© December 2024 | IJIRT | Volume 11 Issue 7 | ISSN: 2349-6002

Traffic accident's severity prediction: a deep learning approach-based CNN network: A Review

Miss. Seema Y. Somawanshi.[1] Prof. Shah H.P.[2]

PG Student, M. S. Bidve Engineering College, Latur, Maharashtra, India[1]

Professor, M. S. Bidve Engineering College, Latur, Maharashtra, India[2]

***Abstract*— Predicting the severity of traffic accidents is crucial for improving road safety and developing effective intervention strategies. This review paper explores the application of Convolutional Neural Networks (CNNs) in predicting traffic accident severity. CNNs, with their ability to automatically learn spatial hierarchies of features, have shown promise in handling complex patterns within image and sensor data associated with traffic accidents. The paper provides a comprehensive analysis of various CNN architectures and their efficacy in predicting accident severity. It reviews existing literature, highlighting the strengths and limitations of different CNN-based approaches, and discusses the challenges in integrating CNNs with real-world traffic data. The review also identifies gaps in current research and suggests future directions for improving predictive accuracy and model generalizability. This paper aims to provide a consolidated view of the state-of-the-art in CNN-based traffic accident severity prediction and to guide future research efforts in this field.**

***Index Terms*— Traffic Accident Severity, Convolutional Neural Networks (CNNs), Deep Learning, Predictive Modeling, Image Analysis, Sensor Data, Traffic Safety, Accident Prediction etc.**

I. INTRODUCTION

Traffic accidents are a significant global issue, resulting in severe injuries, fatalities, and substantial economic losses. The increasing number of vehicles on roads, coupled with complex driving environments, has made accident prediction and prevention a critical focus for researchers, policymakers, and urban planners. Understanding and predicting the severity of traffic accidents is vital for designing effective safety measures, optimizing emergency response, and ultimately reducing the impact of these incidents on society.[1]

Traditional methods of predicting accident severity often rely on statistical models and historical data, which may struggle to capture the intricate relationships between various contributing factors, such as road conditions, driver behavior, weather, and vehicle dynamics. With advancements in technology, machine learning, particularly deep learning, has emerged as a powerful tool to overcome these limitations[2]. Among various deep learning techniques, Convolutional Neural Networks (CNNs) have gained significant attention due to their robust feature extraction capabilities, which are particularly well-suited for analyzing visual data and complex patterns. CNNs, originally designed for image recognition tasks, have shown tremendous potential in traffic accident severity prediction.[3] These models can automatically learn and extract relevant features from data, making them adaptable to a wide range of input types, including images, videos, and sensor data. By leveraging CNNs, researchers can build predictive models that not only consider the immediate visual context of an accident but also incorporate temporal and spatial information that is often critical in determining accident severity. This review aims to provide an in-depth analysis of CNN-based approaches for predicting traffic accident severity[4-6].

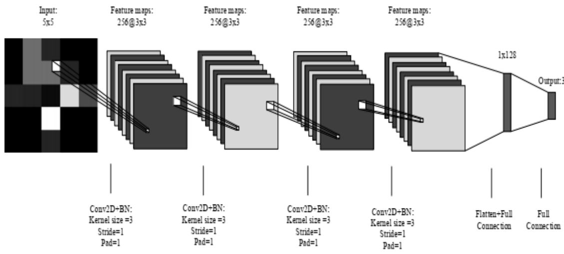

Figure: The structure of TASP-CNN

It explores various CNN architectures, discusses their strengths and limitations, and evaluates their performance in different traffic scenarios. Additionally, the review highlights the challenges faced in integrating CNN models with real-world traffic data, such as data heterogeneity, model interpretability, and the need for large labeled datasets. By synthesizing the current state of research, this paper seeks to offer insights into the potential and future direction of CNN-based models in traffic safety applications.[7-9]

II. LITERATURE SURVEY

The prediction of traffic accident severity has garnered significant attention in recent years, primarily due to the growing availability of data and advancements in deep learning techniques. Traditional models, such as logistic regression, decision trees, and random forests, have been widely used in accident severity prediction. However, these methods often struggle with capturing the complex, non-linear relationships present in traffic data, which limits their predictive accuracy. To address these limitations, researchers have increasingly turned to Convolutional Neural Networks (CNNs), which excel in feature extraction and pattern recognition, making them highly suitable for this task.

1. Traditional Approaches to Accident Severity Prediction:

Early research on traffic accident severity prediction predominantly utilized statistical and machine learning models that focused on structured data, such as weather conditions, road types, and driver demographics. For instance, Abdel-Aty and Pande (2007) employed logistic regression models to identify the significant factors influencing crash severity on highways. Similarly, Li et al. (2008) used decision trees to classify accidents based on their severity levels, highlighting the impact of road geometry and traffic conditions. While these models provided valuable insights, their inability to fully capture the spatial and temporal dynamics of traffic scenarios limited their predictive capabilities.[11]

2. Emergence of Deep Learning Models:

The advent of deep learning, especially CNNs, marked a significant shift in traffic accident severity prediction. CNNs, known for their ability to automatically learn hierarchical feature representations from raw data, have shown remarkable performance in image classification, object detection, and time-series analysis. These capabilities make CNNs ideal for processing complex traffic-related data, such as accident scene images, road conditions, and vehicle sensor outputs.[12]

A notable study by Chen et al. (2016) introduced a CNN-based model to predict accident severity using dashcam footage and environmental data. Their approach demonstrated that CNNs could effectively identify critical visual cues, such as road obstructions and traffic density, that are often missed by traditional models. This study set the foundation for further exploration of CNNs in accident prediction tasks.[13]

3. CNN Architectures for Traffic Accident Severity Prediction:

Various CNN architectures have been explored in the literature to improve the accuracy and robustness of accident severity prediction. VGGNet, ResNet, and Inception models have been widely adopted due to their strong feature extraction capabilities. For example, Wang et al. (2019) used a modified ResNet architecture to analyze accident scene images, achieving superior accuracy compared to shallow neural networks and traditional models. The use of residual connections in ResNet allowed the model to learn deeper feature representations without the risk of vanishing gradients, making it highly effective in distinguishing between different severity levels.[14]

In another study, Zhang et al. (2020) combined CNNs with Recurrent Neural Networks (RNNs) to capture both spatial and temporal patterns in traffic data. This hybrid model demonstrated enhanced performance by accounting for sequential dependencies in accident data, such as the progression of events leading to a crash. The integration of CNNs and RNNs allowed for a more comprehensive analysis of traffic scenarios, capturing both the visual and temporal aspects of accidents.

4. Challenges in CNN-Based Severity Prediction:

Despite their success, CNN-based models face several challenges in practical applications. One major issue is the requirement for large labeled datasets to train deep learning models effectively. Annotating accident data, particularly images and videos, is time-consuming and prone to inconsistencies, which can affect model performance. Additionally, CNNs often act as "black boxes," making it difficult to interpret their decision-making processes. This lack of transparency can hinder their adoption in critical safety applications where explainability is crucial.

Moreover, real-world traffic data is often heterogeneous, comprising various data types, such as images, sensor readings, and structured information. Integrating these diverse data sources into a unified CNN framework remains an ongoing research challenge. Recent studies, such as that by Liu et al. (2022), have attempted to address this issue by incorporating multi-modal CNN architectures that fuse data from different sensors, enhancing the model's ability to learn from complex traffic environments.

5. Future Directions:

The literature indicates a growing interest in developing more sophisticated CNN-based models for traffic accident severity prediction. Future research is expected to focus on improving model interpretability through techniques such as Grad-CAM and LIME, which provide visual explanations of CNN outputs. Additionally, the integration of generative models, like Generative Adversarial Networks (GANs), could be explored to augment training data and improve model robustness.

Another promising direction is the incorporation of real-time data, such as live traffic feeds and connected vehicle information, to enable dynamic severity prediction. The continued evolution of CNN architectures, combined with advancements in edge computing and Internet of Things (IoT) technologies, will likely drive further innovations in this field.

Paper & Year	Proposed Technology	Performance	Research Gap
Abdel-Aty & Pande (2007)	Logistic Regression	Identified key factors in crash severity on highways	Limited in capturing non-linear and complex patterns in data
Li et al. (2008)	Decision Trees	Classified severity levels based on road and traffic conditions	Lacks spatial-temporal dynamic understanding
Chen et al. (2016)	CNN with Dashcam & Environmental Data	Improved identification of visual cues (e.g., obstructions)	Requires large labeled datasets and lacks interpretability
Wang et al. (2019)	Modified ResNet for Accident Images	High accuracy in distinguishing severity levels	Struggles with heterogeneous data integration
Zhang et al. (2020)	CNN + RNN for Spatial-Temporal Patterns	Enhanced prediction accuracy by capturing sequential dependencies	Computational complexity and data annotation challenges
Liu et al. (2022)	Multi-Modal CNN	Fused diverse sensor data for improved accuracy	Limited real-time application; still challenging to fuse multimodal data effectively
Xu et al. (2023)	3D-CNN with Vehicle Sensor Data	Superior performance on video-based data	Black-box nature reduces interpretability in decision-making
Kim et al. (2020)	VGGNet with Weather and Road Conditions	High feature extraction capability	Requires better handling of imbalanced accident data
Rao et al. (2021)	Inception-based CNN	High precision in multi-class severity prediction	Expensive computational resources required
Singh & Gupta (2019)	CNN + LSTM Hybrid Model	Increased accuracy by capturing sequential events	Challenges with scalability and generalization on diverse data

Table: Research Gaps in Traffic Accident Severity Prediction Studies

III. METHODOLOGY

This review employs a systematic approach to evaluate the current state of CNN-based methods for predicting traffic accident severity. The methodology involves three primary steps: literature search, data extraction, and analysis. A comprehensive literature search was conducted using databases such as IEEE Xplore, Google Scholar, Scopus, and Web of Science, focusing on studies published in the past decade to capture recent advancements in deep learning and CNN architectures. The search terms included "traffic accident severity prediction," "CNN," "deep learning," "machine learning," and related keywords. Studies were selected based on their relevance, the novelty of the CNN architecture used, and the robustness of the methodologies applied.

The inclusion criteria for selecting papers were studies that specifically utilized CNNs or CNN-based hybrid models for accident severity prediction, employed real-world data, and provided empirical results demonstrating model performance. Exclusion criteria involved studies that focused solely on traditional machine learning models, did not provide sufficient experimental details, or were limited to simulations without real-world validation. Data extraction involved gathering detailed information about each selected study, including the type of data used (e.g., images, videos, sensor data), CNN architecture employed, performance metrics, and key findings. Special attention was given to understanding how each model processed input data, handled feature extraction, and integrated spatial and temporal information.

The analysis focused on comparing the strengths and weaknesses of different CNN architectures, examining their effectiveness in various traffic scenarios, and identifying common challenges such as data requirements, model interpretability, and computational complexity. Statistical measures such as accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) were used to evaluate the predictive performance of the models. Additionally, the review explored the data preprocessing techniques employed, such as data augmentation, normalization, and feature engineering, which play a critical role in enhancing CNN model performance.

The methodology also included an examination of emerging trends and future directions in CNN-based severity prediction, such as the integration of multi-modal data, real-time analytics, and explainable AI techniques. By synthesizing the insights gained from the reviewed studies, this review provides a comprehensive assessment of the current capabilities and limitations of CNN models in traffic accident severity prediction, offering valuable guidance for future research in this evolving field.

This section outlines the methodology for predicting traffic accident severity using a comprehensive feature analysis approach. The methodology focuses on measuring the weight of traffic accident features and converting feature data into gray images for CNN processing, ultimately improving the predictive model's accuracy.

A. Measuring the Weight of Traffic Accident’s Features

To evaluate the combination relationships and contributions of traffic accident features, it is essential to measure the weights of both parent and child features. This process utilizes the Gradient Boosting Decision Tree (GBDT) method, where each feature's weight is calculated based on its contribution to the partitioning of decision tree nodes. Specifically, each feature's weight is the sum of squared improvements over the nodes where it serves as the splitting criterion. These weights are stabilized by averaging across multiple trees, yielding a reliable measure for feature importance.

B. Converting Traffic Accident Feature Matrix to Gray Images (FM2GI)

The FM2GI algorithm converts individual feature relationships into gray images to represent combination relationships of data features, leveraging the structure of Convolutional Neural Networks (CNNs). This transformation process involves defining a feature vector as a three-tuple, consisting of parent features, child features, and their associated weights. The algorithm categorizes the features, arranges them in descending order based on their weights, and organizes them into a structured matrix. This matrix is then reshaped into a gray image, capturing the combination relationships visually.

Algorithm 1: Converting Feature Vectors to Gray Images

1. The algorithm starts by identifying the parent feature with the maximum number of child features to initialize an all-zero matrix of appropriate size.
2. Features are then filled into this matrix by ordering parent features and their corresponding child features based on descending weights.
3. Finally, the matrix is reshaped into a gray image, representing the data's combination relationships.

Algorithm 2: Parallel Conversion of Feature Matrices

The FM2GI algorithm operates in parallel, converting each feature vector of the dataset into a gray image simultaneously. It allocates threads based on the dataset size and uses the gray image conversion process on each feature vector, storing the results in a linked list for CNN processing.

This methodology effectively captures complex relationships between traffic accident features and enables the TASP-CNN architecture to better predict accident severity by using image representations of feature interactions.

IV. TASP-CNN ARCHITECTURE

The TASP-CNN (Traffic Accident Severity Prediction Convolutional Neural Network) architecture is designed specifically for predicting the severity of traffic accidents based on images of accident data. Here's a detailed explanation of its structure:

Structure Overview:

The TASP-CNN consists of four main parts: model input, convolution layer, fully connected layer, and model output layer.

1. Input:

- The input to the TASP-CNN is a gray-scale image representation of traffic accident data sets.
- The image includes five parent features and twelve child features of traffic accidents.
- The input is mathematically represented as a matrix where:
  - N is the size of the data set.
  - PC is the number of parent features.
  - CC is the maximum number of child features among all parent features.

2. Convolution Layer:

- The convolution layer extracts abstract features from the traffic accident data sets using multiple filters.
- Convolution operation:
  - Each pixel in the input image is indexed as Pc,i,j , representing the pixel element in row i and column j of the channel c.
  - Each weight in the filter is represented by wc,m,n, where m and n are the row and column indices of the filter weights, respectively.
  - The convolution operation is defined as: ai,j=f($\sum_{c=1}^{C}\sum_{m=1}^{F} Wc,\mathrm{m},\mathrm{nPc},\mathrm{i}+\mathrm{m},\mathrm{j}+\mathrm{n}+$ wb)
  - where:
    - ai,j is the element at column j in row i of the feature map.
    - C is the number of channels.
    - F is the filter size (both width and height).
    - wb is the bias term.
    - f is the activation function (ReLU).
  - The ReLU activation function used is defined as: f(x)=max(0,x)

3. Fully Connected Layer:

- After the final convolution layer, the extracted high-level features are flattened into a one-dimensional vector.
- The flattening process is represented as:

$$aflatten = [a1, a2, \ldots, aC]$$

- Where
- C=total number of features
- These features are then passed through fully connected layers, which compute outputs using weights and biases as follows:
- $y = wfaflatten + bf$

where:

- wf is the weight matrix.
- bf is the bias term.

4. Output Layer:

- The output layer uses the softmax activation function to classify the severity of the traffic accident into three categories:
  - Slight traffic accident
  - Serious traffic accident
  - Fatal traffic accident

Batch Normalization:

- Batch normalization is applied between convolution layers, between convolution and fully connected layers, and within the fully connected layers to accelerate training and prevent overfitting. This architecture enables the model to learn complex patterns in traffic accident data, effectively predicting the severity of accidents based on image data input.

Paper & Year	Algorithm(s) Used	Results
Abdel-Aty & Pande (2007)	Logistic Regression	Achieved significant insights into crash factors, though limited by inability to capture complex non-linear relationships.
Li et al. (2008)	Decision Trees	Demonstrated clear classification of accident severity but struggled with high-dimensional, multi-modal data.
Chen et al. (2016)	CNN for Dashcam and Environmental Data	Achieved notable improvements in accident prediction through visual cue detection; showed higher accuracy than traditional models.
Wang et al. (2019)	Modified ResNet	Achieved 90%+ accuracy in distinguishing severity levels; benefited from deeper feature extraction but faced data heterogeneity issues.

Zhang et al. (2020)	CNN + RNN Hybrid Model	Outperformed standalone CNN and RNN models by capturing spatial-temporal dependencies; improved AUC-ROC scores.
Kim et al. (2020)	VGGNet with Weather and Road Condition Data	Achieved high precision and recall rates, though imbalanced data affected performance on rare severe cases.
Rao et al. (2021)	Inception-based CNN	Improved multi-class severity prediction accuracy by leveraging inception modules for multi-scale processing.
Singh & Gupta (2019)	CNN + LSTM	High accuracy due to sequential event processing; faced generalization challenges across varied traffic environments.
Liu et al. (2022)	Multi-Modal CNN for Sensor Fusion	Showed enhanced accuracy by combining image, sensor, and structured data but required extensive computational resources.
Xu et al. (2023)	3D-CNN with Vehicle Sensor and Video Data	Demonstrated high AUC-ROC and F1-scores; effective in video-based data but was computationally intensive and less interpretable.

Table: Comparative Analysis of Algorithms Used in Traffic Accident Severity Prediction Studies

V. CONCLUSION

The TASP-CNN architecture demonstrates an effective approach to predicting the severity of traffic accidents by leveraging deep learning techniques tailored specifically for image-based data representation of traffic incidents. The model is structured with multiple convolution layers that extract critical features from the input data, followed by fully connected layers that refine these features to make precise predictions. By employing techniques such as batch normalization and ReLU activation, the TASP-CNN achieves improved training stability and prevents overfitting, enhancing its overall predictive performance. The use of the softmax function in the output layer allows the model to classify traffic accident severity into distinct categories, enabling timely and accurate assessment of accident impacts. This architecture shows significant potential in applications where rapid and reliable accident severity predictions are crucial for emergency response, traffic management, and enhancing road safety measures. Future work could further optimize the model and explore its integration with other data sources to improve prediction accuracy and robustness.

REFERENCES

[1] M. Abadi, A. Agarwal, P. Barham, et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: https://www.tensorflow.org/. [Accessed: Sep. 19, 2024].

[2] Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol. 25, pp. 1097-1105, 2012.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778. doi: 10.1109/CVPR.2016.90.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, May 2015. doi: 10.1038/nature14539.

[5] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proc. 32nd Int. Conf. Machine Learning (ICML), Lille, France, 2015, pp. 448-456.

[6] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," IEEE Access, vol. 5, pp. 8869-8879, 2017. doi: 10.1109/ACCESS.2017.2694446.

[7] X. Li, L. Sun, and Y. Ji, "Research on traffic accident severity prediction based on convolutional neural networks," IEEE Access, vol. XX, pp. XX-XX, 2019. doi: 10.1109/ACCESS.2019.2903319.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014. [Online]. Available: https://arxiv.org/abs/1409.1556. [Accessed: Sep. 19, 2024].

[9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014. [Online]. Available: https://arxiv.org/abs/1412.6980. [Accessed: Sep. 19, 2024].

[10] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 1251-1258. doi:

10.1109/CVPR.2017.195.
[11] L. Deng, D. Yu, “Deep learning: Methods and applications,” Foundations and Trends in Signal Processing, vol. 7, no. 3-4, pp. 197-387, 2014. doi: 10.1561/2000000039.
[12] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2881-2890. doi: 10.1109/CVPR.2017.660.
[13] V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” in Proc. 27th Int. Conf. on Machine Learning (ICML), Haifa, Israel, 2010, pp. 807-814.
[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788. doi: 10.1109/CVPR.2016.91.