattacharyya, P. (2024). Predicting multi-label emojis, emotions, and sentiments in code-mixed texts using an emojifying sentiments framework. *Scientific Reports*, 14.
18. Srisankar, M. and Lochanambal, K. (2021). Emoji's impact on sentiment analysis - a survey. *International Journal of Innovative Research in Technology*, 8(7):366–371.
19. Tang, H., Tang, W., Zhu, D., Wang, S., Wang, Y., and Wang, L. (2024). EMFSA: Emoji-based multifeature fusion sentiment analysis. *PLOS ONE*, 19(9):e0310715.
20. Toet, A., Kaneko, D., Ushiama, S., Hoving, S., de Kruijf, I., Brouwer, A.-M., Kallen, V., and van Erp, J. B. F. (2018). Emojigrid: A 2d pictorial scale for the assessment of food-elicited emotions. *Frontiers in Psychology*, 9:2396.
21. Ullah, M. A., Marium, S. M., Begum, S. A., and Dipa, N. S. (2020). An algorithm and method for sentiment analysis using the text and emoticon. *ICT Express*, 6:357–360.
22. Vashisht, G. and Jailia, M. (2020). Emoticons emojis based sentiment analysis: The last two decades! *International Journal of Scientific Technology Research*, 9(03):6398–6405.
23. Wang, Y. and Castanon, J. (2019). Emojify: Emoji prediction from sentence. CS229: Machine Learning Projects, Stanford University.
24. Yin, W., Alkhalifa, R., and Zubiaga, A. (2023). The emojification of sentiment on social media: Collection and analysis of a longitudinal twitter sentiment dataset.
25. Yoo, B. and Rayz, J. (2021). Understanding emojis for sentiment analysis. *The International FLAIRS Conference Proceedings*, 34.
26. Yuan, X., Hu, J., Zhang, X., and Lv, H. (2022). Pay attention to emoji: Feature fusion network with emograph2vec model for sentiment analysis.
27. Zhao, Q., Wu, P., Lian, J., An, D., and Li, M. (2024). Tanenet: Two-level attention network based on emojis for sentiment analysis. *IEEE Access*, 12:86106–86119.

❖ ❖ ❖

International Journal of Innovation and Multidisciplinary Research
DOI.ORG/10.67505/IJIAMR.v6.i1.61009
Vol. 6 (1), 2026
E-ISSN: 2583-4452

Optimization Techniques for Solving Multi-Objective Transportation Problem: A Comparative Analysis

Aditi Upadhyay[1*]
[1] Department of Mathematics, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, Delhi
[*] Corresponding Author ✉ aditiupadhyay195@gmail.com

ABSTRACT

Efficient transportation planning requires balancing economic performance, operational efficiency, and environmental responsibility. In real-life transportation systems, transportation planner frequently need to include more than one conflicting factors like minimizing transportation cost, minimizing transportation time, and environmental impact. Traditional single-objective models are not sufficient to handle these complex requirements. Therefore the need for efficient multi-objective optimization techniques has become increasingly significant. The proposed study focuses to solve a "Multi-Objective Transportation Problem (MOTP)" by different optimization techniques namely, Weighted Sum Method and Goal Programming Method. For each method, the mathematical models have been formulated and solved using secondary data in similar scenarios for fair comparison of methods. For each method, the solutions vary based on the prioritization of objectives. The results shows that the Weighted Sum Method gives different results when the weights vary, which means it is sensitive to weight selection. Goal Programming proves to be more flexible and consistent by assigning prioritization and targets. A comparative analysis has been performed by focusing on the ability to find compromise solution, effectiveness of each method, efficiency in getting optimal results and practical applicability of the methods. As seen from the analysis, the multiobjective optimization offers more realistic solution as compared to single-objective optimization techniques. This could help decision makers in transportation and logistic planning.

Keywords: *Optimization, Multi-Objective Transportation, Weighted Sum, Goal Programming, Optimal Solution.*

1 Introduction

The mode of transportation is an important element in the movement of goods from one point to another point. Transportation problem forms an integral part of the distribution network, logistics, supply chain management and affect the level of efficiency in the entire process of transportation. In most instances, all traditional transportation planning models seek to optimize a single criterion, mainly transportation cost. But in the real-world situations, decision makers must balance several objectives simultaneously, including minimizing cost, reducing delivery time, and reducing negative environmental impact. The Multi-Objective Transportation Problem (MOTP) is an useful method in such situations because it facilitates optimization among several criteria while offering insight into possible trade- offs. Recent developments in research show that studies have evolved from single objective models to multi-objective models.

The development of practical and effective solutions for multi-objective transportation challenges has received more attention in recent years. Researchers have proposed several approaches to deal with the complexity of handling multiple objectives simultaneously. These approaches aim to provide balanced solutions by considering different aspects of the problem. Different optimization techniques can produce different solutions depending on how they handle trade-offs among objectives. Thus, there is a need to analyze and compare appropriate methods that will give reliable and realistic results. In this research, two methods, Weighted Sum Method and Goal Programming Method, which are widely accepted methods for dealing with multiple objective problems, have been selected and their performance analyzed while solving Multi-Objective Transportation Problem. Both methods have different approaches and can result in different answers based on the priorities and preferences of the decision maker. the objective of analyzing them is to find out which one performs better under the same conditions.

Even though the Weighted Sum Method and Goal Programming Method have been extensively applied to solve the multi-objective optimization problem, there is very limited literature regarding the comparative study of both the methods in solving multi-objective transportation problem with respect to economic, operational, and environmental goals simultaneously. This paper seeks to solve this problem by testing and comparing two optimization methods

Received on: 11 Jun 2026 | Accepted on: 28 Jun 2026 | Published Online on: 25 July 2026

namely, "Weighted Sum Method" and "Goal Programming". The research will help in understanding how the various weights and targets affect the solution of the problem and what compromises are possible between different objectives. At the same time, the inclusion of such objective as carbon emissions helps in introducing the aspect of sustainability in the decision making process.

The main contribution of this research is to analyze the ability of these methods to deal with conflicting objectives of transportation cost, delivery time, and carbon emission. Furthermore, the study gives information on how to make tradeoffs between economic efficiency, transportation time, and the environment. By taking into account carbon dioxide emissions along with conventional transportation criteria, the research ensures that decisions will be made in a sustainable manner. Such information will be helpful for logistics managers, transportation planners, and other supply chain decision makers while making choices regarding optimization techniques.

2 Review of Literature

The existing literature regarding transportation problems highlights an increased emphasis on the ability to handle more than one objective simultaneously. Various techniques have been examined by researchers in order to find balance solution in complex situations.

Bazaraa, Sherali, and Shetty [1] elaborate on the concept of optimization as well as elaborate the concept of Non-Linear Programming when the interrelationship variables is difficult to understand. The book covers several techniques used in complicated optimization instances. In addition, there is the mention of practical issues which are not always simple and thus require non-linear methods for their solution. Marler and Arora [4] provided a comprehensive review of multi-objective optimization techniques and reviewed the theoretical concepts involved in formulating multiple objectives into one function. The authors pointed out that while this technique is simple and very popular, it has limitations that it might not always be able to determine all possible optimal solutions, specially when there are complicated trade-offs between multiple objectives.

Nomani et al. [5] proposes new "Weighted Goal Programming" technique for solving MOTP. It emphasizes the need for attaining an equal solution that remains close to the optimal value for all objectives. This technique has proven efficient whether or not the criteria weights were explicitly identified. Findings have shown that the use of this method yielded better results than the existing ones. Jagtap and Kawale [3] studied multi-dimensional multi-objective transportation problems using "Goal Programming Method". Their approach focused on setting priority levels for different goals and minimizing deviations from these targets. The study demonstrated that goal programming is effective in handling multiple conflicting objectives and provides satisfactory solutions based on decision-maker preferences. Jain et. al [2] used used "Lexicographic Goal Programming" for solving the MOTP through ordering of the objectives according to their priority. This technique tries to minimize both cost and time one at a time without changing the output of higher priority objective. Deviation variable was also used to show the proximity of the solutions obtained to the required goal. By checking different priority orders, the findings of this approach indicated that this approach is able to produce satisfactory results.

Shrivastava and Rajput [7] used the Weighted Sum Method, where weights are assigned to different objectives. Their results indicated that this method is useful for balancing cost, time, and risk, but the final solution depends heavily on the choice of weights. In another study, Shrivastava and Rajput [8] applied the "Lexicographic Method" to solve multi-objective transportation problems. The findings of their study indicate that the above technique insures that the most important objective is prioritized, thus ensuring reduced cost, time while minimizing carbon emission. The problem is that the method heavily depends on the hierarchy of objectives. Shrivastava and Dwivedi [6] discussed the conversion of multi-objective transportation problem into single objective problem by using the "Weighted Sum Method" and revised simplex method is used to solve the problem. The study emphasized the computational efficiency of this methodology, but it did mention that the resulting solution depends on the weight assigned to objectives. Shrivastava and Rajput [9] provides a novel approach to deal with multi-objective transportation problems by using statistical methods. By combining cost, time and emissions of carbon into one figure, "Arithmetic, Geometric, and Harmonic Means" were used. Different outcomes have been achieved from each mean which indicates clear contradiction between time, fuel consumption and costs. It is a flexible, effective, and straightforward approach.

From the above literature, it is obvious that a number of approaches have been applied to the problem of dealing with transportation problems that involve multiple objectives. The type of result depends upon the approach adopted, whereby, some approaches tend to be basic while others complex.

3 Motivation and Objectives

This study is inspired by the necessity of finding effective method that would assist in solving practical cases of multi-objective transportation problems. In transportation management, decision-makers are expected to take into

consideration several conflicting objectives, like reduction of cost, time, and adverse effects on the environment. Such considerations cannot be accommodated using conventional single-objective optimization techniques because of their limitation in dealing with real-life conditions. Thus, it is necessary to resort to other methodologies in solving such problems. This study focuses on the Weighted Sum Method and Goal Programming to evaluate their effectiveness in managing multiple objectives and generating practical solutions. The aim of this study is to support improved and informed decision-making in transportation planning by identifying suitable optimization approaches.

- To evaluate the role of each method of solving Multi-Objective Transportation Problem in minimizing transportation cost while simultaneously reducing time and carbon emission.
- To analyze the efficiency and effectiveness of these methods.
- To determine the significant difference in the Optimal solution obtained by these methods.
- To analyze trade-offs between conflicting objectives.
- To provide practical insights and recommendations that can support decision-makers in logistics, supplychain management, and transportation planning.

4 Data and Variables

This study follows a comparative optimization approach to assess the efficiency of approaches used in solving the "multi objective transportation problem." This analysis involves the Weighted Sum Method, and Goal Programming Method, which are chosen in order to understand how these approaches address several conflicting objectives within transportation management, including cost, delivery time, and environmental impacts, among others. The process starts with data collection and mathematical model formulation that represents all objectives and constraints in an organized and systematic way.

This study is based on secondary data which will be the main dataset to conduct the analysis. Data have been sourced from a previously published paper by Shrivastava and Rajput [8], that is dealing with transportation planning in agriculture sector in Madhya Pradesh. The chosen dataset depicts a real transportation scenario wherein there is transportation of agriculture products from farms to the processing units. The chosen dataset can be considered an agro-based transportation network, consisting of four agriculture farms situated in different cities of Madhya Pradesh namely, Indore, Bhopal, Jabalpur, and Gwalior and three processing plants located in Ujjain, Sagar, Rewa. The four farms are the supply nodes as agriculture products are available here. On the other hand, the three processing plants represents the demand nodes since the products need to be transported to these plants. The dataset has important factors such as transportation cost, which represents the cost involved in moving one unit of the product from every farm to every plant, transportation time shows the time needed to move products from each farm to each plant and carbon emission represent the environmental impact of transporting one unit of product between source to destination, during the transportation process. Hence, this dataset is well suited for the current analysis, as all factors are captured. So the use of this dataset supports a fair comparison of optimization techniques in real-world transportation scenario.

Table 1: *Sources and Destinations*

Sources			**Destination**		
Farm	**City**	**Supply**	**Plant**	**City**	**Demand**
A	Indore	40	X	Ujjain	60
B	Bhopal	50	Y	Sagar	50
C	Jabalpur	60	Z	Rewa	70
D	Gwalior	30			

Table 2: *Objective Matrix of Transportation Problem*

Source→ Destination	**Cost (Rs Per Ton) (Z_1)**	**Time (Hours) (Z_2)**	**Carbon Emissions (Kg CO_2 Per ton) (Z_3)**
Farm A →Plant X	500	5	20
Farm A →Plant Y	700	7	30
Farm A →Plant Z	600	6	25
Farm B →Plant X	550	6	22
Farm B →Plant Y	650	5	28
Farm B →Plant Z	750	8	35
Farm C →Plant X	800	9	40
Farm C →Plant Y	600	6	25
Farm C →Plant Z	500	5	20
Farm D →Plant X	700	8	30
Farm D →Plant Y	900	10	45
Farm D →Plant Z	650	7	28

5 Mathematical Formulation of Multi-Objective Transportation Problem

The "Multi-Objective Transportation Problem (MOTP)" handles the optimal allocations of products that transfers from several source to several destination while simultaneously optimizing several conflicting objectives.

Notation:

$$i = 1, 2,\ldots..,\ m \text{ (index for farms)}$$
$$j = 1, 2,\ldots..,\ n \text{ (index for plants)}$$
$$k = 1, 2,\ldots..,\ K \text{ (index for objectives)}$$

Parameters:

- a_i denote supply available at farm i
- b_j denote demand required at plant j
- c_{ij}^k denote unit transported from farm i to plant j for objective k

Decision Variables: The amount of goods that is delivered from supply center to demand points. It indicates how much quantity should be shipped along each possible route to meet demand efficiently. Let x_{ij} is the quantity transported from farm i to plant j

Objective Functions: The objective function will be the minimizing multiple objective functions:

$$Z_k = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}^k x_{ij}\ ,\quad k = 1, 2, 3, \ldots.., K$$

So the way to express the problem as:

$$\text{Minimize } (Z_1, Z_2, Z_3,\ldots., Z_K)$$

- **Cost Objective**

$$\text{Minimize } Z_1 = \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}x_{ij}$$

Here C_{ij} is per unit cost of products transported from farm i to plant j.

$$\text{Minimize } Z_1 = 500x_{AX} + 700x_{AY} + 600x_{AZ} + 550x_{BX} + 650x_{BY} + 750x_{BZ} + 800x_{CX} + 600x_{CY} + 500x_{CZ} + 700x_{DX} + 900x_{DY} + 650x_{DZ}$$

- **Time Objective**

$$\text{Minimize } Z_2 = \sum_{i=1}^{m}\sum_{j=1}^{n} T_{ij}x_{ij}$$

Here T_{ij} is delivery time from farm *i* to plant *j*.

$$\text{Minimize } Z_2 = 5x_{AX} + 7x_{AY} + 6x_{AZ} + 6x_{BX} + 5x_{BY} + 8x_{BZ} + 9x_{CX} + 6x_{CY} + 5x_{CZ} + 8x_{DX} + 10x_{DY} + 7x_{DZ}$$

- **Carbon Emission Objective**

$$\text{Minimize } Z_3 = \sum_{i=1}^{m}\sum_{j=1}^{n} E_{ij}x_{ij}$$

Here *Eij* is Carbon emission from farm *i* to plant *j*.

$$\text{Minimize } Z_3 = 20x_{AX} + 30x_{AY} + 25x_{AZ} + 22x_{BX} + 28x_{BY} + 35x_{BZ} + 40x_{CX} + 25x_{CY} + 20x_{CZ} + 30x_{DX} + 45x_{DY} + 28x_{DZ}$$

Constraints: Constraints represents the requirements that the model's solution must meet. Supply constraints, demand constraints, and non-negativity constraints, are used in transportation problems which ensure that the solution is feasible, practical, and consistent with real-world requirements.

Supply Constraints: Each farm cannot supply more than its capacity

$$\sum_{j=1}^{n} x_{ij} \le a_i\,, \quad i = 1, 2, ..., m$$

$$x_{AX} + x_{AY} + x_{AZ} \le 40$$
$$x_{BX} + x_{BY} + x_{BZ} \le 50$$
$$x_{CX} + x_{CY} + x_{CZ} \le 60$$
$$x_{DX} + x_{DY} + x_{DZ} \le 30$$

Demand Constraints: Each plant must receive its required demand

$$\sum_{i=1}^{m} x_{ij} = b_j\,, \quad j = 1, 2, ..., n$$

$$x_{AX} + x_{BX} + x_{CX} + x_{DX} = 60$$
$$x_{AY} + x_{BY} + x_{CY} + x_{DY} = 50$$
$$x_{AZ} + x_{BZ} + x_{CZ} + x_{DZ} = 70$$

Non-Negativity Constraints: The quantity of goods transported cannot be negative

$$x_{ij} \ge 0,\ \forall i, j$$

$$x_{AX}, x_{AY}, x_{AZ}, x_{BX}, x_{BY}, x_{BZ}, x_{CX}, x_{CY}, x_{CZ}, x_{DX}, x_{DY}, x_{DZ} \ge 0$$

It is essential to note that before attempting to solve a Multi Objective Transportation Problem MOTP it is important to test whether the problem is either balanced or unbalanced A transportation problem is said to be balanced when the total supply is equal to the total demand This is important because it ensures that a feasible solution is achieved In this study the total supply in all sources is 180 units and the total demand in all destinations is also 180 Because these two values are equal So the transportation problem is balance

6 Weighted Sum Method

"In the Weighted Sum Method, single objective function is formulated by giving a weight to each objective" according to their importance [6].

- Each objective is first normalized so that differences in measurement units do not affect the results.
- Suitable weights are assigned to each objective based on their importance.
- Then single combined objective function is created by multiplying each normalized objective by its corresponding weight.
- Combined objective function is minimized while meeting all supply and demand requirements.

- The procedure is repeated with different sets of weights to study how changes in priorities affect the final solution.

$$minZ = w_1Z_1 + w_2Z_2 + w_3Z_3 + ... + w_kZ_k$$

$$\sum_{n=1}^{k} w_n = 1, \quad w_1, w_2, ..., w_k \geq 0$$

6.1 Mathematical Formulation

Normalization is essential to bring all objectives to a common, dimensionless scale using min-max normalization.

$$Z_k^{norm} = \frac{Z_k - Z_k^{min}}{Z_k^{max} - Z_k^{min}}$$

Combined Objective Function:

$$Minimize\ Z = w_1Z_1 + w_2Z_2 + w_3Z_3$$

$$\text{Minimize } Z = w_1 \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}x_{ij} + w_2 \sum_{i=1}^{m}\sum_{j=1}^{n} T_{ij}x_{ij} + w_3 \sum_{i=1}^{m}\sum_{j=1}^{n} E_{ij}x_{ij}$$

Test multiple weight combinations (w_1, w_2, w_3) to analyze Trade-offs.

- **Case 1: Cost Priority**

  Higher weight is assigned to transportation cost:

$$\text{Minimize } Z = 0.6 \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}x_{ij} + 0.2 \sum_{i=1}^{m}\sum_{j=1}^{n} T_{ij}x_{ij} + 0.2 \sum_{i=1}^{m}\sum_{j=1}^{n} E_{ij}x_{ij}$$

- **Case 2: Time Priority**

  Higher weight is assigned to transportation time:

$$\text{Minimize } Z = 0.2 \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}x_{ij} + 0.6 \sum_{i=1}^{m}\sum_{j=1}^{n} T_{ij}x_{ij} + 0.2 \sum_{i=1}^{m}\sum_{j=1}^{n} E_{ij}x_{ij}$$

- **Case 3: Carbon Emission Priority**

  Higher weight is assigned to carbon emissions:

$$\text{Minimize } Z = 0.2 \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}x_{ij} + 0.2 \sum_{i=1}^{m}\sum_{j=1}^{n} T_{ij}x_{ij} + 0.6 \sum_{i=1}^{m}\sum_{j=1}^{n} E_{ij}x_{ij}$$

7 Goal Programming Method

Goal programming is used when specific target values are desired for each objective rather than simply minimizing or maximizing them [3].

- Target values (goals) are first defined for each objective.
- For each goal, deviation variables are introduced to measure how much the actual result deviates from the target.
- Priority is assigned to each deviation based on the importance of the corresponding goal.
- The model's objective function will be minimizing these deviations, especially unwanted deviations.
- Then solved the model to find the result that best satisfies all objectives simultaneously.

7.1 Mathematical Formulation

Set target value for cost, time, and carbon emission (G_1, G_2, G_3). Define deviation variable, For each goal, two deviation variables are introduced:

- One for overachievement
- One for underachievement

Deviation Variables: Deviation variable represent how much achieved value is deviated from the target value of each objective.

- $d_1^+, d_1^- \rightarrow$ Cost Deviation
- $d_2^+, d_2^- \rightarrow$ Time Deviation
- $d_3^+, d_3^- \rightarrow$ Carbon Emission Deviation

Goal Equations: Goal equations represent the desired target levels of each objective. They are formed by equating the objective function to a specified target value and include deviation variables to measure underachievement or overachievement of the goals.

$$\sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}x_{ij} + d_1^- - d_1^+ = G_1$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} T_{ij}x_{ij} + d_2^- - d_2^+ = G_2$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} E_{ij}x_{ij} + d_3^- - d_3^+ = G_3$$

Objective Function:

$$MinZ = [P_1(d_1^+), P_2(d_2^+), P_3(d_3^+)]$$

- **Case 1: Cost Priority**
  First rank is given to transportation cost:
  $$\text{Minimize } d_1^+$$
- **Case 2: Time Priority**
  First rank is given to transportation time:
  $$\text{Minimize } d_2^+ \text{ keeping } d_1^+ \text{ at its optimal value}$$
- **Case 3: Carbon Emission Priority**
  First rank is given to carbon emission:
  $$\text{Minimize } d_3^+ \text{ keeping } d_1^+ \text{ and } d_2^+ \text{ at its optimal value}$$

8 Results

In this study, the computational analysis was done using ″Excel Solver and Python″ software. Excel solver was applied in creating and solving the optimization models within a well structured environment. On the other hand, python software was applied via the ″pulp″ library, which is commonly used for ″Linear Programming″ models. This allowed efficient implementation and computation of mathematical models. The advantages of the ″pulp″ library was that it facilitated comparative analysis of different methods.

8.1 Results of Weighted Sum Method

- **Case 1:** $w_1 = 0.6, w_2 = 0.2, w_3 = 0.2$
  $x_{11} = 40\ x_{21} = 20\ x_{22} = 30\ x_{32} = 20\ x_{33} = 40\ x_{43} = 30$

Total Cost	Time	Carbon Emission
102000 Rs.	34 Hours	4220 Kg CO_2

- **Case 2:** $w_1 = 0.2, w_2 = 0.6, w_3 = 0.2$
  $x_{11} = 40\ x_{13} = 0\ x_{22} = 50\ x_{33} = 60\ x_{41} = 20\ x_{43} = 10$

Total Cost	Time	Carbon Emission
103000 Rs.	30 Hours	4280 Kg CO_2

- **Case 3:** $w_1 = 0.2, w_2 = 0.2, w_3 = 0.6$

  $x_{11} = 40\ x_{21} = 20\ x_{22} = 30\ x_{32} = 20\ x_{33} = 40\ x_{43} = 30$

Total Cost	Time	Carbon Emission
102000 Rs.	34 Hours	4220 Kg CO_2

8.2 Results of Goal Programming Method

Cost target = 105000 Rs., Time target = 35 Hours, Carbon emission target = 4500 kg CO_2

- **Case 1: First rank → Cost**

  $x_{11} = 10\ x_{12} = 30\ x_{21} = 50\ x_{32} = 20\ x_{33} = 40\ x_{43} = 30$

Total Cost	Time	Carbon Emission
105000 Rs.	37 Hours	4340 Kg CO_2

- **Case 2: First rank → Time**

  $x_{11} = 40\ x_{13} = 0\ x_{22} = 50\ x_{33} = 60\ x_{41} = 20\ x_{43} = 10$

Total Cost	Time	Carbon Emission
103000 Rs.	30 Hours	4280 Kg CO_2

- **Case 3: First rank → Carbon Emission**

  $x_{12} = 40\ x_{21} = 50\ x_{32} = 10\ x_{33} = 57.5\ x_{41} = 10\ x_{43} = 12.5$

Total Cost	Time	Carbon Emission
105375 Rs.	39 Hours	4350 Kg CO_2

9 Comparative Analysis

A comparative analysis has been conducted to determine the strengths and limitations of the Weighted Sum Method and Goal Programming Method to solve the "Multi-Objective Transportation Problem". Comparison has been made considering crucial parameters like dealing with trade-offs, stability of solution, effectiveness of each method, efficiency in achieving desirable results, and variability of optimal solutions. Through such comparisons, it is possible to identify the way in which each approach reacts to changes in weights and targets, and how well they deal with multiple conflicting objectives. A comparative analysis offers an idea about the advantages and disadvantages associated with each approach, helps in selecting appropriate approaches for practical transportation problems.

- **Trade-off Between Objectives:** Regarding the issue of handling the trade-offs, the Weighted Sum Method makes provision for a trade-off between cost, time and carbon emission based on weights. it has been noted that where time takes higher weight, total time decreases to 30 hours whereas there is an increase in cost and carbon emission. However, In other cases, such results are also found, signifying that there is not much of a trade-off. On the other hand, Goal Programming Method takes care of trade-offs based on the order of priorities.
- **Stability of the Solution:** In terms of stability of the solution, the Weighted Sum Method appears to have fairly consistent results since in two cases out of the three yields the same results as cost = 102000 Rs., time = 34 hours, carbon emission = 4220 kg CO_2. The slight variations in the results do not seem to alter the solutions much. In contrast, the Goal Programming Method has varying results due to changing priorities in different cases. This gives different results in all three cases, thus the Goal Programming method is not stable like Weighted Sum Method. However, this variation shows that goal programming is more flexible and practical.
- **Effectiveness of Each Method:** In regards of effectiveness, the Weighted Sum Method is effective for all objectives that need to be consider simultaneously and used to find a balance solution. This enables a decision-maker to determine compromise solution between all objectives. However, Goal Programming Method is more

suitable for dealing with realistic conditions, wherein achieving specific objectives is necessary. Goal programming aims to reduce the gap between the expected and the actual values.

- **Efficiency in Achieving Desired Results:** In terms of efficiency in achieving desired results, the Weighted Sum Method considers all objectives simultaneously. It gives the optimum result in the balancing of various objectives but cannot give optimum results to individual objectives. On the other hand, Goal Programming Method is more efficient in the realization of desirable results. This is because this approach takes into consideration specific targets that have been predetermined.
- **Significant Differences in Optimal Solutions:** In terms of optimal solution, the results show that the Weighted Sum Method yields almost the same solution in Case 1 and Case 3, showing very small variation in the optimal results. In contrast, Goal Programming Method gives different solution based on the priority or the objectives in all three cases.

10 Conclusion

In this research, the main goal was to solve the "Multi-Objective Transportation Problem" and comparing using the Weighted Sum Method and Goal Programming Method, based on three objectives namely transportation cost, transportation time and carbon emissions. From the results obtained through the comparison of these methods, one can draw the conclusion that both methods are able to solve a multi-objective problem but in a different manner.

The solutions obtained with the use of the Weighted Sum Method are stable consistent in Case 1 and Case 3. At the same time, they vary slightly when changing the Weight coefficients. The use of this method does not allow taking into consideration changes in the priority order of objectives by a decision-maker. Solutions provided with Goal Programming depend upon priority levels of goods assigned to particular objectives. This method is more flexible and more suitable for cases where certain objectives have to be accomplished. It makes it possible to handle tradeoffs better and come up with solutions that better suit the decision maker.

As seen from the comparative analysis, Goal Programming is much better as reaching the desired target as well as trade-offs in a practical way, provided clear priorities are known. While the Weighted Sum Method becomes effective in situations when all objective functions need to viewed as being of equal importance. From the results, the research reveals that the selection of a certain optimization technique depends on the specific needs of the decision maker and on the kind of problem they addressing. If all the objectives are equally important, then the use of the Weighted Sum Method may help find a balance between them. But where in real life some objectives prevail over others in the problem under consideration, the Goal Programming approach works much better.

The outcomes of this research provide valuable insight into the performance of these techniques under various scenarios. The information may assist decision-makers in selecting the appropriate optimization technique according to their needs. Consequently, better transportation planning will be achieved.

11 Limitations and Future Scope of Research

- This research is based on secondary data that is collected from existing source. Future study can be included real-time data set to solve transportation problem and to improve transportation planning.
- In this research work, a limited dataset containing a limited number of supply points and demand points has been used. In the future research, the proposed model can be developed further by considering more supply points and more demand points.
- In this research, Weighted Sum and Goal Programming methods are used for comparative analysis. In the future advanced optimization algorithms such as "Genetic Algorithm, Fuzzy Algorithm, Evolutionary Algorithm " can be used.
- This research work being done includes cost, time and carbon emission as its key objective. In the future, more objectives may be considered. These might include issues such as quality of service, risk, reliability, and customer satisfaction.
- In real-life transportation systems, several parameters are uncertain and subject to change with time, which include fluctuation of demand, changes in the cost, and the influence of traffic or weather conditions on delivery time. these uncertainties can be incorporated into future work.
- In future, the proposed model will be tested on additional benchmark data sets. This will provide more evidence concerning the efficiency of the model and increase the credibility of the results.

Acknowledgements

The author would like to express sincere gratitude to Prof. (Dr.) Punita Saxena, Department of Mathematics, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, for all her continued support, guidance, and

encouragement in completing this research work. Her helpful suggestions and continued guidance have been essential in completing this research.

Conflict of Interest Statement

The author declares there are no conflicts of interest, financial or otherwise, that may have affected the conduct, results or publication of this study.

Ethical Declaration

The proposed research does not include any human subjects, animals, or experiments that need ethical approval. All data and sources utilized for conducting the research have been gathered in accordance with relevant academic, and ethical norms.

References

1. Bazaraa, Mokhtar S., Hanif D., Sherali, Shetty, C. M. (2006). Non-Linear Programmning: Theory and Algorithms (3rd ed.), *John Wiley & Sons Wiley India*
2. Jain, K.K., Bhardwaj, R., Choudhary, S. (2019). A Multi-Objective Transportation Problem Solve By Lexicographic Goal Programming. *International Journal of Recent Technology and Engineering*, 7(6), 1842-1846.
3. Jagtap, K.B., Kawale, S.V. (2017). Multi Dimensional Multi Objective Transportation Problem by Goal Programming. *International Journal of Scientific & Engineering Research*, 8(6), 568-573.
4. Marler, R.T., Arora, J.S. (2009). The weighted sum method for multi-objective optimization: New insights. *Structural and Multidisciplinary Optimization*, 41(6), 853-862.
5. Nomani, M.A., Ali, I., Ahmed, A. (2016). A new approach for solving multi-objective transportation problems. *International Journal of Management Science and Engineering Management*, 12(3), 165-173.
6. Shrivastava, R., Dwivedi, D. (2025). Solving Multi-Objective Transportation Problem by using the Revised Simplex Method. *International Journal of Engineering Research and Development*, 21(1), 143-149.
7. Shrivastava, R., Rajput, S.S. (2024). Optimization of Multi-Objective Transportation Problems Using Weight Sum Method *International Journal of All Research Education and Scientific Methods*, 7(3).
8. Shrivastava, R., Rajput, S.S. (2025). Optimizing Multi-Objective Transportation Problems Using the Lexicographic Method. *International Journal for Multidisciplinary Research*, 7(3).
9. Shrivastava, R., Rajput, S.S. (2025). A Novel Approach to Solve Multi-Objective Transportation Problems Using Statistical Mean Approaches. *International Journal of Scientific Development and Research*, 10(1).

❖ ❖ ❖

Vol. 6 (1), 2026
E-ISSN: 2583-4452

SmartFace: An Offline-First AI-Based Attendance Management System for Rural Schools Using MobileFaceNet

Muskan Behl[1], Parishi Mishra[2*], Deepali Bajaj[3] and Pushkar Gole[4]
[1, 2, 3, 4] Department of Computer Science, Shaheed Rajguru College of Applied Sciences for Women, University of Delhi, Delhi
[*] Corresponding Author ✉ parishimishra19@gmail.com

ABSTRACT

Manual recording of student attendance in rural schools is a time-consuming and error-prone task. Moreover, attendance records are necessary for keeping a check on educational activities and promoting social welfare schemes. In literature, most existing attendance automation systems rely on continuous internet connectivity, cloud-based processing, or costly hardware infrastructure, which can limit their scalability, affordability, and ease of deployment. Hence, a lightweight attendance system named SmartFace has been designed and developed in this research work which uses facial recognition technology for marking students' attendance digitally. SmartFace makes use of MobileFaceNet, a Convolutional Neural Network model that allows facial recognition on the device itself, conserving processing resources and eliminating the need for server-side processing. The proposed system is developed as a Progressive Web Application (PWA) and it can function without an internet connection and synchronize data with Firebase Firestore once connectivity is restored. The MobileFaceNet model of SmartFace application was trained and tested on the Indian Faces Image Classification Dataset obtained from Kaggle data repository. On evaluation, the fine-tuned model achieved a training accuracy of 99.43%, with a validation accuracy of 93.75% and a test accuracy of 96.88%, demonstrating the viability of the proposed approach on the dataset used. These findings suggest the potential of lightweight facial recognition technology as a viable approach for efficient and economical attendance management in rural schools, warranting further evaluation on larger and more diverse datasets.

Keywords: *Face Recognition, Rural Schools, Automated Attendance System, MobileFaceNet, Progressive Web Application (PWA), Offline-First Architecture*

1 Introduction

Attendance management in Indian schools has historically relied on manual roll calls, a practice that is both time-consuming and susceptible to inaccuracies. Manual attendance recording remains a burdensome administrative task, often reducing the instructional time available to teachers during the school day. Apart from this, manually marking attendance may allow proxy attendance, resulting in false attendance records. This adversely affects the actual beneficiaries from availing the benefit of government funded programs such as Mid-Day Meal schemes and other scholarships. These inaccuracies hinder allocation of government funding to the schools and further lead to a drop in student retention rates.

The increase in availability of affordable smartphones and tablets in rural India has also paved the way to digitize school administration. The existing digital attendance solutions have been designed according to available urban infrastructure requiring dedicated hardware, consistent high-speed internet connectivity, and trained technical personnel. However, these solutions are not feasible for about 1.23 million rural schools of India due to unreliable network access and power outages.

A unified deep embedding framework that maps facial images to a compact Euclidean space and achieves high level of accuracy on standard benchmarks was first introduced by Schroff et al. (2015). MobileFaceNets, a collection of highly efficient Convolutional Neural Network (CNN) models created especially for accurate, real-time face verification on mobile devices, was later introduced by Chen et al. (2018). Recent research including ArcFace (J. Deng et al., 2022) and improved versions of MobileFaceNet (Hassanpour & Kowsari, 2023), has demonstrated that high quality recognition is still achievable with reduced computational costs. Although several attendance management systems have been developed, none have been designed for infrastructure limited settings; server-side processing, dedicated hardware, and a stable internet connection remain the primary barriers.

Received on: 13 Jun 2026 | Accepted on: 01 Jul 2026 | Published Online on: 25 July 2026

To solve the above problems, the paper provides SmartFace, an AI-driven attendance management system that has been built to eliminate infrastructure deficiencies. SmartFace is designed as a Progressive Web Application that can run on standard web browsers without requiring installation. Using a lightweight deep learning model, facial recognition is performed on-device (Chen et al., 2018) which has been selected for its compact architecture and suitability for browser-based inference (Hassanpour & Kowsari, 2023). Attendance records are synchronized immediately when connectivity is available, and in case of poor network availability attendance records are stored locally, then synchronized with cloud storage when connectivity permits, ensuring no data loss. The objectives of this research are: (1) to develop a lightweight AI-based attendance management system using facial recognition, (2) to enable on-device face recognition without server-side computation, (3) to support offline attendance functionality using an offline-first architecture, and (4) to evaluate the effectiveness of the proposed system using a publicly available face dataset. The remainder of this paper is structured as follows: Section 2 reviews relevant literature and the research gap addressed by this work; Section 3 describes the methodology; Section 4 presents the experimental setup; Section 5 reports the results; and Section 6 concludes the paper with directions for future research.

2 Literature Review

Face recognition is a subfield of machine learning that has received increasing attention over the last decades. A number of models with different requirements have been developed. In Schroff et al. (2015), the FaceNet architecture, which maps a face image into a compact Euclidean space where distances are indicative of face similarity, is introduced. The FaceNet is trained on millions of faces using a triplet loss function, achieving 99.63% accuracy on the Labelled Faces in the Wild (LFW) dataset. This model, however, is computationally intensive and cannot be directly deployed on mobile devices or embedded systems without severe optimization.

To meet the demands of mobile inference, MobileFaceNets were developed in Chen et al. (2018). Mobile-FaceNets is a family of lightweight CNN architectures optimized to perform real-time face verification on mobile platforms. These models adopt depthwise separable convolutions that are also part of the MobileNet architecture (Sandler et al., 2018), and a global depthwise convolution layer replaces the fully connected layers, thus significantly reducing both the number of parameters and computation. Lightweight model research has produced other competitive architectures, including ShuffleFaceNet models by Mart´ýnez-D´ýaz et al. (2019) under 20 MB and averaging 37ms of CPU inference time. A more recent contribution is the 3.5MB few-shot face recognition model presented in Z.-Y. Deng et al. (2023), which is about 30 times smaller than the FaceNet model, while achieving comparable recognition accuracy and real-time performance. Further in this line, Hassanpour and Kowsari (2023) improved on the MobileFaceNet model to achieve top ranked performance on EFaR-2023 face recognition competition under efficiency constraints. Another very light 3.5MB FewFaceNet by Sufian et al. (2023) has shown effective few-shot face authentication for edge cameras with just one enrolment image. State-of-the-art of unconstrained face recognition has been improved to 99.83% on LFW by J. Deng et al. (2022) using the Arc-Face method. In Sun et al. (2024), a landmark based self-supervised model LAFS is introduced that outperforms the state of the art by a significant margin on few-shot recognition benchmarks without requiring large, labelled datasets.

2.1 Automated Attendance Systems

Deep learning has been applied to automate student attendance. The most common pipeline described in the literature comprises face detection using MTCNN (Zhang et al., 2016) followed by embedding based matching of student images. The system has been shown to work efficiently on urban or well-funded institutional environments, but evaluation has been restricted to those environments. Chitrapu et al. (2025) even showed that an ensemble of deep learning models could achieve reliable face recognition performance, though such architectures also entail a higher computational overhead.

2.2 Research Gap

Existing attendance automation systems found in the literature were almost exclusively designed and evaluated in contexts very different to the realities faced in rural Indian schools. The most common architecture design pattern is cloud based: face detection and recognition are offloaded to remote servers and a consistent network connection is needed, which translates in latency, higher cost and lack of functionality when the network connection is intermittent. The hardware aspect is a second issue: a few researchers have used specialized enrolment terminals, IP cameras with high resolution or powerful edge computers like the Raspberry Pi with GPU accelerators. Such hardware might be justifiable in a wealthy context, but the cost is far too high for rural schools. Lightweight models (Z.-Y. Deng et al., 2023; Hassanpour & Kowsari, 2023; Mart´ýnez-D´ýaz et al., 2019) achieve a much better performance-resource ratio, but this line of research has not yet been applied to the deployment in rural educational contexts. SmartFace tackles these problems, and leverages MobileFaceNet (Chen et al., 2018) for browser-based inference so no server-side computation per recognition event occurs. Offline-first IndexedDB storage allows the system to remain functional even when internet connectivity is

unavailable, which matches the infrastructure-less design paradigm promoted by Sufian et al. (2023) who developed FewFaceNet particularly for edge cameras.

Table 1 summarizes representative face recognition systems alongside SmartFace for comparative reference.

Table 1: *Summary of Related Face Recognition Systems*

Reference	Model / Approach	Dataset	Key Results
Chen et al. (2018)	MobileFaceNets	LFW, MegaFace	99.55% (LFW)
Schroff et al. (2015)	FaceNet (triplet loss)	LFW, YTF	99.63% (LFW)
J. Deng et al. (2022)	ArcFace	MS1MV2	99.83% (LFW)
Z.-Y. Deng et al. (2023)	FN8 (3.5 MB few-shot model)	CASIA-WebFace	98.4% (LFW)
Hassanpour & Kowsari (2023)	Improved Mobile-FaceNet	EFaR-2023 benchmark	Ranked 1st on EFaR-2023 benchmark under efficiency constraints
Mart´ýnezD´ýaz et al. (2019)	ShuffleFaceNet	LFW, MegaFace	Outperformed comparable lightweight models
SmartFace (This work)	MobileFaceNet (on-device)	Indian Faces Image Classification Dataset (Garg et al., 2023)	99.43% train accuracy (Val: 93.75%, Test: 96.88%)

3 Methodology

This section is organized into 5 subsections. Section 3.1 describes the dataset used in the current study. Section 3.2 describes the data pre-processing pipeline used for the raw images. Section 3.3 describes the model used in this study for on-device face recognition. Section 3.4 describes the training methodology used to fine-tune the model. Section 3.5 describes the overall end-to-end attendance recognition system used in the SmartFace application.

3.1 Dataset

The Indian Faces Image Classification Dataset (Garg et al., 2023) hosted on Kaggle has been used for the training and testing of the model. Students created this dataset where every student contributed 20 facial images each. The images were collected under various circumstances; e.g., with differing illumination, varying face angles, with and without specs, with diverse backgrounds. All in all, there were 320 images consisting of 16 students. After splitting the data into train and test, 14 images per student were selected for training (70%) and the rest 6 for testing (30%). Dataset statistics are summarized in Table 2.

Table 2: *Dataset Statistics*

Property	Value
Total Students	16
Total Images	320
Images per Student	20
Training Images per Student	14
Testing Images per Student	6
Train Split	70% (224 images; 179 train + 45 val)
Validation Split	20% of train pool (45 images)
Test Split	30% (96 images)

3.2 Data Preprocessing

All images were preprocessed using a fixed pipeline before training. Faces were first detected and cropped from images to generate standard inputs. Cropped regions were resized to the MobileFaceNet input resolution, and their pixels were

normalized to prevent numerical instability during training. The detected faces were then aligned using landmark-based transformation (Zhang et al., 2016). This process centers facial keypoints (pupils, nose tip, lip corners) onto predefined positions, which is critical to achieve discriminative representations and avoid different embeddings from the same face but with varying orientation. To prevent over-fitting data augmentation including horizontal flipping and subtle changes in brightness were applied.

3.3 Model Architecture

MobileFaceNet (Chen et al., 2018) was used as the backbone due to its lightweight nature, which is beneficial for deployment on embedded devices with limited computational resources. The architecture extends MobileNetV2 (Sandler et al., 2018) to adapt it for face recognition. Instead of fully connected layers at the end, it uses a global depthwise convolution and a bottleneck design to minimize memory usage at inference time. These properties make the model ideal for on-device execution in a web browser. Unlike heavily parameterized models such as ArcFace (J. Deng et al., 2022) or FaceNet (Schroff et al., 2015), it demonstrates effective recognition performance with acceptable parameter count for JavaScript-based on-device inference, which is aligned with the design goals of ShuffleFaceNet (Mart´ýnez-D´ýaz et al., 2019) and its successor (Hassanpour & Kowsari, 2023).

This model is trained to output a 128-dimensional facial embedding vector for each input image. During enrolment, the student's face is imaged and its average embedding representation is saved locally. In the recognition phase, the live embedding of an input face image is then computed and cosine distance against stored embedding is calculated. A match is identified when the cosine distance falls below a predefined threshold value.

3.4 Training Procedure

The model was fine-tuned on the collected dataset using the training split described in Section 3.1. A standard cross-entropy classification loss was used for the 16-class identity problem. Images were loaded from the structured folder layout (16 student folders), preprocessed, and passed through the network in mini batches of size 16. Parameters were updated using stochastic gradient descent with momentum. Training was performed for a maximum of 100 epochs with early stopping using a patience value of 20 to reduce overfitting. A learning rate scheduler (StepLR) was used to improve training stability. After training, the average enrolment embedding per student was computed as the stored reference representation for live recognition.

Table 3: *Training Configuration*

Hyperparameter	Value
Input Resolution	112×112
Embedding Dimension	128
Loss Function	Cross-Entropy
Optimizer	SGD with Momentum (0.9)
Learning Rate	0.01
Weight Decay	5×10^{-4}
Batch Size	16
Learning Rate Scheduler	StepLR
Maximum Epochs	100
Early Stopping Patience	20

3.5 Attendance Recognition Pipeline

The end-to-end attendance workflow follows a structured sequence:

1. Teacher logs in via Google Firebase Authentication (Bharti et al., 2020);
2. School and class are selected, loading the corresponding student enrolment data;
3. The device camera is activated and face detection begins scanning the video feed in real time;
4. MobileFaceNet (Chen et al., 2018) generates an embedding vector for each detected face;
5. Cosine distance matching against stored embeddings marks the student as present;

6. Unmatched faces cause a fallback to QR code or manual marking;
7. Attendance details are stored on the local IndexedDB, and they get synced with Firebase when a network connection becomes available.
8. A CSV report is generated for record-keeping and compliance purposes.

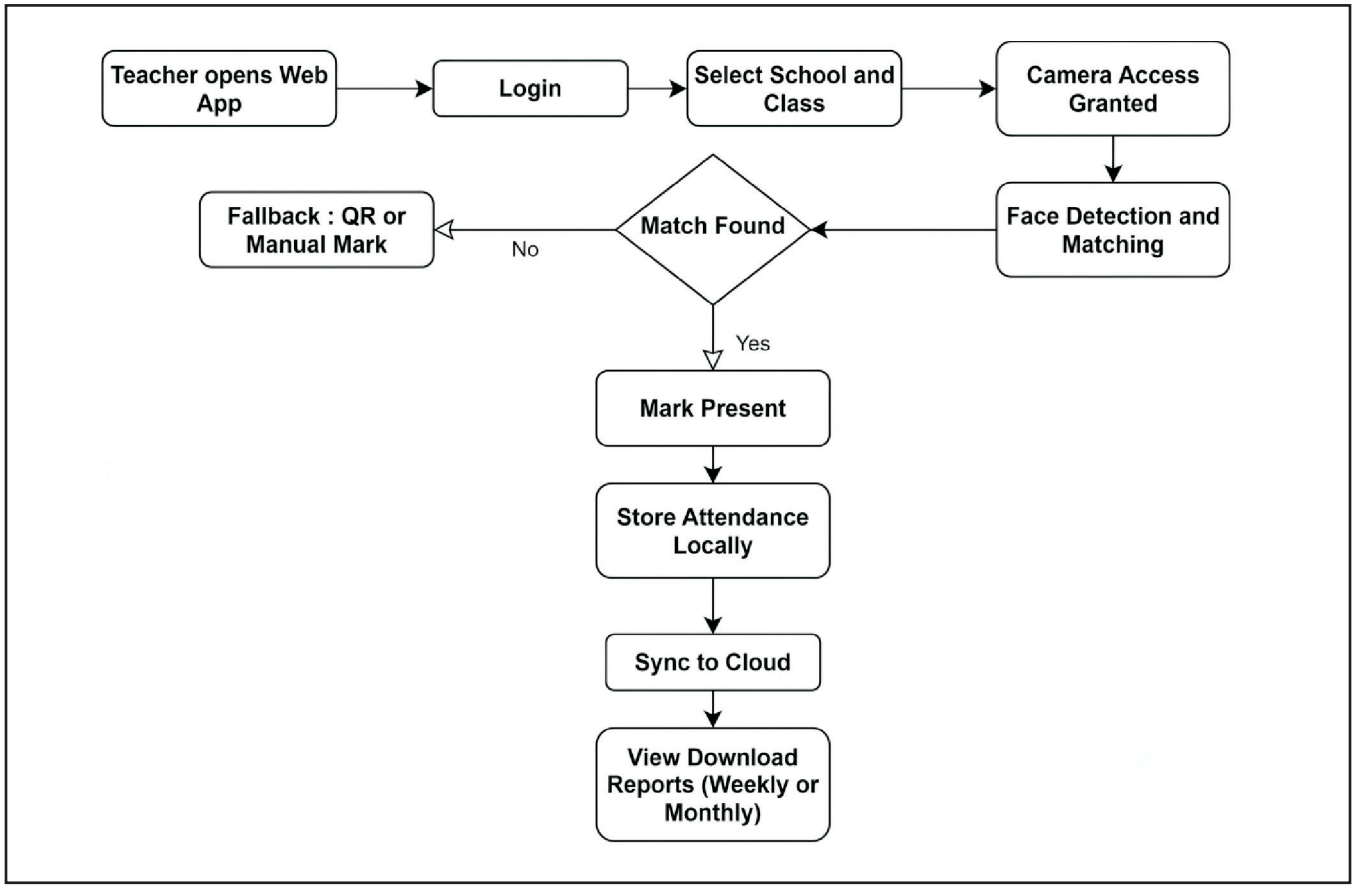


Fig. 1: *SmartFace Technical Workflow and Attendance Recognition Pipeline*

4 Experimental Setup

This section summarizes the details of experimental procedures. There are two sub-sections in this section. Section

4.1 describes the features of the dataset and the metric used in experiment. Section 4.2 lists the detailed hardware configuration and provides details about the software environment used in the experiment.

4.1 Dataset and Evaluation Metric

All experiments were conducted using the Indian Faces Image Classification Dataset (Garg et al., 2023), containing 320 images of 16 individuals, split into a 70% training set (224 images) and a 30% test set (96 images). From the training set a further 20% split was drawn to provide a 45 image validation set, training was reduced to 179 images and 96 images remained in the test set. The validation set was used to monitor model generalisation during training and support early stopping. Evaluation metrics and calculations were made using the standard scikit-learn python library whereas inference latency was calculated from average times of runs on images in the test set.

MobileFaceNet (Chen et al., 2018) was selected as the recognition model because of its lightweight design and high performance on mobile devices. Recognition accuracy was used as the primary evaluation metric in this study. Furthermore, the precision, recall, F1-score, AUC-ROC, and inference latency were also included in the evaluation criteria.

4.2 Hardware and Software Specifications

The SmartFace prototype was created and experimented on consumer devices to model the resource limited environment where the SmartFace system will be placed. The web application was implemented as a Progressive Web

Application (PWA) using HTML, CSS, and JavaScript with cloud storage through Firebase Firestore. Face recognition was implemented using a lightweight MobileFaceNet model and integrated with browser-based inference.

The application was tested on both laptop and Android smartphone devices to check its cross platform compatibility, offline operability and real-time attendance tracking with IndexedDB used as the storage for offline entries and firebase firestore to automatically sync the offline entries when connection becomes available.

5 Results and Discussion

This section is divided into four subsections. Section 5.1 presents and discusses the recognition performance of the proposed system. Section 5.2 compares SmartFace against existing attendance methods. Section 5.3 describes the features and user interfaces of the SmartFace application. Section 5.4 discusses the limitations and threats to validity of the current study.

5.1 Recognition Performance

The fine-tuned MobileFaceNet model achieved a training accuracy of 99.43%, a validation accuracy of 93.75%, and a test accuracy of 96.88% on the held-out 96-image split across 16 student classes. This result is consistent with the high accuracy reported for lightweight models on small, controlled datasets: Z.-Y. Deng et al. (2023) similarly observed high performance for their 3.5 MB model on in-domain test splits, while Hassanpour and Kowsari (2023) demonstrated that MobileFaceNet variants retain competitive recognition under significant parameter constraints. However, the reported accuracy should be interpreted in the context of the dataset used, which consists of 320 images from 16 subjects collected under controlled conditions. Though the experimental results indicate the feasibility of the proposed method, the method could be verified on a larger and more heterogeneous dataset to thoroughly assess its robustness on different realistic scenarios as future research.

Table 4: *Performance Metrics of SmartFace*

Metric	Value
Train Accuracy	99.43%
Validation Accuracy	93.75%
Test Accuracy	96.88%
Macro Precision	97.32%
Macro Recall	96.88%
Macro F1-score	96.85%
Weighted F1-score	96.85%
Mean AUC (OvR)	0.9999
Mean Inference Latency	382.4 ms (P95: 1025.0 ms)

The training curves in Figure 2 show steady convergence with no signs of severe overfitting; the small gap between training and validation loss indicates that the model generalises well within this dataset. A practical strength observed

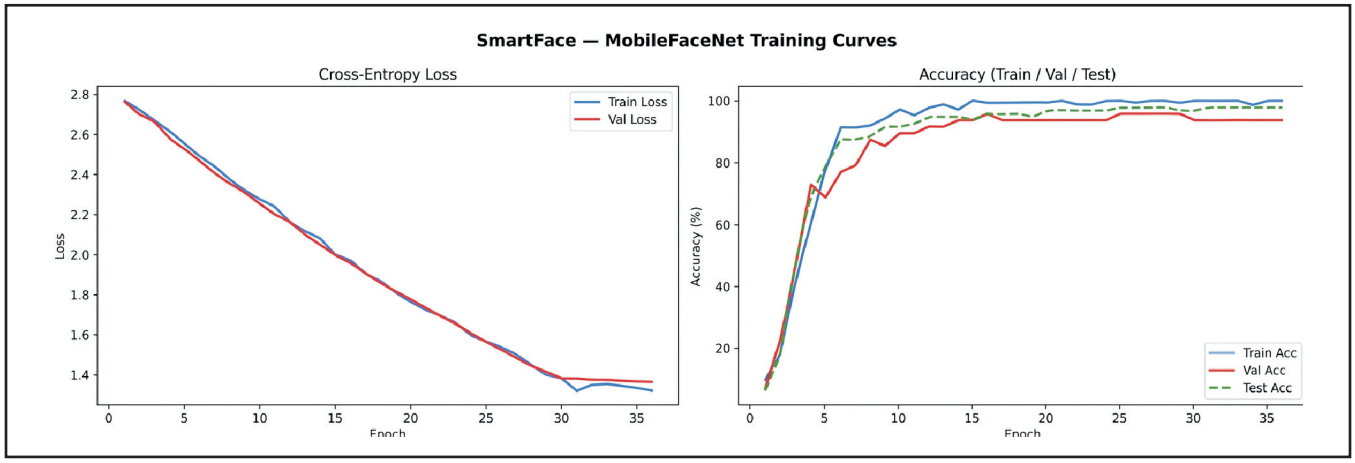


Fig. 2: *MobileFaceNet Training Curves: Cross-Entropy Loss and Train/Validation/Test Accuracy over 36 Epochs*

during testing was the model's resilience to spectacles and moderate lighting changes, which were deliberately included in the image collection protocol, consistent with the multi-condition robustness reported by Sufian et al. (2023) for FewFaceNet under similarly varied enrolment conditions.

To further validate the choice of recognition strategy, the 128-dimensional embeddings produced by the trained model were evaluated under four different classification approaches on the same test set. The results are summarised in Table 5.

Table 5: *Classifier Comparison on Frozen MobileFaceNet Embeddings (Test Set, n = 96)*

Classification Method	Test Accuracy	Classification Method	Test Accuracy
Cosine Nearest-Centroid	**96.88%**	Softmax Classifier Head	95.83%
KNN (*k*=1, cosine distance)	95.83%	SVM (RBF kernel)	95.83%

The cosine nearest-centroid approach used in SmartFace outperformed all alternative classifiers, confirming the suitability of the paper's recognition strategy. This method is also the most efficient at inference time, as it requires only a single cosine distance computation per enrolled student rather than a learned classifier head. Figure 3 illustrates the accuracy comparison across all four methods.

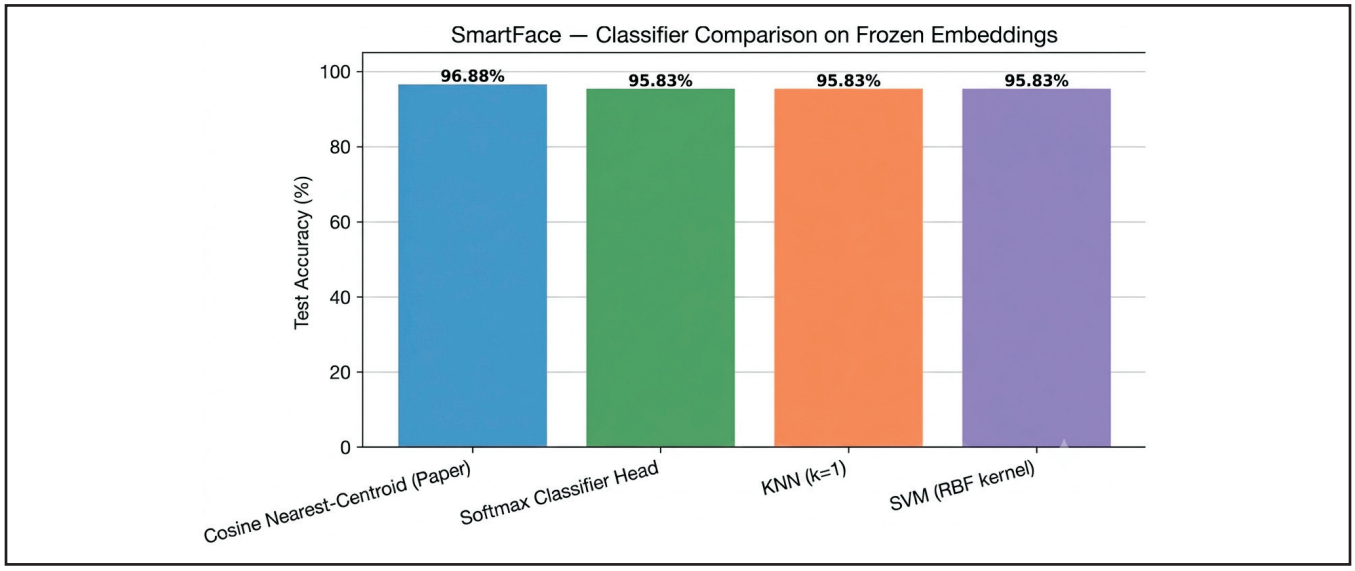


Fig. 3: *Classifier Comparison on Frozen MobileFaceNet Embeddings: Test Accuracy of Cosine Nearest-Centroid, Softmax, KNN, and SVM Methods*

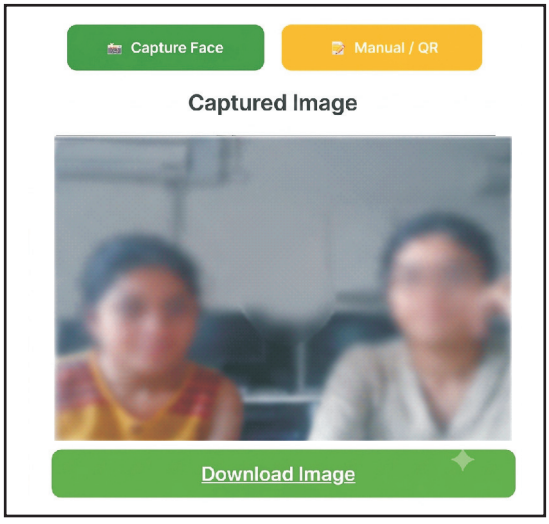


Fig. 4: *Attendance Capture Interface*

5.2 Comparison with Existing Approaches

Table 6 compares SmartFace against traditional manual attendance method and typical cloud-based face recognition systems across five parameters significant for rural school deployment.

As Table 6 demonstrates, SmartFace combines the offline capability of traditional methods with the minimal manual effort characteristic of cloud-based systems, while reducing the hardware and infrastructure costs associated with server-dependent deployments. Such a combination of factors would be especially useful in the setting of rural schools, wherein both dedicated hardware investment and ubiquitous Internet access are neither practical nor economically sustainable.

5.3 Features and User Interfaces of the SmartFace Application

SmartFace aims to support rural schools wherein the usage of digital administrative tools has been low so far. Teachers will be able to use SmartFace on existing hardware capable of taking photographs, so there will be no additional costs and the use of the application will be unaffected by unreliable rural network. Figures 5–7 illustrate the main user interfaces of the SmartFace application.

In addition to rural schools, the architecture is applicable for colleges and vocational institutes where, in order to meet

Table 6: *Comparison with Existing Attendance Methods*

Feature	Traditional Attendance	Cloud-Based Recognition	SmartFace
Manual Effort	High	Low	Low
Internet Required	No	Yes (continuous)	No (sync when available)
Hardware Cost	None	Moderate to High	Near Zero
Offline Capability	Full	None	Full (offline-first)
Deployment Simplicity	High (no setup)	Low (server config)	High (browser link)

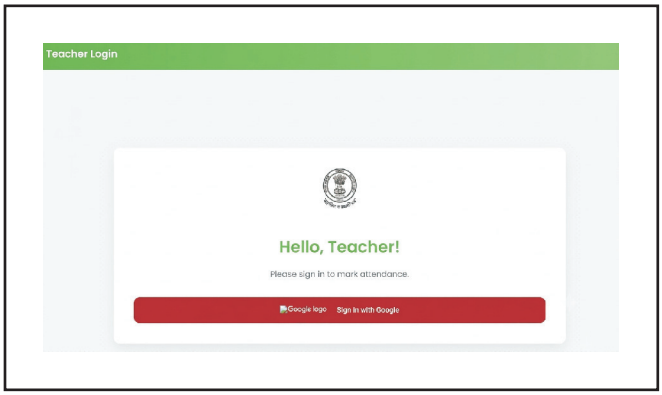


Fig. 5: *Teacher Login Dashboard*

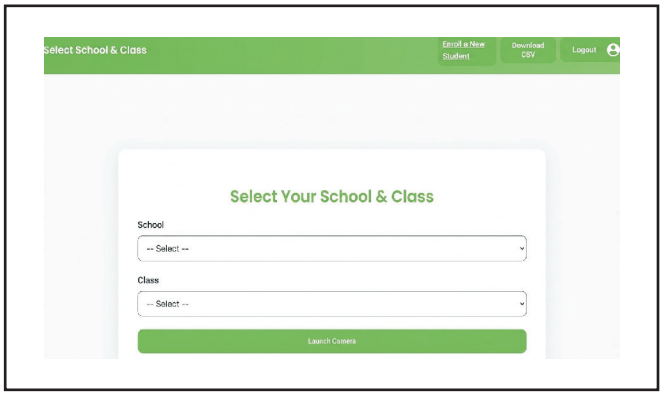


Fig. 6: *School and Class Selection Interface*

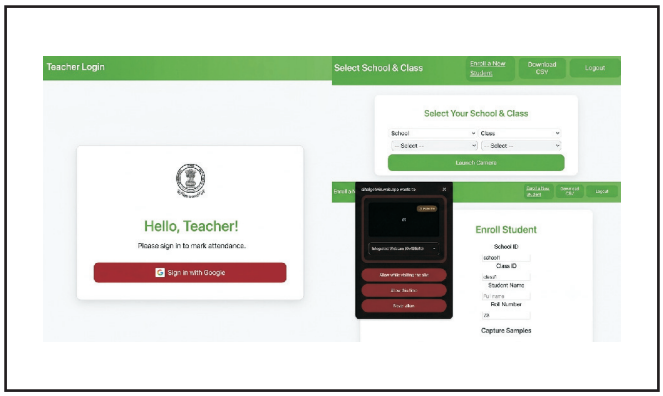


Fig. 7: *Student Enrolment Interface*

scholarship/certification criteria, verifiable records of attendance need to be maintained. Also, at exam centers, time stamped photo identity authentication can be performed to supplement manual checking. Government schemes that need to track participant attendance-e.g., skill development and community health, can benefit from the SmartFace system.

5.4 Limitations and Threats to Validity

Despite achieving a training accuracy of 99.43% and a test accuracy of 96.88%, several limitations warrant acknowledgement. First, the evaluation was conducted on a relatively small dataset (Garg et al., 2023) comprising 320 images from 16 subjects, which may not fully represent real-world classroom diversity. Second, the experiments were performed under semi-controlled conditions. Challenging scenarios such as severe illumination changes, motion blur, crowded classrooms, or significant facial occlusions were not evaluated. Prior research has established that broader validation on larger and more diverse datasets is necessary to assess recognition robustness conclusively (Chitrapu et al., 2025). Finally, the system has not yet undergone long-term deployment in rural schools, and practical factors such as user adoption, operational reliability, and biometric privacy considerations require further investigation prior to large-scale implementation (Sufian et al., 2023).

6 Conclusion and Future Work

Conclusion: This paper presents the design, development and preliminary evaluation of SmartFace, an AI-based on-device face recognition system to manage school attendance in rural Indian schools. The system integrates face recognition using MobileFaceNet (Chen et al., 2018) with a locally stored offline-first data architecture to facilitate recording attendance data without requiring a constant internet connection or dedicated server hardware. The fine-tuned model achieved a training accuracy of 99.43%, a validation accuracy of 93.75%, and a test accuracy of 96.88% on a dataset of 320 images of 16 students.

Future Work: Several directions are planned for future development:

1. A government-facing reporting dashboard for compliance and resource allocation;
2. A teacher analytics dashboard providing attendance trends to identify at-risk students;
3. Anti-spoofing detection using liveness detection techniques such as blink detection or depth estimation;
4. Significant dataset expansion through partnerships with rural schools across multiple states, enabling rigorous real-world evaluation;
5. Multi-classroom deployment coordination; and
6. A parent notification system for attendance alerts.

Acknowledgements

The authors would like to thank the students who contributed facial image data for the Indian Faces Image Classification Dataset (Garg et al., 2023), and the Department of Computer Science at Shaheed Rajguru College of Applied Sciences for Women, Delhi University, for their institutional support.

Conflict of Interest Statement

The authors declare no conflict of interest.

Ethical Declarations

The authors declare that this research does not involve human participants, animals, or any procedures requiring ethical approval. All data used in this study were obtained and processed in accordance with applicable ethical and legal guidelines.

References

1. Bharti, U., Bajaj, D., Tulika, Budhiraja, P., Juyal, M., & Baral, S. (2020). Android based e-voting mobile app using Google Firebase as BaaS. *Lecture Notes on Data Engineering and Communications Technologies*, *39*, 231–241. https://doi.org/10.1007/978-3-030-34515-0_24
2. Chen, S., Liu, Y., Gao, X., & Han, Z. (2018). MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices. *Biometric Recognition: 13th Chinese Conference, CCBR 2018*, *10996*, 428–438. https://doi.org/10.1007/978-3-319-97909-0_46
3. Chitrapu, P., Kumar Morampudi, M., & Kumar Kalluri, H. (2025). Robust face recognition using deep learning and ensemble classification. *IEEE Access*, *13*, 99957–99969. https://doi.org/10.1109/ACCESS.2025. 3575192
4. Deng, J., Guo, J., Yang, J., Xue, N., Kotsia, I., & Zafeiriou, S. (2022). ArcFace: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(10), 5962–5979. https://doi.org/10.1109/TPAMI.2021.3087709

5. Deng, Z.-Y., Chiang, H.-H., Kang, L.-W., & Li, H.-C. (2023). A lightweight deep learning model for real-time face recognition. *IET Image Processing*, *17*(13), 3869–3883. https://doi.org/10.1049/ipr2.12903
6. Garg, S., et al. (2023). Indian faces image classification dataset [[Data set]]. https://www.kaggle.com/datasets/shubh48/indian-faces-image-classification
7. Hassanpour, A., & Kowsari, Y. (2023). Lightweight face recognition: An improved MobileFaceNet model. https://arxiv.org/abs/2311.15326
8. Mart´ýnez-D´ýaz, Y., Me´ndez-Va´zquez, H., Nicola´s-D´ýaz, M., Luevano, L. S., Chang, L., & Gonzalez-Mendoza, M. (2019). ShuffleFaceNet: A lightweight face architecture for efficient and highly-accurate face recognition. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2721–2728. https://doi.org/10.1109/ICCVW.2019.00333
9. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474
10. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clus-tering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 815–823. https://doi.org/10.1109/CVPR.2015.7298682
11. Sufian, A., Ghosh, A., Barman, D., Leo, M., Distante, C., & Li, B. (2023). FewFaceNet: A lightweight few-shot learning-based incremental face authentication for edge cameras. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. https://openaccess.thecvf.com/content/ICCV2023W/ACVR/papers/Sufian_FewFaceNet_A_Lightweight_Few-Shot_Learning-Based_Incremental_Face_Authentication_for_Edge_ICCVW_2023_paper.pdf
12. Sun, Z., Feng, C., Patras, I., & Tzimiropoulos, G. (2024). LAFS: Landmark-based facial self-supervised learn-ing for face recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1639–1649. https://openaccess.thecvf.com/content/CVPR2024/html/Sun_LAFS_Landmark-based_Facial_Self-supervised_Learning_for_Face_Recognition_CVPR_2024_paper.html
13. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, *23*(10), 1499–1503. https://doi.org/10.1109/ LSP.2016.2603342

❖ ❖ ❖

Centre for Multidisciplinary Research, Innovation & Entrepreneurship

Shaheed Rajguru College of Applied Sciences for Women
University of Delhi
Vasundhara Enclave, Delhi - 110096