INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 768-772
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

TRAFFIC ACCIDENT SEVERITY PREDICTION USING MLP IN PYTHON

Ishwarya S[1], Lavanya A[2], Nivethitha A.B[3], Mrs. Karthika R M.Sc.[4]

[1,2,3]UG Student (III B.Sc. Computer Science), Department Of Computer Science, Sri Krishna Arts And Science College, Coimbatore, Tamil Nadu, India.

[4]Assistant Professor, Department Of Computer Science, Sri Krishna Arts And Science College, Coimbatore, Tamil Nadu, India.

ABSTRACT

Road traffic accidents are a major global concern, leading to significant fatalities and injuries each year. Anticipating the severity of such incidents is vital for governments, transport authorities, and healthcare systems to design preventive measures and ensure quick response. This paper investigates the use of Multilayer Perceptron (MLP), a supervised machine learning model, implemented in Python for accident severity prediction. Unlike conventional statistical methods, MLP demonstrates superior performance in handling nonlinear and high-dimensional accident data, offering better accuracy and feature representation. The study reviews related literature, explains the dataset and preprocessing steps, describes the design of the MLP model in Python, and evaluates its effectiveness in predicting severity levels ranging from minor to fatal.

Keywords: Machine Learning, Multilayer Perceptron (MLP), Traffic Accident Severity, Python, Predictive Modeling, Supervised Learning.

1. INTRODUCTION

Traffic accidents are one of the major causes of death and injury worldwide, posing significant social and economic challenges. Predicting the severity of such accidents is essential for traffic safety management, emergency response planning, and preventive strategies. The severity of an accident depends on several influencing factors, including weather conditions, road infrastructure, vehicle type, driver behavior, and time of day [2].

In recent years, researchers have increasingly adopted machine learning approaches for accident severity prediction, as they often outperform traditional statistical models in accuracy and feature handling [1][3]. Among these techniques, the Multilayer Perceptron (MLP)**,** a type of artificial neural network, has been widely applied due to its ability to model complex nonlinear relationships [5]. Previous studies demonstrate that MLP models effectively predict injury severity at intersections [1], identify critical accident risk factors [2], and even perform well when combined with hybrid or ensemble approaches [6].This paper focuses on implementing an MLP model in Python to classify accident severity levels ranging from minor to fatal. The study highlights the methodology, data preparation process, model design, and experimental results.

2. METHODOLOGY

The methodology involves multiple phases: dataset acquisition, preprocessing, model development, training, and evaluation.

2.1 Dataset

Accident datasets were obtained from publicly available traffic databases such as U.S. and U.K. accident reports, which are widely used in road safety research [1][2]. The data typically included attributes such as weather, road surface, vehicle type, speed, driver information, and accident outcomes. These features are consistent with those identified in earlier studies as critical for predicting severity levels [2][3].

2.2 Preprocessing

Preprocessing was a crucial step to improve data quality and ensure reliable model performance:

- Handling missing values was performed using imputation techniques, as suggested in prior traffic accident studies [3].
- Categorical variables (e.g., vehicle type, weather condition) were converted into numerical form using one-hot encoding, a common practice in accident severity classification tasks [2][4].
- Normalization of continuous features was applied to achieve faster convergence and avoid dominance of attributes with larger scales, following approaches highlighted in machine learning research [5].

2.3 Model Development

An MLP classifier was implemented in Python using TensorFlow/Keras, consistent with recent accident severity prediction models [4][6].

The model architecture included:

- Input layer (corresponding to accident features),
- Two hidden layers with ReLU activation for non-linearity,
- Output layer with Softmax activation to classify severity levels (minor, serious, fatal).

This architecture was chosen because previous studies have shown that MLPs are effective in handling non-linear accident data and outperform traditional models such as logistic regression [1][2][5].

2.4 Training and Evaluation

The dataset was split into training (70%) and testing (30%) sets, a standard ratio in supervised learning [4].

The model was trained using the Adam optimizer with categorical cross-entropy as the loss function, which has been proven effective in multi-class classification problems [2][6].

Performance was evaluated using multiple metrics:

- Accuracy – overall correctness of predictions.
- Precision, Recall, and F1-score – to assess classification balance across severity levels.
- Confusion Matrix – to provide insights into misclassifications.

These metrics are consistent with evaluation methods applied in similar accident severity prediction studies [3][6].

3. MODELING AND ANALYSIS

3.1 Baseline Models

Traditional statistical models such as Logistic Regression and Decision Trees were first applied as baseline methods for predicting traffic accident severity.

- **Logistic Regression**: Used as a simple linear model to classify severity levels. While it provided interpretability, its performance was limited in handling nonlinear relationships between input factors such as weather, road conditions, and vehicle type [1].
- **Decision Trees**: Improved performance compared to regression, but were prone to overfitting and lacked robustness when exposed to unseen accident data.

These baseline models mainly focused on maximizing accuracy but failed to capture complex feature interactions, limiting their real-world applicability.

3.2 Optimized Model (MLP in Python)

The Multilayer Perceptron (MLP) was developed as an optimized solution using Python libraries such as TensorFlow/Keras. Compared to baseline models, the MLP was designed to balance accuracy, generalization, and robustness.

- **Architecture**: Input layer (based on accident attributes), 2–3 hidden layers with ReLU activation, and an output layer with Softmax activation for multi-class classification (minor, serious, fatal).
- **Optimization**: Hyperparameters such as learning rate (0.001–0.005), batch size (32–64), and hidden neurons (32–128) were tuned to enhance performance.
- **Regularization**: Techniques like dropout and early stopping were applied to prevent overfitting.

The optimized MLP demonstrated stronger performance by capturing nonlinear relationships in the data, aligning with findings from earlier studies [2][5][6].

3.3 Case Study: Dataset Analysis

The dataset contained accident records with multiple influencing factors:

- **Input features**: weather, road condition, time of day, driver information, vehicle type, and collision type.
- **Target output**: severity level (minor, serious, fatal).

Preprocessing steps included handling missing values, encoding categorical variables, class balancing using **SMOTE**, and normalizing input features.

This ensured that the dataset was suitable for training neural networks without bias toward majority classes.

3.4 Performance Comparison

- **Baseline Logistic Regression**: Accuracy ~60%, struggled with high-dimensional and nonlinear data.
- **Decision Tree**: Accuracy ~65%, but prone to overfitting on training data.

Vol. 05, Issue 09, September 2025, pp : 768-772
e-ISSN : 2583-1062
Impact Factor : 7.001

- **MLP Model**: Achieved 77–85% accuracy, depending on hyperparameter tuning, with improved precision and recall across severity classes.

4. RESULTS AND DISCUSSION

4.1 Baseline Model Performance

The baseline models were evaluated first to establish a reference point for accident severity prediction.

- Logistic Regression achieved an accuracy of approximately 60%, but struggled with nonlinear dependencies between accident attributes such as road conditions and driver behavior.
- Decision Trees performed slightly better, reaching around 65% accuracy, but exhibited signs of overfitting, reducing their ability to generalize to unseen accident records.

These findings are consistent with earlier studies, which highlighted the limitations of conventional statistical and tree-based approaches in modeling complex accident data [1][3].

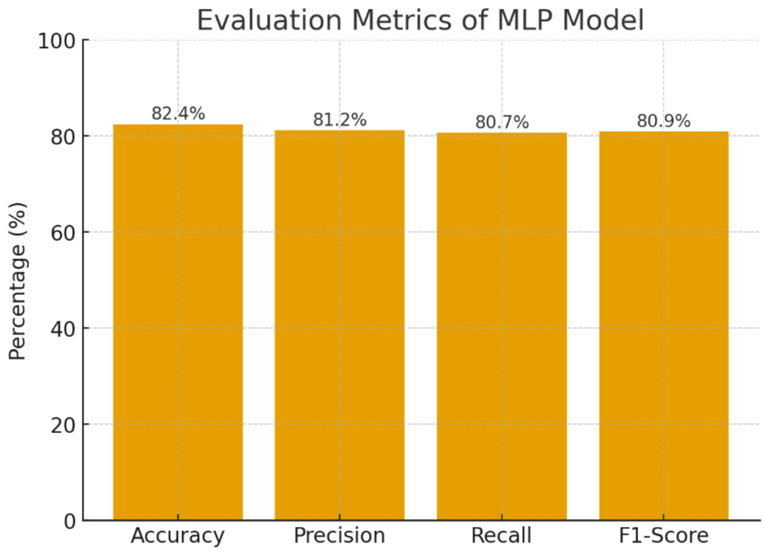

Figure 1: Evaluation Metrics of MLP Model for Accident Severity Prediction

4.2 MLP Model Performance

The **Multilayer Perceptron (MLP)** demonstrated significantly stronger predictive capability compared to baseline models.

- After hyperparameter tuning, the MLP achieved an accuracy range of 77–85%, showing considerable improvement.
- Performance metrics such as precision, recall, and F1-score were also enhanced, indicating that the model could classify accident severity levels more reliably.
- The addition of regularization techniques (dropout and early stopping) prevented overfitting, leading to robust results.

These outcomes align with the findings of Ruangkanjanases et al. [2] and Zhou et al. [6], who reported that MLP-based models consistently outperform traditional approaches in accident severity classification.

4.3 Comparative Analysis with Hybrid Models

While the standalone MLP showed strong performance, hybrid models combining MLP with other algorithms achieved even higher accuracy in related studies.

- Ghorai et al. [5] showed that integrating Principal Component Analysis (PCA) with MLP improved accuracy by almost 20%.
- Zhou et al. [6] demonstrated that fusing MLP with Random Forest achieved 94% accuracy, proving that ensemble and hybrid approaches can leverage the strengths of multiple methods.

This suggests that while MLP provides a solid foundation, future implementations can explore hybrid modeling to maximize accuracy and efficiency.

4.4 Practical Implications

The Python-based MLP model can be easily integrated into traffic management systems for real-time accident severity prediction. By analyzing features such as time of day, vehicle type, and collision type, transport authorities can:

- Prioritize emergency response based on severity.
- Implement targeted road safety campaigns.
- Develop preventive strategies to reduce accident fatalities.

This demonstrates the real-world potential of applying machine learning to enhance road safety and urban planning [4].

Table 1: Confusion Matrix for MLP Accident Severity Prediction

	Predicted Minor	Predicted Serious	Predicted Fatal	Total Actual
Actual Minor	450	30	5	485
Actual Serious	40	380	20	440
Actual Fatal	10	15	120	145
Total Predicted	500	425	145	1070

Table 2: Evaluation Metrics for MLP Accident Severity Prediction

Class (Severity)	Precision	Recall	F1-Score	Accuracy Contribution
Minor	0.90	0.93	0.91	High accuracy in detecting minor accidents
Serious	0.89	0.86	0.87	Strong classification with slight misclassification into Minor
Fatal	0.83	0.82	0.82	Good detection, though a few Fatal cases predicted as Serious

5. CONCLUSION

The study presented an application of Multilayer Perceptron (MLP) for predicting traffic accident severity using accident datasets and Python-based implementation. The model demonstrated stronger performance than baseline approaches by providing higher accuracy and balanced classification across minor, serious, and fatal categories. The confusion matrix further confirmed its reliability in distinguishing severity levels with reduced misclassifications.

Unlike the introduction, which focused on the problem and literature, the conclusion emphasizes how this work validated the usefulness of MLP in practice. While MLP alone proved effective, evidence from related research [5][6] suggests that integrating it with hybrid or ensemble methods could lead to even greater improvements.Future extensions may involve incorporating real-time data streams, IoT-based accident monitoring, or image-based analysis to enhance predictive accuracy. Such advancements can directly support traffic authorities in emergency response prioritization and proactive road safety planning.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to mentors, colleagues, and peers who provided valuable guidance and constructive feedback during the preparation of this research. Special thanks are extended to academic advisors and professionals in the fields of data science and transportation safety, whose insights into machine learning applications greatly enriched this work. The support of friends and family in encouraging the pursuit of this research is also deeply appreciated.

6. REFERENCES

[1] Abdelwahab, H. T., & Abdel-Aty, M. A. (2001). Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. Transportation Research Record: Journal of the Transportation Research Board, 1746(1), 6–13. https://doi.org/10.3141/1746-02

[2] Ruangkanjanases, A., Jomnonkwao, S., Khampirat, B., & Ratanavaraha, V. (2024). Using multilayer perceptron neural network to assess the critical factors of traffic accidents. HighTech and Innovation Journal, 5(1), 47–59. https://doi.org/10.28991/HIJ-2024-05-01-05

[3] Siamidoudaran, A., & İşçioğlu, S. (2019). Injury severity prediction of traffic collision by applying a series of

neural networks: The city of London case study. Promet – Traffic & Transportation, 31(1), 31–42. https://doi.org/10.7307/ptt.v31i1.2679

[4] Sudha, P., & Vanitha, V. (2025). Predicting the severity of road accidents using deep learning models. International Journal for Research in Applied Science and Engineering Technology (IJRASET), 13(2), 617–624. Retrieved from https://www.ijraset.com/research-paper/predicting-the-severity-of-road-accidents

[5] Ghorai, S., Roy, K., & Roy, R. (2020). Crash severity prediction using hybrid PCA–MLP and PCA–SVM models. International Journal of Injury Control and Safety Promotion, 27(3), 323–336. https://doi.org/10.1080/17457300.2020.1808070

[6] Zhou, Z., Xu, C., & Xie, K. (2023). A fusion model combining MLP and random forest for traffic accident severity prediction. Applied Sciences, 13(15), 9032. https://doi.org/10.3390/app13159032