



ARTICLE

ASLP-DL —A Novel Approach Employing Lightweight Deep Learning Framework for Optimizing Accident Severity Level Prediction

Saba Awan^{1,*} and Zahid Mehmood^{2,*}

¹Department of Software Engineering, University of Engineering and Technology, Taxila, 47050, Pakistan

²Department of Computer Engineering, University of Engineering and Technology, Taxila, 47050, Pakistan

*Corresponding Authors: Saba Awan. Email: saba.awan@uettaxila.edu.pk; Zahid Mehmood. Email: zahid.mehmood@uettaxila.edu.pk

Received: 02 November 2023 Accepted: 28 December 2023 Published: 27 February 2024

ABSTRACT

Highway safety researchers focus on crash injury severity, utilizing deep learning—specifically, deep neural networks (DNN), deep convolutional neural networks (D-CNN), and deep recurrent neural networks (D-RNN)—as the preferred method for modeling accident severity. Deep learning's strength lies in handling intricate relationships within extensive datasets, making it popular for accident severity level (ASL) prediction and classification. Despite prior success, there is a need for an efficient system recognizing ASL in diverse road conditions. To address this, we present an innovative Accident Severity Level Prediction Deep Learning (ASLP-DL) framework, incorporating DNN, D-CNN, and D-RNN models fine-tuned through iterative hyperparameter selection with Stochastic Gradient Descent. The framework optimizes hidden layers and integrates data augmentation, Gaussian noise, and dropout regularization for improved generalization. Sensitivity and factor contribution analyses identify influential predictors. Evaluated on three diverse crash record databases—NCDB 2018–2019, UK 2015–2020, and US 2016–2021—the D-RNN model excels with an ACC score of 89.0281%, a Roc Area of 0.751, an F-estimate of 0.941, and a Kappa score of 0.0629 over the NCDB dataset. The proposed framework consistently outperforms traditional methods, existing machine learning, and deep learning techniques.

KEYWORDS

Injury; severity; prediction; deep learning; feature

1 Introduction

Transportation expansion increases highway accidents, emphasizing the need for accurate severity prediction [1]. Severity, categorized into minor, serious, or fatal outcomes, involves interconnected factors like drivers, vehicles, roads, and weather. A 2022 IRTAD Group report estimated 1.2 million annual road fatalities globally [2]. Crash investigations predominantly use analytical measures, neural networks, KNN, SVM [3,4], and LR [5] approaches. but modern approaches lack deep learning's robustness and understanding, limiting recent studies [6,7]. Our study tackles the challenge of deep learning model generalization across diverse regions with variations in environmental conditions, road surfaces, and driver behaviors. To address this, we curated data from three cross-geographical regions



(NCDB, US, and UK), ensuring a comprehensive representation. The proposed scalable framework adapts to different regions, enhancing its applicability. Extensive data preprocessing, transfer learning through Gaussian function, and heterogeneous data fusion techniques created a balanced dataset reflecting various conditions. We addressed data imbalance with Random Under-Sampling (RUS) and Discrete Synthetic Minority Oversampling Technique (D-SMOTE). Feature Engineering involved Grid Search Optimization, Correlation-based Feature Selection using XGBoost, and noise reduction. To boost generalization, robust data augmentation during training expanded the dataset, exposing the model to diverse conditions. The methodology fine-tuned network architectures, optimized hyperparameters with Stochastic Gradient Descent (SGD), incorporated batch normalization, applied dropout regularization, and underwent evaluations using performance metrics on validation and test sets. This approach ensures fair and accurate predictions across diverse conditions, mitigating biases in the dataset. The proposed methodology is evaluated across three crash record databases (NCDB 2018-2019, UK 2015-2020, and US 2016-2021). Results show reasonable accuracy and predictive power in diverse environments. Deep learning architectures (DNN, D-CNN, D-RNN) adapt well to varied datasets, capturing relevant hierarchical and temporal features. Current models require refinement in configuration, data augmentation, and balancing accuracy and recall. Unlike studies focusing on single-region-specific datasets, our approach addresses multiple cross-geographical accidents, enhancing generalization. Comparative analyses reveal the need for improved model performance in similar studies that lack the broader dataset variety and validation performance. Our proposed ASLP-DL framework performs competitively as discussed in the performance evaluation section suggesting that the chosen databases are appropriate for the specific task. While the use of deep learning models for ASL prediction is promising, addressing the concerns about generalization and expanding the dataset coverage can contribute to the robustness and broader applicability of the proposed framework. Our approach minimizes resource needs for training deep models, undergoing a comparative analysis and sensitivity analysis on specific datasets. Employing a profiling approach, we evaluate crash-related parameters' impact on ASL outcomes, advancing previous ASL analysis by addressing significant correlations among input variables. Key contributions of our ASLP-DL framework include:

1. We have introduced ASLP-DL, an optimized and novel framework for predicting ASLs under various conditions. It has been designed with fewer hidden layers, enhancing its lightweight nature compared to current state-of-the-art DL models.
2. To test ASLP-DL's adaptability, we evaluated it on accident datasets from different geographical regions with varying accident frequencies.
3. We improved model robustness and generalization by incorporating data augmentation and Gaussian noise at the time of backpropagation training of the model and applied dropout regularization to prevent overfitting.
4. We optimized predictive performance by fine-tuning hyperparameters using the Stochastic Gradient Descent (SGD) optimizer for each deep learning network.
5. To ensure transparency and interpretability, we conducted feature importance analysis and sensitivity analysis on network parameters using a profiling approach.
6. Integrating multiple deep learning networks in our framework resulted in impressive prediction accuracy when tested on separate datasets, surpassing traditional models.

The article comprises distinct sections: [Section 2](#) evaluates existing methods, [Section 3](#) presents ASLP-DL framework methodology, [Section 4](#) reports comparative analysis and performance assessment, and [Section 5](#) discusses future initiatives in a concise conclusion.

2 Related Work

To predict the monotony and intensity of traffic collision injuries, state-of-the-art statistical and machine learning methods including Support Vector Machine (SVM) [8,9], K-Nearest Neighbours (KNN) [4], as well as Logistic Regression (LR) [10] are being employed. These methodological approaches rely on predefined connections and patterns. Violating these assumptions reduces the algorithm's injury probability prediction accuracy. Computational Intelligence techniques like Decision Trees [11] and Neural Networks [12] are effective for forecasting but require precise design, especially in data-limited scenarios to reduce false positives. The shift to deep learning is transformative, addressing dimensionality and greatly improving injury severity prediction, especially in Speech Recognition [13], Natural Language Processing [14], and much more [15–18]. Our previous research [19] employing the weighted majority voting (WMV) scheme having a multi-model hybrid architecture, which incorporates multinomial logistic regression (MLR) besides multilayer perceptron (MLP) models, is being carried out by utilizing 3 independent crash records, namely IRTAD, NCDB, as well as FARS. The WMV Hybrid technique outperforms the designed models on the IRTD record with higher accuracy and recall scores of 0.894 and 0.996, respectively, and lower MAE and RMSE values of 0.0731 and 0.2705. Deep learning surpasses conventional methods in quantitative assessments, leading to enhanced research in Computational Intelligent studies, incorporating the deep learning framework in the current study. Studies on ASL prediction, statistical approaches, and neural networks have often taken the lead. The study [20] integrated multinomial logit and SVM models to highlight variables affecting accident severity in daytime and night-time passenger-involved accidents. SVM produced higher valid forecast proportions (45.4% for daytime and 53.1% for night-time) compared to the MNL method (37.82% and 41.35%). We come up with a more robust method for extensive data analysis with deep learning techniques to overcome the constrained capability of state-of-the-art methods. [Table 1](#) below presents the comparative analysis between the existing approaches and the proposed ASLP-DL.

3 Methodology

In this section, we have discussed our ASLP-DL framework, which aims to predict accident severity levels (ASL) in three distinct geographical regions, enhancing road safety. The primary objective is to analyze the impact of highway, weather, and transportation factors to improve road safety and efficient traffic management. [Fig. 1](#) illustrates the various stages of the framework, along with the algorithmic steps used for training and testing.

1. Data Extraction: Integration of driver, vehicle, road, and environmental factors.
2. Data Preprocessing:
 - a. Feature selection using CFS and XGBoost from NCDB, UK, and US datasets.
 - b. Addressing data imbalance through DSMOTE and RUS techniques, Handling incomplete data with a substitution filter, and Probabilistic resampling via k-fold cross-validation for training and testing data.
3. Network Development and Configuration: with optimized hyperparameters and settings
 - a. DNN
 - b. D-CNN
 - c. D-RNN

4. Network Assessment:

- a. Hyperparameter sensitivity analysis, Complex input feature representations using the Profiling method, and Hyperparameter Adaptation.
 - b. Comparative analysis of computational performance.
5. Accurate prediction of ASL as the target attribute for a specific instance.
6. Iteration: Continuing to stage 2 to analyze forthcoming accident records.

Detailed discussions of these phases follow in the subsequent subsections. The research utilizes the Weka Deep Learning4j [21] Package on a Core i7 desktop PC with 16 GB of RAM to implement and refine DNN, D-CNN, and D-RNN networks.

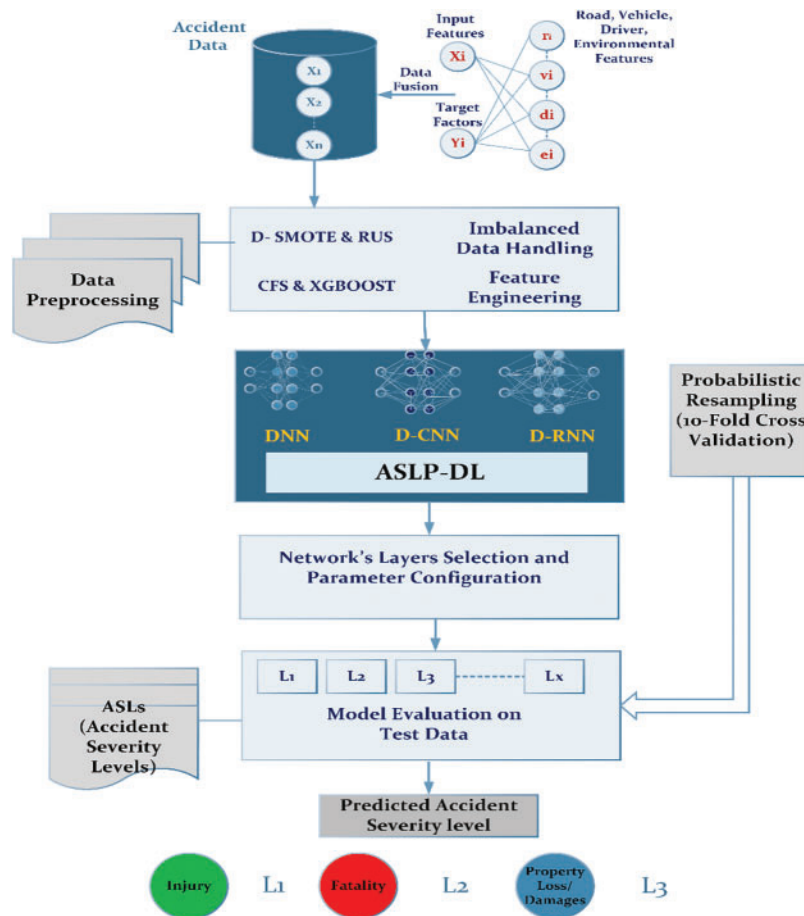


Figure 1: Schematic layout showing the proposed framework ASLP-DL

Table 1: A comparative analysis between the existing approaches and the proposed ASLP-DL

Research work	Dataset	Data preprocessing	Backpropagation training	Feature engineering and selection	Factor contribution analysis	Hyperparameter sensitivity analysis	Computational complexity comparison	Classifier/ Approach	Research gap
[22]	Single dataset: UK traffic accident dataset 2011–2016	None	None	Loss function	Feature importance ranking using TabNet model	None	None	Multiple deep models, including 1. MLP using Keras, 2. MLP with embedding layers 3. TabNet	Limited, diverse accident datasets constrain neural network architecture evaluation, which performs best with broader data.
[23]	Single dataset: Louisiana Work zone Accident Record from 2014–2018.	Converting nominal and ordinal level data to dummy variables.	None	Cart Mars	None	None	None	Single deep learning model EfficientNet and SVM model	Single PEC Approach having comparison between deep learning and the SVM models without any complexity comparison and factor contribution analysis.
[24]	Single dataset: UK traffic accident dataset 2005–2016	Simple data profiling and SMOTE	None	Catboost shapley	None	None	None	Single deep learning model: stacked sparse auto-encoder (SSAE)	Single PEC Approach involving comparison between deep learning and two baseline models DNN and AE with limited evaluation matrices.
[25]	Samples data of self-labeled sequences recorded from the cameras for recording the vehicles and pedestrians	Label Encoding	Data augmentation	None	None	None	Yes	Single deep learning model: Deep convolutional neural network RCNN	In the RCNN, the selective search consumes most of the efficiency and causes slow detection speed.

(Continued)

Table 1 (continued)

Research work	Dataset	Data preprocessing	Backpropagation training	Feature engineering and selection	Factor contribution analysis	Hyperparameter sensitivity analysis	Computational complexity comparison	Classifier/Approach	Research gap
Proposed model	Three cross-geographical crash databases: 1. NCDB 2018–2019 2. UK 2015–2020 3. US 2016–2021	D-SMOTE RUS	1. Data augmentation 2. Gaussian noise	CFS XGboost	Profile-based method	1. Optimization method 2. Sample size 3. Learning rate 4. Dropout parameters	Yes	ASLP-DL: Streamlined and network architecture for deep ASL classification with Lightweight deep learning models using: D-RNN, D-CNN and DNN have reduced hidden layers and enhanced generalization capabilities.	While the model excelled compared to existing methods, improving validation accuracy for real-world RTA databases remains essential.

3.1 The Acquisition of Accident Datasets

3.1.1 NCDB (National Collision Database)

Firstly, we chose NCDB [26] which includes, 28984 accident reports containing 20 unique vehicles, drivers, and climatic factors. It covers all police-reported motor vehicle crashes for the year 2020. It contains the Severity Category target feature, which has a number between 1 and 3, where 1 indicates Injury severity, 2 indicates Fatality, and 3 indicates Property damage only which has a significant impact on traffic.

3.1.2 Road Safety Data UK

Secondly, the proposed research selects the UK accident dataset [27] which comprises detailed road safety information on the situation of GB road accidents from the year 2019–2020. Road type (expressways, urban areas, country highways), road users (pedestrians, bicyclists, vehicle passengers, and many others), age, gender, and sitting posture, in addition to the environmental elements and climate conditions at the time of the collision.

3.1.3 US Accidents Record

The US nationwide traffic collision record [28], which contains data from each of the 49 US states, is being selected as the final dataset. From February 2016 to December 2021, accident data was acquired, accumulating more than 15000 occurrences with 47 distinct accident variables and a target attribute.

3.2 Data Description

The crash dataset includes driver information, vehicle details, roadway characteristics, and environmental attributes. The main target variable is Accident Severity Level (ASL), categorized into injuries, fatalities, and property loss. There are also 12 independent variables related to driver, vehicle, road, and environmental features.

3.3 Data Pre-Processing

3.3.1 Heterogeneous Data Fusion

Accident data acquisition involves compiling diverse datasets, from highway conditions to weather, timing, transportation, and psychological factors. This comprehensive approach aims to effectively predict and prevent road accidents. To ensure robust analyses, we employ data fusion, synthesizing insights from various sources into a unified accident matrix. Before utilizing deep learning models, organizing and consolidating pertinent attributes is a crucial preliminary step, enhancing our understanding of accident-influencing factors visually represented in the following equation:

Accident Matrix = CONCAT (VI (Drivers Data, Weather Conditions) :

$$\text{CN (Vehicle Information, Transportation Details, Environmental Factors)} \quad (1)$$

$$X_n = \text{CONCAT (VI (d}_i, r_i) : \text{CN (v}_i, e_i)) \quad (2)$$

where, we have n accident records, each with intricate details about driver attributes (d_i), vehicle attributes (v_i), road attributes (r_i), and environmental factors (e_i). To consolidate this wealth of information, a meticulous merging process is undertaken. The key to this integration lies in *Crash Number* and *Vehicle Index*, serving as linking keys across the datasets. Imagine a complex network

where each accident record is a node, and the merging process involves establishing connections between these nodes based on *Crash Number* and *Vehicle Index*. Through this interconnected structure, casualties are associated with specific vehicles, extending to include pedestrians involved in the accidents. The dataset is expanded by matching it with additional accident data, using *Crash Number* as the common identifier. This final step ensures the creation of a comprehensive dataset that encapsulates a holistic view of each accident event. The resulting dataset becomes a powerful tool, allowing for in-depth analysis and insights into the multifaceted factors contributing to road accidents. This equation underscores our commitment to distilling complex, multidimensional data into a structured format, laying the foundation for more effective predictive modeling and analysis in the realm of road safety.

3.3.2 Handling Imbalanced Data Using D-SMOTE and RUS

In the preprocessing phase, we balance the crash severity data by employing both under-sampling and over-sampling methods. We use Discrete Synthetic Minority Over-sampling (D-SMOTE) [29] and Random Under-Sampling (RUS) [30] for under-sampling. This ensures a balanced dataset with equal proportions for all class labels.

3.3.3 Filter to Incorporate Missing Entries

Missing data is a prevalent issue in real-world predictive models, as seen in accident records from databases like the UK, NCDB, and the US. To handle this, we use mean substitution-based imputation, replacing missing values with estimates calculated from neighboring data points by computing the mean of the features.

3.3.4 Feature Engineering and Selection

To optimize resource usage and improve model performance, we conduct feature engineering. This step involves removing ineffective accident features and creating more predictive attributes through nonlinear data transformations, aiming to produce a structured dataset with accident predictors related to severity levels and ASL.

Correlation-Based Feature Selection

Additionally, the CFS approach and the Greedy-Stepwise-Search method are used after the preprocessing stage. The CFS is employed to identify and remove unnecessary, improper, and repetitive information from the crash data. To forecast the target attribute label, CFS finds the features that are more critical and prospective predictors. This is how the CFS assessment is described:

$$\text{CFS} = \max_{s_k} \left[\frac{r_c f_1 + r_c f_2 + \dots + r_c f_k}{\sqrt{k + 2 (r f_1 f_2 + \dots + r f_i f_j + \dots + r f_k f_{k-1})}} \right] \quad (3)$$

where, the mean quantity of all related attributes classification correspondence is represented by the r_{cf_k} . While the mean score of all attribute-to-attribute correspondence is represented by $r_{f_i f_j}$.

XGBOOST

The XGBoost method significantly improved the features chosen by CFS by using them as input data in its assessment of characteristic significance with severity category. The impact of every attribute on the XGBoost algorithm's predictive accuracy is then evaluated. The following is a description of the XGBoost Log function:

$$L^{(t)} = \sum_{i=1}^n l(y_i^i, y_i^{\wedge(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (4)$$

where, y^i is the predicted value, n is the number of samples, and l is represented by the loss method $f_t(x_i)$, whereas the preceding phase projected value ($t-1$) is indicated by a and the latest learner which we have to include in stage t is Δx . The CART learning mechanism has l as the summation of the cumulative trees from the current and the past. For the XGBoost algorithm to determine feature relevance relative to the target ASL a multitude of accident attributes have been utilized as input. The impact of every characteristic on the XGBoost algorithm's prediction score is then estimated. The feature relevance factor with the peak score is computed for the feature: weather condition as (1.168).

3.3.5 Probabilistic Resampling via K-Fold Cross-Validation Technique

Further, a cross-validation procedure consisting of 10 repetitions is employed to prevent overfitting and precisely evaluate multiple network models. The process is used to divide the supplied feature set into test and preparation data. Pre-processed data is employed to train the samples, and test data remains functioning for the evaluation of the model to be trained. We have randomly split the source dataset across ten equivalent simple partitions to employ a Cross-Validation process involving 10 passes. In addition, whenever k-fold cross-validation is performed, normally single examination is utilized for confirmation evaluating /testing evidence, and the remaining k-1 checks of data are employed as training evidence. The algebraic equation for probability testing is:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - Y_i^{\wedge})^2 \quad (5)$$

$$CV_k = \frac{1}{k} \sum_{i=1}^k MSE_i \quad (6)$$

where, n is the data points count, MSE is the mean squared error, Y_i is observed, and Y_i^{\wedge} is predicted. The procedure repeats for k epochs (usually 5–10), computing MSE, and assessing scenarios where $k(i) = i$, excluding the i_n assessment.

4 Model Development and Prediction

Deep learning algorithms transform input sequences into output sequences using linear combinations. This section trains and tests DNN, D-RNN, and D-CNN networks to categorize accident severity levels based on driver behavior, road specs, vehicle details, and environmental factors through a 10-fold cross-validation approach with selected input features. The model's effectiveness relies on precise hyperparameter tuning, achieved through iterative selection using the Stochastic Gradient Descent (SGD) optimizer in the (ASLP-DL) framework. Recognizing the impact of data characteristics on deep learning performance, we move beyond generic parameters and tailor the network architecture, refining D-NN, D-CNN, and D-RNN networks.

Employing the SGD optimizer within WekaDeeplearning4j ensures efficiency, especially with large datasets. SGD's noise-handling ability, adaptive learning rates, and flexibility in complex parameter spaces contribute to faster convergence, making it effective for optimizing model performance in intricate scenarios. The rationale behind the chosen hyperparameters with the SGD optimizer stems from its adaptability to large, noisy datasets and its efficiency in navigating complex parameter spaces, ultimately improving overall model performance.

4.1 Implementation of D-RNN Model of Proposed ASLP-DL Framework

In the network development phase, we introduce recurrent neural networks (D-RNNs) with feedback connections to model non-linear accident patterns. D-RNNs offer computational capacity and the ability to leverage previous information through recurrent linkages. To address gradient issues, we employ Long Short-Term Memory (LSTM), featuring specialized units with input, output, and forget gates for controlled functionality and memory cell operations. A typical LSTM cell begins by deciding whether to retain or discard data from the previous time step. The equations for this process are as follows:

$$f_t = \sigma [x_t * U_f + H_{t-1} * W_f] \quad (7)$$

$$i_t = \sigma [x_t * U_i + H_{t-1} * W_i] \quad (8)$$

$$C_t = f_t * C_{t-1} + i_t * N_t \quad (9)$$

$$H_t = O_t * \tanh(C_t) \quad (10)$$

$$O_t = \sigma [x_t * U_o + H_{t-1} * W_o] \quad (11)$$

where, at timestamp t (x_t), U_f , U_i , and W_f are weight matrices for input and hidden state connections. Gates f_t , i_t , and O_t manage information flow, and σ calculates the updated cell state (\tanh) using the present hidden state ($H(t-1)$). To prevent over-generalization, the D-RNN model employed three key strategies including the recurrent approach to capture temporal associations in road accident data, optimizing hidden layers with batch normalization, and incorporating data augmentation for enhanced generalization. Overfitting is mitigated by emphasizing core connections with dropout regularization, including techniques like adding Gaussian noise, initializing hidden units with ReLU, and applying a 30% dropout probability. The D-RNN, guided by a 12-characteristic input vector, is optimized for road crash ASL prediction through a continuous grid search algorithm with cross-validation.

4.1.1 Proposed D-RNN Model's Hyperparameter Tuning

The D-RNN model's hyperparameters are optimized through a methodical 100-epoch grid search, resulting in significant improvements. Diverse parametric combinations are explored, and cross-validation with 10 epochs assesses each predictor, identifying the most effective parameters. A network model is constructed using the optimal hyperparameters from [Table 2](#), featuring LSTM architecture, dense layers, and a Softmax layer. To reduce complexity, three dropout layers with a 0.3 probability are employed. The training utilizes Stochastic Gradient Descent (SGD) with a batch size of 8 and a learning rate of 0.01 as presented in [Table 2](#) below. Grid search and 10-fold cross-validation select network settings, and a sensitivity analysis evaluates their impact on crash severity outcomes. [Fig. 2](#) below illustrates the high-level design of the D-RNN model of the proposed ASLP-DL framework. In the context of accident severity prediction, overfitting poses a challenge when models become excessively complex and perform well on training data but struggle to generalize to new data. This can lead to inaccurate predictions in real-world scenarios. Dropout regularization is a preferred solution to address overfitting. It involves randomly "dropping out" a proportion of neurons during training, preventing any single neuron from becoming overly specialized. This promotes a more robust and generalized learning by forcing the model to rely on a broader set of features. In accident severity prediction, where diverse and unpredictable factors can influence outcomes, dropout helps prevent the model from memorizing noise in the training data, enhancing its ability to make accurate predictions on new and unseen data.

Table 2: Proposed D-RNN optimized hyper-parameters

Name of parameter	Representation	Search range	Pre-defined value	Optimized value
Nb hidden	Hidden neurons	[16, 32, 64, 128, 256, 512]	32	64
Batch size	Training instances	[4, 8, 16, 32, 64, 128, 256]	32	8
Optimization algorithm	Weight adjustment	[SGD, Adam, Nadam, Adagrd]	Nadam	SGD
L2	Regularization factor	[0.001, 0.01, 0.1, 0.3]	0.01	0.3
RL	Rate of learning	[0.0001, 0.001, 0.01, 0.3]	0.001	0.01
Dropout	Neuron removal	[0.1, 0.2, 0.3, 0.4, 0.5]	0.2	0.3

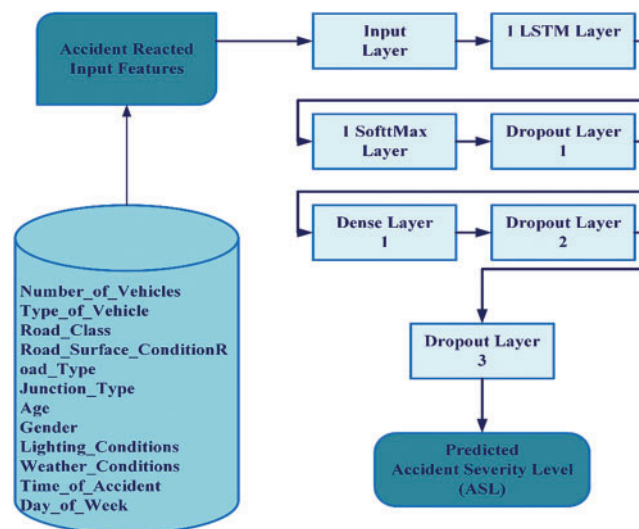


Figure 2: High-level design of the D-RNN model of the proposed ASLP-DL framework

In accident severity prediction, where diverse and unpredictable factors can influence outcomes, dropout helps prevent the model from memorizing noise in the training data, enhancing its ability to make accurate predictions on new and unseen data. The challenge lies in finding the optimal dropout rate, which we have addressed through the sensitivity analyses step and find an optimal dropout rate for each model thus balancing regularization benefits with model performance on the specific accident severity prediction task.

4.2 Implementation of D-CNN Model of Proposed ASLP-DL Framework

The Convolutional Neural Network (D-CNN) is a powerful choice in ASLP-DL development, especially in computer imaging and classification. AlexNet’s introduction in 2012 established its significance in computer vision and pattern recognition with eight layers, including convolution, pooling, and fully connected layers. D-CNN processes one-dimensional vectors representing traffic

accident data and predicts features within the 0 to 1 range. It employs convolution, pooling, fully connected layers with weight matrices, and activation functions for this purpose.

$$y_{jk}(x) = f\left(\sum_{i=1}^{n_H} w_{jk}x_i + w_{j0}\right) \quad (12)$$

where, a dot product between the weights matrix W and the input vector x is being taken. The bias term ($W0$) can be added inside the non-linear function f . All outputs (scaled by the filtered elements) from the units of the preceding layer must be added up to calculate the pre-nonlinearity, supplied to a certain unit x_i in the current tier.

4.2.1 Proposed D-CNN Model's Hyperparameter Tuning

In continuation with the model development phase, the second developed network, D-CNN, uses pooling and convolution activities to restructure the pre-refined input variables into a distinct characteristic presentation as part of the model construction and classification process. It starts with a set of pre-processed and selected accident features. To accommodate the sequential crash data, a single-dimensional convolution procedure is used. To isolate the restored features, the greatest number of pooling activities are used. The characteristics are then made usable by flattening them. To find the precise configuration for an ideal network system for estimating the overall severity of road crashes, multiple settings of hyperparameter combinations of D-CNN are being tried and refined by grid search over a designated search area employing cross-validation. Optimized D-CNN hyperparameters are summarized in [Table 3](#) below. The network uses 1D convolution, max-pooling, and softmax layers for ASL prediction. A 0.3 dropout layer reduces complexity and overfitting. Training with backpropagation, Nadam optimizer (batch 16, learning rate 0.001), and three dropout layers to mitigate overfitting. Parameters were selected via grid search and 10-fold cross-validation. Sensitivity analysis assesses refined parameter impact on injury severity outcomes.

Table 3: Proposed D-CNN optimized hyper-parameters

Name of parameter	Representation	Search range	Pre-defined value	Optimized value
Nb hidden	Hidden neurons	[16, 32, 64, 128, 256, 512]	32	128
Batch size	Training instances	[4, 8, 16, 32, 64, 128, 256]	32	16
Optimization algorithm	Weight adjustment	[SGD, Adam, Adagrad, Nadam, RMSprop]	Adam	Nadam
L2	Regularization factor	[0.001, 0.01, 0.1, 0.3]	0.01	0.3
RL	Rate of learning	[0.01, 0.01, 0.01, 0.3]	0.001	0.02
Dropout	Neuron removal	[0.1, 0.2, 0.3, 0.4, 0.5]	0.2	0.3

4.3 Implementation of the DNN Model of the Proposed ASLP-DL Framework

Feedforward neural networks constitute a collection of scientific learning techniques used in machine learning. Inputs, hidden layers, and output layers make up the three layers that comprise a basic deep-feed forward neural architecture, which is an arrangement of neurons or nodes. The anticipated research illustrates a continuous-time connection across the input data (accident variables)

and the output factors ASL by the model architecture. Fig. 3 below illustrates the high-level design of the D-CNN model of the proposed ASLP-DL framework.

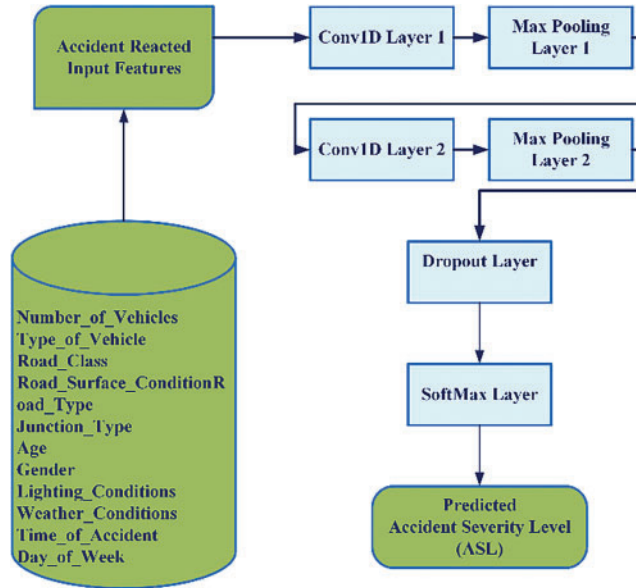


Figure 3: High-level design of the D-CNN model of the proposed ASLP-DL framework

Weight vectors are connected systematically in neurons, which are frequently structured into tiers with complete connections between one layer and the next. The sources towards the node are adjusted by a standard initiation function, which determines the resultant signal. The mean square error cost function for DNN is defined as follows:

$$C(w, o) = \frac{1}{n} \sum_{i=1}^n [y_i - o_i] \quad (13)$$

where, in this context, “y” denotes true labels, “n” represents training data count, and “o” is the network’s predictions. The DNN effectively classifies road crash severity but faces challenges like high operational costs and remote sensing limitations.

4.3.1 Proposed DNN Model's Hyperparameter Tuning

To determine the best network for road crash severity assessment, DNN hyperparameters undergo a grid search with 100 iterations and cross-validation. The optimized DNN model, featuring two fully connected layers, a Softmax layer, and a Long-Short Term Memory (LSTM) layer, is applied to all three accident datasets. Network complexity is reduced with two dropout layers (0.3 probability) to prevent overfitting. Backpropagation using SGD optimization (batch size 8, learning rate 0.001) guides the network. With a grid search along with a cross-validation analysis of 10 epochs, the model’s settings are chosen as presented in Table 4 below. To find out how these variables affected the results of injury severity, the DNN model is undergoing a risk sensitivity assessment. Interpreting the predictions of deep learning models, including D-NN (Deep Neural Network), D-CNN (Deep Convolutional Neural Network), and D-RNN (Deep Recurrent Neural Network), poses a challenge due to their complex and nonlinear nature. However, there are approaches and considerations for understanding how these models arrive at their predictions. Firstly, for the D-NN model, analysis of intermediate layers reveals the hierarchical feature representations. For the D-CNN approach, the Visualization of feature maps

in convolutional layers provides insight into learned accident patterns. Activation maps and filters help identify which regions of input data are crucial for predictions. Lastly, we examined the hidden states in recurrent layers of the D-RNN model to understand temporal dependencies. Fig. 4 below illustrates the high-level design of the DNN model of the proposed ASLP-DL framework.

Table 4: Proposed DNN optimized hyper-parameters

Name of parameter	Representation	Search range	Pre-defined value	Optimized value
Nb hidden	Hidden neurons	[16, 32, 128, 512]	32	128
Batch size	Training instances	[4, 8, 32, 128, 256]	32	8
Optimization algorithm	Weight adjustment	[SGD, Nadam]	Nadam	SGD
L2	Regularization factor	[0.01, 0.01, 0.3]	0.001	0.02
RL	Learning rate	[0.001, 0.1, 0.3]	0.01	0.3
Dropout	Neuron removal	[0.1, 0.2, 0.4, 0.5]	0.2	0.3

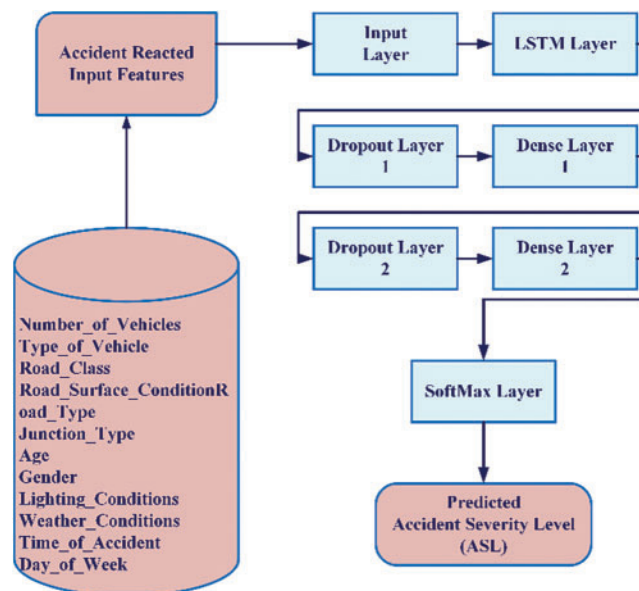


Figure 4: The high-level design of the DNN model of the proposed ASLP-DL framework

5 Proposed Framework's Performance Evaluation and Discussions

The ASLP-DL framework's developed networks undergo validation via 10-fold cross-validation using three distinct test accident records: NCDB, UK, and US. After constructing, configuring, and optimizing the networks with refined hyperparameters, a quantitative analysis of the results is conducted. This study focuses on performance evaluation and comparison between the ASLP-DL framework and separately built deep learning techniques (DNN, D-RNN, and D-CNN) using diverse crash data sources.

5.1 Quantifiable Evaluation of Results

Choosing the most appropriate evaluation metric depends on the nature of the task, the dataset, and the specific goals of the evaluation. It is often recommended to consider multiple metrics to gain a comprehensive understanding of model performance. In our experimental evaluation phase, a comprehensive comparison of different evaluation metrics is utilized to assess the specific strengths and weaknesses in different aspects of prediction performance. In the first stage of the evaluation comparison phase, we used robust evaluation metrics like Precision, Recall, mean absolute error (MAE), and root mean squared error (RMSE) to accurately assess our techniques beyond prediction accuracy. All three networks in the ASLP-DL framework are evaluated with these metrics. Among all three accident records NCDB, UK, and US, the D-RNN achieved the highest precision (0.895) and recall (0.996) over the NCDB dataset. It also achieved the lowest MAE and RMSE, both at 0.0731, among the developed networks for the NCDB dataset as shown in [Table 5](#) below.

Table 5: Performance assessment of the proposed ASLP-DL scheme over selected datasets

General evaluation profile					
Proposed ASLP-DL method	Accident dataset	Precision	Recall	MAE	RMSE
DNN	NCDB	0.892	0.994	0.1176	0.3439
	UK	0.712	0.814	0.2710	0.3110
	US	0.793	0.894	0.2176	0.2839
D-CNN	NCDB	0.894	0.892	0.2249	0.3312
	UK	0.811	0.901	0.1420	0.2715
	US	0.601	0.881	0.1120	0.2922
D-RNN	NCDB	0.895	0.996	0.0731	0.2705
	UK	0.810	0.921	0.1320	0.2850
	US	0.884	0.794	0.1752	0.2809

5.2 Performance Analyzation and Comparisons

In the second stage of the evaluation comparison phase, we employed a confusion matrix analysis, along with metrics like F-measure, ROC Area, and Kappa, to assess prediction accuracy. Kappa Rate indicates agreement between predictions and actual outcomes, with a score above 0 suggesting the model outperforms random chance and individual classifiers for each target class. ROC Area helps identify better classifiers, with a perfect model approaching a score of 1, indicating high accuracy compared to random chance. The optimized D-RNN model outperforms D-CNN and DNN in predictive accuracy, achieving high F-Measure and ROC Area values, with an accuracy score of 89.03% as shown in [Table 6](#) below. However, D-CNN slightly outperforms D-RNN in F-Measure across the accident records.

Table 6: Efficiency analysis of the proposed framework according to performance measures

Proposed ASLP-DL	Crash record	ACC	F-Estimate	ROC region	Kappa coefficient
DNN	NCDB	88.7971%	0.940	0.553	0.0211
	UK	87.591%	0.850	0.511	0.0143
	US	87.247%	0.912	0.652	0.0202
D-CNN	NCDB	88.8149%	0.942	0.542	0.0329
	UK	88.8012%	0.752	0.421	0.0451
	US	85.214%	0.779	0.665	0.0321
D-RNN	NCBD	89.0281%	0.941	0.751	0.0629
	UK	89.0132%	0.889	0.701	0.0612
	US	88.8241%	0.901	0.712	0.0541

5.3 Comparative Evaluation of the Proposed and Prevailing Approaches

In the third and last stage of empirical assessment, a comparison between the proposed ASLP-DL framework and standard methodologies for crash severity assessment is conducted. Table 7 below demonstrates that the proposed ASLP-DL framework outperforms the standard methodologies when predicting ASLs using deep learning approaches, achieving the highest Precision and ACC scores of 89.0281% and 0.88. These assessment statistics establish the superiority of the proposed research mechanism, underscoring the framework's significance in accurately predicting road crash severity.

Table 7: Comparative assessment of the proposed ASLP-DL framework and existing approaches

Empirical methodology	Model	Execution assessment approach	Precision	ACC
DL [21]	MLP, TabNet	Training and validation accuracies	0.69	71.77%
DL [22]	EfficientNet and SVM model	BEV-Statistics	0.187	80.00%
DL [23]	Stacked sparse auto-encoder (SSAE)	Average accuracy precision	0.51	76.76%
DL [24]	Lightweight D-CNN	Average accuracy precision	0.83	83.60%
Proposed ASLP-DL	Lightweight DNN, D-RNN D-CNN for ASL prediction	Goodness-of-fit Measures	0.88	89.0281%

5.4 Computational Complexity Comparison of the Proposed and Prevailing Approaches

The study compared the computational complexity of the ASLP-DL methodology with state-of-the-art approaches. Model time complexity has been calculated by adding learning and evaluation time for each iteration. Table 8 below shows the training and validation times for each iteration with a group size of 32. On average, the proposed D-RNN network takes 149.14 milliseconds for learning and 13 milliseconds for evaluation of new cases. While this highlights the model's computational efficiency, it's worth noting that the learning time can be increased by reducing group or pattern size or increasing the number of training examples.

Table 8: Comparative assessment of computational complexity of the proposed ASLP-DL framework and existing approaches

Cutting-edge techniques	Computing duration per iteration (milliseconds) of model
DL [21]	159.41
DL [22]	165.52
DL [23]	150.23
DL [24]	N/A
Proposed ASLP-DL	149.14

5.5 Assessment of Sensitivity Analysis and Effectiveness of Various Hyperparameters

Customization of deep learning algorithms goes beyond default settings due to input variations and computational methods. This study optimizes D-RNN, D-CNN, and DNN models using a grid search for enhanced ASL estimation accuracy with twelve predictors. Optimization analysis favors SGD, producing precision values of 0.891, 0.887, and 0.881 for D-RNN, D-CNN, and DNN models. Sample size and learning rate impact were explored, highlighting DNN and D-RNN excellence with a sample size of 8, while D-CNN performed best with a batch size of 16. Sensitivity analysis on dropout parameters revealed optimal values of 0.3, 0.2, and 0.5 for DNN, D-RNN, and D-CNN, respectively. Dropout emerges as crucial in preventing overfitting, particularly in CNN and RNN models with a substantial number of parameters. Given dropout rates’ substantial influence on model accuracy and their dependence on parameters, selecting appropriate keep probability requires a dataset-specific and task-dependent grid search. Fig. 5 below illustrates the optimization algorithm and batch size impact on ASLP-DL accuracy. Fig. 6 below illustrates the learning rate and dropout complexity impact on ASLP-DL accuracy.

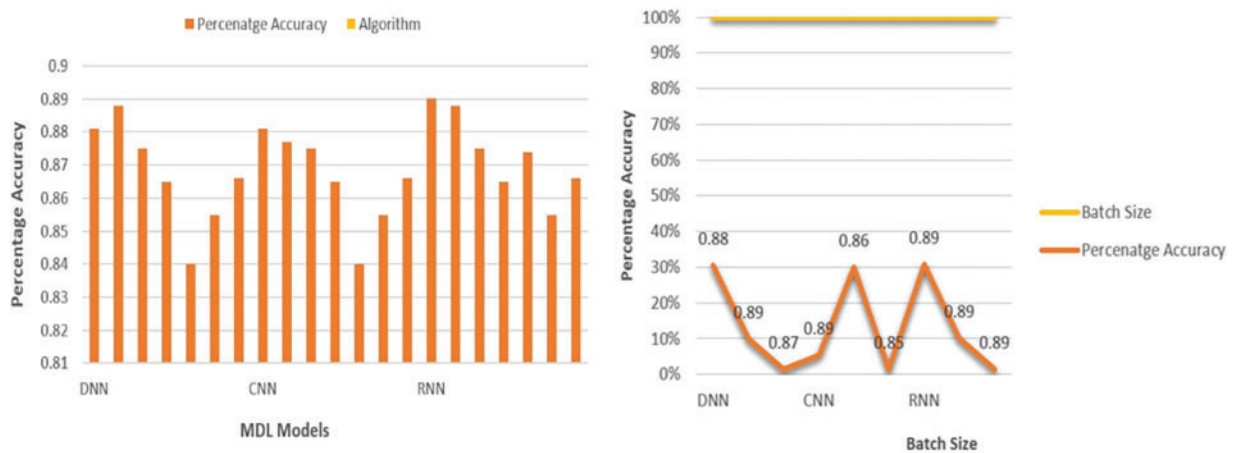


Figure 5: Optimization algorithm and batch size impact on ASLP-DL accuracy

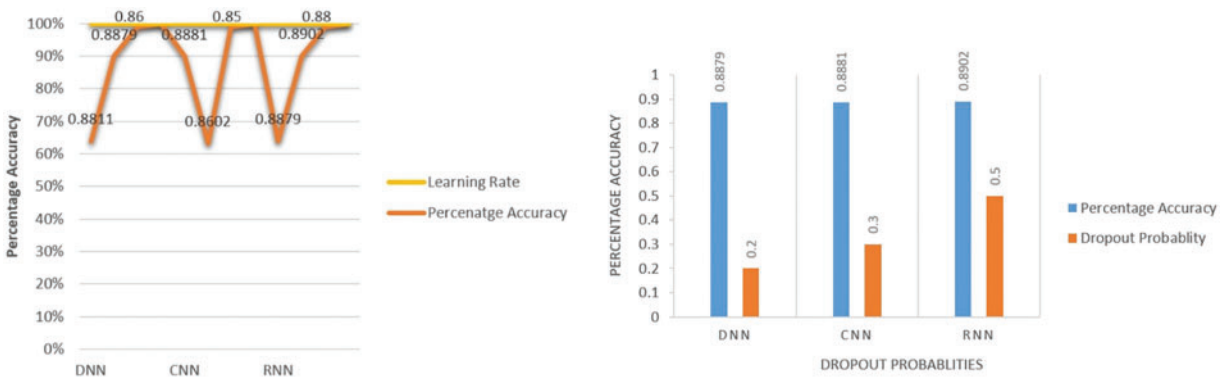


Figure 6: Learning rate and dropout complexity impact on ASLP-DL accuracy

5.6 Factor Contribution Analysis and Knowledge Discussion

Using profiling, we assessed the impact of individual factors on road crash severity in ten intervals, with results shown in [Table 9](#). Understanding the influential predictors identified through sensitivity analysis and factor contribution analysis is crucial for real-world application and acceptance, particularly in domains like highway safety. Road_Surface_Condition and Time_of_Accident scored the highest (1.168 and 1.145), highlighting the significance of slippery road surfaces and specific weather conditions. Ages 18 to 30 were prone to severe crashes on major highways, while cars and motorcycles posed higher risks than buses. Female drivers had a lower risk and crashes on entrance/exit routes, toll stations, and major roadways carried higher risks, particularly in dark conditions. Among deep networks, DNN, D-CNN, and D-RNN excel in forecasting accident severity, with D-RNN's focus on temporal factors being particularly noteworthy. D-CNN is effective for two-dimensional data and surpasses DNN in prediction accuracy. D-RNN outperforms D-CNN by considering temporal factors and incorporating information related to traffic conditions, vehicle speed, and weather. Its capacity to leverage historical data is instrumental in identifying complex accident patterns. Accurate predictions of crash probabilities on specific road segments contribute to more informed highway design. Multi-model deep learning approaches surpass traditional methods and neural networks in handling unevenly distributed data. Our D-CNN model, trained on US accident databases, falls short in cross-record evaluation for recognition accuracy ([Table 7](#)). Conversely, the D-RNN model excels in individual database experiments, demonstrating superior evaluation metrics ([Table 6](#)). Notably, D-CNN outperforms D-RNN in F-Measure over the NCDB database, due to its proficiency in learning spatial relationships. D-CNN excels in recognizing patterns in 2D arrays, effective for analyzing accident event features. In contrast, RNN models focus on temporal patterns, showing higher accuracy in predicting traffic accidents. D-RNN's memory capabilities automate feature identification, advantageous for accident forecasting. However, their complex training algorithms may limit applications, especially with limited datasets lacking temporal features. These findings underscore the importance of training deep learning models on realistic databases for enhanced generalization.

Table 9: The estimated weights of crash-related factors

Feature	Score
Number_of_vehicles	1.032
Type_of_vehicle	1.063
Road_class	1.168
Road_surface_condition	1.130
Road_type	1.115
Junction_type	1.019
Age	1.029
Gender	1.022
Lighting_conditions	1.032
Weather_conditions	1.019
Time_of_accident	1.029
Day_of_week	1.145

6 Conclusion and Future Directions

The ASLP-DL framework delves into factors influencing accidents of varying severity. Male drivers are more associated with severe incidents, while female drivers with minor ones. Key attributes for determining ASLs are Time-of-Accident and Surface-Condition-of-the-Road. Precise accident severity prediction enhances highway network management and road safety. The study employs three deep learning models (DNN, D-RNN, and D-CNN). D-RNN outperforms D-CNN and NN with 89.0281% accuracy using SGD optimization. Optimal batch sizes range from 4 to 8, and dropout rates of 0.2 to 0.5 are crucial for D-CNN and D-RNN. Further research is essential for comprehensive tuning, particularly on extensive datasets, to adapt deep learning for practical applications and improved highway safety, benefiting state agencies and organizations.

Acknowledgement: We would like to thank the “anonymous” reviewers and editors for their thoughtful insights. We are also immensely grateful for their comments on an earlier version of the manuscript and they have kindly assisted as research volunteers.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm their contribution to the paper as follows: study conception and design: Saba data collection: Saba; analysis and interpretation of results: Saba, Zahid; draft manuscript preparation: Saba, Zahid, All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author Saba, upon reasonable request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] U. Sirisha and B. S. Chandana, "Privacy-preserving image encryption with optimal deep transfer learning based accident severity classification model," *Sensors*, vol. 23, no. 1, pp. 519, 2023. doi: [10.3390/s23010519](https://doi.org/10.3390/s23010519).
- [2] ITF, "Road safety annual report 2022," OECD Publications, vol. 1, no. 1, pp. 1–67, 2022.
- [3] S. Sahu, B. Maram, V. Gampala, and T. Daniya, "Analysis of road accidents prediction and interpretation using KNN classification model," in *Proc. IEMIS*, Singapore, Springer, 2022, vol. 2, pp. 163–172.
- [4] B. Princess, P. J., S. Silas, and E. B. Rajsingh, "Classification of road accidents using SVM and KNN," in *Advances in Artificial Intelligence and Data Engineering Book Series*, Springer, 2021, pp. 57–73.
- [5] K. Santos, J. P. Dias, and C. Amado, "A literature review of machine learning algorithms for crash injury severity prediction," *J. Saf. Res.*, vol. 80, no. 1, pp. 254–269, 2022. doi: [10.1016/j.jsr.2021.12.007](https://doi.org/10.1016/j.jsr.2021.12.007).
- [6] R. O. Mujalli, H. Al-Masaeid, and S. Alamoush, "Modeling traffic crashes on rural and suburban highways using ensemble machine learning methods," *KSCE J. Civ. Eng.*, vol. 27, no. 2, pp. 814–825, 2022. doi: [10.1007/s12205-022-0658-4](https://doi.org/10.1007/s12205-022-0658-4).
- [7] M. A. Rahim and H. M. Hassan, "A deep learning based traffic crash severity prediction framework," *Accid. Anal. Prev.*, vol. 154, no. 1, pp. 106090, 2021. doi: [10.1016/j.aap.2021.106090](https://doi.org/10.1016/j.aap.2021.106090).
- [8] A. EKitali, S. Mokhtarimousavi, C. Kadeha, and P. Alluri, "Severity analysis of crashes on express lane facilities using support vector machine model trained by firefly algorithm," *Traffic Inj. Prev.: Taylor & Francis*, vol. 22, no. 1, pp. 79–84, 2021. doi: [10.1080/15389588.2020.1840563](https://doi.org/10.1080/15389588.2020.1840563).
- [9] J. Li, J. Guo, J. S. Wijnands, R. Yu, C. Xu, and M. Stevenson, "Assessing injury severity of secondary incidents using support vector machines," *J. Transp. Saf. Secur.: Taylor & Francis*, vol. 14, no. 2, pp. 197–216, 2022. doi: [10.1080/19439962.2020.1754983](https://doi.org/10.1080/19439962.2020.1754983).
- [10] R. E. Al Mamlook, T. Z. Abdulhameed, R. Hasan, H. I. Al-Shaikhli, I. Mohammed and S. Tabatabai, "Utilizing machine learning models to predict the car crash injury severity among elderly drivers," in *Proc. of IEEE IEEE Int. Conf. Electro Inf. Technol. (EIT)*, Chicago, IL, USA, 2020, pp. 105–111.
- [11] V. G. Costa and C. E. Pedreira, "Recent advances in decision trees: An updated survey," *Artif. Intell. Rev.*, vol. 56, no. 5, pp. 4765–4800, 2022. doi: [10.1007/s10462-022-10275-5](https://doi.org/10.1007/s10462-022-10275-5).
- [12] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti, and D. De, "Fundamental concepts of convolutional neural network," in *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, New York City: Springer, 2020, vol. 1, pp. 519–567.
- [13] K. B. Bhangale and M. Kothandaraman, "Survey of deep learning paradigms for speech processing," *Wirel. Pers. Commun.*, vol. 125, no. 2, pp. 1913–1949, 2022. doi: [10.1007/s11277-022-09640-y](https://doi.org/10.1007/s11277-022-09640-y).
- [14] S. Wu *et al.*, "Deep learning in clinical natural language processing: A methodical review," *J. Am. Med. Inform. Assoc.*, vol. 27, no. 3, pp. 457–470, 2020. doi: [10.1093/jamia/ocz200](https://doi.org/10.1093/jamia/ocz200).
- [15] M. Soori, B. Arezoo, and R. Dastres, "Artificial Intelligence, machine learning and deep learning in advanced robotics, A review," *Cognitive Robotics*, vol. 3, no. 3, pp. 54–70, 2023. doi: [10.1016/j.cogr.2023.04.001](https://doi.org/10.1016/j.cogr.2023.04.001).
- [16] A. Ferrari, D. Micucci, M. Mobilio, and P. Napolitano, "Deep learning and model personalization in sensor-based human activity recognition," *J. Reliab. Intell. Environ.: Springer*, vol. 9, no. 1, pp. 27–39, 2023. doi: [10.1007/s40860-021-00167-w](https://doi.org/10.1007/s40860-021-00167-w).
- [17] V. Singh, P. Gangsar, R. Porwal, and A. Atulkar, "Artificial intelligence application in fault diagnostics of rotating industrial machines: A state-of-the-art review," *J. Intell. Manuf.*, vol. 34, no. 3, pp. 931–960, 2023. doi: [10.1007/s10845-021-01861-5](https://doi.org/10.1007/s10845-021-01861-5).
- [18] A. Entezari, A. Aslani, R. Zahedi, and Y. Noorollahi, "Artificial Intelligence and machine learning in energy systems: A bibliographic perspective," *Energy Strategy Rev.*, vol. 45, no. 1, pp. 101017, 2023. doi: [10.1016/j.esr.2022.101017](https://doi.org/10.1016/j.esr.2022.101017).
- [19] S. Awan *et al.*, "Profiling casualty severity levels of road accident using weighted majority voting," *Comput. Mater. Contin.*, vol. 71, no. 3, pp. 4609–4626, 2022. doi: [10.32604/cmc.2022.019404](https://doi.org/10.32604/cmc.2022.019404).
- [20] S. Mokhtarimousavi, "A time of day analysis of pedestrian-involved crashes in California: Investigation of injury severity, a logistic regression and machine learning approach using HSIS data," *Inst. Transp. Eng.*, vol. 89, no. 10, pp. 25–33, 2019.

- [21] WEKA, “WekaDeeplearning4j 3.9. Machine learning software in java,” 2023. Accessed: Jan. 01, 2022. [Online]. Available: <https://www.cs.waikato.ac.nz/ml/weka/>
- [22] K. Sattar, F. C. Oughali, K. Assi, N. Ratrou, A. Jamal and S. M. Rahman, “Transparent deep machine learning framework for predicting traffic crash severity,” *Neural. Comput. Appl.*, vol. 35, no. 2, pp. 1535–1547, 2023. doi: [10.1007/s00521-022-07769-2](https://doi.org/10.1007/s00521-022-07769-2).
- [23] M. Rahim, M. Adilur, and H. M. Hassan, “A deep learning based traffic crash severity prediction framework,” *Accid. Anal. Prev.*, vol. 154, no. 1, pp. 106090, 2021. doi: [10.1016/j.aap.2021.106090](https://doi.org/10.1016/j.aap.2021.106090).
- [24] Z. Ma, G. Mei, and S. Cuomo, “An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors,” *Accid. Anal. Prev.*, vol. 160, no. 1, pp. 106322, 2021. doi: [10.1016/j.aap.2021.106322](https://doi.org/10.1016/j.aap.2021.106322).
- [25] K. F. Lee *et al.*, “An intelligent driving assistance system based on lightweight deep learning models,” *IEEE Access*, vol. 10, no. 1, pp. 111888–111900, 2022. doi: [10.1109/ACCESS.2022.3213328](https://doi.org/10.1109/ACCESS.2022.3213328).
- [26] O. G. Canada, *National Collision Database (NCDB): A Database Containing All Police-Reported Motor Vehicle Collisions on Public Roads in Canada*. Transport Canada, 2020.
- [27] D. F. T. UK, *Road Safety Data-Casualties 2019–2010–Provisional Mid Year Unvalidated Data*. UK, 2019.
- [28] S. Moosavi, M. M. Samavatian, S. Parthasarathy, and R. Ramnath, “US Accidents (2016–2023): Country wide car accident dataset,” 2023. Accessed: Jan. 01, 2022. [Online]. Available: <https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents/>
- [29] A. M. Carvalho and R. C. Prati, “DTO-SMOTE: Delaunay tessellation oversampling for imbalanced data sets,” *Information*, vol. 11, no. 12, pp. 557, 2020. doi: [10.3390/info11120557](https://doi.org/10.3390/info11120557).
- [30] S. M. Liu, J. H. Chen, and Z. Liu, “An empirical study of dynamic selection and random under-sampling for the class imbalance problem,” *Expert Syst. Appl.*, vol. 221, no. 1, pp. 119703, 2023. doi: [10.1016/j.eswa.2023.119703](https://doi.org/10.1016/j.eswa.2023.119703).