Graphical-Based Neural Network Model for Determining the Level of Roadway Breakdown Injuries.

[1] Bindu S V, [2] Dr.Geetha M

[1]Student, Department of MCA, BIET, Davanagere, India

[2]Associate Professor, Department of MCA, BIET, Davanagere, India

ABSTRACT: Predicting injury severity in road accidents is vital for improving transportation safety and emergency response. Traditional methods using factors like weather, vehicle type, and road conditions often fail to capture complex interactions. This study presents a Graph Neural Network (GNN) model that treats crash data as a graph, revealing intricate relationships among influencing factors. The model was tested against XGBoost, Random Forest, and Artificial Neural Networks using UK transport data. Uniform preprocessing ensured fairness across models. Evaluation used metrics like accuracy, precision, recall, F1-score, and MCC. GNN achieved the highest accuracy at 85.55%, outperforming XGBoost (83.36%), RF (83.18%), and ANN (83.27%). Its advantage lies in modeling relational patterns more effectively. The results demonstrate GNN's potential in real-world traffic systems for more accurate injury predictions.

Keywords: Graph Neural Networks (GNNs), Road Traffic Accident Severity, Machine Learning Models, Crash Data Analysis, Emergency Response Optimization.

I. INTRODUCTION

Agricultural Road traffic accidents continue to be a global public safety issue, leading to significant human casualties, severe injuries, and economic losses. Accurate prediction of crash injury severity is essential for improving emergency response strategies, road safety policies, and healthcare resource allocation. Traditional machine learning (ML) models have been widely used for this purpose, utilizing structured tabular data with features like weather conditions, road layouts, vehicle types, and traffic signals. However, these approaches often fall short when it comes to identifying complex relationships between different crash instances. To address these limitations, the current research explores the capabilities of Graph Neural Networks (GNNs)—a class of deep learning models that work on graph-structured data. By constructing a graph from accident records using similarity-based metrics such as *k-nearest neighbors (kNN)*, the proposed model uncovers hidden patterns and interdependencies that traditional methods might miss. Unlike conventional algorithms that treat records independently, GNNs allow for joint learning over interconnected data points. In this study, we propose a GraphSAGE-based GNN framework for predicting the severity level of road crash injuries using real-world data from the United Kingdom. The performance of this model is compared against widely accepted ensemble classifiers such as Random Forest, Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANNs). Results indicate that the graph-based approach provides superior predictive performance across multiple evaluation metrics, suggesting its potential for practical deployment in smart traffic management and emergency response systems. In addition to these limitations, the dynamic nature of real-world traffic scenarios necessitates models that can generalize beyond static feature sets. Traditional approaches often fail to account for the temporal and spatial dependencies that exist between crash events occurring under similar environmental or infrastructural conditions. Furthermore, as urban road networks become increasingly complex, there is a growing need to understand not only individual crash characteristics but also how these events are interconnected within the broader traffic ecosystem. By adopting a graph-based perspective, this research leverages the power of relational learning, enabling the model to infer patterns not just from isolated data points, but from their interactions within a structured graph. This marks a significant shift from purely statistical inference to contextual, structure-aware prediction, providing deeper insight into the causes and potential outcomes of roadway incidents.

Moreover, the integration of advanced graph learning techniques like GraphSAGE brings scalability and adaptability to the model, allowing it to handle unseen crash records effectively through inductive learning. This is especially important for real-time applications, where new accident data continuously streams in and immediate predictions are required. The ability to learn from both individual crash features and their surrounding context gives Graph Neural Networks a strategic edge over conventional methods. This positions the proposed system as a promising tool for intelligent transportation systems (ITS), enabling smarter decision-making in road safety, accident mitigation planning, and emergency resource allocation.

II. RELATED WORK

A Study on Road Accident Prediction and Contributing Factors Using Explainable Machine Learning Models.

Authors: S. Ahmed, M. A. Hossain, S. K. Ray, M. M. I. Bhuiyan, and S. R. Sabuj

This study explores the application of explainable machine learning models to predict road accidents and identify contributing factors. By leveraging interpretability techniques such as SHAP and LIME, the authors highlight key variables influencing accident likelihood. The analysis combines diverse traffic, environmental, and driver behavior data, aiming to provide stakeholders with actionable insights. The results demonstrate the effectiveness of models like XGBoost and Random Forest in producing accurate predictions while also offering transparency. This dual focus on prediction and interpretability makes the approach valuable for both policymakers and road safety engineers.[1]

Identifying the Factors Contributing to the Severity of Truck-Involved Crashes in Shanghai River-Crossing Tunnel

Authors: S. Chen, S. Zhang, Y. Xing, and J. Lu

The paper examines the severity of truck-related crashes in a critical urban infrastructure—the Shanghai river-crossing tunnel. Using logistic regression and statistical analysis, the study pinpoints variables such as tunnel lighting, time of day, vehicle load, and driver demographics as key influencers of crash outcomes. The findings provide essential implications for targeted traffic management and tunnel safety interventions, suggesting the need for adaptive control systems and stricter monitoring protocols.[2]

Several Issues About Urbanization and Urban Safety

Authors: J. Li, Q. Liu, and Y. Sang

This paper investigates the interplay between rapid urbanization and urban safety, particularly in the context of transportation systems. It discusses infrastructural stress, congestion, and inadequate urban planning as core issues exacerbating safety risks. Through a multidisciplinary perspective, the authors advocate for integrated planning that aligns urban development with safety-enhancing strategies. The paper is an important contribution to the discussion on sustainable urban growth and risk mitigation.[3]

Accident Prediction Modelling and Crash Scene Investigation

Authors: S. Naznin, P. H. Sumayya, L. S. Panackel, S. Zaviar, and S. Babu

This study combines predictive modeling with on-site crash investigation to enhance the understanding of accident causality. Utilizing historical crash data, machine learning models are developed to forecast accident-prone zones, while field investigations validate the model's findings. The integration of quantitative and qualitative analysis helps bridge the gap between model predictions and real-world road conditions, providing a comprehensive tool for traffic authorities and safety analysts.[4]

A Hybrid Machine Learning Model for Predicting Real-Time Secondary Crash Likelihood

Authors: P. Li and M. Abdel-Aty

Addressing the challenge of secondary crashes, this paper presents a hybrid model that combines real-time traffic data with machine learning techniques. By fusing Support Vector Machines and decision trees, the model predicts crash likelihoods under evolving traffic conditions. The real-time capability is especially beneficial for freeway management systems aiming to prevent cascading accidents. The study confirms the model's accuracy and highlights its potential for integration into intelligent transportation systems.[5]

A Novel Approach for Real-Time Crash Prediction at Signalized Intersections

Authors: L. Zheng and T. Sayed

This research proposes a real-time crash prediction method tailored to signalized intersections using advanced video analytics and machine learning. By extracting traffic flow and vehicle movement patterns, the model anticipates potential conflicts and crash risks. The approach is validated using real-world data, achieving high accuracy and rapid response time. It demonstrates promise for urban traffic control centers seeking to enhance intersection safety with predictive capabilities.[6]

A New Econometric Approach for Modeling Several Count Variables: A Case Study of Crash Frequency Analysis by Crash Type and Severity

Authors: T. Bhowmik, S. Yasmin, and N. Eluru Introducing a multivariate econometric modeling technique, this paper tackles the challenge of jointly analyzing multiple crash types and severities. Using a dataset of roadway crashes, the model captures correlations among different count variables, improving prediction accuracy and interpretability. The method supports more nuanced safety planning by addressing the heterogeneity of crash characteristics, making it a valuable tool for policymakers and researchers in transportation safety.[7]

Crash Frequency and Severity Modeling Using Clustered Data from Washington State

Authors: J. Ma and K. Kockelman

This study utilizes clustered traffic data to model both crash frequency and severity across Washington State. By applying Poisson and ordered probit models, the paper identifies road characteristics, traffic volume, and environmental factors as influential. The clustered approach ensures that spatial and temporal dependencies are accounted for, increasing the robustness of the results. The findings contribute to a deeper understanding of roadway risk patterns, aiding in targeted safety improvements.[8]

III. METHODOLOGY

The methodology outlines the step-by-step process used to build and evaluate the injury severity prediction system. It includes data collection, preprocessing, algorithm design, interface development, and system integration using graph-based learning.

Data Collection: Accurate and relevant data is crucial for predictive modeling. In this study, crash data is sourced from the UK Department for Transport's open repository. This dataset includes detailed records of road accidents, vehicle types, weather conditions, road layouts, and injury severity. These records form the foundational input for building the predictive model.

Preprocessing: Raw data is often inconsistent, containing missing values, noise, or irrelevant attributes. Preprocessing involves cleaning the dataset, handling missing values, encoding categorical features, and normalizing numerical values. This step ensures uniformity across the data and prepares it for graph construction. Clean, structured data is essential for model accuracy.

Information Retrieval: Feature engineering and selection are done to extract meaningful attributes that contribute to prediction. Important features like time of accident, road surface conditions, lighting, and vehicle maneuver are prioritized. These attributes are retrieved and organized to maintain relevance and reduce dimensionality. This refined information improves learning efficiency.

User Interface Design: A simple and intuitive user interface is developed to ensure user-friendly interaction. Users can input crash-related information and receive real-time predictions. The interface displays outputs such as injury severity level and model confidence score. It also allows users to review past predictions through a clean dashboard.

Integration and Testing: All components data preprocessing, model inference, and UI are integrated into a unified platform. The system is thoroughly tested using both functional and nonfunctional testing methods. Testing ensures prediction accuracy, robustness, and responsiveness of the application. The integration phase ensures smooth operation from user input to result display.

3.1 Dataset Used

The dataset utilized in this research is obtained from the UK Department for Transport's Road Safety Data portal. It comprises thousands of crash records, each containing detailed fields such as accident severity (slight, serious, fatal), number of vehicles involved, road types, weather and lighting conditions, and driver characteristics. This government-backed dataset is comprehensive, reliable, and suitable for training advanced machine learning models due to its diversity and volume. It provides a realistic foundation for building predictive systems aimed at assessing injury severity across various traffic scenarios.

3.2 Data Preprocessing

The preprocessing pipeline is essential to transform raw crash data into a format suitable for model consumption. First, missing values are handled through imputation techniques or by removing incomplete records based on severity. Categorical variables (e.g., vehicle type, lighting) are encoded using one-hot or label encoding, while numerical features are standardized or normalized to ensure uniform scale. Outlier detection is applied to eliminate anomalies that may skew predictions. Finally, the clean dataset is used to construct a k-nearest neighbors (kNN) graph, where each node represents a crash and edges represent similarity, setting the foundation for Graph Neural Network learning.

3.3 Algorithm Used

The core algorithm used in this study is the GraphSAGE (Graph Sample and Aggregate), an inductive Graph Neural Network architecture. GraphSAGE works by sampling and aggregating features from a node's neighbors to generate embeddings that preserve both local and global graph structures. This enables the model to generalize well to unseen crash events. GraphSAGE is compared with traditional algorithms such as Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANN) to highlight its superior performance. Each algorithm is fine-tuned using GridSearchCV for hyperparameter optimization and fairness in evaluation.

3.4 Techniques

Several advanced techniques are integrated into this research to enhance prediction performance. Graph construction is done using kNN and Spearman correlation, ensuring meaningful connections between crash records. Feature engineering focuses on extracting high-impact variables influencing severity. Hyperparameter tuning is executed through GridSearchCV, optimizing learning rate, depth, and layer size for each algorithm. Furthermore, performance metrics such as accuracy, precision, recall, F1-score, and Matthews Correlation Coefficient (MCC) are used for evaluation. The use of embedding techniques like attribute-weighted isometric embedding (AWIE) further improves the representation of categorical features in graph-based learning.

System Architecture

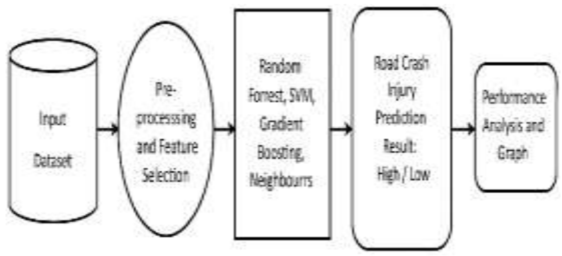

Fig 3.4.1: System Architecture

IV. MODULE DESCRIPTION

The proposed system is structured into two primary modules: the Remote User module and the Service Provider module. Each of these plays a distinct role in ensuring the smooth functionality and usability of the Graph Neural Network-based prediction platform. The modular design of the system allows for role-specific functionalities while maintaining a cohesive architecture that supports both predictive intelligence and administrative control.

Remote User module is designed for individuals who wish to utilize the system to predict the severity of injuries resulting from road crashes. These users are required to register on the platform, following which they can log in and access their personalized profile. The user profile serves as a central hub for managing personal information and reviewing past prediction activities. It ensures that users have a secure and tailored experience, with access to their usage history and system interactions. Once logged in, users are directed to the prediction interface, which is the core feature of this module. The prediction page accepts various crash-related parameters, such as vehicle type, road and environmental conditions, traffic control features, and possibly driver-related inputs. This data, once submitted, is internally transformed into a graph structure using a similarity-based approach like the k-nearest neighbors method. This graph captures the interdependencies among crash records based on their similarity in features.

The GraphSAGE Graph Neural Network model then processes the constructed graph. Unlike traditional models that evaluate each record independently, the GNN model learns to aggregate information from neighboring nodes, thereby understanding hidden patterns and relationships within the crash data. This enables the system to provide a more accurate prediction of the injury severity level—classified typically as minor, moderate, or severe—along with a model confidence score. The interaction remains seamless for the user, who receives the prediction in a comprehensible format. This advanced integration of GNN in the user-facing interface enhances the predictive power while ensuring a user-friendly experience. **Service Provider module** is focused on administrative functionalities and overall system oversight. This module is restricted to authenticated administrative personnel who are responsible for managing the entire ecosystem. One of the primary responsibilities under this module is managing the remote users. The service provider can access all registered user profiles, monitor their activities, and perform actions such as updating credentials, activating or deactivating accounts, and ensuring compliance with system policies. In addition to user management, the service provider can access the entire repository of prediction results. This includes the inputs provided by users, the corresponding model predictions, the timestamps of prediction requests, and additional metadata that helps in tracking model performance and user engagement.

To support analytical capabilities, the system includes a built-in data visualization component. This feature presents the prediction trends in graphical formats such as pie charts, bar diagrams, and line graphs. These visualizations help the service provider in understanding the distribution of injury severities, evaluating the effectiveness of the model over time, and making informed decisions regarding model updates or data acquisition strategies. The division into two modules reflects a well-balanced system that supports both end-user interaction and administrative governance. By combining a graph-based learning approach with a structured user interface, the system ensures high accuracy in injury severity prediction while remaining accessible and manageable. The use of GraphSAGE in particular enhances the system's ability to generalize across unseen data and extract valuable patterns from relational crash data, which traditional machine learning models often fail to capture. This modular design contributes significantly to the system's scalability, robustness, and potential for real-world application intrafficmanagement and accident response systems.

V. RESULT

The results of this study demonstrate that the proposed GraphSAGE-based Graph Neural Network significantly outperforms traditional machine learning models such as Random Forest, XGBoost, and Artificial Neural Networks in predicting roadway injury severity. Using real-world UK crash data, the GNN model achieved an accuracy of 85.55%, compared to 83.36%, 83.18%, and 83.27% for XGBoost, RF, and ANN, respectively. It also recorded higher precision, recall, and F1-scores, indicating better overall performance.

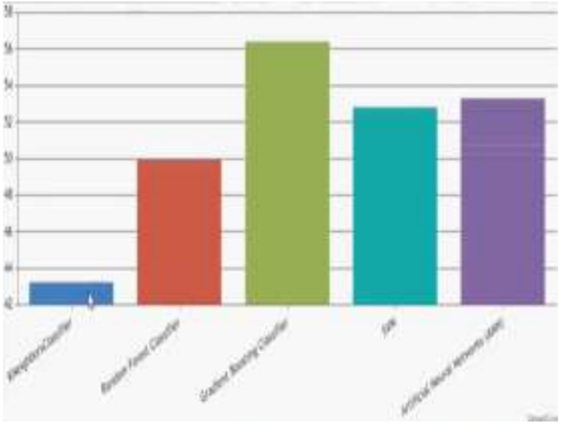

Fig 5.1: Graph Neural Network

The model's ability to aggregate information from similar crash records allowed it to capture complex relationships, making it particularly effective in identifying severe cases. Despite the complexity of graph-based computation, the GraphSAGE framework enabled efficient processing, supporting its potential for real-time predictive systems.

VI. CONCLUSION

In conclusion, this research presents a graph-based deep learning framework capable of accurately predicting the severity of injuries in road accidents. By utilizing the GraphSAGE model, the system effectively captures the relational and contextual patterns among crash records, offering a clear improvement over conventional ML approaches. The comparative analysis confirmed the GNN model's superiority in both accuracy and generalization. The system's modular design—supporting both end-users and administrators—

ensures usability and scalability. This approach offers promising applications in traffic safety management and emergency response planning, and it sets the groundwork for future enhancements involving real-time data, temporal features, and larger datasets.

VII. REFERENCES

[1] S. Ahmed, M. A. Hossain, S. K. Ray, M. M. I. Bhuiyan, and S. R. Sabuj, "A study on road accident prediction and contributing factors using explainable machine learning models: Analysis and performance," *Transp. Res. Interdiscipl. Perspect.*, vol. 19, May 2023, Art. no. 100814.

[2] S. Chen, S. Zhang, Y. Xing, and J. Lu, "Identifying the factors contributing to the severity of truck-involved crashes in Shanghai river-crossing tunnel," *Int. J. Environ. Res. Public Health*, vol. 17, no. 9, p. 3155, May 2020.

[3] J. Li, Q. Liu, and Y. Sang, "Several issues about urbanization and urban safety," *Proc. Eng.*, vol. 43, pp. 615–621, Jan. 2012.

[4] S. Naznin, P. H. Sumayya, L. S. Panackel, S. Zaviar, and S. Babu, "Accident prediction modelling and crash scene investigation," in *Proc. Int. Conf. Struct. Eng. Construct. Manage.*, 2022, pp. 1121–1138.

[5] P. Li and M. Abdel-Aty, "A hybrid machine learning model for predicting real-time secondary crash likelihood," *Accident Anal. Prevention*, vol. 165, Feb. 2022, Art. no. 106504.

[6] L. Zheng and T. Sayed, "A novel approach for real time crash prediction at signalized intersections," *Transp. Res. C, Emerg. Technol.*, vol. 117, Aug. 2020, Art. no. 102683.

[7] T. Bhowmik, S. Yasmin, and N. Eluru, "A new econometric approach for modeling several count variables: A case study of crash frequency analysis by crash type and severity," *Transp. Res. B, Methodol.*, vol. 153, pp. 172–203, Nov. 2021.

[8] J. Ma and K. Kockelman, "Crash frequency and severity modeling using clustered data from Washington state," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Jun. 2006, pp. 1621–1626.

******