URBAN TRAFFIC CRASH ANALYSIS USING DEEP LEARNING TECHNIQUES

Mummaneni Sobhana[1], Nihitha Vemulapalli[1], Gnana Siva Sai Venkatesh Mendu[1], Naga Deepika Ginjupalli[1], Pragathi Dodda[1], Rayanoothala Bala Venkata Subramanyam[2]

[1]Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Vijayawada, India, [2]National Institute of Technology Warangal, Department of CSE, Warangal, India

Abstract. Road accidents are concerningly increasing in Andhra Pradesh. In 2021, Andhra Pradesh experienced a 20 percent upsurge in road accidents. The state's unfortunate position of being ranked eighth in terms of fatalities, with 8,946 lives lost in 22,311 traffic accidents, underscores the urgent nature of the problem. The significant financial impact on the victims and their families stresses the necessity for effective actions to reduce road accidents. This study proposes a framework that collects accident data from regions, namely Patamata, Penamaluru, Mylavaram, Krishnalanka, Ibrahimpatnam, and Gandhinagar in Vijayawada (India) from 2019 to 2021. The dataset comprises over 12,000 records of accident data. Deep learning techniques are applied to classify the severity of road accidents into Fatal, Grievous, and Severe Injuries. The classification procedure leverages advanced neural network models, including the Multilayer Perceptron, Long-Short Term Memory, Recurrent Neural Network, and Gated Recurrent Unit. These models are trained on the collected data to accurately predict the severity of road accidents. The project study to make important contributions for suggesting proactive measures and policies to reduce the severity and frequency of road accidents in Andhra Pradesh.

Keywords: classification, gated recurrent unit, long-short term memory, multilayer perceptron, recurrent neural network, road accidents

ANALIZA KOLIZJI W RUCHU MIEJSKIM Z WYKORZYSTANIEM TECHNIK GŁĘBOKIEGO UCZENIA

Streszczenie. Liczba wypadków drogowych w Andhra Pradesh niepokojąco rośnie. W 2021 r. stan Andhra Pradesh odnotował 20% wzrost liczby wypadków drogowych. Niefortunna pozycja stanu, który zajmuje ósme miejsce pod względem liczby ofiar śmiertelnych, z 8 946 ofiarami śmiertelnymi w 22 311 wypadkach drogowych, podkreśla pilny charakter problemu. Znaczący wymiar finansowy dla ofiar i ich rodziny podkreśla konieczność podjęcia skutecznych działań w celu ograniczenia liczby wypadków drogowych. W niniejszym badaniu zaproponowano system gromadzenia danych o wypadkach z regionów Patamata, Penamaluru, Mylavaram, Krishnalanka, Ibrahimpatnam i Gandhinagar w Vijayawada (India) w latach 2019–2021. Zbiór danych obejmuje ponad 12 000 rekordów danych o wypadkach. Techniki głębokiego uczenia są stosowane do klasyfikowania wagi wypadków drogowych na śmiertelne, poważne i ciężkie obrażenia. Procedura klasyfikacji wykorzystuje zaawansowane modele sieci neuronowych, w tym wielowarstwowy perceptron, pamięć długoterminową i krótkoterminową, rekurencyjną sieć neuronową i Gated Recurrent Unit. Modele te są trenowane na zebranych danych w celu dokładnego przewidywania wagi wypadków drogowych. Projekt ma wnieść istotny wkład w sugerowanie proaktywnych środków i polityk mających na celu zmniejszenie dotkliwości i częstotliwości wypadków drogowych w Andhra Pradesh.

Słowa kluczowe: klasyfikacja, gated recurrent unit, pamięć długotrwała i krótkotrwała, perceptron wielowarstwowy, rekurencyjna sieć neuronowa, wypadki drogowe

Introduction

Road accidents on public roads involve vehicles, pedestrians, and cyclists. Factors like driving errors, mechanical malfunctions, weather conditions, and road infrastructure lead to accidents. Negligent behaviors such as speeding, distracted driving, and fatigue worsen the situation. Accidents result in injuries, fatalities, property damage, and emotional distress [5]. Addressing the issue requires strict regulations, speed enforcement, awareness campaigns, driver education, and improved infrastructure.

Prevention, awareness, and technology can reduce accidents, making roads safer for everyone. Various types of road accidents can occur. Rear-end collisions happen when one vehicle hits the back of another. Intersection accidents occur when vehicles collide while crossing paths at intersections. Single-vehicle accidents involve a vehicle colliding with a fixed object like a tree or a pole [29]. Head-on collisions happen when two vehicles collide front to front. When two vehicles crash side by side, the result is a side-impact collision, sometimes known as a T-bone accident. Rollover accidents involve a vehicle flipping over onto its side or roof. Pedestrian and cyclist accidents involve collisions with non-motorized road users.

Road accidents happen for different reasons. When drivers get distracted by using their phones or eating, they lose focus on the road [19]. Driving too fast because of impatience can make it hard to stay in control. If people drive after drinking alcohol or taking drugs, their judgment and coordination are affected. Reckless driving, like taking dangerous actions and ignoring traffic rules, makes accidents more likely [23]. Bad roads, bad weather, tired drivers, problems with vehicles, inexperienced drivers, and not following traffic laws can all cause accidents too. Road accident severity pertains to the extent of seriousness and impact of an accident, which encompasses injuries, fatalities,

and property damage [6]. Factors that contribute to this severity include speed, collision type, road conditions, weather, and human behavior. Through the analysis of accident data, patterns and causes can be identified to develop targeted strategies, such as improving infrastructure, enforcing regulations, promoting responsible driving, and enhancing emergency response. The focus is on ensuring road safety and minimizing the occurrence of accidents, with efforts aimed at reducing severity, saving lives, and minimizing harm.

Controlling road accidents demands a comprehensive approach involving various strategies and stakeholders. One of the fundamental pillars is road safety education and awareness, disseminating knowledge about responsible driving practices and the potential consequences of reckless behavior through campaigns and outreach programs. Equally critical is strict enforcement of traffic laws, employing increased police presence and technology like speed cameras to deter dangerous driving habits. Driver education and training programs play a pivotal role in equipping motorists with essential skills and knowledge for safe driving. Continuous infrastructure improvements, including well-designed roads, clear signage, and pedestrian-friendly facilities, contribute to accident prevention. Advanced Driver Assistance Systems (ADAS) empowered by technologies like deep learning can assist drivers in avoiding collisions and maintaining safe driving behavior [18]. Adapting speed limits based on road conditions and implementing weather monitoring and warning systems add additional layers of safety. Involving the community in road safety initiatives fosters shared responsibility for safer roads, while international cooperation enables the exchange of best practices and innovative solutions. Integrating these measures, societies can make significant strides towards minimizing road accidents, protecting lives, and ensuring safer road environments for everyone.

1. Literature review

Mohammad Zarei et al. [27] proposed a non-parametric Empirical Bayes (EB) procedure using CGAN to model crash frequency data. The CGAN-EB method surpasses the conventional NB-EB method concerning hotspot identification and prediction. Compared to the negative binomial (NB) model, it offers superior accuracy for fitting accident data and making predictions. When it comes to network screening and crash modeling, CGAN-EB outperforms the standard NB-EB methodology by a significant margin. The study demonstrates the benefit of the approach through the analysis of datasets obtained from both real-life and hypothetical accident scenarios. It examines various aspects, including the model's fit to the data, the expected accuracy of predictions, and the outputs of network screening. This comprehensive evaluation sheds light on the advantages and effectiveness of the method.

Kaffash Charandabi et al. [14] analyzed a road accident risk mapping approach to reduce road accidents by identifying perilous areas. The proposed hybrid model combines a GRNN and a self-organizing map, considering 22 predictor attributes. An average of 90.74% in the testing on Iran's Tabriz Marand dual carriageway demonstrated a high level of accuracy in estimating accident probability. The study emphasized that areas prone to accidents were notably influenced by factors such as the presence of traffic surveillance cameras, the day of the week, driver age, local climate conditions, altitude variations, and even the specific type of vehicle traversing the area.

Zheng et al. [28] explored a hybrid deep-learning model for short-term traffic flow prediction in smart transportation systems. The framework merges a Bi-Directional Long Short-Term Memory (Bi-LSTM) model, which captures daily and weekly periodicities, with an attention-based Convolutional LSTM (Conv-LSTM) module that encodes spatial and short-term temporal data. Through empirical validation, their approach exhibited remarkable prediction performance surpassing existing methods, and underscoring the efficacy of integrating spatial-temporal features and an attention mechanism. The study not only advances traffic forecasting but also contributes to the foundation of intelligent transportation systems, hinting at the potential of their holistic approach in shaping the future of urban mobility.

Komol et al. [15] examined the effectiveness of advanced warning systems in detecting drivers' intended movements at crossings. Through in-depth analysis, they scrutinized predictors including vehicle speed, acceleration, and yaw rate, leveraging data from the Ipswich Connected Vehicle Pilot project across varying warning distances. Notably, the study demonstrated enhanced prediction accuracy by adopting a strategy of training distinct prediction models for each junction. This innovative approach not only refines prediction precision but also underscores the significance of context-sensitive models in bolstering the functionality of advanced driver assistance systems.

Dr. Rahman et al. [20] studied the significance of considering both the causes and effects of traffic accidents in road safety legislation. They examined purposeful and unintentional driving behaviors and their links to different types and severity levels of accidents using a BBN model. According to the study, which looked at the connections between driving habits and accidents in Al-Ahsa, the chance of occurrence of an accident is significantly increased by speeding alone. It doubles when brake failure is present. The complicated relationships between driving habits and accident causes are effectively captured by the BBN model.

Jie Yan et al. [25] proposed a novel approach for analyzing accident risk associated with highway geometric alignment using satellite maps and clustering algorithms. The method was validated using real-world data from the Nanfu highway in Chongqing, China. The study emphasized the significance of input variables such as section unit length, curve radius, and slope gradient. The MLP model outperformed the negative binomial model in predicting accident risk, revealing complex nonlinear relationships between variables. This research provides insights for enhancing traffic safety in mountainous regions.

Vani Suthamathi Saravanarajan et al. [22] proposed a study, that emphasizes the vital role of accident detection models in Autonomous Vehicles (AVs) safety. It highlights the need for robust deep learning methods to identify various accident scenarios, including isolated vehicle incidents. The study presents an innovative system using three deep learning models, achieving an 86.25% Accident Detection Rate (ADR) and a 33.00% False Alarm Rate (FAR). These results underscore the substantial potential of these models in enhancing AV safety and dependability.

Camilo Gutierrez-Osorio et al. [11] suggested an ensemble Deep Learning Model using GRU and CNN to predict traffic accidents. They leverage social media and open data sources to access detailed information, including unreported incidents. The model undergoes data quality and feature engineering processes. Results show that the ensemble model gave a better output than base models. The model's information can assist traffic control agencies in planning accident prevention activities, especially in high-traffic regions and intersections.

Dragan Gatarić et al. [9] to predict traffic accidents in Serbia and Bosnia & Herzegovina, two ANN models are suggested. These models forecast the number and severity of accidents using objective variables such as the length of the type of terrain, width of the road, volume of traffic, and speed threshold. The ANN models successfully predict outcomes and show good generalization skills. The study emphasizes how ANN might help with accident prediction and transportation planning. Accidents are found to be significantly influenced by section length, indicating that shortening it could lessen accidents and their severity.

Vishal Mandal et al. [17] proposed a novel approach for real-time traffic monitoring using CNNs and a GUI. This cutting-edge system employs advanced deep-learning algorithms and object tracking to identify lines, track stationary vehicles, and perform vehicle counting. Notably adaptable, it excels even in challenging environmental conditions, attesting to its robustness. The presented framework stands as a testament to its potential to revolutionize real-time traffic analysis, offering a reliable and comprehensive solution for varied scenarios.

Carlos M. Ferreira-Vanegas et al. [8] proposed a systematic literature review that investigates the utilization of statistical analysis (SA), machine learning (ML), smart city technologies (TESC), and Geographic Information Systems (GIS) for analyzing road traffic accidents (RTAs). Analyzing a dataset of 3,888 papers published between 2000 and 2021, the review identifies important papers, authors, and journals, as well as highlights emerging regression models and advanced technologies. Addressing the issue of unobserved heterogeneity and the significance of human factors in RTA analysis, the study emphasizes the value of computational algorithms, data visualization, and bibliometric tools for future research.

A. Comi et al. [7] proposed a system that utilizes data mining techniques to analyze road accident data in Rome Municipality. Through clustering and descriptive analysis, significant causes and patterns of accidents are identified, highlighting the influence of vehicle types and road infrastructure conditions on accident severity. The findings emphasize the potential of data mining in planning measures to reduce accidents and predict accident-prone areas. However, limitations are identified, suggesting the need for further research to investigate the impact of focusing on fatal accidents and exploring hybrid prediction approaches combining statistical and machine learning models.

K. Athiappan et al. [4] studied identifying accident-prone areas and the key factors causing accidents in the medium-sized Indian city of Tirunelveli. It employs geospatial analysis to pinpoint these "blackspots" and ranks the most influential factors based on recurrent accidents. The research provides crucial insights for transportation planners to mitigate casualties during

road construction and in existing conditions. The analysis identifies 30 accident-causing factors, with the top 10 including issues like volume-to-capacity ratio, lack of sight distance, uncontrolled intersections, and driver behaviors such as drunk driving and speeding. This comprehensive approach highlights the complexity of addressing accident causes in India, emphasizing the need for proactive measures to significantly reduce road accidents in the future.

Brunna de Sousa Pereira Amorim et al. [3] proposed study employs machine learning, analyzing attributes like spatial data, weather, and accident type. The study focused on Brazilian federal highways, using supervised algorithms. A neural network yielded the best results, with 83% accuracy, 84% precision, 83% recall, and 82% F1-score. Future work involves a smartphone app leveraging this model to alert drivers of high-risk highway sections, incorporating real-time external data for enhanced accuracy.
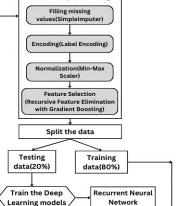
Piotr Gorzelanczyk et al. [10] studied that increasing vehicular traffic poses a significant and escalating threat of traffic accidents, resulting in loss of lives and economic resources. Limited dataset size challenges effective analysis. This study presents a methodology, utilizing multi-criteria optimization, to identify key factors influencing road accidents in Poland. Weather conditions, province, road type, and even weekdays were found to significantly impact accident rates. This versatile algorithm holds the potential for optimizing accident-reduction strategies, with plans for future research in accident forecasting and synergy-based solutions.

2. Proposed methodology

The proposed system architecture aims to ensure scalability, adaptability, and security. The Process Flow Diagram visually illustrates the sequential steps involved in the system's operation. The model consists of essential modules: Road accident data collection, Pre-processing Dataset, Training the deep learning models, classification of road accidents based on severity, and performance evaluation of the models. This provides valuable insights for analysis and decision-making. Fig. 1 shows the proposed methodology, emphasizing the importance of a well-defined approach in addressing road accidents. Incorporating advanced deep learning techniques and rigorous evaluation, the system aims to improve accuracy and efficiency in predicting the severity of road accidents. This contributes to suggesting effective measures for reducing the severity of road accidents.

2.1. Dataset description

The data utilized in this study has been gathered from the Vijayawada traffic police department and encompasses information from the years 2019, 2020, and 2021. It encompasses various regions within Vijayawada, namely Penamanaluru, Ibrahimpatnam, Gandhinagar, Krishnalanka, Patamata, and Mylavaram. The dataset consists of 13 features that provide valuable insights into road accidents, including Time, Location, P Vehicle, P Vehicle Age, Month, Day, S Vehicle, S Vehicle Age, Victim Gender, Road condition, Victim Age, and Alcohol Consumption. With a total of 12,712 rows, this dataset offers a substantial amount of data for analysis and evaluation. Each row represents a unique observation related to road accidents in the specified regions. These features and data points will serve as crucial inputs for the different modules and tasks involved in the study, facilitating comprehensive exploration and analysis of road accidents in the Vijayawada region. By leveraging this dataset, the study aims to uncover patterns, trends, and factors contributing to road accidents, ultimately assisting in developing effective measures to enhance road safety and reduce accidents in the region.

Fig. 1. Process Flow diagram of Road Accident Severity Classification model

2.2. Data pre-processing

Dealing with missing values

Dealing with missing values in a road accident dataset is crucial to maintaining data integrity and completeness for accurate analysis and modeling. A simple imputer algorithm can be used to address this issue. The algorithm identifies columns with missing values, finds the most common category in each column, and fills in the missing values accordingly. By replacing the original missing values with the most common category, the dataset becomes more reliable for further analysis. It is recommended to review the filled-in values to ensure their accuracy and representation of the missing data [2]. Employing a simple imputer algorithm ensures the reliability of the road accident dataset by filling in missing values and enabling accurate analysis and modeling.

Encoding the data

Encoding categorical data is a crucial step in data analysis as it enables numerical representation, algorithm compatibility, relationship identification, and improved model performance. The algorithm for encoding categorical data involves several steps. The algorithm identifies the columns in the dataset that contain categorical values. For each categorical column, a unique numerical label is assigned to each distinct category. A mapping or dictionary is created to keep track of the assigned labels for each category. The categorical values in each column are replaced with their corresponding numerical labels [13].

The analysis can then proceed using the encoded numerical values. The mapping can be retained for future reference or decoding purposes. The process of encoding categorical data facilitates efficient data analysis and enhances the performance of models by transforming categorical values into numerical representations.

Data normalization

Normalization is a crucial step in data analysis as it ensures that all features in a dataset are on a similar scale, preventing the dominance of certain features and enabling fair comparisons. The algorithm for normalization involves several steps. The algorithm identifies the numerical features in the road accident dataset. An instance of the Min-MaxScaler is created [12]. The scaler is fitted to the numerical features, calculating the minimum and maximum values. The numerical features are transformed to scale the values between 0 and 1. The original numerical features are then replaced with the normalized values. This allows for the continuation of the analysis using the normalized data. By normalizing the dataset, the algorithm ensures that all features are treated equally, facilitating unbiased analysis and comparisons.

Feature selection

Feature selection is an essential step in data analysis, and it can be performed using Recursive Feature Elimination (RFE) with Gradient Boosting. This method involves training a gradient-boosting model on the full feature set, ranking the importance of each feature, and repeatedly eliminating the least important features till the desired number of features is obtained. The purpose of this process is to optimize model performance and interpretability. The algorithm takes a normalized dataset as input and outputs a list of selected features. The steps involve loading the dataset, determining the desired number of features, training a gradient boosting model, evaluating feature importance, and repeating the process until the desired number of features is obtained [24]. The selected features are used to build the final model. By applying RFE with Gradient Boosting, the algorithm identifies the most relevant features, enhancing model performance and facilitating better interpretation of the results.

2.3. Classification of road accidents severity

Multi-layer perceptron

The classification of road accident severity is addressed using MLP. The MLP consists of interconnected neurons that process accident-related features to predict severity levels. Training the MLP requires a labeled dataset, divided into training and testing sets. The MLP learns patterns and relationships during training, optimizing its weights and biases. The trained MLP is evaluated using the testing set. Accurate severity classification aids in identifying high-risk situations, enabling targeted interventions for accident prevention [1].

Recurrent neural network

Classification of road accident severity involves utilizing RNNs to process accident data as a sequential time series. RNNs capture temporal relationships and patterns, making predictions based on accident-related features [21]. The parameters of the model are optimized using the training set, and performance is evaluated using the testing set. By leveraging the temporal dynamics and long-term dependencies in the data, RNNs improve the accuracy of severity predictions. This enables timely interventions and targeted safety measures to mitigate the impact of road accidents.

Long-short term memory

The classification of road accident severity can be effectively performed using LSTM networks, a type of RNN architecture. LSTMs capture long-term dependencies in sequential accident data, extract relevant features, and make accurate severity predictions. Proper data preprocessing, architecture design, and hyperparameter tuning are critical for optimal accuracy [16].

Gated recurrent unit

Classification of road accident severity using GRU is an effective approach for modeling temporal dependencies in accident data. GRUs, a variant of RNN, utilize gating mechanisms to selectively remember or forget information from previous time steps [26]. By processing accident data through GRU layers, the model captures sequential patterns and makes accurate predictions. Training involves a labeled dataset, split into training and testing sets. Proper data preprocessing, architecture design and hyperparameter tuning are crucial for optimal classification performance.

2.4. Training the deep learning models

Train deep learning models (MLP, RNN, LSTM, GRU) on labeled road accident data to classify accident severity, leveraging their respective architectures for capturing patterns and temporal dependencies. Algorithm 1 describes the training of deep learning algorithms on the Road accident dataset.

Algorithm 1 ROAD ACCIDENT SEVERITY CLASSIFICATION

Input: Pre-processed Road accident dataset

Output: Classification of road accident severity

1: Split the dataset into training and testing sets using an 80-20 ratio
2: Train an MLP model on the training data
3: Train an RNN model on the training data
4: Train an LSTM model on the training data
5: Train a GRU model on the training data
6: Evaluate the accuracy of each model using the testing data
7: Select the model with the highest accuracy for predicting the severity of road accidents

---

2.5. Performance evaluation

Evaluate the performance of deep learning models (MLP, RNN, LSTM, GRU) for road accident severity classification using metrics like accuracy, precision, recall, and F1-score to assess their effectiveness in predicting accident severity. Algorithm 2 describes the performance evaluation of deep learning models.

Algorithm 2 PERFORMANCE EVALUATION OF Deep Learning Models

Input: Actual labels y_{true}, Predicted labels y_{pred}

Output: Confusion matrix, Accuracy, Precision, Recall, F-score, ROC-AUC

1: $C \leftarrow confusion\_matrix(y_{true}, y_{pred})$

2: $Accuracy \leftarrow \dfrac{True\ Positives + True\ Negatives}{Total\ Samples}$

3: $Precision \leftarrow \dfrac{True\ Positives}{Total\ Samples}$

4: $Recall \leftarrow \dfrac{True\ Positives}{True\ Positives + False\ Positives}$

5: $F - score \leftarrow 2 \times \dfrac{Precision\ \times Recall}{Precision + Recall}$

6: Calculate ROC-AUC Score.

---

3. Results and discussion

3.1. Confusion matrix and ROC-AUC evaluation for deep learning models in predicting road accident severity

This study rigorously assesses four deep learning models: MLP, RNN, LSTM, and GRU, for their predictive capabilities in road accident severity prediction.

Confusion matrix analysis

For each model, confusion matrices are utilized to visually represent classification accuracy. These matrices provide an exhaustive breakdown of metrics such as True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The presentation includes clear titles, labeled axes, color-coded cells, and informative annotations, facilitating a precise understanding of each model's effectiveness in accident severity categorization. This approach enables a comprehensive scrutiny of model performance nuances. Fig. 2, Fig. 3, Fig. 4, and Fig. 5 respectively show the confusion matrices for MLP, RNN, LSTM, and GRU.
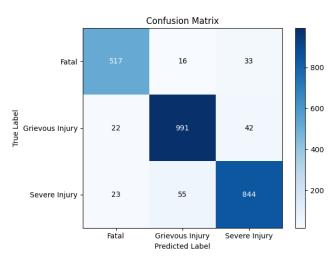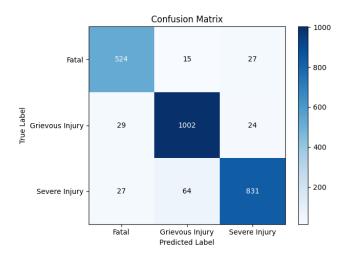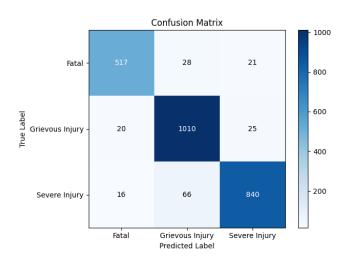

Fig. 2. Confusion Matrix of MLP

Fig. 3. Confusion Matrix of RNN
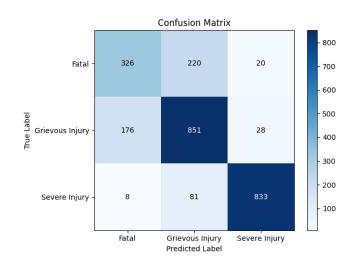

Fig. 4. Confusion Matrix of LSTM

Fig. 5. Confusion Matrix of GRU

ROC-AUC curve assessment

In addition to the confusion matrix analysis, Receiver Operating Characteristic (ROC) curves and corresponding Area Under the Curve (AUC) scores are employed. These metrics offer nuanced insights into the models' discriminatory capabilities. ROC curves plot True Positive Rate (TPR) against False Positive Rate (FPR) at varying classification thresholds, with higher AUC scores indicating superior discriminatory prowess. Fig. 6, 7, 8, and 9 respectively show the ROC-AUC curves for MLP, RNN, LSTM, and GRU.

3.2. Evaluation metrics of MLP, RNN, LSTM, and GRU

This study explores the effectiveness of various deep learning models, including Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU), in predicting road accident severity. The models' performance is evaluated using essential metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The objective is to assess these models' capability to accurately classify the severity of road accidents, providing valuable insights into their suitability for such tasks. Table 1 presents the comprehensive evaluation of these deep learning models.
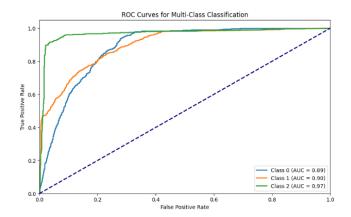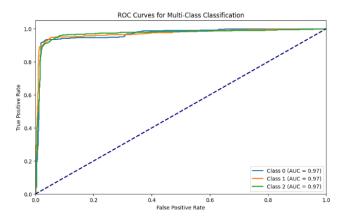

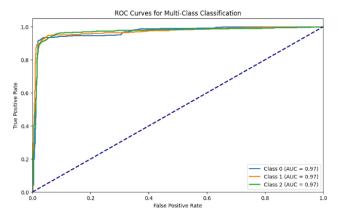*Fig. 6. ROC-AUC of MLP*

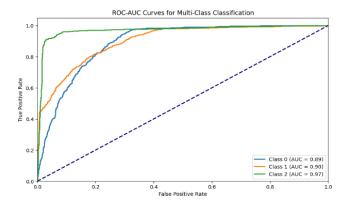*Fig. 8. ROC-AUC of LSTM*

Fig. 7. ROC-AUC of RNN

Fig. 9. ROC-AUC of GRU

Table 1. Evaluation metrics of MLP, RNN, LSTM, and GRU

Metrics	MLP	RNN	LSTM	GRU
Precision (%)	92.38	92.41	93.26	77.44
Accuracy (%)	92.48	92.68	93.07	79.04
Recall (%)	92.27	92.56	92.72	76.20
F1-Score (%)	92.32	92.46	92.96	76.70
ROC-AUC (%)	97.67	97.03	97.66	92.14

Among the evaluated deep learning models, LSTM emerges as the most accurate model, achieving an impressive accuracy of 93.07 percent. It is closely followed by the RNN model, which attains an accuracy of 92.68 percent. The MLP model demonstrates an accuracy of 92.48 percent, while the GRU model achieves 79.04 percent accuracy. Fig. 10 visually depicts the accuracy scores obtained by MLP, RNN, LSTM, and GRU, underscoring LSTM's superior performance in accurately classifying road accident severity. These results affirm LSTM as the most promising and effective model for precisely predicting and classifying road accident severity.

3.3. Accuracies of MLP, RNN, LSTM, and GRU for the classification of road accident severity

This study provides valuable insights into the potential of deep learning models, particularly LSTM, for enhancing the prediction and classification of road accident severity. LSTM's remarkable accuracy underscores its suitability for real-world applications in traffic safety and accident management.
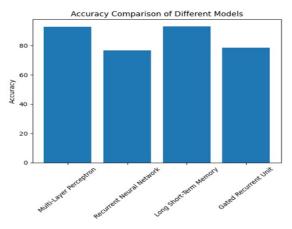

Fig. 10. Accuracies of MLP, RNN, LSTM, and GRU for the Classification of The Severity of Road Accidents

3.4. User interface

The user interface receives various inputs, including weather conditions, details of the primary and secondary vehicles, road conditions, the gender of the victim, the month and day of the incident, alcohol consumption, time, and the ages of the primary and secondary vehicles, as well as the victim's age. It then classifies the severity of the accident into one of three categories: fatal, severe, or grievous. To exemplify the model's performance, we present a specific use case: A scenario where the input parameters correspond to 'hot' weather conditions, a 'three-wheeler' as the primary vehicle, a 'pedestrian' as the secondary vehicle, 'dry' road conditions, a 'female' victim, an incident occurring in the 'month of January' on a 'Saturday' at '8:45 PM,' alcohol consumption, both the primary and secondary vehicles aged '3 years,' and the victim's age noted as '50 years.' The model, when presented with these input values, consistently predicts the accident's severity as 'Fatal' with a high degree of confidence. Consequently, if an actual accident instance mirrors these specific parameter values, our system unequivocally classifies it as 'Fatal.' For enhanced understanding and visualization of this classification process, we incorporate graphical representations into the user interface. Fig. 11 and Fig. 12 visually depict the UI effectively categorizing the road accident instance as 'Fatal.'

Fig. 11. Prediction of road accident instance as Fatal

Fig. 12. Prediction of road accident instance as Fatal

4. Conclusion

The scope of this work is confined to the Vijayawada region, a city with its own unique set of road accident challenges. Focusing on this specific geographical area, the study aims to provide targeted insights and solutions to address the severity of the road accident. Analyzing road accident data and utilizing deep learning algorithms, the system aims to categorize the seriousness of road accidents into three distinct classes: Fatal, Grievous, and Severe injuries. This classification model serves as a crucial tool in understanding the magnitude of accidents and enables stakeholders to take proactive measures in terms of policymaking, infrastructure improvements, and awareness campaigns. To accomplish the objectives of the study, future work entails refining the models used for classification. This involves fine-tuning the algorithms, optimizing their performance, and enhancing their accuracy.

References

[1] Al Bataineh A., Kaur D., Jalali S. M. J.: Multi-layer perceptron training optimization using nature-inspired computing. IEEE Access 10, 2022, 36963–36977.

[2] Alghamdi T.A., Javaid N.: A survey of preprocessing methods used for analysis of big data originated from smart grids. IEEE Access 10, 2022, 29149–29171.

[3] Amorim B. d. S.P., et al.: A Machine Learning Approach for Classifying Road Accident Hotspots. ISPRS International Journal of Geo-Information 12(6), 2023, 227.

[4] Athiappan K., et al.: Identifying Influencing Factors of Road Accidents in Emerging Road Accident Blackspots. Advances in Civil Engineering, 2022.

[5] Cai Q.: Cause analysis of traffic accidents on urban roads based on an improved association rule mining algorithm. IEEE Access 8, 2020, 75607–75615.

[6] Chen M.-M., Chen M.-Ch.: Modeling road accident severity with comparisons of logistic regression, decision tree, and random forest. Information 11(5), 2020, 270.

[7] Comi A., Polimeni A., Balsamo Ch.: Road accident analysis with data mining approach: evidence from Rome. Transportation research procedia 62, 2022, 798–805.

[8] Ferreira-Vanegas C. M., Vélez J. I., García-Llinás G. A.: Analytical methods and determinants of frequency and severity of road accidents: a 20-year systematic literature review. Journal of Advanced Transportation, 2022.

[9] Gatarić D., et al.: Predicting Road Traffic Accidents - Artificial Neural Network Approach. Algorithms 16(5), 2023, 257.

[10] Gorzelanczyk P., Tylicki H.: Methodology for Optimizing Factors Affecting Road Accidents in Poland. Forecasting 5(1), 2023, 336–350.

[11] Gutierrez-Osorio C., González F. A., Pedraza C. A.: Deep Learning Ensemble Model for the Prediction of Traffic Accidents Using Social Media Data. Computers 11(9), 2022, 126.

[12] Islam M. J., et al.: Application of min-max normalization on subject-invariant EMG pattern recognition. IEEE Transactions on Instrumentation and Measurement 71, 2022, 1–12.

[13] Jia B.-B., Zhang M.-L.: Multi-dimensional classification via decomposed label encoding. IEEE Transactions on Knowledge and Data Engineering, 2021.

[14] Kaffash Charandabi N., Gholami A., Abdollahzadeh Bina A.: Road accident risk prediction using generalized regression neural network optimized with self-organizing map. Neural Computing and Applications 34(11), 2022, 8511–8524.

[15] Komol, M.M.R., et al.: Deep RNN Based Prediction of Driver's Intended Movements at Intersection Using Cooperative Awareness Messages. IEEE Transactions on Intelligent Transportation Systems 24(7), 2023, 6902–6921.

[16] Le X.-H., et al.: Application of long short-term memory (LSTM) neural network for flood forecasting. Water 11(7), 2019, 1387.

[17] Mandal V., et al.: Artificial intelligence-enabled traffic monitoring system. Sustainability 12(21), 2020, 9177.

[18] Novikov A., Shevtsova A., Vasilieva V.: Development of an approach to reduce the number of accidents caused by drivers. Transportation research procedia 50, 2020, 491–498.

[19] Östh J., et al.: Driver kinematic and muscle responses in braking events with standard and reversible pre-tensioned restraints: validation data for human models. SAE Technical Paper, 2013, 2013-22-0001.

[20] Rahman M.M., et al.: Towards sustainable road safety in Saudi Arabia: Exploring traffic accident causes associated with driving behavior using a Bayesian belief network. Sustainability 14(10), 2022, 6315.

[21] Rezk N. M., et al.: Recurrent neural networks: An embedded computing perspective. IEEE Access 8, 2020, 57967–57996.

[22] Saravanarajan V.S., et al.: Car crash detection using ensemble deep learning. Multimedia Tools and Applications, 2023, 1–19.

[23] Sobhana M., et al.: A Hybrid Machine Learning Approach for Performing Predictive Analytics on Road Accidents. 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), 2022.

[24] Upadhyay D., et al.: Intrusion detection in SCADA based power grids: Recursive feature elimination model with majority vote ensemble algorithm. IEEE Transactions on Network Science and Engineering 8(3), 2021, 2559–2574.

[25] Yan J., et al.: Relationship between Highway Geometric Characteristics and Accident Risk: A Multilayer Perceptron Model (MLP) Approach. Sustainability 15(3), 2023, 1893.

[26] Yin Y., et al.: SE-GRU: Structure Embedded Gated Recurrent Unit Neural Networks for Temporal Link Prediction. IEEE Transactions on Network Science and Engineering 9(4), 2022, 2495–2509.

[27] Zarei M., Hellinga B., Izadpanah P.: CGAN-EB: A non-parametric empirical Bayes method for crash frequency modeling using conditional generative adversarial networks as safety performance functions. International Journal of Transportation Science and Technology 12(3), 2023, 753–764.

[28] Zheng H., et al.: A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction. IEEE Transactions on Intelligent Transportation Systems 22(11), 2020, 6910–6920.

[29] Road Accidents in Malaysia: Top 10 Causes & Prevention. Kurnia, 21 Sept. 2022 [http://www.kurnia.com/blog/road-accidents-causes].

Prof. Mummaneni Sobhana
e-mail: sobhana@vrsiddhartha.ac.in

Dr. M. Sobhana is currently working as an associate professor in the Department of Computer Science and Engineering, V. R. Siddhartha Engineering College, Vijayawada, India. She received Ph.D. degree in Computer Science and Engineering in 2018 from Krishna University. She has 16 years of teaching experience. Her research interests lie in areas such as Artificial Intelligence, Machine Learning, Data Analytics, Cyber Security, and Software Engineering. She published 35 papers in National and International journals and published 7 patents.

http://orcid.org/0000-0001-5938-5740

Eng. Nihitha Vemulapalli
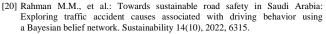e-mail: nihithavemulapalli@gmail.com

Nihitha Vemulapalli is a final-year B. Tech student specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College, Vijayawada, India. She is passionate about Deep Learning and has been recognized as a Google Women Engineers Scholar. Additionally, she holds the position of IEEE Chair at the WIE (Women in Engineering) student chapter.

http://orcid.org/0009-0007-8626-0012

Eng. Gnana Siva Sai Venkatesh Mendu
e-mail: sivasaivenkatesh.m@gmail.com

Gnana Siva Sai Venkatesh Mendu is a final-year B. Tech. student, specializing in Computer Science and Engineering at V. R. Siddhartha Engineering College, Vijayawada, India. He is passionate about Artificial Intelligence and Machine Learning. He achieved Bronze in the Cisco NetAcad Riders 2023 competition. He holds the position of IEEE Chair of the GRSS Student Chapter.

http://orcid.org/0009-0004-6406-000X

Eng. Naga Deepika Ginjupalli
e-mail: nagadeepikaginjupalli@gmail.com

Ginjupalli Naga Deepika is currently pursuing her final year of graduation in computer science and engineering from V. R. Siddhartha Engineering College, Vijayawada, India. She is interested in doing projects in the machine learning domain. She holds the position of webmaster of the Computer Society student chapter.

http://orcid.org/0009-0005-0436-0008

Eng. Pragathi Dodda
e-mail: pragathidodda@gmail.com

Pragathi Dodda is currently pursuing B. Tech in computer science and engineering at V. R. Siddhartha Engineering College, Vijayawada, India. Her interests are in Artificial Intelligence and machine learning applied to detecting individuals with Lupus symptoms.

http://orcid.org/0009-0003-4879-6298

Prof. Rayanoothala Bala Venkata Subramanyam
e-mail: rbvs66@nitw.ac.in

R.B.V. Subramanyam received his M. Tech. and Ph.D. from the Indian Institute of Technology Kharagpur, India. Currently, He is working at the National Institute of Technology Warangal, India. He has published many journal and conference papers in the areas of Data Mining. Some of his research interests include Data Mining, Distributed Data Mining, Fuzzy Data Mining, and Big Data Analytics. He is one of the reviewers for IEEE Transactions on Fuzzy Systems and the Journal of Information and Knowledge Management. He is a member in IEEE and The Institution of Engineers (India).

http://orcid.org/0009-0005-8907-1984