



Forecasting road traffic accident using deep artificial neural network approach in case of Oromia Special Zone

Kannaiya Raja¹ · Karthikeyan Kaliyaperumal² · L. Velmurugan¹ · Subash Thanappan³

Accepted: 2 March 2023 / Published online: 16 March 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Millions of people are dying, and billions of properties are damaged by road traffic accidents each year worldwide. In the case of our country Ethiopia, the effect of traffic accidents is even more by causing injuries, death, and property damage. Forecasting road traffic accident and predicting the severity of road traffic accident contributes a role indirectly in reducing road traffic accidents. This study deals with forecasting the number of accident and prediction of the severity of an accident in the Oromia Special Zone using deep artificial neural network models. Around 6170 road traffic accidents data are collected from Oromia Police Commission Excel data and Oromia Special Zone Traffic Police Department; the dataset consists of accidents in the Special Zone of Oromia Districts (Woredas) from 2005 to 2012 with 15 accidents attributes. 5928 or (80%) of the dataset was used for the training model, and 1482 or (20%) of the dataset was used for the testing model. This study proposed six different neural network architectures such as backpropagation neural network (BPNN), feedforward neural network (FFNN), multilayer perceptron neural network (MLPNN), recurrent neural networks (RNN), radial basis function neural network (RBFNN) and long short-term memory (LSTM) models for accident severity prediction and the LSTM model for a time serious forecasting of number accidents within specified years. The models will take input data, classify accidents, and predict the severity of an accident. Accident predictor GUI has been created using Python Tkinter library for easy accident severity prediction. According to the model performance results, RNN model showed the best prediction accuracy of 97.18%, whereas MLP, LTSM, RBFNN, FFNN, and BPNN models showed the accuracy of 97.13%, 91.00%, 87.00%, 80.56%, and 77.26%, respectively. LTSM model forecasted accident for three years which is 3555 where the actual accident number is 3561. The prediction and forecast result obtained from the model will be helpful in planning and management of road traffic accidents.

Keywords Accidents · Severity · Forecasting · Artificial neural networks · Backpropagation neural network (BPNN) · Feedforward neural network (FFNN) · Multilayer perceptron neural network (MLPNN) · Recurrent neural networks (RNN) · Radial basis function neural network (RBFNN) · Long short-term memory (LSTM) models

1 Introduction

Globally, the rate of is urbanization increasing at an alarming rate. A high urbanization rate in combination with the intense desire for car ownership causes rapid growth of motorization. The main purpose of transportation is to provide efficient movement from one place to another place; there have been many vehicles on the road such as cars, motorcycles, buses, and others that help to move from one destination to another. Such movements face many challenges, which are mainly caused by rapid urbanization and an increase in car ownership, which then influence both the flow of traffic and the environment.

✉ Kannaiya Raja
kannaiya55raja@gmail.com

¹ Department of Computer Science, School of Informatics and Electrical Engineering, IoT-HH Campus, Ambo University, Ambo, Ethiopia

² Department of Information Technology, School of Informatics and Electrical Engineering, IoT-HH Campus, Ambo University, Ambo, Ethiopia

³ Department of Civil Engineering, School of Civil and Environmental Engineering, IoT-HH Campus, Ambo University, Ambo, Ethiopia

Before cars were invented, road traffic injuries occurred involving people, animals, and carts. The numbers grew exponentially as cars, buses, and other motor vehicles were invented and became more common. A cyclist in New York City was the first recorded injury case on 30 May 1896, and a Pedestrian in London was the first recorded motor vehicle death on 17 August of the same year (World's first road death 2003).

The total of road traffic deaths had reached an estimated 25 million by 1997. In 2002, an estimated 1.18 million people died from road traffic accidents; on average, 3242 people are dying per day. Road traffic accident accounted 2.1% of all deaths; this makes them the 11th leading cause of global deaths.

Road traffic injuries place a heavy burden on household finances, not only on global and national economies. Many families are driven deeper into poverty by the loss of breadwinners and the added burden of caring for members disabled by road traffic injuries.

The global status report on road safety, launched by the World Health Organization (WHO) in December 2018, highlights that the number of road traffic deaths has reached 1.35 million annually (World Health Organization 2018). And road traffic crashes cost most countries 3% of their gross domestic product. It is supposed that 30–50 million people are exposed to physical disability annually by traffic accidents in the world according to the World Health Organization (WHO) reported in 2012. This report included that more than 600-billion-dollar property can be damaged by traffic accidents annually.

Road traffic injuries are currently estimated to be the 9th leading cause of death across all age groups globally and are predicted to become the 7th leading cause of death by 2030. Following current trends, about 2 million people could be expected to be killed in vehicle accidents each year. Road traffic deaths occur among young males under the age of 25 years who are almost three times as young females (Global Health Estimates 2014).

In a middle- and low-income countries, they particularly affect active age group, or those who contribute to family, society, and the workforce in general. Many families are susceptible poverty by the expenses of prolonged medical care, or the added burden of caring for a family who is disabled from a road traffic accident. Different documents verify that the traffic accident is in 7th place among different causes of death in low-income countries in 2019 ().

Every day, about 3700 people are died globally in vehicle accidents. More than 50% of people who are killed are pedestrians and motorcyclists. Accidents are estimated to be the 8th leading cause of death globally for all age groups and the leading cause of death for children and young people five to twenty-nine (5–29) years of age (World Health Organization 2018). More people now are

dying in traffic accident than from HIV/AIDS. It is estimated that accident will cost the world economy approximately 1.8 trillion dollars from (2015–2030). That's equivalent to a yearly tax of 0.12% on world GDP (Chen et al. 2019).

The WHO Global Burden of Disease study predicts road traffic injuries from 1990 to 2020 (The Global Burden of Disease 1990).

- Road traffic injuries will rise in rank to 6th place as a major cause of death worldwide.
- Road traffic deaths will increase worldwide, from 0.99 million to 2.34 million (Representing 3.4% of all deaths).
- Road traffic deaths will increase on average by over 80% in low-income and middle-income countries and decline by almost 30% in high-income countries.
- Road traffic injuries will become the second leading cause of disability-adjusted life years (DALYs) lost for low-income and middle-income countries
- In high-income countries, the annual number of road traffic deaths is projected to decrease by 27% from 2000 to 2020

More than 90% of road traffic deaths occur in middle and low-income countries. Even though these countries have 60% of the world's vehicles. In developing countries, this situation is made worse by unplanned and rapid urbanization. The inadequate infrastructure in cities, together with the lack of a legal frameworks, make the exponential rise in the number of road accidents. The African continent has the highest road death rates of all the world's regions that is (26.6 per 100,000) population relative to the global rate of (17.5 per 100,000) population. The continent owns only 2% of the world's vehicles; it contributes 16% to worldwide deaths. The region will continue to have the highest road traffic death rates due to the high rate of urbanization and motorization but lagging road infrastructural development as well as poor road and vehicles' safety (World Report on Road Traffic Injury Prevention 2015; AFRO factsheet 2013).

The rates of death and serious injury are many times higher in car-to-truck collisions than in car-to-car collisions (Mackay and Wodzin 2002). Other problems are failure of the passengers use a protective material, lack of features to stop passengers from being ejected from the car and lack of other safety features, such as high-mounted stop lamps in the rear. Pedestrians involved in accidents more often sustain multiple injuries resulting in death or disability than do car occupants.

In our country Ethiopia, the number of deaths due to traffic accidents is reported to be among the highest in the world. According to the WHO, in 2013 the road accident fatality rate in Ethiopia was 4984.3 deaths per 100,000

vehicles annually, compared to 574 across sub-Saharan African countries. Ethiopia ranks 24th in the world. The number of people injured or killed in one crash is about 30 times higher than crashes in the United States (Mackay and Wodzin 2002). According to the WHO statistics report in 2015, the cause-specific mortality due to accidents in Ethiopia is 94 per 100,000 population which is more than 8% of all deaths. Ethiopia is among the countries experiencing the highest road traffic accidents in the world with a 79% fatality rate of injuries related to traffic accidents (Global Status Report and on Road Safety 2015).

The Sustainable Development Goals (SDGs) target to reduce road traffic deaths by 50% in 2020 (Loewe and Rippin 2015). The Ethiopian Health Sector Development Plan gives attention to an accident and aims to reduce fatalities related to it. In a country like Ethiopia where the majority of the deaths is occurring among the economically active population (MoH 2015).

Estimates of the magnitude of road traffic incidents and traffic injuries in low-income countries are primarily obtained from police records and sometimes hospital registry data. In Ethiopia, traffic police reports are the source of data for road-related incidents. Though death related to road traffic accident is one of the major health challenges in Ethiopia, it is highly underreported (Abegaz et al. 2014).

According to a review undertaken by the Economic Commission for Africa and the Secretariat of the UN Secretary-General's Special Envoy for Road Safety, Mr. Jean Todt, in collaboration with the Ministry of Transport of Ethiopia, in our country Ethiopia on average, about 13 people die in a road traffic accident each day, according to a report. About 85% of all road traffic accidents occur in Oromia, Amhara, and SNNP regional states as well as Addis Ababa city administration (Economic Commission for Africa (ECA) xxxx).

According to WHO data published in 2017, road traffic accident deaths in Ethiopia reached 27,140 or 4.27% of total deaths. The age-adjusted death rate is 36.36 per 100,000 population ranking Ethiopia 22nd in the world (Yohannis 2019).

Factors contributing to the high incidence of RTAs in Ethiopia include drivers' and pedestrians' lack of awareness, poor technical inspections of vehicles, older vehicles, driving license problems, weather conditions, failure to enforce traffic laws, and poor road quality (Mackay and Wodzin 2002).

Oromia is the largest region among the regions in Ethiopia. Oromia is the most populated region and highest road traffic movement next to Addis Ababa (Finfinnee), which is the capital city of Ethiopia and Oromia regional State, accounting for 58% of all fatalities and two-third of all injuries. Many cross-country roads from another region cross this region to enter Addis Ababa city, and the region

is mostly subject to traffic accidents due to this congested traffic flow.

The Oromia region has 22 zonal administrations, special zone of Oromia is among the zones in Oromia. This special zone includes Woredas like Mulo, Sululta, Sabeta Hawas, Barak, Chancho, Sendafa, Welmera, Akaki, Burayu, Holeta, and Sebeta Town. Since this zone surrounds the capital city Addis Ababa, the traffic flow is very overcrowded and many traffic accidents are frequently occurring in this area. The number of registered vehicles is increasing from time to time, due to this increase in the number of road traffic the road accidents rate is increasing.

Many factors may cause accidents including geographical factors such as road conditions, weather conditions and human factors such as age, gender, alcohol, or drugs, and also vehicle technical problems such as vehicle age and body type. Therefore, traffic accident prediction and forecasting are recommended to reduce road accidents in future.

Traffic accident prediction and forecasting involves a complex relationship between variables or parameters of road traffic accidents; the advantage of artificial neural networks (ANNs) over machine learning algorithms depends on their ability to solve complex and non-algorithmic problems. ANN learning is a good tool in solving prediction and classification problems, and ANN allows the inclusion of a large number of variables. Neural networks can relate input with output and are also error-tolerant. Therefore, the primary interest of this study is to analyze the current road traffic accidents, predict the accident severity depending on the input parameters, and forecast accident numbers using different artificial neural network (ANN) models.

1.1 Statement of the problem

Economic development is strongly related to the availability of transportation. The increase in the number of vehicles created a major social problem through traffic accidents. A traffic accident has a major impact on a country's economic growth by destroying infrastructures, harming and killing youths who are working-age groups. However, the technique or process of analyzing traffic accidents is poor due to the manual way of accident monitoring and controlling mechanisms. Factors contributing to the high incidence of RTAs in Ethiopia include rampant reckless driving behaviors, poor road network, substandard road conditions, failure to enforce traffic laws, and poor conditions of vehicles (Road traffic accidents in Ethiopia 2008). Many factors can contribute to traffic accidents, but only certain factors can appear on certain roads according to statistics from Oromia Region Police Commission and Oromia Special Zone.

In general, the major factors of road traffic accident in Oromia Special Zone are:

- Driver Faults
- Poor quality of roads, signals invisibility
- Lack of lane for pedestrian and bicycles
- Lack of knowledge of road safety
- Poor vehicle maintenance tradition
- Animal drawn carts and animals frequently using on main highways
- Lack of safety-conscious design and planning of road network
- Disrespect for traffic rules and regulations

The problems above made me go for research to predict severity of traffic accident and forecast the number of accidents using machine learning which indirectly contribute a role in reducing road traffic accident. It is expected that neural networks will be effective in correctly classifying and forecasting the accident severity.

1.2 Objectives

1.2.1 General objective

The main objective of this study is to forecast road traffic accidents and predict the severity of accidents in Oromia Special Zone using ANN-based models.

1.2.2 Specific objectives

- To collect data from the Oromia Police Office in Addis Ababa and preprocess it with feature extraction and feature selection using Embedded Model
- To build artificial neural network (ANN) models such as BPNN, FFNN, MLPNN, RNN, RBFNN, and LSTM for accident severity prediction.
- To compare the performance of BPNN, FFNN, MLPNN, RNN, RBFNN, and LSTM models and identify the best model which is suitable for accident severity prediction
- To construct the forecasting models using LSTM model.
- To develop an interactive Graphical User Interface prototype (GUI) for the accident severity prediction and an email notification system to notify the stakeholders about the future number of accidents.

2 Literature review

2.1 Introduction

This subsection includes road traffic accident in Ethiopia level, common factors of road traffic accidents, machine learning applications, and review of local and global researches that were done on road traffic accident predictions (Tables 1, 2, 3, 4).

2.2 Traffic accident in Ethiopia

In Ethiopia road traffic accident is a common public health problem and the most common causes of death and injury of human; according to World Health Organization (WHO) report, Ethiopia is considered one of the worst countries in the world where road traffic accident kills and injures a large number of road users; every year nearly 2000 people die due to road traffic accidents where 48% are pedestrians, 45% passengers and 7% drivers, and over 400 to 500 million ETB are lost yearly, as a result of road traffic accident (Amdeslasie et al. 2016).

According to a report of the Ministry of Transport National Road Safety Council, pedestrians account 51% and passengers account 46% of fatality and of the total traffic accidents; 93% of the cases were associated with human factors, 5% accounted for vehicle factors, and 2% were associated with road-related problems. Similarly, the Ethiopian Federal Police report of 2011 stated that the road accident level of the country is one of the worst in the world, as expressed by per 10,000 vehicles. The average traffic accident record from 2003 to 2007 of the Federal Police Commission shows that, of the total fatal accidents, 76% were caused due to drivers' error, 6% due to vehicle technical problems, 5% due to pedestrian error, 2% due to road defects and 12% due to other factors. This indicates that drivers' error has been contributing to the majority of road accidents in Ethiopia. The report also shows that 82% of the accidents occur on straight and level roads with a good surface, while 83% of all the accidents occurred on good asphalt roads (Minister of Transport 2011). As per the same report, the underlying reasons for road traffic accidents were:

- a. Improper behavior or low skill of drivers, which is associated with:
 - Drivers are not respecting the pedestrian priority
 - Overspeeding
 - Overloading of passenger
- b. Vehicle's defect (technique problem)
- c. Animals and carts using the highways
- d. Pedestrians not taking proper precautions

Table 1 Total road accident in Oromia region from 2011 to 2012 EC

| Year | Types of accidents | | | | Property damage estimation |
|-----------------|--------------------|----------------|---------------|-------------|----------------------------|
| | Fatality | Serious injury | Slight injury | Damage only | |
| 2011 | 1968 | 1860 | 2278 | 2041 | 234,492,016 |
| 2012 | 1728 | 1574 | 2358 | 3975 | 256,561,482 |
| Total accidents | 3696 | 3434 | 4636 | 6016 | 491,053,498 |

Table 2 Total road accident in Oromia Special Zone Woredas from 2011 to 2012 EC

| Year | Types of accidents | | | | Property damage estimation |
|-----------------|--------------------|---------------|---------------|-------------|----------------------------|
| | Fatality | Series injury | Slight injury | Damage only | |
| 2011 | 51 | 102 | 447 | 182 | 18,580,900.00 |
| 2012 | 44 | 94 | 504 | 157 | 17,067,070.00 |
| Total accidents | 95 | 196 | 951 | 339 | 35,647,970.00 |

Table 3 Dataset parameters/ attributes and their descriptions

| S. No | Parameter name | Data type | Description |
|-------|-------------------|-------------|--|
| 1 | District | Categorical | District at which accident has occurred |
| 2 | Area | Categorical | Specific location at which accident has occurred |
| 3 | Accident date | Date | Date of accident |
| 4 | Time | Time | Exact hour at which the accident has occurred |
| 5 | Time of accident | Categorical | (Morning, afternoon, evening, night) |
| 6 | Cause of accident | Categorical | The reason behind the accident |
| 7 | Road type | Categorical | Type of road (asphalt, gravel) |
| 8 | Road surface | Categorical | Road condition during the collision (wet, dry) |
| 9 | Light condition | Categorical | Light condition on the road (daylight, dark) |
| 10 | Weather condition | Categorical | Weather condition (smoke, rainy, fine) |
| 11 | Casualty | Categorical | Person injured (driver, passenger, pedestrian) |

Table 4 Data resampling results

| Accident severity class | Data before resampling | After SMOTE | After Random Under Sampler | After Random Over Sampler |
|-------------------------|------------------------|-------------|----------------------------|---------------------------|
| Slight | 3122 | 3122 | 539 | 3122 |
| Serious | 2470 | 2470 | 539 | 3122 |
| Fatal | 539 | 3122 | 539 | 3122 |

- e. Poor traffic law enforcement
- f. Low safety consideration during roads developments

Ethiopia is one of the 50 countries with the deadliest roads in the world according to a WHO report in 2015. Principally, injured people have occupied 30 to 70 percent of orthopedic beds in developing countries' hospitals. Road traffic accident-related causalities are very high in Ethiopia. This makes an urgent need for bringing road safety to the country's public health agenda. Pedestrian fatality in Ethiopia is common and high. Accordingly, about 74% of the fatal, serious, and slight injury accidents happened when pedestrians tried to crossroads (Abagaz 2018).

2.3 Paradigms of machine learning

2.3.1 Supervised learning

Supervised learning is machine learning in which the network is trained (learn a function from training data) by providing it with input and matching output patterns. These input–output pairs can be provided by the system or an external teacher which contains the network self-supervised. There are two kinds of supervised learning tasks: regression and classification. Classification predicts discrete responses. The algorithm labels or classify by choosing two or more classes for each example. If a

classification is done between two classes, then it is called *binary classification*, and else if it is done between two or more classes, then it is called *multi-class classification*. Regression predicts continuous responses. Here, the algorithms return a statistical value. Some of the most common algorithms are support vector machine (SVM), decision tree (DT), and artificial neural network (ANN).

2.3.2 Unsupervised learning

Unsupervised learning is machine learning in which an output is trained to respond to clusters groups of patterns within the input. There is no category into which the patterns are to be classified rather the system develops its representation of the input. It mostly discovers the unknown patterns in the data, but most of the time these approximations are weak when compared with supervised learning. There are two unsupervised learning tasks: *Association* and *Clustering*. In clustering, the input data are formed into clusters or groupings. Some of the most common algorithms are K-means clustering and association rule learning algorithm.

2.3.3 Reinforcement learning

Reinforcement learning is a kind of machine learning method concerned with how software agents automatically determine the ideal behavior within a specific context, to maximize its performance. It consists of two learning tasks: *Classification* and *Control*. Some applications of reinforcement learning are computer-played board games, robotic hands, and self-driving cars. The most commonly used algorithms are Q-learning, temporal difference, and deep adversarial networks (Haug 2018).

2.3.4 Neural networks

In 1943, McCulloch and Pitts (McCulloch and Pitts 1943) were the first to introduce the concept of an artificial neuron. The motivation for the development of neural network technology is a desire to develop an artificial intelligence system that could perform intelligent tasks like human brain. Their research work showed that the simple class of network can perform and analyze highly complex problems. Also, it was concluded that this system of neurons could be able to implement any complex function consisting of a system of neurons interconnected by the synaptic weights. Neural networks (NN) represent deep learning using artificial intelligence. ANNs are computational modeling tools that have recently emerged and extensive acceptance in many disciplines for modeling complex real-world problems. They are composed of simple elements operating in parallel inspired by biological

nervous systems. It is a structure that consists of simple processors trying to imitate the basic abilities of human brains and consists of artificial neurons (nodes) that are capable of performing parallel computations for data processing and knowledge representation.

ANN is similar to the human brain in two ways:

- They acquire knowledge through learning
 - Their knowledge is stored within interneuron connection strengths known as synaptic weights. The nodes are stacked next to each other in three layers: the *input layer*, the *hidden layer(s)*, and the *output layer*.
- *Input layers* are a bottom layer that takes input from the dataset which is also called the visible layer, because it is the exposed part of the network. A neural network is drawn with a visible layer with one neuron per input value (column in the dataset).
 - *Hidden layers* are layers after the input layer which is called hidden layers because they are not directly exposed to the input. It is a place where all the action is performed. The simplest network structure has a single neuron in the hidden layer that directly outputs the value. *As the number of neurons in hidden layers increases, a very deep neural network can be constructed.*
 - *Output layer* is the final hidden layer that is responsible for outputting a value or vector of values that correspond to the format required for the problem. A classification problem with multi-class may have multiple neurons in the output layer.

The attractiveness of ANNs comes from the remarkable information processing characteristics of the biological system such as nonlinearity, high parallelism, robustness, fault and failure tolerance, learning, and their ability to handle imprecise and fuzzy information. Nonlinearity allows fit to the data, noise insensitivity provides accurate prediction in the presence of uncertain data and measurement errors, and hardware failure tolerance, adaptivity and learning allow the system to modify its internal structure in according to changing environment. ANN models increased problem-solving capabilities by increasing the number of hidden layers or by increasing the number of neurons in any given layer and the number of paths between neurons.

In ANN the independent variables are called *inputs*, and the dependent variables are named as *output*. Each neuron receives a multiplied inputs and weights which are then added with static bias value, then passed to an appropriate activation function which decides the final value to be given out of the neuron. There are different types of activation functions available as per the nature of input values.

Once the output is generated from the final neural net layer, the loss function is calculated, and backpropagation is performed where the weights are adjusted to make the loss minimum. Weights are numeric values that are multiplied with inputs. In backpropagation, they are modified to reduce the loss and self-adjust depending on the difference between predicted outputs vs training inputs. Activation function is a mathematical formula that helps the neuron to switch ON/OFF. The input layer represents the dimensions of the input vector. The hidden layer takes a set of weighted inputs and produces output through an activation function. The output layer represents the output of the neural network (Figs. 1, 2).

2.3.5 Backpropagation neural network (BPNN)

In this study, a neural net was used to create a model and forecast visibility. The goal was to create a statewide model rather than one for each location or a set of locations by category. As shown in Fig. 3, a neural network has an input layer, a hidden layer, and an output layer. Each layer is made up of a group of neurons. Synapses connect the neurons of the input layer to those of the hidden layer. Synapses are links with numerical weights. At the hidden layer level, the input layer neurons are graded by their respective numeric weights and summed.

Applied superficially thresholds or biases are then employed to decrease the cumulative input to an activation function and generate the output in the output layer (Arqub et al. 2021). This study’s BPNN visibility model has multiple layers and employs the feedforward backpropagation learning algorithm. Tangent sigmoid function (Arqub et al. 2021) feature are the techniques commonly for the output and hidden layers, respectively. As the learning algorithm, the Bayesian regulation function is



Fig. 2 Sample pictures of traffic accidents in Oromia Special Zone

employed to minimize the linear combination of squared errors and weights while avoiding costly cross validation (Abu Arqub 2021). The BPNN visibility model was trained with the objective of reducing overall prediction error.

2.3.6 Multilayer perceptron neural network

A multilayer perceptron neural network is the most common type of feedforward neural network. The framework for multilayer perceptron neural networks is the same as for single layer neural networks, but with more hidden layers. An MLP has at least three node layers: an input layer, a hidden layer, and an output layer. In general, MLP trains the model using a supervised learning method known as backpropagation. Among the various types of NNs, the multilayer perceptron neural network has become the most common, universal, basic, and required for the majority of tasks.

MLPNN was developed and compared to other neural network models for the prediction of serious level of accident. Both of these studies investigated MLPNN’s ability to predict accident seriousness more accurately than other conventional models.

The ANN with MLP architecture was used to predict the severity of traffic accidents using twelve input variables and three serious levels: fatal, slight, and minor (Abu Arqub 2021). The ANN provided the best prediction performance in this study in Addis Ababa Special Zone, Ethiopia, with an R-value of around 0.87.

2.3.7 Radial basis function neural network

The radial basis function neural network model explains the more specific interactions between crash frequency and risk factors (Abu Arqub 2021). This model allows for the rapid replacement of redundant nodes (Momani et al. 2011). The RBFNN has good approximation accuracy and a fast convergence speed, which can effectively address the shortcomings of BP neural networks (Momani et al. 2011). RBFNN has been widely used in the fields of function approximation, problem solving, computer vision, system identification, and so on due to its deep physiology foundation, simple network structure, fast learning speed, and

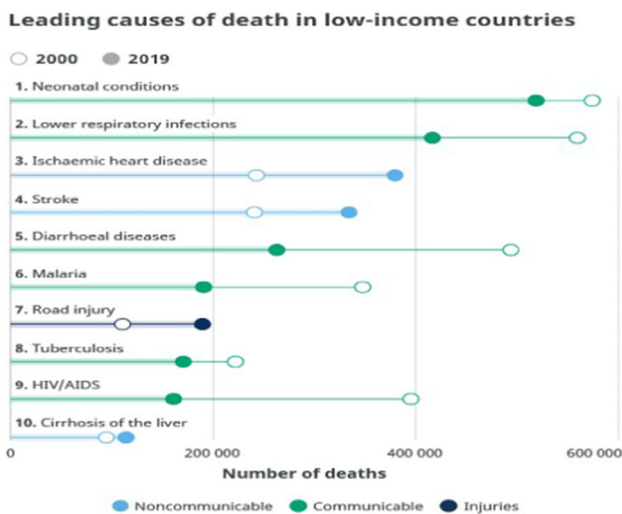
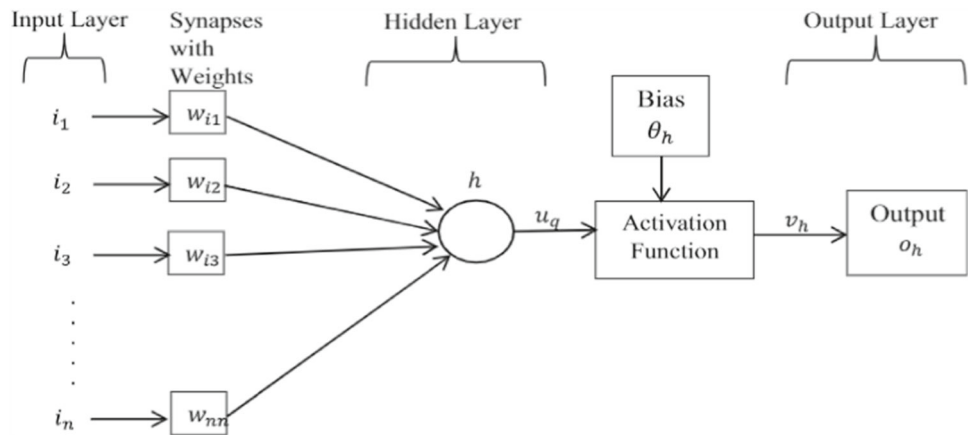


Fig. 1 Causes of death in low-income countries

Fig. 3 Network architecture for ANN



outstanding approximate solution abilities (Song and Li 2011). RBFNN has been widely used in a variety of applications due to its deep physiology foundation, simple network structure,

According to the findings of this study, the RNN neural network was the best model for determining the severity of driver injuries. Huang et al. 2016b developed a method for predicting accident frequency and determining the risk parameters that were more significant for road accidents to occur. While optimizing the RBFNN, the efficiency of its estimation increases, despite the fact that other variables are found to have little effect on the frequency of occurrence for the data.

2.3.8 Long short-term memory

Short-term memory that is long (Hochreiter and Schmidhuber 1997) [66] proposed the 2e LSTM (Yohannis 2019) model, which is a recurrent neural network variant (RNN). It creates a specialized memory storage unit that uses a time backpropagation algorithm to train the data. It can address the issue of the RNN's lack of long-term dependence.

The standard LSTM is written as follows. Each step t and its corresponding input sequence are $X \times 1, \times 2, \dots, x_t$, where t is the input gate, it is the forget gate, and f_t is the output gate. Memory cell state c_t regulates data memory and oblivion via various gates.

Different architectures are feasible for neural networks. The direction of information flow is set by the connections between their neurons. They are classified as feedforward or recurrent based on their network connections (backpropagating).

2.3.8.1 Feedforward neural networks Information flows in only one direction in these neural networks, from input layer to output layer. When the weights are determined, they are rarely changed. Weights can be ascertained

explicitly or using functions such as the radial basis function. The nodes in this area perform their functions without regard for whether the results produced are accurate or not (i.e., they do not re-adjust based on the results produced). There is no response from the layers ahead.

2.3.8.2 Recurrent neural networks (backpropagating)

Recurrent neural networks are neural networks that are generally designed to describe functions and improve with feedback links. They perform better and are more technically acceptable than feedforward networks. Hochreiter and Schmidhuber (1997) and [66] proposed a solution to the traditional RNNs algorithm's vanishing or exploding gradient problem (1997).

To produce a result, information is passed from the input layer to the output layer. Errors in the result are now communicated back to previous layers. Nodes learn how much they contributed to the incorrect answer. The weights have been re-adjusted. The neural network has been enhanced. It grows up. There is information flow in both directions. This essentially implements both algorithms, feedforward and backpropagation.

2.3.9 Researches on road traffic accident prediction

Nowadays machine learning is one of the most advanced scientific fields of artificial intelligence that can be applied to get a better result. Many types of research are conducted on accident data analysis and tried to show major factors causing accidents and also used different kinds of techniques to predict severity of road traffic accidents. Some of them are:

Maria LÍgia Chuerubim, Alan Valejo, Barbara Stolte Bezerra, Irineu Da Silva (2019), classified the severity of road accidents using artificial neural network (ANN) modeling with MLP. The traffic accident database used in this research consists of accidents that occurred between 2009 and 2012 from km 125 to km 145.5 of Dom Pedro I

Highway (SP-065), located in the city of Campinas (Brazil). During the period considered, 2824 accidents occurred, excluding missing observations, incomplete data. Eleven (11) variables were selected to classify the accident severity with fatal and non-fatal. 1964 (70%) of the data were used for ANN training and 860 (30%) of the data were used for the ANN test. ANN model with MPL classified each category of injury accidents being accidents with no injury (NI) and accidents with injury (WI). ANN ranked fatalities (WI) with an accuracy of 2.6%, while non-fatal accidents (NI) were classified with high accuracy (97.4%). Generally, the results obtained with the ANN had an accuracy of 77.9% and with mean AUC (area under curve) values of 0.618.

Meisam Siamidoudaran and Ersun Iscioglu (2019) applied MLP neural network (that can be used to categorize various road safety problems), SVM neural network (applied for the classification of collision injury severity), and hybrid MLP-SVM (The combination of MLP and SVM network). The data used were obtained from the road accident data in the city of London. SVM and MLPNN models were applied to predict the accident severity by using MATLAB software. All the factors were normalized between zero and one. The dataset was randomly separated into training data of 70% and test data of 30%. In the MLP model, the accuracy measure for the prediction in the training and test phase was found to be 84.41% and 83.82%, respectively. In the SVM model, the amount of accuracy obtained was 81.63% for the training phase and 80.88% for the test phase. The hybrid model provided a superior fitting of the model compared to the MLP and SVM models, and the obtained results have been very satisfactory. The hybrid model was able to maximize the accuracy rate for predicting the injury severity by 10%. The testing classification accuracy of the hybrid model was 90.58%.

3 Methodology

3.1 Data collection

Dataset used is data of road traffic accidents collected from Oromia Police Commission Excel Data and Oromia Special Zone Police Department hardcopy data; the dataset consists of accidents in a special zone of Oromia districts (Woredas) from 2005 to 2012. Most of the data are available in hardcopy, so first, the data are filled on Excel and some of the data are retrieved from Excel and then converted to CSV (comma separated values) format; these data can be viewed using a spreadsheet software such as MS Excel. CSV format data are easy to import using the pandas library.

3.2 Study area

The Oromia Special Zone Surrounding Finfinnee (in Afaan Oromo: *Godina Addaa Oromiyaa Nannawaa Finfinnee*) is one of the zones in the Oromia Region in Ethiopia that surrounds Addis Ababa. It was created in 2003 E.C from parts of North Shewa Zone, East Shewa Zone, Southwest Shewa Zone, and West Shewa Zones. The zone was created to support the cooperation and development of the surrounding areas of Addis Ababa. The administrative center of this zone is in Addis Ababa the capital of the Oromia region. The districts and town in this zone include Akaki, Berek, Burayu, Holeta Town, Mulo, Sebeta Hawas, Sebeta Town, and Sendafa.

3.3 Dataset parameters

The dataset contains person-related parameters (such as casualty, gender, and age), accident-related parameters (such as accident time, day of accident, cause of accident, and location of accident), environmental characteristics (such as weather condition and light condition), vehicle-related characteristics (such as vehicle type), and road-related parameters (such as road type and road surface).

Today's real-world datasets are susceptible to missing, noisy, and redundant, contain errors and inconsistent data. Working with low-quality data leads to low-quality results. Therefore, the first step in every machine learning application is to explore the dataset and understand its characteristics to make it ready for the modeling stage. This process is commonly known as data preprocessing. Data preprocessing is a technique to transform raw data into a simple and understandable format. Real-world data is often incomplete, inconsistent, or likely to contain errors. Data preprocessing includes data integration, data cleaning, normalization, transformation, data balancing, feature extraction, and selection. Data preprocessing prepares data for further processing by correcting outliers, handling missing values, normalization, and scaling of data (Abu Arqub 2021). Data preprocessing helps the neural network in learning the important patterns, which subsequently improves the data fitting and prediction accuracy.

3.4 Data preprocessing techniques

3.4.1 Data integration

Road traffic accident data are collected from hardcopy and Excel Softcopy files. These data need to be integrated common file and the file converted to a common file format which is comma-separated values (CSV) format.

3.4.2 Data cleaning

Cleaning data is the removal or fixing of missing features. There may be data parameters that are incomplete and do not carry the data you believe you need to address the problem. These instances may need to be removed. Additionally, there may be sensitive information in the attributes and these features may need to be removed from the data entirely. The dataset may contain many unnecessary entries that took up large memory size. This is solved by applying the required parameters that directly factor in road accidents.

- *Set Null (Missing) Values* For various reasons, many real-world datasets contain missing values, often encoded as blanks, NaNs. Such datasets are incompatible with scikit-learn estimators' library which assumes that all values in an array are numerical and that all have and hold meaning. A better strategy is to impute or fill the missing values. Different procedures are used to estimate missing values.
- *Univariate Imputation* is a simple strategy to fill missing value, and missing value will be imputed using different statistical operations such as *mean* (Substitute a feature's mean value computed from entire column to fill in missing data values on the remaining cases. A smarter solution is using feature mean for all samples belonging to the same class to fill in the missing value) and *mode* (the value of the feature that occurs most often is selected to be the value for all the unknown values of the feature) each column where missing value exists than the general feature means.
- *Dealing with outliers* Outliers are data points that are far from other data points. They are problematic for many analyses because they can cause tests to either miss significant findings or distort real results. Outliers can be treated like missing data and be either removed or replaced by techniques similar to missing data, e.g., with *mean*, *mode*, or *median*.
- *Noisy Data Correction* Noise is a random error or variance in a measured variable. Therefore, the data need to be smoothed. The following data are smoothing techniques.
- *Binning methods* Smooth a sorted data value by consulting the values around it. The sorted values are distributed into several buckets or bins. This method consults the neighborhood of values; they perform local smoothing values around it.

3.4.3 Feature extraction

Feature extraction is a dimensionality reduction technique. Unlike feature selection, which selects and retains the most significant attributes, feature extraction transforms the attributes. The transformed attributes, or features, are linear combinations of the original attributes.

Rules for Feature Extraction

Rules for extracting fatal accidents

Condition 1:

If cause of accident is overspeeding or head collision or rear collision or angular collision and

If the road surface is wet or snow or mud and

If weather conditions are raining or fog

If lighting conditions are dark

Condition 2:

If cause of accident is vehicle defect and

If the road surface is wet or snow or mud and

If weather conditions are raining or fog

Condition 3:

If cause of accident is roll over or fall and

If the road surface is wet or snow or mud and

If weather conditions are raining or fog

Condition 4:

If cause of accident is pedestrian fault and

If the road surface is wet or snow or mud and

If weather conditions are raining or fog or snowing

3.4.4 Rules for extracting serious accidents

Condition 1:

If cause of accident is vehicle defect and

If weather conditions are raining or fog or snowing

Condition 2:

If cause of accident is overspeeding or side collision or sliding

If weather conditions are fine

Condition 3:

If cause of accident is driver fault or pedestrian fault and

If the road surface is wet or snow or mud and

If weather conditions are fine

Condition 4:

If cause of accident is fall down or roll over

If the road surface is dry

Condition 5:

If cause of accident is hitting animals or poor light and

If the road surface is wet, snow, or mud and

If weather conditions are raining or fog or snowing

3.4.5 Rules for extracting slight accidents

Condition 1:

If cause of accident is hitting animals or driver fault or pedestrian fault)

If the road surface is dry

Condition 2:

If cause of accident is poor light, weather condition or road condition

If weather conditions are fine

If the road surface is dry

Condition 3:

If the cause of accident is overspeeding or vehicle defect and

If weather conditions are fine

If the road surface is dry

3.4.6 Feature selection

Working with a large number of parameters may affect the performance of models because training time increases exponentially with the number of features. Therefore, feature selection methods are used for selecting a subset of the input features for training the model depending on their contribution to identifying the features (inputs) important for classification and ignoring the irrelevant or redundant ones. So, feature selection is a critical factor for reducing *model* overfitting, reducing training time, reducing the complexity of a model and making it easier to interpret, and getting a more accurate prediction. Sklearn is a python library that has been used to diverge the feature with less importance. To obtain the most essential features, we operate an experiment by applying three different algorithms of feature selection.

4 Result and discussion

As presented in the previous chapter, the data are well described, explored, selected, and clean enough to be used for model building. This chapter presents the detailed activities carried out in selecting a modeling technique, implementation of the technique selected using the most appropriate algorithms, and evaluation of the models to select the best one for prediction. The study focuses on identifying determinant factors of drivers and road that leads to traffic accidents and building a prediction model. Various classification models have been built by using ANN. The models have been tested on different numbers of the selected attributes, and finally, the model with the best performance is selected.

4.1 Data distributions of collected dataset

The dataset contains a total of 6170 instances, and a total of fourteen variables (*District, Area, Accident Date, Time, Time of Accident, Cause of Accident, Road Type, Road Surface, Lighting Conditions, Weather Conditions, Casualty, Sex of Casualty, Age of Casualty, and Type of Vehicle*) are used as input variables (Figs. 4, 5).

4.2 Exploratory data analysis

4.2.1 Accident severity distribution

Figure 6.

4.2.2 Accident severity distribution by year

Figure 7.

4.2.3 Accident-prone areas

Figure 8.

4.3 Data preprocessing

A total of 6170 entries are present in the dataset. The categorical attributes (*District, Time of Accident, Cause of Accident, Road Type, Road Surface, Lighting Conditions, Weather Conditions, Casualty, Sex of Casualty, Type of Vehicle, and Accident Severity*) are converted into numeric using *Data Frame. Replace ()*. The Accident Date attribute is slit into new attributes like day, month, and year which can be used as a feature for the LSTM model to forecast accident number.

Fig. 4 The capital of the Oromia region

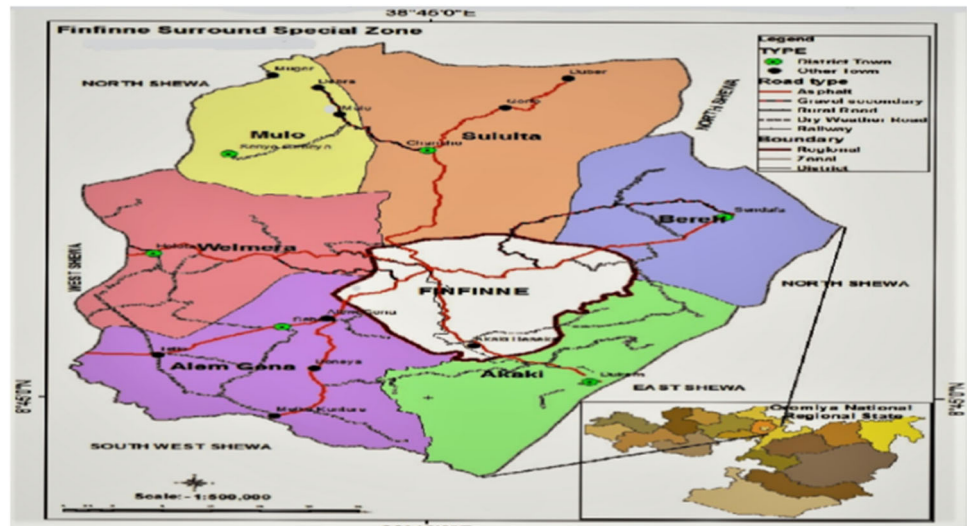


Fig. 5 Low-quality data lead to low-quality results

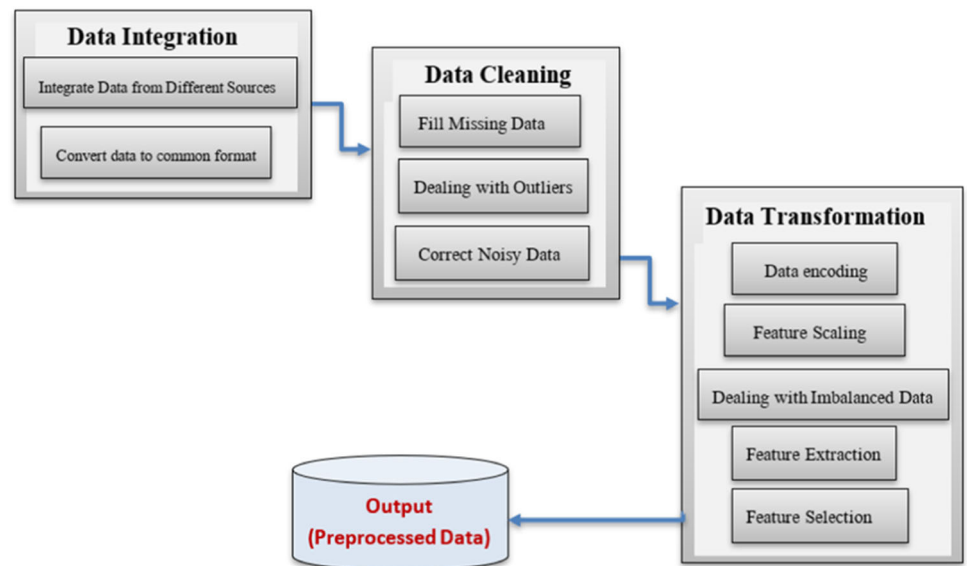
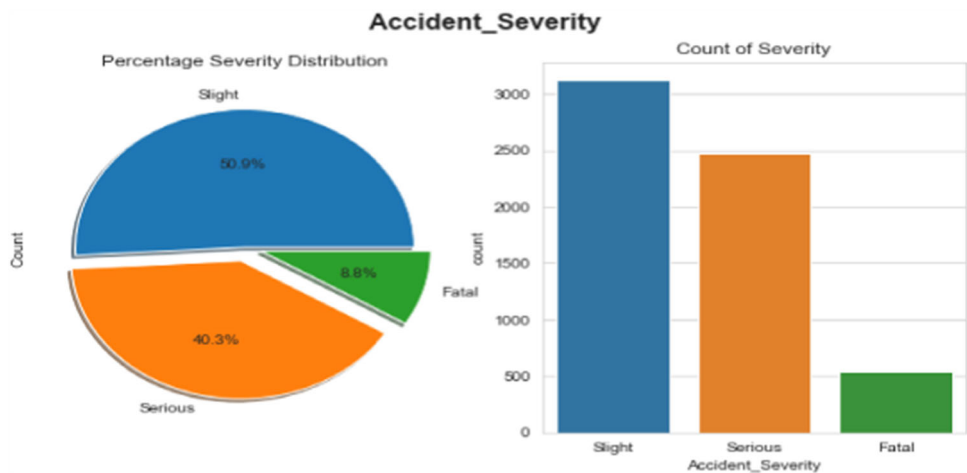


Fig. 6 Accident severity distribution



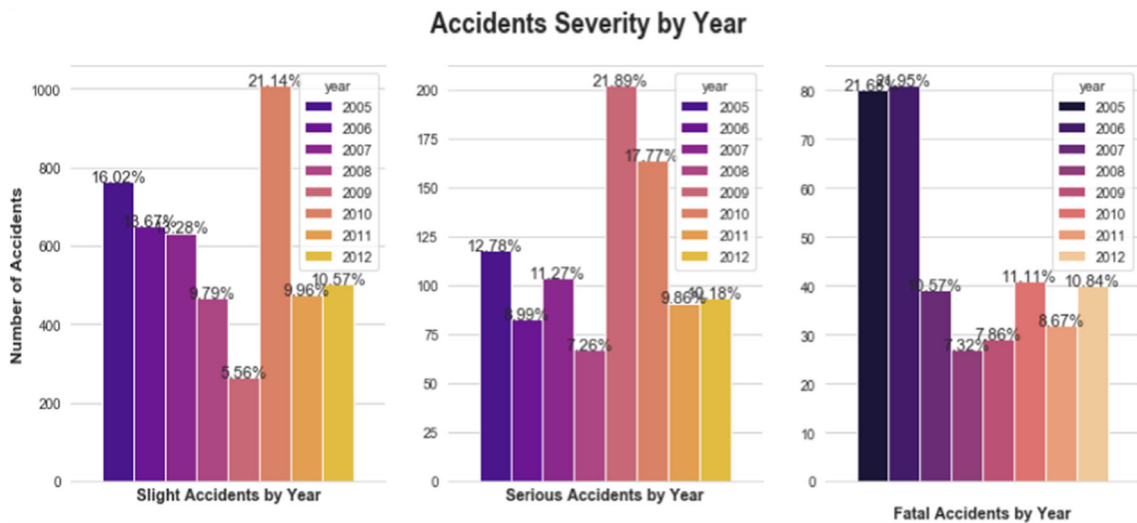
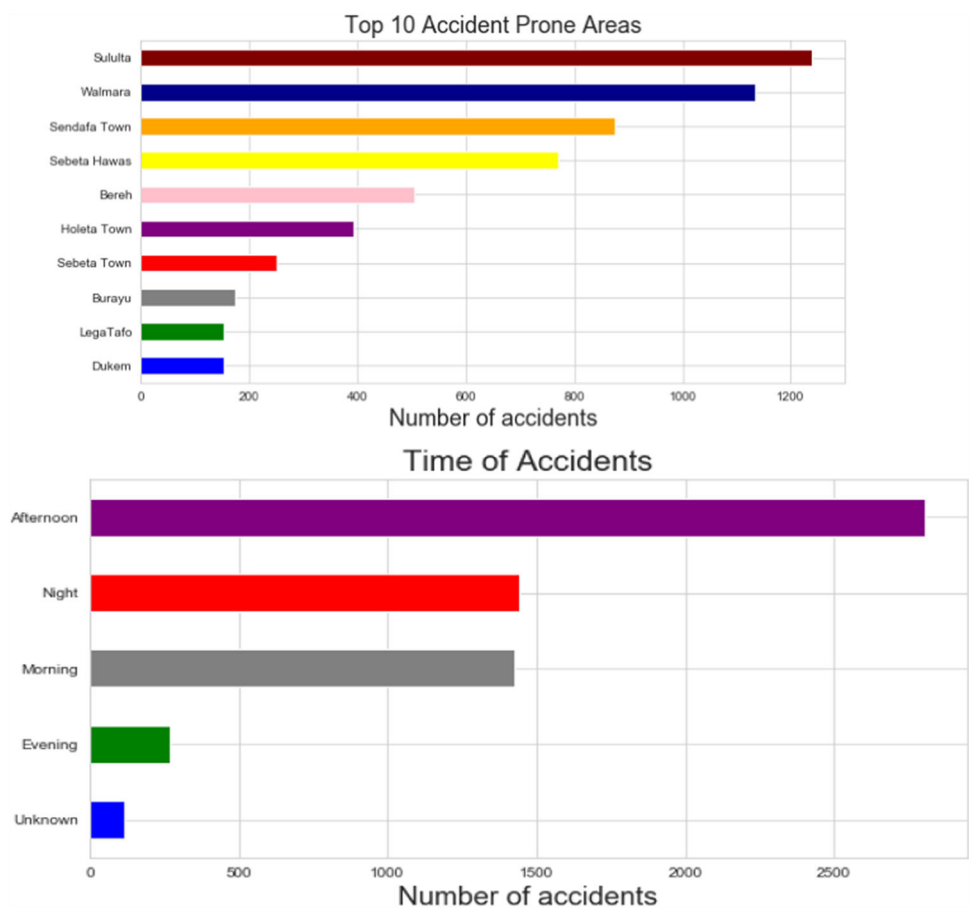


Fig. 7 Accident severity by year

Fig. 8 Accident distribution by time



4.4 Fill null values

Statistical operations mean is used to fill a null row in attribute (*Substitute a feature's mean value computed from the entire column to fill in missing data values*).

4.4.1 Feature scaling

Features are scaled between 0 and 1 using the Min Max Scaler library.

4.5 Resampling unbalanced data

The imbalanced dataset is balanced using SMOTE, Random Over Sampler, and Random Under Sampler from the 'imblearn' module of the python library. SMOTE, Random Over Sampler is an up-sampling technique that balances the data by increasing the number of data points for the minority class. Random Under Sampler balances the data by decreasing the number of data points for the minority class (Figs. 9, 10).

4.6 Feature selection

For determining the most significant factors, we have applied two approaches: embedding method and feature importance.

4.6.1 LASSO regularization (L1) result

In Fig. 11.

4.6.2 Feature importance results

Figure 12 shows the results of different attributes for feature importance. Feature importance assigns a score to input features based on how useful they are at predicting a target variable.

4.6.3 Permutation importance results

Figure 13 shows the results of different attributes for permutation importance. The method permutes a parameters value and then measures the increase in the model error.

4.7 Accident severity prediction

Six neural network algorithms have been used for accident severity classification in this study. They are feedforward neural network, backpropagation neural network, multi-layer perceptron neural network, radial basis function neural network, recurrent neural network, and long short-


| Age_of_Casualty | Age_of_Casualty |
|---|-----------------|
| 21 | 21 |
| 34 | 34 |
| 34 | 34 |
|  | <u>33.8</u> |
| 34 | 34 |
| 29 | 29 |
| 24 | 24 |
| 15 | 15 |

Fig. 9 Original rows with a null value and substituted row after applying mean operation

term memory neural network. The outcome of models and the performance evaluations metrics are discussed below.

4.8 The outcome of feedforward neural network model

Feedforward neural network model was trained with hidden layers of 200 neurons and ReLu activation function. FFNN has scored an accuracy of 80.46%, and the precision, recall, and F-measure are 0.804, 0.804, and 0.804, respectively. *Cohen's Kappa statistic* is 0.706. Mean absolute error and root mean squared error are 0.198 and 0.453, respectively (Figs. 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25).

4.9 The outcome of backpropagation neural network model

BPNN has scored an accuracy of 77.26%, and the precision, recall, and F-measure are 0.756, 0.756, and 0.756, respectively. *Cohen's Kappa statistic* is 0.556.

4.10 The outcome of multilayer perceptron neural network model

MLPNN model is built using 50 and 15 neurons hidden layers, *logistic* activation function, and *adam* solver. MLPNN has scored an accuracy of 97.13%, and the precision, recall, and F-score are 0.979, 0.979, and 0.979, respectively. *Cohen's Kappa statistic* is 0.968. Mean absolute error and root mean squared error are 0.021 and 0.151, respectively.

4.11 The outcome of radial basis function neural network model

RBFNN has scored an accuracy of 87.00%, and the precision, recall, and F-score are 0.717, 0.717, and 0.717, respectively. *Cohen's Kappa statistic* is 0.581. Mean absolute error and root mean squared error are 0.318 and 0.598, respectively.

4.12 The outcome of the recurrent neural network model

RNN model is constructed with 3 cells and 150 neurons in each cell for 100 epochs. RNN has scored an accuracy of 97.18%, and the precision, recall, and F-measure are 0.9718, 0.9718, and 0.9718, respectively. *Cohen's Kappa statistic* is 0.957. Mean absolute error and root mean squared error are 0.03 and 0.18, respectively.

Fig. 10 Output of dataset after applying feature scaling

```

=====Feature Scaling:=====
District  Time_of_Accident  Cause_of_Accident  Road_Type  Road_Surface  \
0  0.428571          0.50          0.111111      0.0          0.000000
1  0.142857          0.25          0.000000      1.0          0.333333
2  0.071429          0.00          0.777778      0.0          0.000000

Lighting_Conditions  Weather_Conditions  Casualty  Sex_of_Casualty  \
0          0.25          0.25          0.5          0.0
1          0.00          0.00          0.0          0.0
2          0.00          0.00          0.5          0.0

Age_of_Casualty  Type_of_Vehicle  year  month  day
0          0.271739      0.125  0.428571  0.727273  0.433333
1          0.250000      0.250  1.000000  0.545455  0.866667
2          0.228261      0.125  0.857143  0.727273  0.633333
    
```

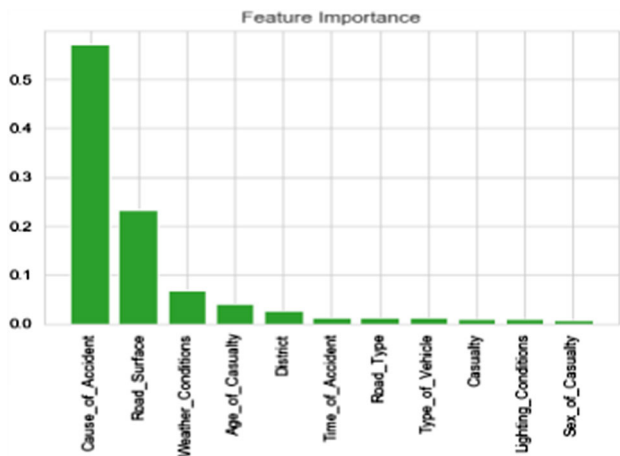


Fig. 11 Lasso regularization result

| feature | importance |
|---------------------|------------|
| Cause_of_Accident | 0.571 |
| Road_Surface | 0.233 |
| Weather_Conditions | 0.068 |
| Age_of_Casualty | 0.040 |
| District | 0.026 |
| Time_of_Accident | 0.013 |
| Road_Type | 0.013 |
| Type_of_Vehicle | 0.012 |
| Lighting_Conditions | 0.009 |
| Casualty | 0.009 |
| Sex_of_Casualty | 0.006 |

Fig. 12 Feature importance

4.13 The outcome of long short-term memory neural network model

LSTM model has been trained using *ReLU* activation function, *adam* solver for 100 epochs and 300 batches. LSTM has scored an accuracy of 91.00%, and the precision, recall, and F-measure are 0.894, 0.894, and 0.894, respectively. *Cohen's Kappa statistic* is 0.808. Mean

```

===== Feature Importance =====
Parameters          Weight
=====
Road_Surface:       0.432 +/- 0.014
Weather_Conditions: 0.411 +/- 0.011
Sex_of_Casualty:    0.134 +/- 0.007
Road_Type:          0.033 +/- 0.003
Casualty:           0.011 +/- 0.002
Cause_of_Accident: 0.010 +/- 0.003
Time_of_Accident:   0.008 +/- 0.002
Lighting_Conditions: 0.007 +/- 0.002
Age_of_Casualty:    0.007 +/- 0.002
Type_of_Vehicle:    0.007 +/- 0.002
District:           0.007 +/- 0.002
    
```

Fig. 13 Permutation importance scores

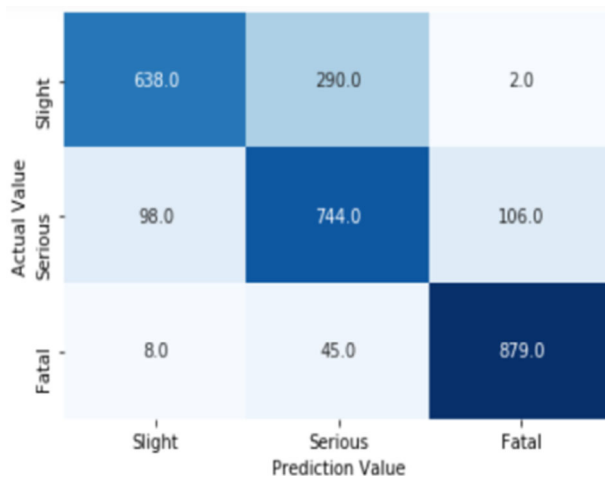


Fig. 14 Confusion matrix of FFNN

absolute error and root mean squared error are 0.105 and 0.327, respectively.

4.14 Comparison of the performance between algorithms

Comparison between the classification algorithms is shown in Table 5 where the highest accuracy 97.18% have been

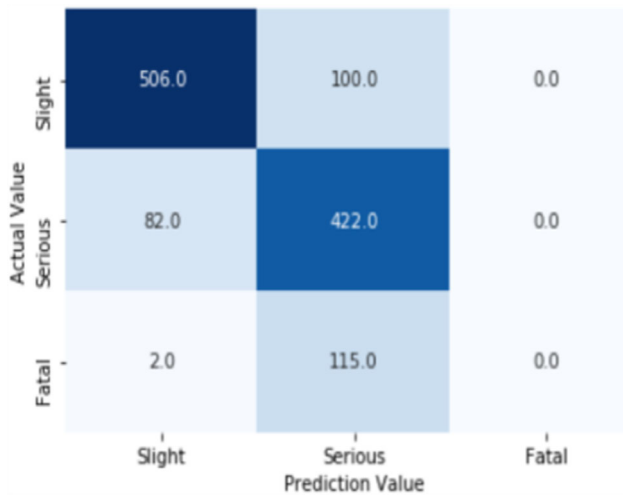


Fig. 15 Confusion matrix of BPNN

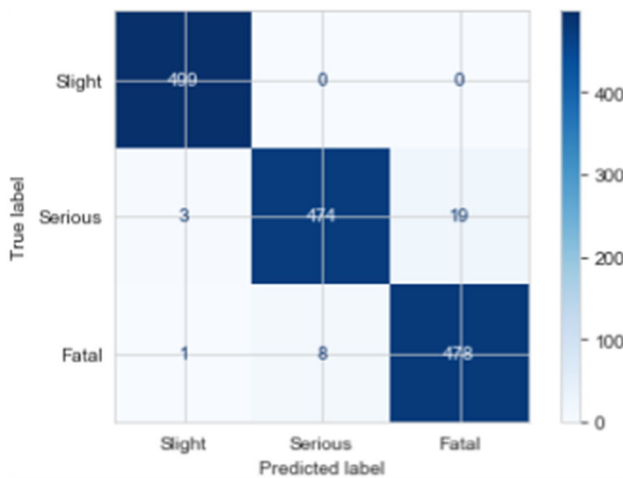


Fig. 16 Confusion matrix of MLPNN

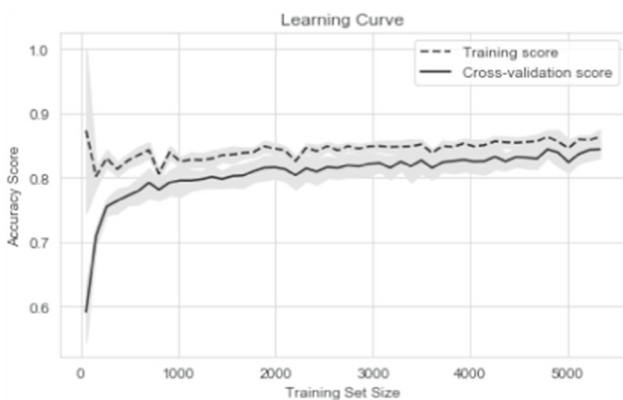


Fig. 17 Learning curve of MLPNN

obtained in the RNN model and the accuracies of MLP, LSTM, RBFNN, FFNN, and BPNN are 97.13%, 91.00%, 87.00%, 80.56%, and 77.26%, respectively.

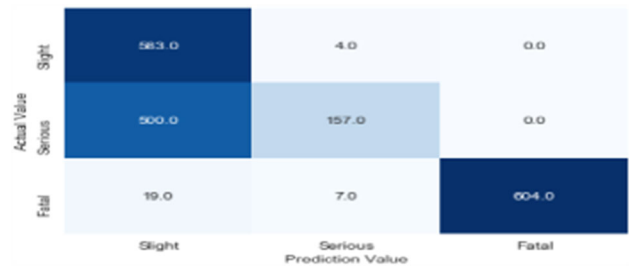


Fig. 18 Confusion matrix of RBFNN

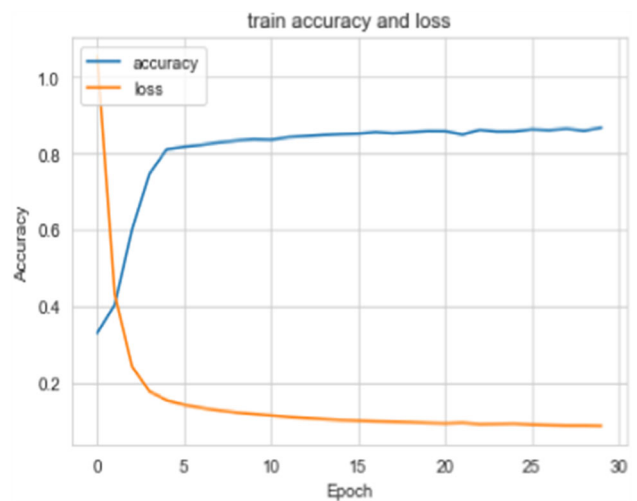


Fig. 19 RBFNN model accuracy and loss

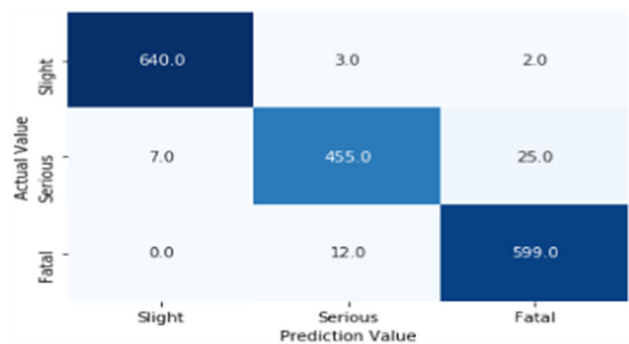


Fig. 20 Confusion matrix of RNN

4.15 Accident forecasting

4.15.1 The outcome of long short-term memory neural network model

LSTM model has trained using *ReLU* activation function, *adam* solver for 18 epochs and 300 batches and the model forecasted accidents number 3555 for three years. LSTM has scored a mean absolute error of 1.81, a root mean squared error of 2.455 and a mean percentage error of - 50.79.

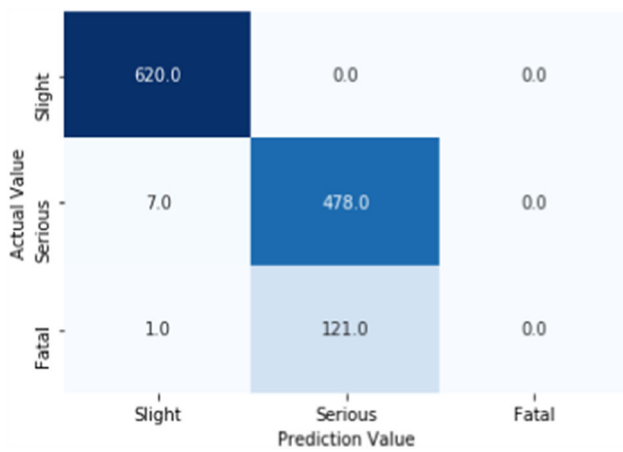


Fig. 21 Confusion matrix of LSTM

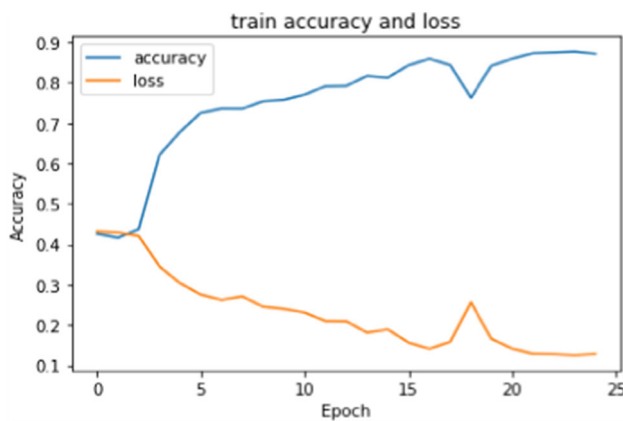


Fig. 22 LSTM loss and accuracy

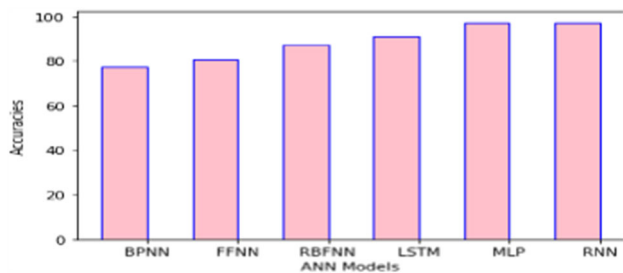


Fig. 23 Models accuracy comparisons

4.16 Identifying accident-prone areas

Figure 26 shows areas with accident severity low, mid, and high. The *red spot* shows accident areas with a high accident severity number, *green spot* shows areas with mild accident severity numbers, and the light green spot shows areas with a low accident severity number (Table 6).

5 Conclusion and future work

5.1 Conclusions

Road traffic accidents have become a major cause of death and injury nowadays. With increasing urbanization and growing populations, the volume of vehicles has increased exponentially and it became one of the most problems that drain the material and human resources. Road traffic accidents usually cause damage and injuries, ranging from minor to death or permanent disability. As a result, road traffic accident forecasting and identification of accident-prone areas can help in reducing the risks of traffic accidents. The objective of this research has been to explore the use of artificial neural networks as a technique for classifying severity levels in accidents, to identify the most significant factors that contribute to the severity of such crashes, and to forecast accident numbers. The data about the road traffic accidents were obtained from the Oromia Police Commission and Oromia Special Zone Police Department. The output or the accident severity class was divided into three major categories namely: slight accident, a serious accident, and death or fatal accident.

The major finding of this research is that ANN models perform well on the objectives above. In this paper, we have used six deep learning algorithms for prediction and LSTM for forecasting accident numbers. From this study, we can conclude the following points:

1. In testing the performance of the ANNs models developed in this study, the successful classification was found as follows:
 - a. BPNN with a classification accuracy of 77.26%
 - b. FFNN with a classification accuracy of 80.46%
 - c. RBFNN with a classification accuracy of 87.00%
 - d. RNN with a classification accuracy of 91.00%
 - e. MLP with a classification accuracy of 97.13%
 - f. RNN with a classification accuracy of 97.18%

The results have been compared based on different parameters of the algorithms such as kappa statistics, mean absolute error, root mean squared error, F-measure, etc. By comparing all these statistical metrics, it was found that the recurrent neural network model has given the best performance among the models.

The above classification performance indicates that the ANN was successful in capturing the relationship between the different parameters and the severity of accident cases used in testing the performance of the network.

2. During the process of data reduction and classification, some factors were removed because they seemed to

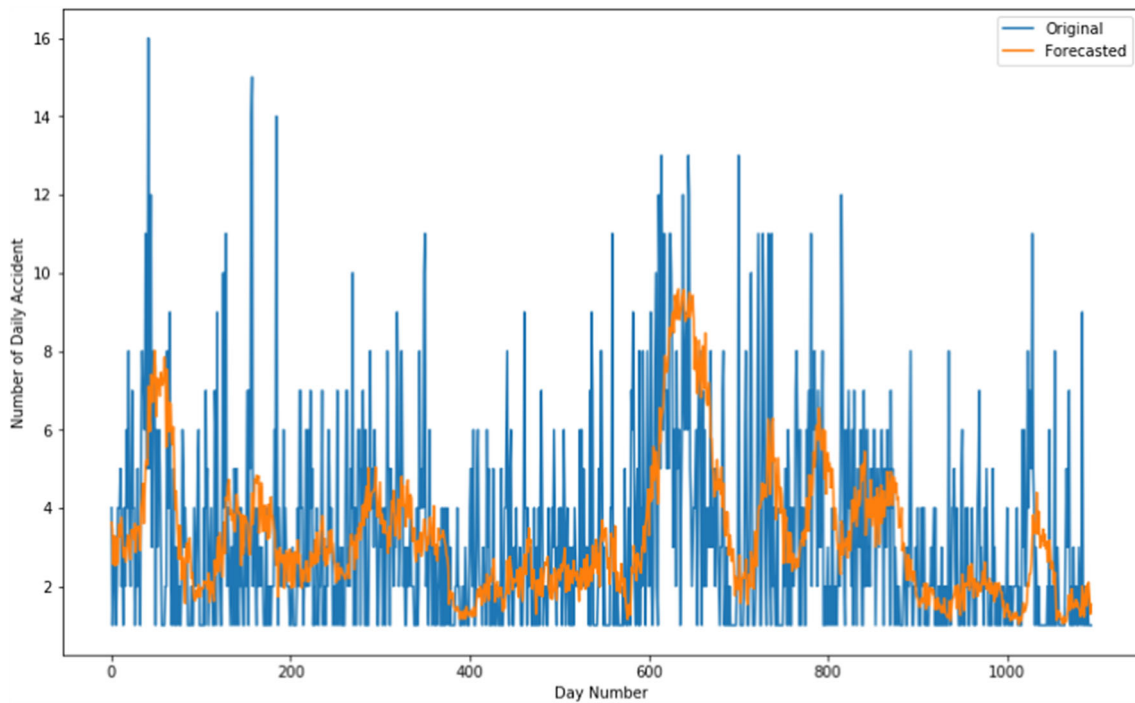
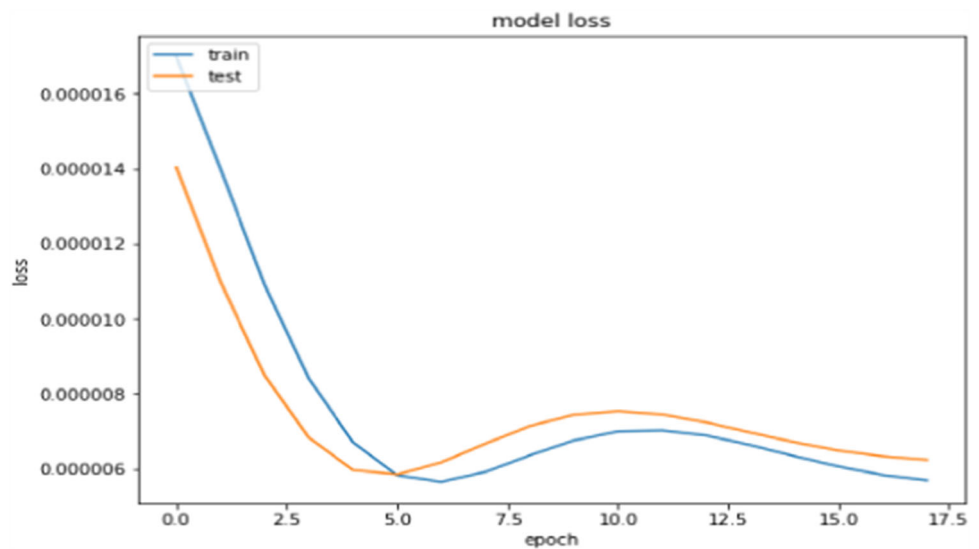


Fig. 24 Accident number forecast

Fig. 25 LSTM model loss on train and test



have no effect or little effect on the accident severity level.

- Road surface, weather conditions, and cause of accident are among the most important factors that influence the severity of the accident.

Although a significant improvement in accuracy has been observed, this study has some limitation with accident dataset which contains limited number of data samples from the year 2005 to 2012 since the zone was established recently. This study provides a way to analyze the severity

of road accidents, predict accident severity type, forecast severity number, and analyze the factors that lead to them. It was observed that factors such as road surface, weather conditions, and cause of accident had a high effect on the severity of an accident. Although the occurrence of an accident cannot be controlled, the analysis of this data can enable the government and its citizens to take precautionary steps toward keeping themselves safer.

Table 5 Performance comparison between neural network algorithms

| Evaluation metrics | Algorithms names | | | | | |
|--------------------------------|------------------|--------|---------|--------|---------|--------|
| | FFNN | BPNN | RBFNN | LSTM | MLP | RNN |
| Classification accuracy (%) | 80.56% | 77.26% | 87.00% | 91.00% | 97.13% | 97.18% |
| Precision | 0.804 | 0.756 | 0.717 | 0.894 | 0.979 | 0.971 |
| Recall/sensitivity | 0.804 | 0.756 | 0.717 | 0.894 | 0.979 | 0.971 |
| F-Score | 0.804 | 0.756 | 0.717 | 0.894 | 0.979 | 0.971 |
| <i>Cohen's Kappa statistic</i> | 0.706 | 0.556 | 0.581 | 0.808 | 0.968 | 0.957 |
| Mean absolute error | 0.198 | 0.245 | 0.029 | 0.105 | 0.021 | 0.033 |
| Root mean squared error | 0.453 | 0.248 | 0.177 | 0.327 | 0.151 | 0.183 |
| Mean percentage error (%) | - 1.046 | | - 4.016 | 5.58 | - 0.742 | 0.458 |
| Log Los | 4.793 | 5.123 | 9.28 | 0.197 | 6.571 | 0.237 |
| R-squared | 0.689 | 0.422 | 0.517 | 0.755 | 0.965 | 0.956 |
| Time taken to classify | 16 s | 1.93 s | 130 s | 99 s | 300 s | 103 s |

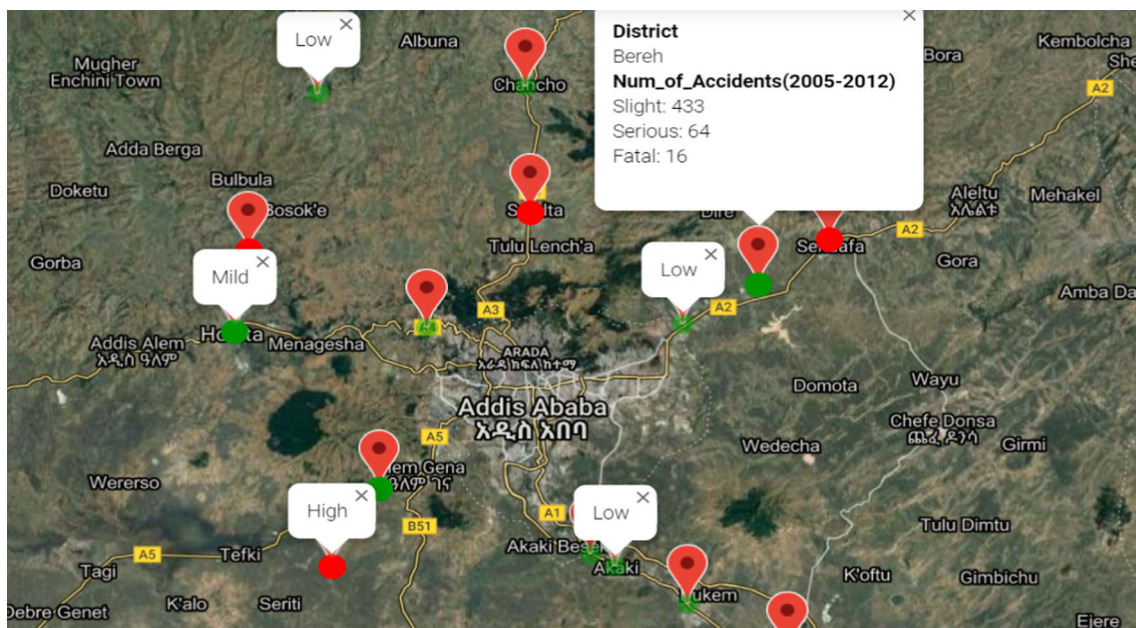


Fig. 26 Accident-prone areas

Table 6 Comparing actual accident number and forecasted accident number

| Number of days | Actual accident number | Forecasted accident number | No. of neurons | Epochs | Batch size |
|----------------|------------------------|----------------------------|----------------|--------|------------|
| After 10 days | 29 | 31 | 15 | 8 | 290 |
| After 50 days | 230 | 170 | 15 | 20 | 320 |
| After 100 days | 419 | 382 | 10 | 8 | 300 |
| After 150 days | 598 | 573 | 10 | 8 | 300 |
| After 1 year | 1486 | 1391 | 15 | 8 | 300 |
| After 2 years | 2405 | 2377 | 15 | 8 | 300 |
| After 3 years | 3561 | 3555 | 100 | 18 | 300 |

5.2 Recommendations

From the study performed in this thesis, we recommend the following:

- The number of instances in the dataset needs to be increased to make a more effective predictive model.
- There are some useful techniques such as regression analysis, Bayesian belief, Naïve Bayes, etc., that can be used for data preprocessing and there are several algorithms such as ensemble learning algorithms, time-domain neural network (TDNN) and convolution neural network (CNN) with multi-class that will be implemented. Finally, an expert system will be developed based on the best model achieved from this research study.

Funding No fund received for this project.

Data availability statement Not Applicable.

Code availability Not Applicable.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Ethical approval and human participation No ethics approval is required.

Consent to participate Not Applicable.

Consent for publication Not Applicable.

References

- Abegaz T et al (2014) Road traffic deaths and injuries are under-reported in ethiopia: a capture-recapture method. *PLoS ONE* 9:7
- Abu Arqub O, Singh J, Alhodaly M (2021) Adaptation of kernel functions-based approach with Atangana–Baleanu–Caputo distributed order derivative for solutions of fuzzy fractional Volterra and Fredholm integrodifferential equations. *Math Methods Appl Sci*. <https://doi.org/10.1002/mma.7228>
- Abu Arqub O, Singh J, Maayah B, Alhodaly M (2021) Reproducing kernel approach for numerical solutions of fuzzy fractional initial value problems under the Mittag-Leffler kernel differential operator. *Math Methods Appl Sci*. <https://doi.org/10.1002/mma.7305>
- AFRO factsheet (2013)
- Al-Khalifa K, Hamouda AMS (2012) Prediction of road accidents in Qatar 2022. Qatar foundation annual research forum proceedings. AHP31. <https://doi.org/10.5339/qfarf.2012.AHP31>
- Alkheder S, Taamneh M, Taamneh S (2017) Severity prediction of traffic accidents using an artificial neural network. *J Forecast* 36(1):100–108
- Almamlook R, Kwayu K, Prefer A (2019) Comparison of machine learning algorithms for predicting traffic accident severity. In: 2019 IEEE Jordan international joint conference on electrical engineering and information technology (JEEIT), pp 272–276. IEEE. <https://doi.org/10.1109/JEEIT.2019.8717393>
- Al-Moqri T, Haijun X, Namahoro JP, Alfalahi EN, Alwesabi I (2020) Exploiting machine learning algorithms for predicting crash injury severity in Yemen: hospital case study. *Appl Comput Math* 9(5):155–164. <https://doi.org/10.11648/j.acm.20200905.12>
- Al-Zuhairi M, Pradhan B, Shafri H, Hamid H (2019) Applications of deep learning in severity prediction of traffic accidents. In: GCEC 2017: Proceedings of the 1st global civil engineering conference 1, pp 793–808, Springer https://doi.org/10.1007/978-981-10-8016-6_58
- Al-Zuhairi M, Pradhan B (2017) Severity prediction of traffic accidents with recurrent neural networks. *Appl Sci*. <https://doi.org/10.3390/app7060476>
- Ameslasie F, Kidanu M, Lerebo W, Ali D (2016) Patterns of trauma in patient seen at the emergency clinics of public hospitals in Mekelle, northern Ethiopia. *Ethiop Med J* 54(2):63–68
- Arhin S, Gatiba A (2019) Predicting injury severity of angle crashes involving two vehicles at unsignalized intersections using artificial neural networks. *Eng Technol Appl Sci Res* 9:3871–3880. <https://doi.org/10.48084/etasr.2551>
- Assi K (2020) Traffic crash severity prediction-a synergy by hybrid principal component analysis and machine learning models. *Int J Environ Res Public Health* 17(20):7598. <https://doi.org/10.3390/ijerph17207598>
- Chen S, Kuhn M, Prettnner K, Bloom DE (2019) The global macroeconomic burden of road injuries: estimates and projections for 166 countries. *Lancet Planet Health* 3(9):e390–398
- Chong M, Abraham A, Paprzycki M (2005) Traffic accident analysis using machine learning paradigms. *Inform (slovenia)* 29:89–98
- Chuerubim ML, Valejo A, Bezerra BS, da Silva I (2019) Artificial neural networks restriction for road accidents severity classification in unbalanced database. *Sigma J Eng Nat Sci* 37(3):927–940
- Çodur M, Tortum A (2015) An artificial neural network model for highway accident prediction: a case study of Erzurum, Turkey. *PROMET Traffic Transp*. <https://doi.org/10.7307/ptt.v27i3.1551>
- Doğan E, Akgüngör A (2008) Estimating road accidents of turkey based on regression analysis and ANN approach. *Adv Transp Stud Int J* 16:11–22
- Economic Commission for Africa (ECA) and the Secretariat of the UN Secretary-General’s Special Envoy for Road Safety, Mr. Jean Todt. URL: <https://www.fanabc.com/english/about-13-people-die-in-road-traffic-accident-in-ethiopia-each-day/>
- Abagaz T, Gebremedhin S (2018) Magnitude of road traffic accident-related injuries and fatalities in Ethiopia. *PLoS ONE* 14(1):e0202240. <https://doi.org/10.1101/382333>
- Farhat Z, Karouni A, Chauvet P, Daya B, Hamadeh N (2020) Traffic accidents severity prediction using SVM models. *J Eng Des Technol* 9:1345–1350. <https://doi.org/10.35940/ijitee.F4393.059720>
- Farhat Z, Karouni A, Daya B, Chauvet P (2019) Comparative study between decision trees and neural networks to predict fatal road accidents in Lebanon. In: 5th International conference on computer science, information technology, Aircc Publishing Corporation, pp 01–14 <https://doi.org/10.5121/csit.2019.91101>
- García de Soto B, Bumbacher A, Deublein M, Adey BT (2018) Predicting road traffic accidents using artificial neural network models. *Infrastruct Asset Manage* 5(4):132–144. <https://doi.org/10.1680/jinam.17.00028>
- Global health estimates. Geneva: World Health Organization (2014) (http://www.who.int/healthinfo/global_burden_disease/)

- [estimates/en/index1.html](#), http://www.who.int/healthinfo/global_burden_disease/projections/en, Accessed 15 September 2015).
- Hartika HA, Ramli MZ, Abidin MZZ, Zawawi MH (2017) Study of road accident prediction model at accident blackspot area at Selangor. *Int J Sci Res Sci Eng Technol (IJSRSET)* 3(5):466–470
- Haug S (2018) <https://towardsdatascience.com/Introduction-To-Various-Reinforcement-Learning-Algorithms-I-Q-Learning-Sarsa-Dqn-Ddpg-72a5e0cb6287>
- Hochreiter S, Schmidhuber J (1997a) Long short-term memory. *Neural Comput* 9:1735–1780
- Hochreiter S, Schmidhuber J (1997b) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Huang H, Zeng Q, Pei X, Wong SC, Xu P (2016a) Predicting crash frequency using an optimized radial basis function neural network model. *Transp A Transp Sci* 12:1–24. <https://doi.org/10.1080/23249935.2015.1136008>
- Huang H, Zeng Q, Pei X, Wong SC, Xu P (2016b) Predicting crash frequency using an optimised radial basis function neural network model. *Transp Metrica Transp Sci* 12(4):330–345
- Jadaan K, Alkhaledi Q, Najjar A. Development of An Accident Prediction Model Using Artificial Neural Network. Available at: <https://www.psd.gov.jo/images/jti/docs/>
- Kotsiantis S, Kanellopoulos D, Pintelas P (2006) Data preprocessing for supervised learning. *Int J Comput Sci* 1:111–117
- Kunt M, Aghayan I, Noii N (2011) Prediction of traffic accident severity: Comparing the ANN, genetic algorithm, combined genetic algorithm, and pattern search methods. *Transport* 26:353–366. <https://doi.org/10.3846/16484142.2011.635465>
- Ławrynowicz A, Tresp V (2014) Introducing machine learning. *Perspect Ontol Learn* 12:19
- Lee J, Yoon T, Kwon S, Lee J (2019) Model evaluation for forecasting traffic accident severity in rainy seasons using machine learning algorithms in Seoul City. *Appl Sci* 10:129. <https://doi.org/10.3390/app10010129>
- Loewe M, Rippin N (2015) The Sustainable Development Goals of the Post-2015 Agenda: Comments on the OWG and SDSN Proposals, 2015
- Mackay GM, Wodzin E. Global priorities for vehicle safety. International conference on vehicle safety 2002: IMechE conference transactions. London, Institution of Mechanical Engineers, 2002:3–9
- Persson (2008), Road traffic accidents in Ethiopia: Magnitude, causes and possible interventions
- Marius P, Balas V, Perescu-Popescu L, Mastorakis N (2009) Multilayer perceptron and neural networks. *WSEAS Trans Circ Syst* 8(7):579–588
- Mashhadi R, Mahdi M, Nazneen S, Ksaibati K (2020) Application of deep learning techniques in predicting motorcycle crash severity. *Eng Rep.* <https://doi.org/10.1002/eng2.12175>
- McCulloch WS, Pitts W (1943) Logical Calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 5(4):115–133
- Minister of Transport NR (2011) National Road Safety strategic plan of Ethiopia 2011 to 2020. AA
- Moghaddam FR, Afandizadeh S, Ziyadi M (2011) Prediction of accident severity using ANN. *IJCE.* 9(1):41–48
- FDRE MoH (2015) HSTP: Health Sector Transformation Plan, 2015/16–2019/20, 2015, The Federal Democratic Republic of Ethiopia Ministry of Health (FDRE MoH): Addis Ababa, Ethiopia
- Momani S, Abo-Hammour ZS, Alsmadi OM (2016) Solution of inverse kinematics problem using genetic algorithms. *Appl Math Inform Sci* 10(1):225. <https://doi.org/10.1016/j.ins.2014.03.128>
- Moor J (2006) The dartmouth college artificial intelligence conference: the next fifty years. *Artif Intell Mag* 27(4):87–91
- Murray CJL, Lopez AD (eds) (2020) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Boston
- Odhiambo JN, Wanjoya AK, Waititu AG (2015) Modeling road traffic accident injuries in nairobi county: model comparison approach. *Am J Theor Appl Stat* 4(3):178–184. <https://doi.org/10.11648/j.ajtas.20150403.24>
- Olutayo VA, Eludire AA (2014) Traffic accident analysis using decision trees and neural networks. *Int J Inf Technol Comput Sci (IJITCS)* 6(2):22–28. <https://doi.org/10.5815/ijitcs.2014.02.03>
- Road traffic accidents in Ethiopia (2008) Magnitude, causes and possible interventions. *Adv Transp Stud* 15:5–16
- Sazli M (2006) A brief review of feed-forward neural networks. *Communications, Faculty of Science, University of Ankara.* 50: 11–17. <https://doi.org/10.1501/0003168>
- Schwenker F, Kestler HA, Palm G (2001) Three learning phases for radial-basis-function networks. *Neural Netw* 14(4–5):439–458
- Shaik M, Hossain, Q (2018) An artificial neural network model for road accident prediction: a case study of Khulna Metropolitan City
- Siamoudouaran M, İşçioğlu E (2019) Injury Severity prediction of traffic collision by applying a series of neural networks: The City of London case study. *PROMET* 31(6):643–654
- Song C, Li Q (2011) The prediction model of macro-road traffic accident basing on radial basis function. *Appl Mech Mater* 97–98:981–984
- Urbanization in Small Cities and Their Significant Implications on Landscape Structures: The Case in Ethiopia, retrieved February 8, 2020
- Walczak B, Massart DL (2000) Local modelling with radial basis function networks. *Chemom Intell Lab Syst* 50(2):179–198
- Werbos PJ (1988) Generalization of backpropagation with application to a recurrent gas market model. *NeuralNetwork* 1:339–356
- Werbos PJ (1990) Backpropagation Through time: what it does and how to do it. *Proc IEEE* 78(10):1550–1560
- WHO Global Status Report on Road Safety (2015) 20 avenue Appia, 1211 Geneva 27: World Health Organization; 2015
- World Health Organization (WHO). Global Status Report on Road Safety (2018) https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/
- World Health Organization (2019) Report on causes of death. Available from URL: <https://www.who.int/news-room/factsheets/detail/the-top-10-causes-of-death>
- World Report on Road Traffic Injury Prevention (2015)
- World's first road death. London, Road Peace (2003) (<http://www.roadpeace.org/articles/WorldFirstDeath.html>, Accessed on 17 November 2003)
- Yohannis D (2019) The Impact of Road Traffic Accident on Economic Growth in Ethiopia, Munich, GRIN Verlag, <https://www.grin.com/document/507875>
- Yousif J, AlRababaa S (2013) Neural technique for predicting traffic accidents in Jordan. *J Am Sci* 9:347–358

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.