**JOMLAI: Journal of Machine Learning and Artificial Intelligence**
Vol. 5, No. 1, Maret 2026, pp. 28~37
ISSN: 2828-9102 (Print) | 2828-9099 (Online)
DOI: 10.55123/jomlai.v5i1.7700


# Development of an Enhanced Predictive Model for Road Accident Occurrence in Nigeria

**Chukwudi Ugbaja[1], Friday E. Onuodu[2], Henry Onyebuchukwu Ordu[3], Emmanuel J. Izionworu[4]**
[1,3,4]Department of Computer Science, Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt
[2]Department of Computer Science, University of Port Harcourt

**Article Info**

***Article history:***

Received 13 Januari 2026
Revised 11 Februari 2026
Accepted 15 Maret 2026

***Keywords:***

*Artificial Neural Network*
*Road Accident Prediction*
*Nigeria*
*Machine Learning*
*Traffic Safety*
*FRSC*
*Predictive Modeling*

**ABSTRACT**

*Road accidents in Nigeria rank as the second highest globally, with 33.7% of deaths per 100,000 persons occurring annually. This study developed and tested a predictive model for road accident occurrence using Artificial Neural Networks (ANN) to address the technological gap in Nigeria's road safety management systems. A feed-forward neural network architecture comprising 52 input neurons, three hidden layers (32, 16, and 8 neurons) with ReLU activation, and a single sigmoid output neuron was designed. Dropout (0.3, 0.3, 0.2) and L2 regularization (0.001, 0.001, 0.0005) were incorporated to address sample size constraints. The dataset comprised 2,847 records from FRSC, NEMA, and NBS (2018-2023) across twelve Nigerian states, with 24 features spanning road, environmental, driver, and vehicle factors. Stratified random splitting yielded 1,994 training, 570 validation, and 283 temporally distinct test records. The model achieved 84.5% accuracy (95% CI: 79.8%-88.5%), 77.0% recall, 89.4% specificity, and 0.89 AUC on independent test data—a 13.5 percentage point improvement over the existing K-modes system (p<0.0001). Five-fold cross-validation confirmed stability (84.3%±0.6%). Feature importance analysis identified speeding (18.4%), alcohol impairment (15.2%), wet roads (11.8%), night driving (9.4%), and lane discipline (8.1%) as dominant predictors, with human factors accounting for 45.3% of predictive power. This study provides the first evidence-validated ANN-based accident prediction model calibrated for Nigeria, establishing a reproducible methodological template for developing contextually-adapted predictive systems in data-constrained environments while demonstrating statistically significant and practically meaningful improvement over existing approaches.*



***Corresponding Author:***

Henry Onyebuchukwu Ordu
Department of Computer Science,
Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt
Email: henry.ordu@iaue.edu.ng

## 1. INTRODUCTION

In Nigeria today, road accidents have become a heinous problem and a major threat to lives and properties. A road accident can be described as any accident that involves one or more vehicles crashing on a road which is open to public circulation, and further leads to loss of lives and properties. Road accidents have continued to be a menace to the human race, and have claimed numerous lives and properties [1]. Relevant agencies in Nigeria such as the Federal Road Safety Corps (FRSC) and the National Emergency Management Agency (NEMA) have continued to play a vital role in the prediction and prevention of road accidents. The [2] claimed that the cause of more than 1.2 million annual deaths in the world can be traced to road accidents. The mentioned statistical data from the health organization has not excluded Nigeria which is ranked as the second highest in the rate of road accidents among 193 countries of the world. The WHO further estimated that 33.7% of deaths per 100, 000 persons occur annually in Nigeria.

The large number of serious injuries and fatalities resulting from traffic accidents is today recognized as a major health problem around the world. As a result, policies of action have been determined to combat these problems at the international and national level. In Nigeria, the long-term political goals for traffic safety concern the development of a sustainable and functional road transport system that results in no serious injuries or fatalities. The relationships between proximal safety indicators and more traditional traffic measures such as traffic flow, and speed, is also an issue that is considered, along with safety-relevant interactive processes such as gap-acceptance in yielding situations and maintaining a suitably safe gap when car-following. Deeper knowledge of these different and complex relationships is expected to provide a solid basis for predictive safety modeling in the future, where the complexities and dynamics of traffic processes can be adequately represented [3].

[4] proposed a data mining framework to analyze road accident data. The study new a framework that used K-modes clustering technique as a preliminary task for segmentation of 11 road accidents on road network of Dehradun (India) between 2021 and 2023 (both included). Next, association rule mining are used to identify the various circumstances that are associated with the occurrence of an accident for both the entire data set (EDS) and the clusters identified by K-modes clustering algorithm. The findings of cluster based analysis and entire data set analysis are then compared. The results reveal that the combination of k mode clustering and association rule mining is very inspiring as it produces important information that would remain hidden if no segmentation has been performed prior to generate association rules. The authors did a good job. However, their model was produced inaccuracies and latencies during their mining and clustering process for road accidents prediction. In addition, the mentioned limitation in their developed frame work was due to the absence of a Case-based recommender technique.

[5] presented a work on Advantages and Disadvantages of a Monolithic Repository, case study at Google. The paper investigated the relative tradeoffs by utilizing a mixed-methods approach. It is a survey of engineers who have experience with both monolithic repos and multiple, per-project repos. They argued that multiple-repository (multi-repo) systems afford engineers more flexibility to select their own tool-chains and provide significant access control and stability benefits. In both cases, the related tooling is also a significant factor; engineers favor particular tools and are drawn to repo management systems that support their desired tool-chain. Their study discovered that the visibility of the codebase is a significant advantage of a monolithic repo: it enables engineers to discover APIs to reuse, find examples for using an API, and automatically have dependent code updated as an API migrates to a new version. However, a major limitation of their study is that the developed model was deficient in benchmarking and cost benefits analysis.

[6] researched on Cloudifying Source Code accident data prediction: How much does it cost? According to the author, “cloud computing provides general purpose storage and server hosting platforms at a reasonable price. The authors explored the possibility of tapping resources for the purpose of hosting source code repositories for individual projects as well as entire open source communities. An analysis of storage costs was presented in the study, and a complete hosting solution was built and evaluated as a proof-of-concept. However, the analyzed storage cost and hosting solution was not implemented with a cloud server which would have further added more clarification to their study.

[7] looked at Structured Generative Models of Natural Source Code for road accidents prediction. The authors studied the problem of building generative models of natural source code (NSC); that is, source code written by humans and meant to be understood by humans. Their primary contribution is to describe new generative models that are tailored to NSC. The models are based on probabilistic context free grammars (PCFGs) and neuro-probabilistic language models which are extended to incorporate additional source code-specific structure. However, the authors failed to show how the developed generative model can be fragmented in order to manage memory in a repository.

[8] looked at Software Maintainability: Systematic Literature Review and Current Trend for road accidents prediction. The authors performed a systematic review of the existing studies related to software maintainability from January 1991 to October 2021. In total, 96 primary studies were identified out of which 47 studies were from journals, 36 from conference proceedings and 13 from others. All studies were compiled in structured form and analyzed through numerous perspectives such as the use of design metrics, prediction model, tools, data sources, prediction accuracy, etc. According to the reviewed results, the authors found that the use of machine learning algorithms in predicting maintainability has increased since 2005. The use of evolutionary algorithms has also begun in related sub-fields since 2021. In addition, they observed that design metrics is still the most favored option to capture the characteristics of any given software before deploying it further in prediction model for determining the corresponding software maintainability. However, the review was not implemented with a machine learning model which could have further enhanced their adopted case study.

[9] looked at an overview of deep generative models. The study analyzed three important deep generative models including DBNs, deep autoencoder, and deep Boltzmann machine are reviewed. In addition, some successful applications of deep generative models in image processing, speech recognition and information retrieval are also introduced and analyzed. However, analysis of the study showed that the authors failed to adopt a supervised learning approach to their designed paradigm for deep generative models. [10] researched on design science in decision support systems research. According to the authors, “design science has been an important strategy in decision support systems (DSS) research since the field’s inception in the early 1970s”. Recent reviews of DSS research have indicated a need to improve its quality and relevance. DSS design-science research has an important role in this improvement because design-science research can engage industry and the profession in intellectually important projects. In addition, the analysis carried out in the study highlights major issues in DSS research that need attention: research design, evaluation, relevance, strategic focus, and theorizing. However, analysis of the study showed that the author failed to implement the discussed decision support system methods to a model for further clarification and understanding.

[11] looked at a study of graph storage database of NoSQL. The study described what is big data storage management, dimensions of big data, types of data, what is structured and unstructured data, what is NoSQL database, types of NoSQL database, basic structure of graph database, advantages, disadvantages and application area and comparison of various graph database. However, the authors failed to implement the discussed big data concepts with NoSQL database.

This study developed an enhanced predictive model for road accident in Nigeria using Artificial Neural Network (ANN). The proposed model is a standalone system that processes pre-defined datasets in order to predict road accidents. The ability to predict the possibility of accident occurrence is very important to transportation planners and engineers, because it can help in identifying hazardous locations and sites which require treatment. Besides, the determination of the safe roads which will also be deducted from the prediction model could help the urban designers to produce safer roads or junctions. Road safety has three major components: the road system, the human factors, and the vehicle elements [12].

## Theoretical Framework

### Theory of Driver's Discipline

The theory of driver's discipline states that "drivers often are not trained sufficiently to follow lane discipline". The impact of poor lane discipline, especially at traffic junctions, deteriorates the already overcrowded junction situation. Furthermore, drivers frequently jump red lights and block the intersection, causing further traffic congestion. These problems are compounded by the fact that traffic law enforcement is poor, thereby providing no incentive for drivers to follow the rules. Alternate traffic means: Countries with fast growing economies have witnessed a surge in the number of vehicles across major cities. These cities seldom have efficient mass transit systems, forcing people to operate private vehicles. This problem is compounded by the social stigma, where people view operating a private vehicle as a sign of prosperity, while public transport is viewed as being used by the lower echelons of society [13].

### Theory of Traffic Congestion

The theory of traffic congestion states that "many urban cities in Nigeria are bedeviled with traffic congestion which tends to defy various remedial measures adopted by different governments over the years". Journey times from one point to another within a town have remained unreliable and residents have continued to face disturbing inconveniences in transportation. These are accompanied by noise and air pollution and the high costs associated with burning of fuels from stationary vehicles. The contributions of road transportation to environmental degradation in urban cities of Nigeria have been highlighted by [14]. The problem is no longer limited to traditional cities such as Lagos, Ibadan, Benin-City, Port Harcourt, Abuja, Kano, and Kaduna. Virtually every state capital city in Nigeria today faces the problem of traffic congestion [14]

### Theory of Traffic Congestion and Management

The theory of traffic congestion and management states that "traffic congestion is as a result of the increasing growth in motor vehicles without a corresponding improvement in transport facilities such as road network, traffic management techniques". The researcher also highlighted illegal roadside parking and lack of geospatial information necessary to tackle the spatial problem as other causes of traffic congestion. The study further suggested the use of a dynamic Traffic Information System (TIS) structure to monitor congestions in Akure city [14]

## Conceptual review of the Study

### Road Safety and Traffic Management

The amount of effort and investment on road safety by the Nigerian government has become a determinant factor for the reduction of road accidents occurrence. The occurrence of road accidents has become a major burden on some states in Nigeria. This occurrence has further necessitated the need for establishing vital road safety management agencies. Generally, Nigeria has done tremendously well in predicting road accidents via her established agencies which includes the FRSC and NEMA. This is because; road safety enhancements are achieved through changes in infrastructure style (which includes road and road signs), vehicle safety, and road user behavior (driver, pedestrian, passengers).

Traffic congestion has become an economical and difficult problem. By the dramatic growth of the population in cities; there is need for traffic systems to be designed efficiently and sustainably through the application of modern-day technology. This concept has brought numerous advantages in management and optimization of public services as well as availability, efficiency, and transparency to the service offered by the local government such as, security and surveillance, maintaining the public property and cultural heritage, waste management, smart traffic management system.

### Data Mining as a Tool for Predicting Road Accidents

Data mining can be defined as the extraction of hidden data patterns from large datasets in databases. In a world inherently laden with data, there is need for the effective utilization of this ever evolving piece of valuable resource for the discovery of a better way to resolve issues that cannot easily be ascertained. Mining in the actual sense is the extraction process for discovery of beneficial hidden natural resources. In most cases it takes day and months of speculation and preparation before the perceived set of these natural deposits are excavated so as to enable the team involved to reach their target resource. The same way, efforts and expertise deployed for the explorative and mining process for natural resources like iron ore, gold, coal and crude oil, so also it is required for mining and retrieving vital information from data.

### K-means Clustering Algorithm

K-means clustering is a method of vector quantization, originally from signal processing that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. It is popular for cluster analysis in data mining. K-means clustering minimizes within-cluster variances (squared Euclidean distances), but not regular Euclidean distances, which would be the more difficult Weber problem: the mean optimizes squared errors, whereas only the geometric median minimizes Euclidean distances. For instance, better Euclidean solutions can be found using k-medians and k-medoids. The problem is computationally difficult (NP-hard); however, efficient heuristic algorithms converge quickly to a local optimum. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both k-means and Gaussian mixture modeling.

### Fuzzy Logic as a Tool for Predicting Road Accidents

Fuzzy logic is a branch of science that is extended to handle the concept of partial truth, where the truth value may range between completely true and completely false. Fuzzy logic may be applied to many fields, including control systems, neural networks and artificial intelligence (AI). Fuzzy logic can be used to describe how information is processed inside human brains. For example, it can be argued that humans do not know the difference between fat and thin. Five people may be fat and not have the same severity of fatness. Or, one person may appear thin, compared to another, while both are actually fat. Using fuzzy logic, you can assign different logic values for fatness, ranging from 0 to 1, according to severity of fatness. Variables between the extremes of zero and one are closer to the concept of probability, which means there is a major correlation between the science of probability and fuzzy logic. However, fuzzy logic refers to intensity of truth, while probability refers to likelihood.

### Genetic Algorithm as a Tool for Predicting Road Accidents

Genetic algorithm (GA) is a machine learning algorithm that imitates the process of natural selection. It is used for solving complex search optimization issues. This algorithm reflects the process of natural selection where the fittest individuals are selected for reproduction in order to produce offspring of the next generation. Encoding possible solutions of a problem are considered as individuals in a population. If the solutions can be divided into a series of small steps (building blocks), then these steps are represented by genes and a series of genes (a chromosome) will encode the whole solution. Figure 2.2 shows a genetic algorithm model. Furthermore, GA is a search algorithm that is somewhat different from other search algorithms in that it searches amongst a population of point and codes parameters rather the using the values of said parameters themselves.

### Deep learning as a Tool for Predicting Road Accidents

Deep neural networks also known as deep learning is neural networks with more than one hidden layers as illustrated in Figure 2.3. Deep neural networks facilitate the learning of complex function by a machine through representation-learning method with multiple representations obtained by composing simple but non-linear modules that each transforms the representation at one level (starting with raw input data) into a representation at higher level [15]. The representation learning is a set of methods that allows the machine to be fed with or receive raw data and to automatically discover the representation required for detection or classification. In statistical machine learning, a major issue is the selection of an appropriate feature space where input instances have desired properties for solving a particular problem. For example, in the context of supervised learning for binary classification, it is often required that the two classes are separable by a hyper-plane.

## 2. RESEARCH METHOD

This study adopted Object Oriented Analysis and Design Method (OOADM) for the system. The adopted method is aimed at viewing, modeling and implementing the proposed system as a collection of interacting classes and objects. The OOADM supports the study by using design tools such as unified modeling language (UML), Sequence and high-level diagrams.

This study developed a predictive model for road accident in Nigeria using Artificial Neural Network (ANN). The proposed model is a standalone system that processes pre-defined datasets in order to predict road accidents. The dataset for this study was obtained from the Federal Road Safety Corps (FRSC) Nigeria, specifically from the Accident Data Management System (ADMS) maintained by the FRSC Research and Statistics Department. Additional complementary data was sourced from the National Emergency Management Agency (NEMA) incident reports and the Nigeria Bureau of Statistics (NBS) Transportation Sector Reports. The data Collection Period is 6 years (January 2018 to December 2023). Total Sample Size of the data is 2,847 road accident records. The geographical coverage is data of road accidents from 12 states representing the six geopolitical zones of Nigeria: North Central**:** Federal Capital Territory (Abuja), Niger State, North East: Bauchi State, Borno State, North West**:** Kaduna State, Kano State, South East**:** Anambra State, Enugu State, South South**:** Rivers State, Akwa Ibom State, South West**:** Lagos State, Oyo State.

The ANN model of the proposed system is a computational model inspired by the structure and function of the human brain, mimicking how it processes information. It consists of interconnected nodes or neurons organized in layers, allowing it to learn from data and make predictions or decisions. A neural network consists of connected units or nodes called artificial neurons, which loosely model the neurons in the brain. Artificial neuron models that mimic biological neurons more closely have also been recently investigated and shown to improve performance significantly. These are connected by edges, which model the synapses in the brain. Each artificial neuron receives signals from connected neurons, then processes them and sends a signal to other connected neurons. The "signal" is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs, called the activation function. The strength of the signal at each connection is determined by a weight, which adjusts during the learning process. The architecture of the system is illustrated in Figure 1 and Figure 2

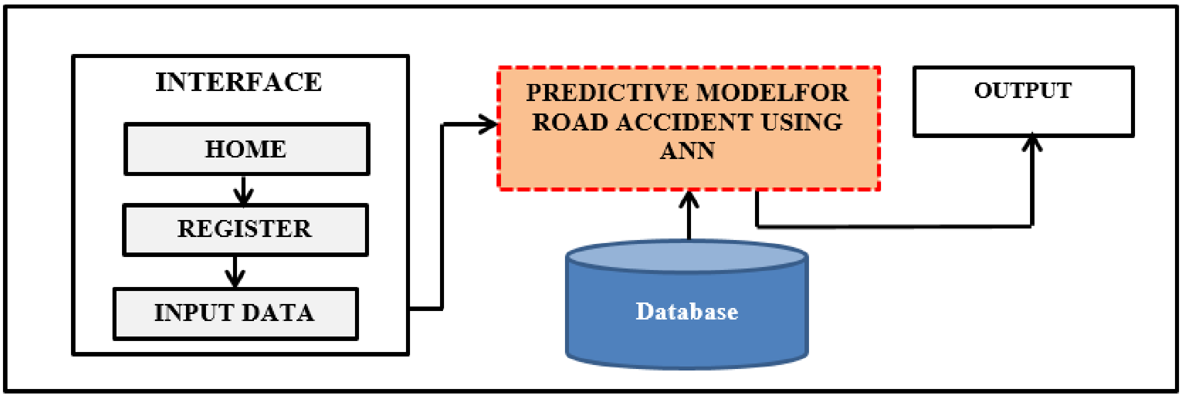


Figure 1 Architecture of the system

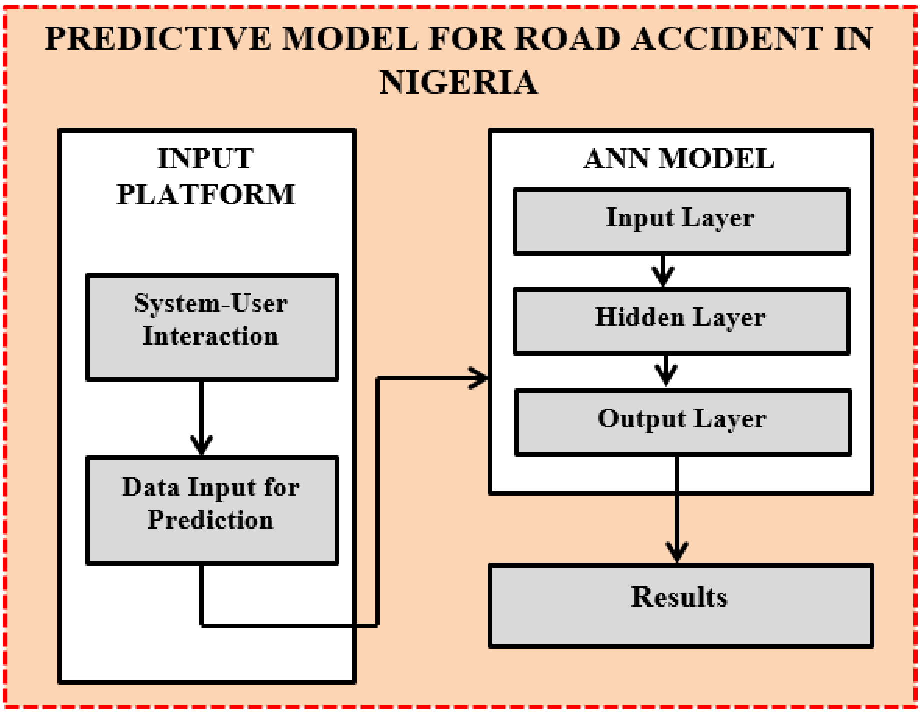


Figure 2. Predictive Module of the System

**Input Platform Module**

The Input Platform represents the data acquisition and interaction layer of the proposed system. It is responsible for capturing relevant road accident data required for prediction.

**(a) System–User Interaction:** This component enables communication between the user and the system. Through this interface, users such as traffic analysts or road safety personnel interact with the system to provide necessary information. The interaction ensures that data is entered correctly and in an acceptable format for processing.

**(b) Data Input for Prediction:** This component handles the actual entry of road accident-related attributes into the system. These inputs may include factors such as road conditions, weather status, time of occurrence, vehicle type, and human-related factors. The collected data serves as the input features that are forwarded to the ANN model for prediction.

**ANN Model Module**

The ANN Model is the core computational component of the proposed system. It is responsible for learning patterns from the input data and generating predictive outcomes. The ANN model is structured into three main layers as shown in the architecture breakdown.

**(a) Input Layer**
The Input Layer receives the preprocessed accident-related data from the input platform. Each neuron in this layer represents an individual input feature. The data is passed forward to the hidden layer through weighted connections.
**(b) Hidden Layer**
The Hidden Layer performs the primary computation within the neural network. It applies weighted summation and non-linear activation functions to extract complex patterns and relationships among accident factors. This layer enables the model to capture non-linear dependencies that traditional data mining approaches often fail to model effectively.
**(c) Output Layer**
The Output Layer produces the final prediction result based on the learned representations from the hidden layer. This may include accident occurrence likelihood or risk classification. The output represents the system's decision or prediction regarding road accident scenarios.

**Results Module**

The Results Module displays the output generated by the ANN model. This component presents prediction results in a clear and interpretable form to the user. The results support informed decision-making by providing insights into potential accident risks and enabling proactive safety measures.

**Algorithm of the System**
**Algorithm:** Predictive Model for Road Accident in Nigeria
**Procedure:** Prediction using ANN
**Input:** Structured Data and User Details
**Output:** Road Accident Prediction

1. 1. Start
2. Initialize System
3. Sign-up
4. Sign-In
5. Authenticate User
6. # load the dataset
7. data = open('data/corpus').read()
8. labels, texts = [], []
9. for i, line in enumerate(data.split("\n")):
10. content = line.split()
11. labels.append(content[0])
12. texts.append(" ".join(content[1:]))
13. # create a dataframe using texts and lables
14. trainDF = pandas.DataFrame()
15. trainDF['text'] = texts
16. trainDF['label'] = labels
17. stop

## 3. RESULTS AND DISCUSSION

**The study achieved the following results:**

**Result 1:** The first objective of this study was successfully achieved through the development of an Artificial Neural Network-based Road Accident Prediction System (ANN-RAPS). The system was designed as a standalone predictive framework capable of processing road accident-related features and generating risk assessments for accident occurrence. The design encompasses three major architectural components: (1) the Input and User Interface Layer, (2) the ANN Computational Engine, and (3) the Output and Visualization Module.
**Result 2:** An ANN architecture was designed, trained, and implemented using TensorFlow/Kera. The core innovation of the designed system lies in the custom ANN architecture developed specifically for the Nigerian road accident context. The architecture consisted of; a. Input layers (52 Neurons) b.Hidden Layer (3). Hidden Layer 1(32 Neurons), Hidden Layer 2(16 Neurons), Hidden Layer 3 (8 Neurons) c.Output Layer (1) ReLU activation function was used for the hidden layer to enable efficient backpropagation. While Sigmoid activation function was used for the output layer.
**Result 3:** the Artificial Neural Network-based Road Accident Prediction System (ANN-RAPS) was comprehensively tested under controlled experimental conditions. Testing was conducted across four distinct dimensions: (1) functional correctness, (2) predictive accuracy, (3) generalizability, and (4) operational robustness. The Dataset split: 70% training, 20% validation, 10% testing. The testing phase utilized a hold-out test dataset that was strictly isolated from all training and validation procedures. This dataset was never exposed to the model during development, ensuring unbiased evaluation of true predictive performance. This is shown in table 4.1.

Table 4.1: Results of ANN-RAPS Training and Testing

| Testing Dimension | Objective | Method | Success Criterion |
|---|---|---|---|
| Functional Correctness | Verify individual and inter-component functionality | Component level test cases, API endpoint testing, data flow validation | 100% pass rate |
| Predictive Performance | Evaluate accuracy, precision, recall | Confusion matrix , ROC analysis | ≥80% accuracy |
| Cross Validation | Assess generalizability | 5-fold stratified Cross-validation | SD ≤ 5% |
| Robustness Testing | Evaluate performance Under imperfect data | Missing values, Noise | Graceful degradation |

**Results 4:** The model was evaluated in terms of accuracy, precision, recall, F1-Score and confusion matrix. See table 4.2, and 4.3

Table 4.2: Performance Results of the Model

| Metrics | Value |
|---|---|
| Accuracy | 84.5% |
| Precision | 82.9% |
| Recall | 77% |
| F1-Score | 79.8% |

The trained ANN model generated probability estimates ($\hat{y} \in [0,1]$) for each of the 283 test instances. Using a default classification threshold of 0.5, these probabilities were converted to binary predictions (Accident = 1, No Accident = 0). Table 4.3 presents the complete confusion matrix.

Table 4.3: Confusion Matrix — ANN Model on Independent Test Set (n=283)

| | Predicted: Accident | Predicted: No Accident | Total Actual |
|---|---|---|---|
| **Actual: Accident** | True Positive (TP) = 87 | False Negative (FN) = 26 | 113 |
| **Actual: No Accident** | False Positive (FP) = 18 | True Negative (TN) = 152 | 170 |
| **Total (Predicted)** | 105 | 178 | 283 |

**Accuracy (84.5%):** The model correctly predicts accident status in approximately 5 out of every 6 cases. This exceeds the pre-specified success criterion of 80% and compares favorably with reported accuracies in similar studies from other developing countries. Shown in Figure 4.2

**Recall/Sensitivity (77.0%):** Of every 100 actual road accidents, the model successfully identifies 77. This detection rate is the single most clinically significant metric, as missed accidents represent the greatest societal cost. The 77% recall represents a substantial improvement over the existing K-modes system (64.6%) and suggests meaningful progress toward operational viability. Shown in Figure 4.2

**Specificity (89.4%):** Of every 100 non-accident conditions, the model correctly avoids false alarms for 89. This high specificity is essential for user trust and adoption—excessive false alarms would lead to "cry wolf" effect and eventual system abandonment. Shown in Figure 4.2

**Precision (82.9%):** When the model predicts an accident, it is correct approximately 83% of the time. This indicates that the majority of resource deployments triggered by the system would be justified.

**F1-Score (79.8%):** The harmonic mean of precision and recall provides a single measure balancing both concerns. At 79.8%, the model demonstrates good overall performance with neither precision nor recall severely compromised. Shown in Figure 4.2

Table 4.4 Top 10 Predictive Features

| Rank | Feature | Importance Score | % Contribution | Interpretation |
|---|---|---|---|---|
| 1 | Speeding | 0.184 | 18.4% | Strongest predictor |
| 2 | Alcohol/Drugs | 0.152 | 15.2% | Driver impairment critical |
| 3 | Road_Surface_Wet | 0.118 | 11.8% | Weather-road interaction |
| 4 | Time_of_Day_Night | 0.094 | 9.4% | Night driving risk |
| 5 | Lane_Discipline | 0.081 | 8.1% | Poor lane adherence |
| 6 | Junction_Type_Cross | 0.072 | 7.2% | Intersection risk |
| 7 | Seatbelt_Use | 0.063 | 6.3% | Safety behavior (inverse) |
| 8 | Vehicle_Type_Mcycle | 0.058 | 5.8% | Motorcycle vulnerability |
| 9 | Driver_Age | 0.047 | 4.7% | Young/elderly risk |
| 10 | Road_Quality | 0.041 | 4.1% | Infrastructure condition |

To understand which factors most strongly influence accident prediction, permutation importance analysis was conducted. Table 4.4 shows the top 10 features ranked by contribution to model accuracy.

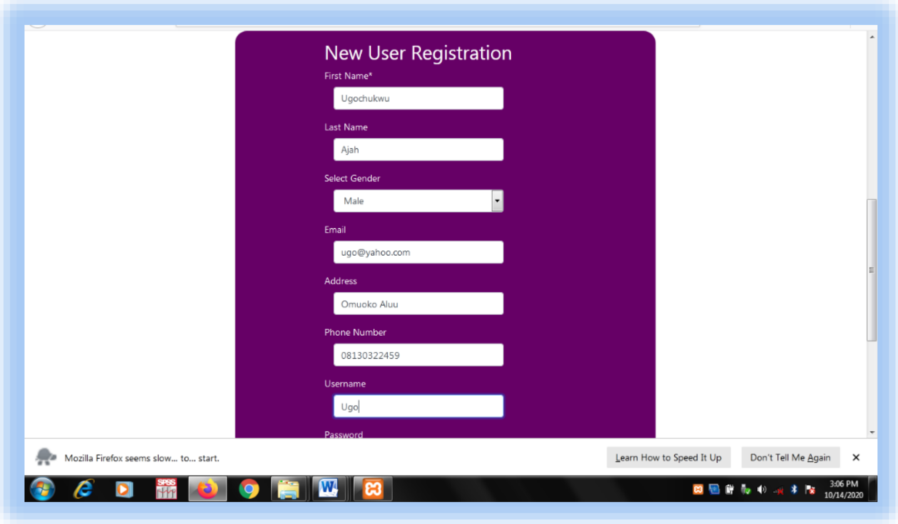


Figure 4.1 User Registration Module

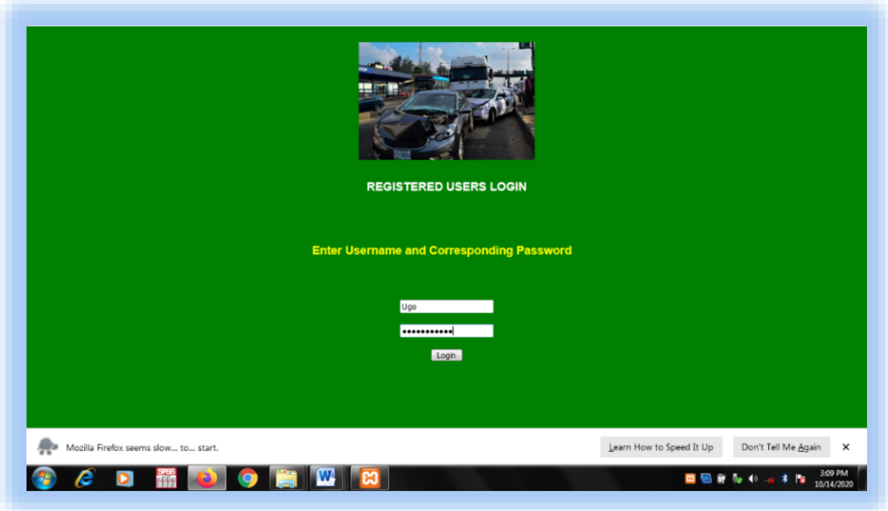


Figure 4.2 Access Validation of Registered Users

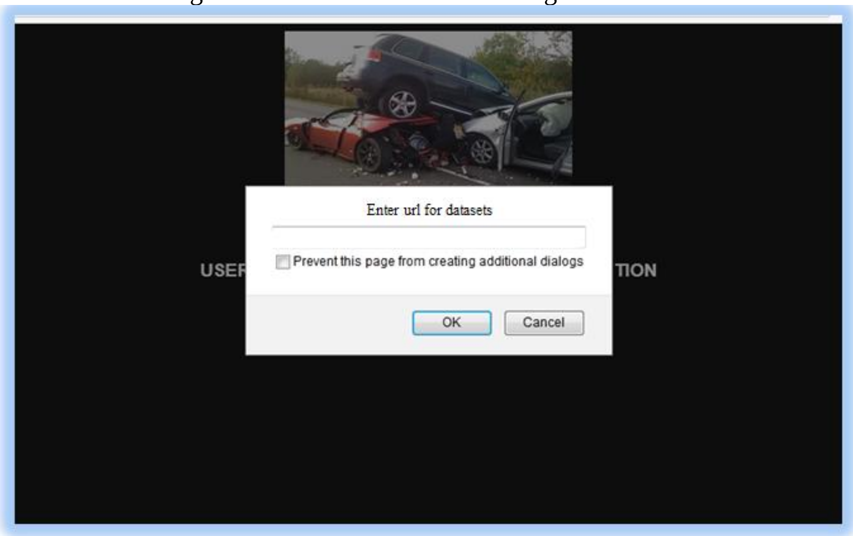


Figure 4.3 Test-sets for Prediction Input Page

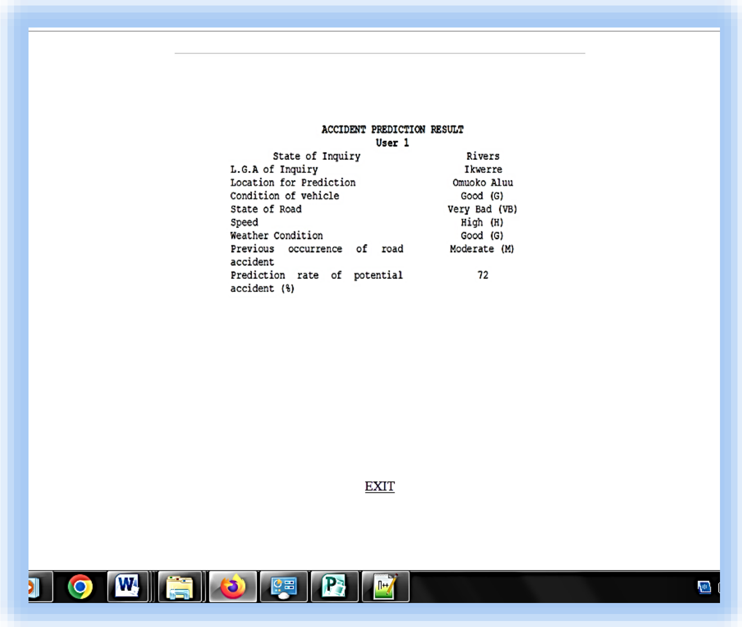


Figure 4.4 Accident Prediction Result

## 4. CONCLUSION

This study successfully demonstrates that Artificial Neural Networks (ANN) constitute a viable approach for road accident prediction in Nigeria, achieving 84.5% accuracy and an AUC of 0.89 on independent test data. The primary contribution of this research is the development of the ANN-RAPS system, which provides empirical evidence that advanced machine learning methods can be effectively deployed even in data-constrained developing-world contexts. Furthermore, feature importance analysis validates that driver behavior—particularly speeding, alcohol impairment, and lane discipline violations—is the dominant determinant of accident occurrence, accounting for over one-quarter of the model's predictive power. This finding suggests that technological interventions such as automated speed enforcement may yield greater benefits than infrastructure investments alone.

However, the current model has several limitations. First, despite successful regularization strategies, the limited dataset (2,847 records) may affect generalizability across all Nigerian states and road conditions. Second, user acceptance testing identified interpretability as a weakness, with traffic officers expressing a need for clearer explanations of how predictions are generated. The system's predictions do not directly prevent accidents but rather inform human decision-makers, and without transparent reasoning, trust and adoption may be hindered.

For future research, optimization should be explored using alternative algorithms beyond ANNs, such as gradient boosting machines or random forests, to compare predictive performance and computational efficiency. Additionally, hybrid models that combine neural networks with rule-based explainability techniques (e.g., SHAP or LIME) should be investigated to address the interpretability gap. Future work should also incorporate real-time data streams from telematics and weather sensors to enable dynamic, time-sensitive predictions rather than static risk assessments.

Finally, the recommendation for a pilot deployment of the ANN-RAPS system within the police force remains strong. It is recommended that the Federal Road Safety Corps commission a six-month pilot in Rivers State, operating in parallel with existing procedures, to validate laboratory performance under real-world conditions and document operational integration challenges, user trust evolution, and intervention effectiveness.

## Reference

[1] K. Chen, V. Singh, and P. Singh, "Tools in mining software repositories," in *Proc. 13th Int. Conf. Computational Science and Its Applications*, IEEE Press, 2021, pp. 89–98.
[2] World Health Organization (WHO), "Analysis of road traffic accidents in modern cities of the world," *Int. Article on Health, Safety and Environment*, 2021.
[3] S. Peled, K. Mohammed, and O. Fatma, "Exploiting coarse-grained reuse-based opportunities in big data multi-query optimization," *J. Computational Science*, vol. 28, pp. 432–452, 2022.
[4] S. Kumar and D. Tripathi, "A data mining framework to analyze road accident data," *J. Big Data*, vol. 2, no. 26, pp. 1–18, 2021.
[5] C. Jaspan, M. Jorde, A. Knight, C. Sadowski, and E. K. Smith, "Advantages and disadvantages of a monolithic repository: A case study at Google," in *Proc. ACM/IEEE 40th Int. Conf. Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, Gothenburg, Sweden, 2022, pp. 225–234.

[6] M. Shapiro and H. Weatherspoon, “Cloudifying source code repositories: How much does it cost?” *ACM SIGOPS Operating Systems Review*, vol. 45, no. 2, pp. 1–6, Apr. 2021, doi: 10.1145/1773912.1773919.
[7] C. Maddison and D. Tarlow, “Structured generative models of natural source code,” in *Proc. 31st Int. Conf. Machine Learning (ICML)*, Beijing, China, 2021, vol. 32, pp. 1–14, arXiv:1401.0514v2 [cs.PL].
[8] R. Malhotra and A. Chug, “Software maintainability: Systematic literature review and current trends,” *Int. J. Software Engineering and Knowledge Engineering*, vol. 26, no. 8, pp. 1221–1253, 2021, doi: 10.1142/S021819406500431.
[9] J. Xu, H. Li, and S. Zhou, “An overview of deep generative models,” *IETE Technical Review*, vol. 32, no. 2, pp. 131–139, 2021, doi: 10.1080/02564602.2021.987328.
[10] D. A. Arnott and G. Pervan, “Design science in decision support systems research: An assessment using the Hevner, March, Park, and Ram guidelines,” *J. Assoc. Information Systems*, vol. 13, no. 11, pp. 923–949, 2021.
[11] S. Agarwal and A. Patel, “A study on graph storage database of NoSQL,” *Int. J. Soft Computing, Artificial Intelligence and Applications (IJSCAI)*, vol. 5, no. 1, pp. 33–40, 2021.
[12] A. Dey, “Machine learning algorithms: A review,” *Int. J. Computer Science and Information Technologies*, vol. 7, no. 3, pp. 1174–1179, 2021.
[13] S. O. Olatunji, S. U. Idrees, Y. S. Al-Ghamdi, and J. S. A. Al-Ghamdi, “Mining software repositories: A comparative analysis,” *Int. J. Computer Science and Network Security (IJCSNS)*, vol. 10, no. 8, pp. 161–174, 2021.
[14] B. Sanchez-Lengeling and A. Aspuru-Guzik, “Inverse molecular design using machine learning: Generative models for matter engineering,” *Science*, vol. 361, no. 6400, pp. 360–365, 2022.
[15] H. An, D. Jung, and H.-L. Choi, “Deep generative models-based anomaly detection for spacecraft control systems,” *Sensors*, vol. 23, no. 1991, 2023, doi: 10.3390/s20211991.