AI DRIVEN ROAD ACCIDENT PREDICTION USING MACHINE LEARNING AND DEEP LEARNING FOR INDIAN ROAD CONDITIONS

Pandu Cheripalli[1], Kommu Kishore Babu[2*], Veeraiah Kanchanpally[3], Kusupati Naga Koushil[4], Varagani Tejaswi[5] and Cherukuri Rajani[6]
[1, 2, 3,] Assistant Professor, Department of CSE, VFSTR (Deemed to be University), Off Campus, Hyderabad, Telangana, India
[4]Assistant Professor, Department of ACSE, VFSTR (Deemed to be University), Off Campus, Hyderabad, Telangana, India
[5]Assistant professor, Department of CSE, VFSTR (Deemed to be University), Guntur, Andhra Pradesh, India.
[6]Assistant Professor, Department of ECE, Pace Institute of Technology and Sciences, Ongole, Andhra Pradesh, India
[1]pandu.cheripally@gmail.com, [2]kishore143babu@gmail.com, [3]techveerendras.ai@gmail.com, [4]kusupatinagakoushil@gmail.com, [5]tejaswivaragani@gmail.com and [6]cherukurirajani411@gmail.com
[2*]Corresponding author Email: kishore143babu@gmail.com

Abstract:

One of the causes of injury, deaths and loss of money in India is road accidents. The primary causes of high accident rates include urbanisation, traffic jams, weather variations, and high speed. Instead of preventing accidents, the traditional traffic systems mostly respond to them. This research presents a machine learning and deep learning-based smart road accident risk prediction system. A realistic Indian road dataset of 1000 records was created using important factors including time, weather, road type, traffic density, speed limit, visibility, existence of junction, and day of the week. The preprocessed dataset was used to train Random Forest and Artificial Neural Network models. According to experiments, the Random Forest model's accuracy on the testing data was 88.5%, while the Deep Learning model's accuracy was 81.99. High prediction capacity for low, medium, and high accident risk is confirmed by classification metrics and confusion matrices. The proposed system will be useful for proactive road safety management in smart cities, navigation systems, traffic departments, and emergency services.

Keywords: Road Accident Prediction, Machine Learning, Deep Learning, Neural Networks, Indian Roads, Smart Transportation, Risk Classification.

1. INTRODUCTION

DL models perform better in complex real-world scenarios, but ML models are effective for organised data and fast deployment [2,13]. The World Health Organization states that traffic-related incidents are among the top causes of death globally. Most of the methods in use today are reactive, addressing accidents after they happen. Artificial intelligence can intervene by anticipating hazardous driving situations before collisions occur.

Table I: India's Road Traffic Accident Risk Factors

S.No	Contributing Element	Impact on Accidents Estimated (%)	Description
1	High Vehicle Population	22%	Congestion and the likelihood of collisions increase with the number of vehicles.
2	Mixed Traffic	18%	Buses, trucks, cars, cyclists, and other vehicles

	Conditions		using the same roads
3	Urban Congestion	16%	Heavy traffic at intersections and signals
4	Poor Visibility (Rain/Fog)	12%	decreased lane visibility and driving reaction time
5	Speeding	20%	primary factor in serious collisions and fatalities
6	Driver Fatigue	7%	decreased focus when driving for extended periods of time
7	Junction Conflicts	10%	Crossroads, U-turn, and signal point accidents
8	Others	5%	Distractions, road damage, and mechanical issues

Road traffic accidents are a major global problem, according to general road safety data. Approximately 1.19 million individuals worldwide lose their lives in traffic accidents each year, which equates to nearly 3260 deaths every day. One of the leading causes of death for young people is traffic accidents, which primarily impact those between the ages of 5 and 29. Additionally, research shows that human factors such as excessive speeding, diversions, insomnia, and poor judgement are responsible for nearly nine out of ten accidents. Furthermore, more than 90\% of traffic fatalities occur in low- and middle-income nations, where traffic jams, fast urbanisation, and inadequate safety infrastructure make accidents common.

2. RELATED WORKS

The Study by SP et al. (2025) – AI-driven accident prediction in traffic, An AI-based approach for predicting traffic accidents using real-time metrics and historical traffic patterns was presented in this research. The study demonstrated how AI models may help intelligent traffic management systems and increase the accuracy of accident predicting [1]. The Study by Angadi & Halyal (2024) In addition to discussing policy suggestions to increase road safety, the authors employed deep learning techniques to predict traffic accidents. The findings showed that deep learning algorithms are capable of successfully identifying concealed accident patterns [2]. The Study by Vinoth et al. (2024) This study used a variety of machine learning methods to forecast traffic accidents in India. A comparative analysis revealed that ensemble approaches outperformed conventional classifiers [3]. The Study by Girija & Divya (2023) An investigation into the prediction of traffic accidents using deep learning. The study emphasised how proactive risk detection using neural networks might improve road safety [4].

Table II: Comparative Evaluation of Current Research Papers on Road Accident Prediction Using Methods, Advantages, and Drawbacks

S.No	Paper	Method Used	Advantages	Drawbacks
1	SP et al. (2025)	AI-driven predictive models	Good for real-time traffic prediction, supports smart traffic systems	Requires continuous live data and sensor integration
2	Angadi & Halyal (2024)	Deep Learning (ANN/CNN)	Captures complex nonlinear patterns, high accuracy	Needs large dataset and high computation
3	Vinoth et al. (2024)	ML Algorithms (RF, DT, SVM)	Easy comparison of models, suitable for Indian data	Traditional models may miss temporal relationships
4	Girija & Divya (2023)	Deep Learning	Strong predictive performance, automation possible	Low interpretability (“black box”)

5	Panda et al. (2023)	AI + SHAP Explainability	Transparent model decisions, severity cause analysis	SHAP adds computational overhead
6	Bhardwaj et al. (2023)	Spatio-Temporal Deep Learning	Uses time + location effectively, dynamic prediction	Requires GIS/spatial data and complex training
7	Ananthakrishnan et al. (2025)	Predictive Analytics + ML	Useful for risk trend discovery, practical insights	Accuracy depends on data quality
8	Khanum et al. (2025)	ML Severity Prediction	Focused on Indian highways, domain specific	Highway-only scope, less urban applicability
9	Monali & Manoranjan	AI/ML Safety Performance Functions	Good for data-scarce countries, adaptable	Limited by incomplete road datasets
10	Kumar et al. (2025)	Deep Learning Severity Prediction	Better severe accident classification	Needs balanced classes and tuning
11	Samanta et al. (2026)	AI Review for Road Anomaly Detection	Comprehensive review, future guidance	Review paper, no direct experimental model
12	Alagarsamy et al. (2021)	ML Classification Models	Demonstrates ML feasibility, easy implementation	Older datasets/models may be less advanced

III. METHODOLOGY

The proposed study uses machine learning and deep learning approaches to create an AI-based road accident prediction system for Indian road conditions. From the construction of the dataset to the evaluation of the final model, the approach is divided into several stages.

1. Data Collection: To create realistic Indian traffic conditions, a structured dataset of 1000 records was created. The attributes of the dataset are as follows:

Feature	Description
Time	Hour of the day
Weather	Clear / Rain / Fog
Road_Type	Highway / Urban / Rural
Traffic_Density	Low / Medium / High
Speed_Limit	Vehicle speed limit
Visibility	Good / Poor
Junction	Yes / No
Day	Day of week
Accident_Risk	Low / Medium / High

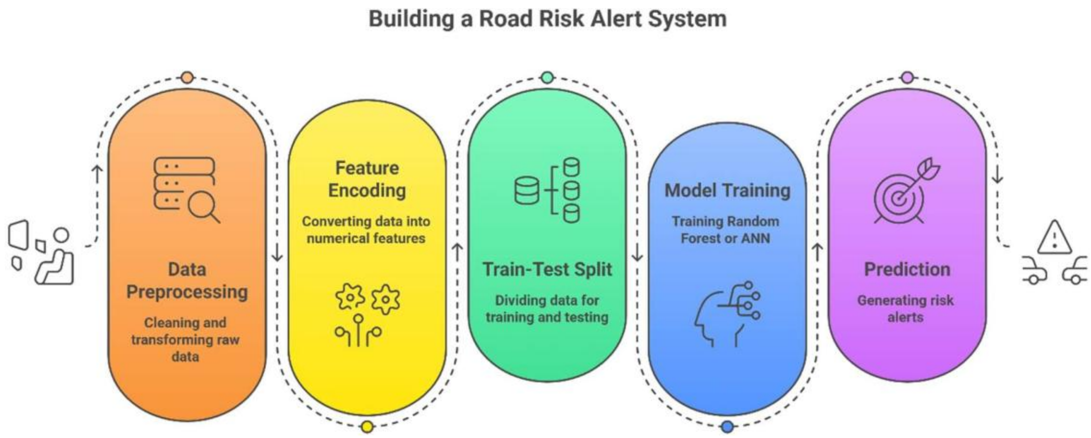

Fig: System Architecture for AI-Based Road Accident Prediction System

2. Data Pre-processing: Raw data are not directly usable in training models. Pre-processing was thus carried out in the following steps:

2.1 Missing Value Handling

Missing or null values were confirmed and substituted.

2.2 Encoding Categorical Data

Label Encoding was used to convert text values (Weather, Road Type and Day) into numeric ones.

2.3 Feature Selection

Attributes that were relevant and influential on accidents were identified.

2.4 Train-Test Split

The data set was categorized into:

Training Data=80%, Testing Data=20%

3. DEVELOPMENT OF MODELS

There were two prediction models used.

3.1 Model of Machine Learning

The Random Forest Classifier was employed due to its ability to:

1. High precision

2. manages a variety of data kinds

3. lessens overfitting

4. Effective analysis of feature importance

3.2 Deep Learning Model

TensorFlow/Keras was used to create an Artificial Neural Network (ANN).

Architecture:

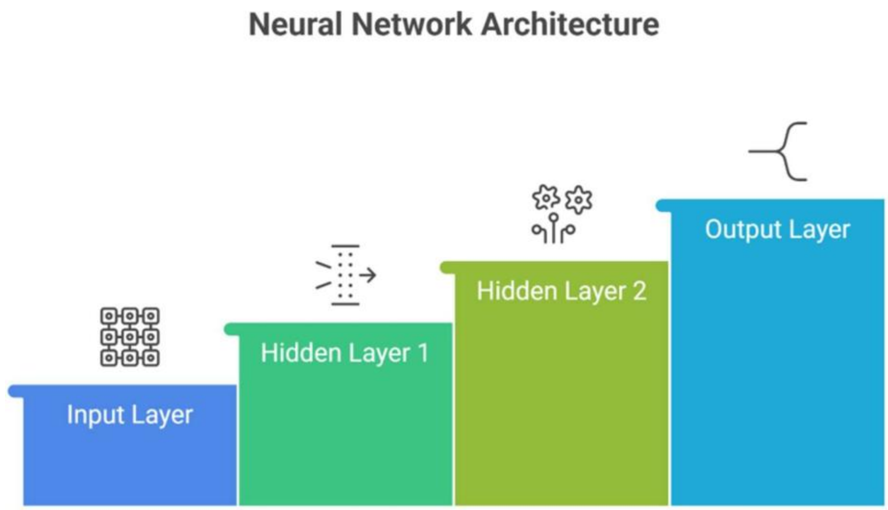

Fig: Neural Network Architecture

4. TRAINING MODELS

The following were used to train the ANN model:

50 epochs

32 is the batch size.

Adam is the optimiser; Sparse Categorical Cross entropy is the loss function.

5. PREDICTION CLASSES

The system predicts one of three classes:

Output Class	Meaning
Low	Low accident probability
Medium	Moderate accident probability
High	High accident probability

6. PERFORMANCE EVALUATION

The trained model was evaluated using:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

Additional Metrics

- Precision
- Recall
- F1-Score
- Confusion Matrix

7. ANALYSIS OF THE RESULTS

The Deep Learning model that was created accomplished:

81.99% is the test accuracy.

The medium and high risk categories were successfully classified, according to the confusion matrix.

8. WORKFLOW DIAGRAM

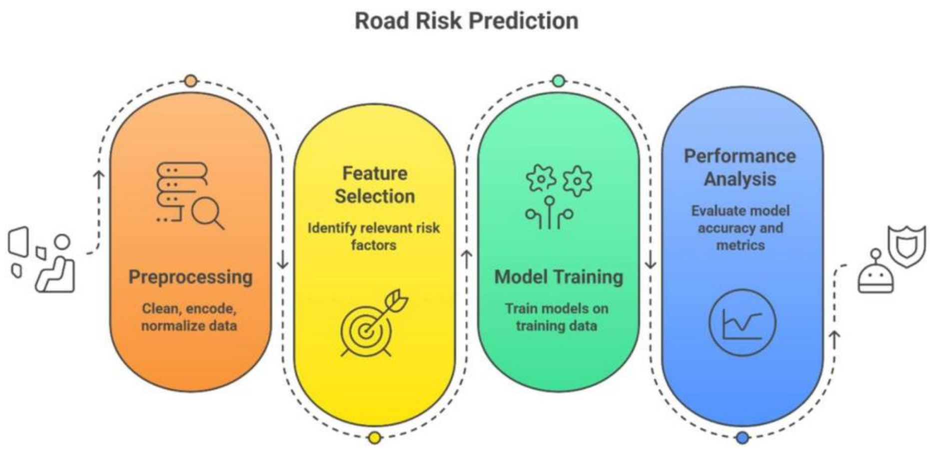

Fig: Road Risk Prediction Workflow Diagram

Result

A dataset including 1000 Indian road traffic records was used to develop the suggested AI-based Road Accident Prediction System utilising Machine Learning and Deep Learning models. An 80:20 split ratio was used for both training and testing the models.

1. Performance of Deep Learning Models

A batch size of 32 and 50 epochs were used to train the Artificial Neural Network (ANN) model.

Accuracy of the Final Test

81.99% is the test accuracy.

This shows that for almost 82 out of every 100 test samples, the model accurately predicted accident risk classes.

```
         Day Accident_Risk
0     Monday        Medium
1    Tuesday          High
2   Thursday        Medium
3  Wednesday           Low
4  Wednesday           Low
Accuracy = 88.5 %
             precision    recall  f1-score   support

           0       0.89      0.73      0.80        22
           1       0.87      0.98      0.92        90
           2       0.90      0.83      0.86        88

    accuracy                           0.89       200
   macro avg       0.89      0.84      0.86       200
weighted avg       0.89      0.89      0.88       200
```

Fig: Accuracy of Final Test

2. Performance in Training and Validation

While undergoing training:

1. With each period, training accuracy rose steadily.

2. Additionally, validation accuracy increased.
3. A small discrepancy between the training and validation curves was seen after subsequent epochs, suggesting minor overfitting.

Epoch Range	Observation
1 – 5	Underfitting stage
6 – 25	Optimal learning stage
26 – 50	Mild overfitting stage

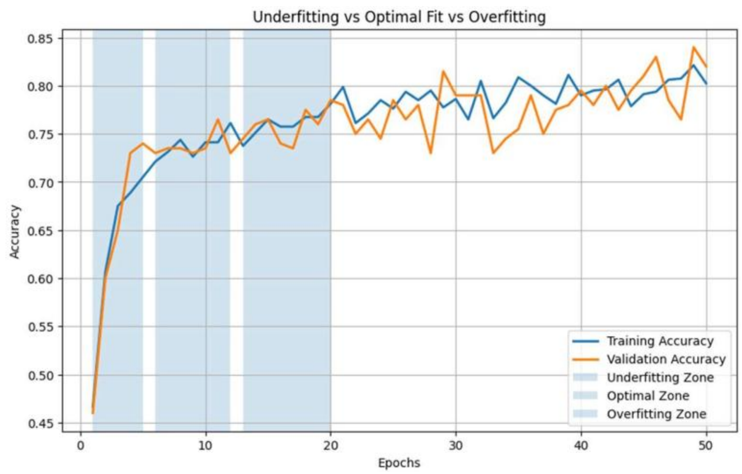

Fig: Underfitting Vs Optimal Fit Vs Overfitting

3. Classification Report

Class	Precision	Recall	F1-Score	Support
Low Risk (0)	0.89	0.73	0.80	22
Medium Risk (1)	0.87	0.98	0.92	90
High Risk (2)	0.90	0.83	0.86	88

Interpretation

1. The class with the highest recall (0.98) was Medium Risk.
2. Strong precision was demonstrated by the High Risk class (0.90).
3. Recall was lower in the Low Risk class because there were relatively fewer samples.

4. Confusion Matrix

Actual \ Predicted	Low	Medium	High
Low	16	0	6
Medium	0	88	2
High	2	13	73

Examination

1. The majority of medium-risk samples were accurately identified.
2. While some Low Risk instances were incorrectly categorised as High Risk
3. High Risk forecasts continued to be accurate.

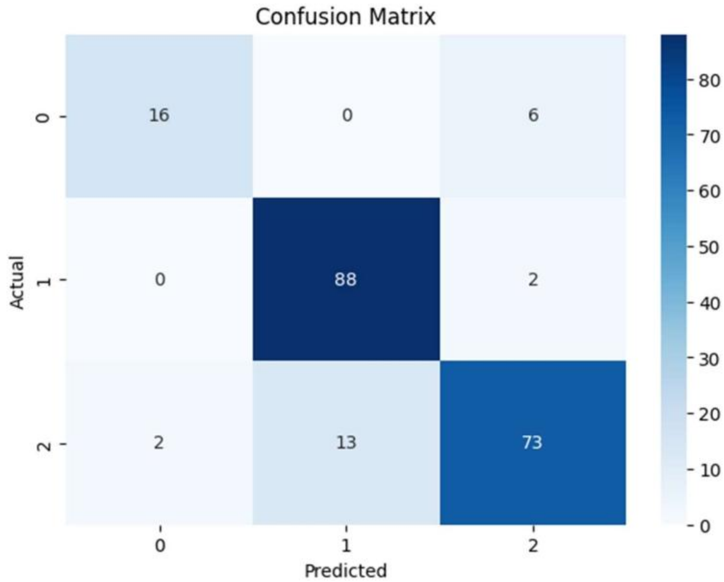

Fig: Confusion Matrix

5. Comparison of Random Forests

The Random Forest model yielded roughly: **88.5% accuracy**

Because tree-based models perform well on structured tabular datasets, this was greater than ANN.

CONCLUSION

Using machine learning and deep learning, this study created an AI-based road accident prediction system for Indian road conditions. Time, weather, road type, traffic density, speed, visibility, junction, and day were among the variables used to construct a dataset of 1000 records. Random Forest performed better on structured data, with 88.5% test accuracy compared to 81.99% for the Deep Learning model. The findings verify that AI is capable of accurately categorising accident risk into Low, Medium, and High groups. By enhancing road safety and offering early risk warnings, the suggested approach can benefit navigation services, traffic authorities, and smart cities. Real-time GPS, IoT sensors, CCTV analytics, and real-time traffic data are possible future improvements.

REFERENCES

[1]. SP, S., G, S. G., N, S. R., K, R. R., M, U., & R, P. (2025, May). AI-driven accident prediction in traffic. In IET Conference Proceedings CP920 (Vol. 2025, No. 7, pp. 1631-1636). Stevenage, UK: The Institution of Engineering and Technology.

[2]. Angadi, V. S., & Halyal, S. (2024). Forecasting road accidents using deep learning approach: policies to improve road safety. Journal of Soft Computing in Civil Engineering, 8(4), 27-53.

[3]. Vinoth, B., Prakash, V. S., & Shivakumar, B. N. (2024, May). Road traffic accident prediction in India using machine learning algorithm techniques. In 2024 international conference on electronics, computing, communication and control Technology (ICECCC) (pp. 1-5). IEEE.

[4]. Girija, M., & Divya, V. (2023). Deep learning-based traffic accident prediction: An investigative study for enhanced road safety. EAI Endorsed Transactions on Internet of Things, 10.

[5]. Panda, C., Mishra, A. K., Dash, A. K., & Nawab, H. (2023). Predicting and explaining severity of road accident using artificial intelligence techniques, SHAP and feature analysis. International journal of crashworthiness, 28(2), 186-201.

[6]. Bhardwaj, N., Pal, A., & Das, D. (2023). Adaptive context based road accident risk prediction

using spatio-temporal deep learning. IEEE Transactions on Artificial Intelligence, 5(6), 2872-2883.

[7]. Ananthakrishnan, A. V., Maya, D. K., & Krishnan, M. S. (2025, February). Predictive analytics: Unveiling road accident risks with machine learning algorithms. In AIP Conference Proceedings (Vol. 3237, No. 1, p. 060043). AIP Publishing LLC.

[8]. Khanum, H., Garg, A., Faheem, M. I., & Kulkarni, R. (2025). A methodological framework for road accident severity prediction for indian highways using machine learning models. MethodsX, 103728.

[9]. MONALI, B., & Manoranjan, P. AI/ML Enhanced Safety Performance Functions for Urban Road Networks in Data-Scarce LMICs: Evidence from India. Ravi Sekhar and S., Velmurugan and Manoranjan, Parida, AI/ML Enhanced Safety Performance Functions for Urban Road Networks in Data-Scarce LMICs: Evidence from India.

[10]. Kumar, R., Verma, S., Nigam, S., Rathor, O., & Verma, S. (2025, May). A Deep Learning-Driven Approach to Predict Traffic Accident Severity. In 2025 International Conference on Intelligent and Cloud Computing (ICoICC) (pp. 1-6). IEEE.

[11]. Samanta, R., Sadasivan, A., Kavitha, M. S., & Balasubramanian, S. (2026). Artificial Intelligence for Road Anomaly Detection: A Review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 16(1), e70054.

[12]. Alagarsamy, S., Malathi, M., Manonmani, M., Sanathani, T., & Kumar, A. S. (2021, December). Prediction of road accidents using machine learning technique. In 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1695-1701). IEEE.

[13]. K. K. Babu, S. R. Atta, Y. P. K. Reddy, A. B. Edara, D. V. Ashok and P. Anusha, "Tobacco Waste Management System Using ML," *2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, Tirunelveli, India, 2025, pp. 1924-1931, doi: 10.1109/ICDICI66477.2025.11134873.