2020-08-10

CRIME ANALYSIS AND PREDICTION USING HYBRID DEEP LEARNING ALGORITHMS

Meron, Tamiru Shibiru

BAHIR DAR UNIVERSITY

BAHIR DAR INSTITUTE OF TECHNOLOGY

SCHOOL OF RESEARCH AND POSTGRADUATE STUDIES

COMPUTING FACULTY

CRIME ANALYSIS AND PREDICTION USING HYBRID DEEP LEARNING ALGORITHMS

MSc. Thesis

By

Meron Tamiru Shibiru

Program: Software Engineering

Main Advisor: Ayalew Belay Habtie(PhD.)

Co-Advisor: Elias Wondmagegn Muchie (MSc.)

August 10, 2020

Bahir Dar, Ethiopia

CRIME ANALYSIS AND PREDICTION USING HYBRID DEEP LEARNING ALGORITHMS

Meron Tamiru Shibiru

A Thesis submitted to the school of Research and Graduate Studies of Bahir Dar Institute of Technology, BDU in partial fulfillment of the requirements for the degree of Masters in the Software Engineering in the Computing Faculty.

Main Advisor: Ayalew Belay Habtie(PhD.)

Co-Advisor: Elias Wondmagegn Muchie(MSc.)

August 10, 2020

Bahir Dar, Ethiopia

DECLARATION

This is to certify that the thesis entitled "**Crime Analysis and Prediction Using Hybrid Deep Learning Algorithms**", submitted in partial fulfillment of the requirements for the degree of Master of Science in **Software Engineering** under **Computing Faculty**, Bahir Dar Institute of Technology, is a record of original work carried out by me and has never been submitted to this or any other institution to get any other degree or certificates. The assistance and help I received during the course of this investigation have been duly acknowledged.

Name: Meron Tamiru Shibiru Signature:  Date: August 10, 2020

BAHIR DAR UNIVERSITY
BAHIR DAR INSTITUTE OF TECHNOLOGY
SCHOOL OF RESEARCH AND GRADUATE STUDIES
COMPUTING FACULTY
Approval of thesis for defense result

I hereby confirm that the changes required by the examiners have been carried out and incorporated in the final thesis.

Name: Meron Tamiru Shibiru Signature: Date: August 9, 2020

As members of the board of examiners, we examined this thesis entitled 'Crime Analysis and Prediction Using Hybrid Deep Learning Algorithms' by Meron Tamiru. We hereby certify that the thesis is accepted for fulfilling the requirements for the award of Degree of Master of Science in Software Engineering.

Board of examiners:

Name of Main Advisor	Signature	Date
Ayalew Belay Habtie (PhD.)		August 9, 2020
Name of Co-Advisor	Signature	Date
Elias Wondemagegn Muchie (MSc.)		August 9, 2020
Name of External Examiner	Signature	Date
Adane Letta (PhD.)		August 9, 2020
Name of Internal Examiner	Signature	Date
Gebeyehu Belay (PhD.)		August 9, 2020
Name of Chairperson	Signature	Date
Mekonnen Wagaw (PhD.)		August 9, 2020
Name of Chair Holder	Signature	Date
Gebeyehu Belay (PhD.)		August 9, 2020
Name of Faculty Dean	Signature	Date
Belete Biazen (MSc.)		August 9, 2020

v

ACKNOWLEDGEMENTS

First and above all I want to thank my GOD. Without his mercy, grace, help and blessings, I am nothing. Thanks to GOD for protecting my families and helping me in my whole life.

I would like to thank my thesis advisors Dr. Ayalew Belay and Mr. Elias Wondemagegn for their guidance, and support. I want to thanks my beloved families for their affection towards me. They mean a lot to me. I would like to thank Bahir Dar city police commission and police stations for their cooperation in gathering crime data that I need for this research. I am also thankful to my friends for always being there in case of any difficulty in my research.

ABSTRACT

Crime is the act of offending and doing illegal activity. Previous studies have shown crime should be predicted using demography, location and time features. This thesis attempt to predict crime types using hybrid of deep learning algorithms. 11,363 records (September 2005-April 2011 E.C.) was collected from Bahir Dar city. The data was in paper form and which has been converted to electronic format. After preprocessing the data, we left with 10,270 records. The crime record dataset has been analyzed and tested using the hybrid of feedforward artificial neural network and long short-term memory recurrent neural network. We have compared the performance of FFANN, LSTM-RNN and the hybrid model. MAE, MSE and accuracy is used to measure the model performance. The result for MAE is 0.249, 0.181, and 0.212 for FFANN, LSTM-RNN and the hybrid model respectively. MSE is measured as 0.097, 0.051, and 0.049 for FFANN, LSTM-RNN and the hybrid model respectively. Whereas the accuracy is 90.20%, 94.80%, and 95.07% for FFANN, LSTM-RNN and the hybrid model respectively in which the proposed model outperforms the two algorithms. The study result show that applying the hybrid of FFANN and LSTM-RNN enables to come up with more accurate prediction. Predicting crimes accurately helps to improve crime prevention and to optimize resource allocation in police departments. This crime prediction result can be improved by applying attribute selection by computing information gain and correlation.

Keywords: Criminology, crime prediction, Deep learning, FFANN, LSTM-RNN

TABLE OF CONTENTS

ABBREVIATIONS

ANN	Artificial neural network
CNN	Convolutional neural network
CNSL	Conditional neural sequence learner
CPU	Central processing unit
DNN	Deep neural network
E.C.	Ethiopian calendar
FFANN	Feedforward artificial neural network
G.C.	Gregorian calendar
GPU	Graphics processing unit
GRU	Gated recurrent unit
ISCED	International standard classification of education
ISCO	International standard classification of occupation
LSTM	Long-short-term-memory
MAE	Mean absolute error
MSE	Mean squared error
OS	Operating system
PC	Property crime
RNN	Recurrent neural network
SC	State crime
SVM	Support vector machine
USA	United states of America
VC	Violent crime
VCC	Victimless crime

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1. Background

Crime is the act of offence that has a consequences of community condemnation and punishment. As of the (Natarajan, 2016) study there are serious crimes in developing countries, such as Ethiopia. The rate of crime is increasing from time to time. ToppiReddy et al., (2018) described that crimes are the major social problems that can affect individual's life and economic growth of the country. Crime affects peoples' life directly or indirectly. It is one of the major variables, which mainly affects the development of the country.

Crime analysis is a systematic approach that identifies and analyzes patterns and trends in crime (Sathyadevan & Devan, 2014). Crime analysis includes exploring and detecting crimes and their relationships with criminals. It needs data mining techniques to extract hidden knowledge from huge collection of crime scene and criminal records (Jabar et al., 2013). Generally, crime analysis is the act of analyzing crime. The objective of crime analysis is to get meaningful information from large dataset and to assist law enforcement bodies as well as to reduce crimes (Al-janabi, 2011).

The law enforcement bodies need to allocate police force to reduce crimes in area where crime rate is high, this prevents occurrence of crimes. The difficulty is they are not aware of crime-dense area. If we are able to generate a pattern from the past crime data that the same event is likely to happen in the future then, we can predict crime. Crime is a predictable event and can be done by analyzing and generating a pattern from the past crime dataset that will aid us to make prediction on a future crime event (ToppiReddy et al., 2018).

In the process of crime prediction, type of crime, when and where the crime possibly will happen can be predicted. Predicting the crime only will not be enough for the law enforcement bodies but also prediction of the crime in line with possible suspect criminals will help to get detail information of the crime to be happened (Jadhav et al., 2017).

In relation to these different attributes, demography is also used in researches. According to Roth et al., (2009), demographic factors are social and materialistic features of a population in specific area. It includes income level, age, marital status, occupation, gender, education level, and related issues. Criminals follow constant or linear behavior patterns to perform the crime. Therefore, analyzing demography and criminal behaviors can help to produce the crime pattern in specific area.

Nevertheless, analyzing the demography of criminals cannot help to predict the location where the crime will happen. Researches in this area are conducted by analyzing the map of city using grid methods (Lin et al., 2018). Accordingly, both demographic data and location-based approach helps to have better prediction (Hilbert, 2016).

Several previous scholars like Mcclendon & Meghanathan (2015), Lin & Chen (2017), Bharati & RA.K (2018) have used machine learning algorithms for crime prediction and analysis. Even though, they have used different machine learning algorithms researches like Bharati & RA.K (2018), Stec & Klabjan (2018), Lin et al. (2018) focused to predict crime types based on spatial and temporal features of crime records.

Whereas Sathyadevan & Devan (2014) and other researchers have employed data mining to extract previously unknown pattern from unstructured data. Sreedevi et al., (2018) also used data mining to extract hidden patterns by focusing on crime factors rather than on causes of crime occurrences.

Applying the crime prediction models on Ethiopian crime record dataset is not efficient as of the other countries (Endalew, 2017). The researchers have used 3000 crime records with 7 features. Prediction algorithms need many records for making accurate prediction. Crime records available on UCI repository have 128 attributes which is not available in our country.

Findings of Lin et al. (2018) needs latitude and longitude attributes. These attributes did not exist in Ethiopian crime record. Converting the recorded location to longitude and latitude points will reduce the accuracy of crime analysis, since the recorded location is in general location name. Location data analysis is useful for identifying crime hotspot areas and predicting where the next crime will happen. Whereas time data need for forecasting the crime occurrence time.

According to Kalantari et al. (2016) crime spatial analysis is influenced by space and urban characteristics as well as social and economic factors. The authors have described that countries differ by their social, economic, structural, demographic and cultural issues. Crime variables are the main subjects of crime analysis. There are 3 general types of crime variables: spatial-temporal, crime natural specification, and offender profile (Keyvanpour et al., 2010). Meti (2016) described that social and economic factors can highly contribute for committing crimes. They have described that there is scarcity of crime investigation in Ethiopia. The researchers have conducted their research on analyzing socio-economic factors towards committing crime.

Crime rate can be reduced by predicting, so as to allocate resources like police force and patrol. For making crime prediction, crime pattern extraction is the base. Analyzing demographic and location attributes helps to extract crime patterns and criminal networks (Kalantari et al., 2016).

Artificial neural networks are the popular class of deep learning and applied for prediction of categorical data. The researcher of (Aitelbour et al., 2018), (Nguyen et al., 2017) have deployed ANN for crime prediction. Feed forward neural network is a type of ANN which have input layer as a first layer and output layer as last layer. It consists hidden layer(s) which do not have connection to the external world. Aitelbour et al. (2018) compared the performance of Naïve Bayes, Decision Tree, Random Forest, Neural Network, and SVM. The neural network outperforms the rest machine learning algorithms with accuracy 81%.

RNN is a type of deep learning. It usually applied for time-series forecasting. LSTM is a type of RNN which overcome the problem of vanishing gradient descent by having long memory. Tsion (2019) have applied LSTM-RNN for crime prediction. Other researches like Ajao et al., (2018), Stec & Klabjan (2018) used LSTM by combing with CNN.

Our dataset consists demographic, date, and location type features. ANN used for categorical type of data with better performance (Nguyen et al., 2017), (Aitelbour et al., 2018). LSTM-RNN is a recommended algorithm for location and time-series forecasting (Tsion, 2019). We have combined FFANN and LSTM-RNN. Kaliakatsos-papakostas et al., (2017), Makris et al., (2019) have also deployed this approach for predicting the drum sequence based on music rhythm.

1.2. Statement of the problem

Crime affects humans' life, economic growth, and general development (ToppiReddy et al., 2018). The trend shows that crime rate is increasing from time to time. Considering the number of crime scenes, crime coverage, and the sophistication of the crime makes it difficult for analyzing effectively with the currently available resources in police stations. Crime is a major issue which needs high priority by all government (Jabar et al., 2013). Crime can be reduced by prediction and taking preventive actions. According to Jabar et al. (2013) understanding the crime domain and having prior knowledge for the crime data will have a big view for crime analysis.

Crime variables are parameters that can describe characteristics of the crime. They are the main subject of crime prediction model (Keyvanpour et al., 2010). Crime record variables are different among countries even if some of them are common. A number of attempts has been made to design a model to empower the crime analysis and prediction process. As of the previous scholars mentioned, the availability of crime record and crime record variables variances makes it a challenge to have a suitable and efficient crime analysis model for Ethiopian crime records. There is a lack of modern approaches to predict and analyze crimes in Ethiopia (Tsion, 2019).

The crime prediction model's accuracy differs based on the attributes of the dataset (Endalew, 2017). There are researches that have developed crime prediction using

Ethiopian crime record dataset (Leul, 2003), (Endalew, 2017), (Tsion, 2019). The research done by (Leul, 2003) did not consider location-based data. But, according to ToppiReddy et al., (2018), location is a critical attribute for analyzing crime data. Tsion (2019) have studied the feasibility of LSTM-RNN model on crime dataset. They have used time and location data to predict the next crime type to be committed.

Kalantari et al. (2016) described that countries differ by their social, economic, demographic and cultural aspects. The aim of the research was finding the proper correlation between crime and urban land use. The correlation of crime and demographic factors differ from country to country. Accordingly, conducting a research to determine the correlation between crime and demography in Ethiopia is highly crucial.

Meti, (2016) conducted their research on whether socio-economic factors have impact on committing crime or not. The result of the study shows demography has high relation with committing crimes. We need to analyze criminals' demography to predict the crime type.

A research of Tsion (2019) showed that crime is time-based incident and can be predicted using time-series forecasting algorithms. A social study research by Meti (2016) which analyze the relation of criminals' demography with crime showed that crime is highly related with criminals' demography. The result showed that committing crime is highly influenced by demographic of criminals. So, we need to combine demographical data, time-series feature (date, month, year), and location in order to predict what type of crime will happen in the future.

LSTM algorithms are capable of predicting sequential data (Tsion, 2019), whereas FFANNs are good for categorical features (Nguyen et al., 2017). Combining FFANN and LSTM will be used to analyze both nature of attributes (Kaliakatsos-papakostas et al., 2017). The authors of Shermila et al., (2018) have recommended to improve the accuracy using complex Neural Networks and Recurrent Neural Networks. The following research questions are formulated and addressed in this research.

- How to develop crime prediction model for crime data records using demography, time and location factors?

- ➢ How to augment deep learning algorithms for crime data processing and structuring?
- ➢ How to integrate deep learning algorithms to enhance crime prediction model performance?

1.3. Objectives of the study

The general and specific objectives of this research will be described in this section.

1.3.1. General objective

The general objective of this research is to design and develop crime analysis and prediction model using demography, location and time factors.

1.3.2. Specific objectives

In order to achieve the main objective of the research work, we have formulated the following specific objectives:

- ✓ To design and develop crime prediction model.
- ✓ To augment deep learning algorithms for crime data analysis and structure.
- ✓ To enhance crime data analysis.
- ✓ To identify the determinant factors of crime.

1.4. Methodology

Design science is a type of information technology research methodology which focus on evaluating the performance of the outcome. It is a research paradigm where the creation of new artifact and evaluation of the artifact is a key contribution. In this research we have used process model designed by (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007) which has six phases. These are; problem identification, defining the objective, designing and development, demonstration, evaluation and communication. In order to achieve the objective of our research and answer the research questions formulated in the statement of the problem section, this research methodology is used.

1.4.1. Problem Identification

In this first phase of design science methodology, the research problem is defined and the value of a solution is justified. It is used to develop an artifact for the solution. Identifying the problem helps to have deep insight for the domain. We have reviewed different literatures and analyze our day to day activities to acquire knowledge about the state of the problem and the importance of the solution. Research works that have been done to predict crimes, so as to prevent before it happened.

For proper and deep understanding of the crime analysis and prediction process, we have employed systematic literature review in this research. Systematic literature review helps to understand and identify the problems in a clear manner. Systematic literature review is a type of literature review which uses systematic approach to identify and evaluate previous research studies (Tranfield et al., 2003). We have also reviewed several previous related work journals, articles, books, crime related documents and materials. Relevant documentation about tools and techniques for the model design and development have been reviewed and analyzed. Systematic literature review solves problems by identifying, critically evaluating, and integrating the findings of all relevant, and high-quality studies by addressing one or more research questions (Cruz-benito, 2016)

1.4.2. Defining Objective

The objectives of a solution are inferred from the problem specification. Different literatures have been reviewed to understand the state of the problem, the state of current solutions and state of the art. The objective of the research is to solve the problem mentioned by developing crime prediction model.

1.4.3. Designing and Development

In this section, the solution design is created and developed. This activity includes determining the artifacts of desired functionality and its architecture and then developing the actual model. Keras (using TensorFlow as backend) is used for designing the FFANN, LSTM and their hybrid model. Python is used for writing the required source codes.

1.4.4. Demonstration

The developed system is demonstrated by ability to predict crime types based on a given time, location and demographic features. We have used Anaconda tool, Spyder editor with Python language to develop the model.

1.4.5. Evaluation

The developed system is evaluated to measure how well it supports a solution to the problem. To evaluate the system in a rational method, testing datasets were fed into the developed model. Subsequently, the model was evaluated by comparing its output against the observed data using accuracy, MAE and MSE.

1.4.6. Communication

In this section, the problems, the designed and developed solution, the result and other related information are communicated to audiences using different mechanisms.

1.5. Scope

This research is about designing and developing a model for crime analysis and prediction. Crime record dataset of Bahir Dar city from September 2005 up to April 2011 of E.C was used for building and evaluating the model. We have evaluated feed forward artificial neural network, LSTM-RNN, and the hybrid of feed forward artificial neural network and LSTM-RNN.

According to Ethiopian case, there are three major types of crimes. Traffic crime, crimes that can be treated with social law, and crimes which treated by criminal law. In this research we have considered those crimes stated under criminal law.

The research did not consider forensic crime data and social network analysis of the criminals.

1.6. Significance of the Study

This research will help law enforcement bodies by providing knowledge about the pattern of crimes to know what will be the next move of the criminal. It will help them by predicting what type of crime will happen in the future. Law enforcement bodies will also be beneficiary from the crime analysis and prediction. It will assist the police department by optimizing the resource allocation like police force and patrol. Identifying crime hotspot area helps to take preventive actions. Analyzing crime helps to reduce crime occurrences.

Having a city with reduced number of crimes will be comfortable for living and working. This helps the country by increasing investment and economy flow.

The output of this research work can contribute to other future scholars who want to do their research on this domain and act as a base to do further improvement on the model or the techniques this research has followed.

1.7. Thesis Organization

The remaining part of the thesis is prepared as follows. Chapter two presents literature review on definition of crime, crime types, crime analysis, crime prevention and prediction techniques. Chapter three the methodologies that we have used to accomplish this thesis is discussed. It includes data collection, data preprocessing and prediction methods. Chapter four discuss the results of the designed model. Results of the experiment are also analyzed and interpreted. Conclusion and recommendation for future is presented in chapter five.

CHAPTER TWO
LITERATURE REVIEW

This section provides the detail literature review work, which consists of the aim and focus of the current research. Crime and its types have discussed followed by the process of crime investigation and analysis approaches followed. Finally, we provide some of the related research works done by the previous scholars.

2.1. Introduction

Usually, law enforcements, citizens and property protection, and control civil disorder has been entrusted to the police commissions that is empowered by the state governments. For such commission to be more effective in its duties, Information is the vital component regarding catching the criminals and investigating what did they do wrong. And again, this information also helps to in force law and impeachment. Through the process the police force had the power to legitimize the use of force.

Especially law enforcement involves mainly on the policing activities, which may include developing strategies to answer civil society request call for help, identify suspects, search for criminals who are at large, collect evidence on the crime scene, investigate the crime and finally up to make the case to the court. A suspect here means a person who is accused or suspected that he/she committed the crime. Common problems faced by police officers is predicting the future crime so as to prevent the crime before it occurred. Criminal investigation and law enforcement bodies should be enabled with technology to bring solutions for their struggles and prevention activities of organized crime by identifying and targeting players of a crime done with complex network.

To this end, different approaches have been applied for crime prevention that focus in the intervention, types of activities that are organized, and the mechanisms that are applied to deal with environmental, social and economic, and justice system of a country.

2.2. Criminology

Criminology is the logical study of crimes including its characteristics, causes, rectification, anticipation, behavior of the criminal, and the social impact of crime (Yamuna & Bhuvaneswari, 2012). Wrongdoing investigation is an critical portion of criminology (Mowafy et al., 2018). Crime prediction is also the central portion of criminology. It is the scientific study of the nature, extent, causes, and control of criminal behavior in both the individual and in society (Thotakura, 2011).

Crime is an offense which violates the law of state and is un acceptable act by the society. The crime rate is increasing from time to time alarmingly. This increase in crime rate may be due to various causes and social problems. Crime is a public wrong. It is an act of offense which violates the law of the state and is strongly disapproved by the community. It can be defined as acts or omissions forbidden by law that can be punished by imprisonment or other type of punishment. It is derived from the Latin word “Crimen”, It represents offence and also a wrong-doer. Crime is considered as an anti-social behavior (Thotakura, 2011).

According to Thotakura (2011) definition, for the crime to be accomplished there should be an intention of committing crime, guilty mind, guilty act and an injury which can be physical, mental or property that violates a law of state. Intention, preparing for committing crimes, attempt to commit crime, and commission of crime are the four stages of committing crime. The definition of crime has been expanded and also contracted over time because its nature is more dynamic, complex, depends on social construction and contingent reality. Crime happens everywhere, but it varies over time and place.

Basically, the criminal justice and law enforcement bodies of the state departments are empowered to use any legitimate approach on solving a crime. As the crime rates and crime sophistication increases, these bodies should also consider all other options on the table, not only solving the already happened crime but also to predict the upcoming crimes with the support of technology. Technologies such as artificial intelligence brings concept and tools so that the police force within the state can be assisted to protect citizens and property theft or damage before it happened at all. For this, it is crucial to apply criminology techniques to assess the occurrence of crime. Criminology is the process of finding crime and understanding criminal characteristics. Understanding and assessing characteristics of crimes and criminals is a step in the process of crime analysis.

2.2.1. Types of Crime

Crime types differ among different disciplines and countries. According to Ethiopian law basically there are two laws for punishing crimes. These are; criminal law and social law(Hassen, 2014), there is additional type of crime which is traffic crime. social law is issues which is related to property. A person who accused in social law is not expected to be punished in prison or other types of punishments rather he/she will pay for the property. Criminal law deals with crimes which make the criminals to be punished in prison or other punishments. In addition to those two crimes there is a traffic crime which has subcategories under it. Traffic rule violation and traffic accident are types of traffic crime. Our research is based on crime types which is listed under criminal law.

According to Chen et al. (2004) there are three general types of crimes. These are violent or local crime, global crime, and internet crime. Violent crimes include crimes like murder, robbery, forcible rape, and aggravated assault. Whereas terrorism is a type of global crime. The Internet's pervasiveness likewise makes identity theft, network intrusion, and cyberpiracy.

There are eight general types of crimes (Thotakura, 2011). These are Personal crime, Property crime, Victimless crime, White collar crime, Organized crime, Juvenile delinquency, Computer crime, and Violation of public safety crime. Personal crimes are crimes which target an individual person. It includes assault, homicide, and sexual assault. The target of property crime is materialistic property. Burglary, theft, arson fires, automobile theft, and vandalism are some of the examples of property crime. Victimless crimes are acts which is against moral values of an individual. It includes crimes like prostitution, illegal gambling, illegal drug use, and the like. These crimes do not have an identifiable victim. Embezzlement, fraud, identity theft, and corruption are types of white-collar crimes. Organized crimes defined as acts which are committed by two or more criminals in an organized manner. It includes kidnapping, dacoities, marketing of illegal or prohibited goods, money laundering, trafficking people, buying votes, etc, Juvenile delinquency is also known as youth crime. It is the crime committed by an individual who is under the age of 18 years. A crime which uses computer and network is known as computer crime. Violation of public safety is violation of laws which threaten public safety. Terrorism is the major example.

Meti (2016) Have described that there are four types of crimes. These are crime of violence, crime against property, crime against the state, and victimless crimes. As of the researchers' description violence crimes includes murder, aggravated assault, forcible rape, abduction, kidnapping, armed robbery and burglary. Crimes against property are crimes which are committed to get wealth in illegal ways. Shoplifting, corruption, embezzlement, stealing and gambling are the common example of property crimes. Crimes against the state are crimes which can be done against the government and crime by the government. Protesting the structure of the government, treason, sedition, and the like are crimes against the government. Whereas violation of citizens' right and civil liberties are crimes that can be done by the government. Victimless crimes are crimes which do not have victims. It includes drug abuse, prostitution, gambling, homosexuality, drunkenness (alcoholism) and Vagrancy.

According to Hagen (2010) there are 7 general types of crimes. These are violent crimes, property crimes, white-collar crimes, political crimes and terrorism, organized crimes, public order crime, and computer crime. Violent crimes are crimes which harms the victim with violence. It includes rape, sexual assault, assault, robbery and murder. Property crimes are the victims' property is stolen or destroyed by the offenders without physical force on the victims. It includes theft, burglary, vandalism, shoplifting and arson. White-collar crimes are financial motivated crimes and they are non-violent crimes. Fraud, embezzlement, tax evasion and money laundering are types of white-collar crimes. Organized crimes are crimes which is committed by a group of criminals. Kidnaping for ransom, and gambling are examples of organized crimes. Public order crimes are wrongdoing which includes acts that meddled with the operations of society and the capacity of individuals to operate efficiently. Computer crime is also known as cyber-crime, which includes utilizing computer and arrange to assault the casualties. Cracking, creating malware, cybersquatting, denial of service, identity theft, phishing, and spam are examples of computer crime. Political crime and terrorism are an offense including plain acts or exclusions (where there's obligation to act), which partiality the interface of the state, its government, or the political framework. It is to be distinguished from state crime, in which it is the states that break both their own criminal laws or public international law.

2.2.2. Crime analysis

Crime investigation is a set of orderly and expository forms at giving coordinated data and analytic processes directed at providing timely and pertinent information about crime patterns and trends. Crime analysis supports police operations through different strategy planning, manpower deployment and investigation assistance. It supports police and criminal justice planning effort, provides data for program and tactical evaluation.

Crime analysis provides benefits to the police through in different circumstances. Crime analysis should:

- Assist in the establishing, screening and ordering lists of suspects for individual crimes
- Assist in assembling and ordering the case and specific crimes which may involve a suspect already in custody.
- Assist in the proper assignment and deployment of preventive patrols and other police functions in order to assure police observation of crime in progress.

Among different types of crime analysis, tactical, strategic, and administrative crime analysis are the popular (Gottlieb & Arenberg, 2000). There are also other types of crime analysis such as linkage analysis, statistical analysis, profiling analysis, and spatial analysis. During suspect investigation, the predefined behaviors may vary, some are included and some others are excluded from the list. Tactical crime analysis gives data to help operations work force within the distinguishing proof of particular wrongdoing issues and within the capture of criminal offenders. Strategic crime analysis examination outfits data concerning long-range issues. It also used to have statistical summary. Administrative crime analysis focuses on the provision of economic, geographic or social information to administrators.

2.3. Deep Learning

Deep learning can be defined as neural network with more than two layers. It has some evolution on simple neural network such as; consisting more neurons, more complex ways of connecting layers, and automatic feature extraction (Patterson & Gibson, 2017).

Generally deep learning is a sub-disciplines of machine learning that uses a hierarchical level of artificial neural networks to carry out the process of problem solving. The artificial neural networks are built like the human brain, with neuron nodes associated together like a web. Traditional programs analyze data in a linear way, the hierarchical function of deep learning systems empowers machines to process data with a nonlinear approach.

2.3.1. Artificial Neural Network

ANN is a computing system which is motivated by the biological neural networks of animal brains in which any simple units are working in parallel with no centralized control unit. The weights between the units helps to store information with in the neural networks. Neural networks will learn new information and patterns by updating their weights (Patterson & Gibson, 2017). ANNs learn to perform tasks by considering examples, generally without being programmed with task-specific rules.

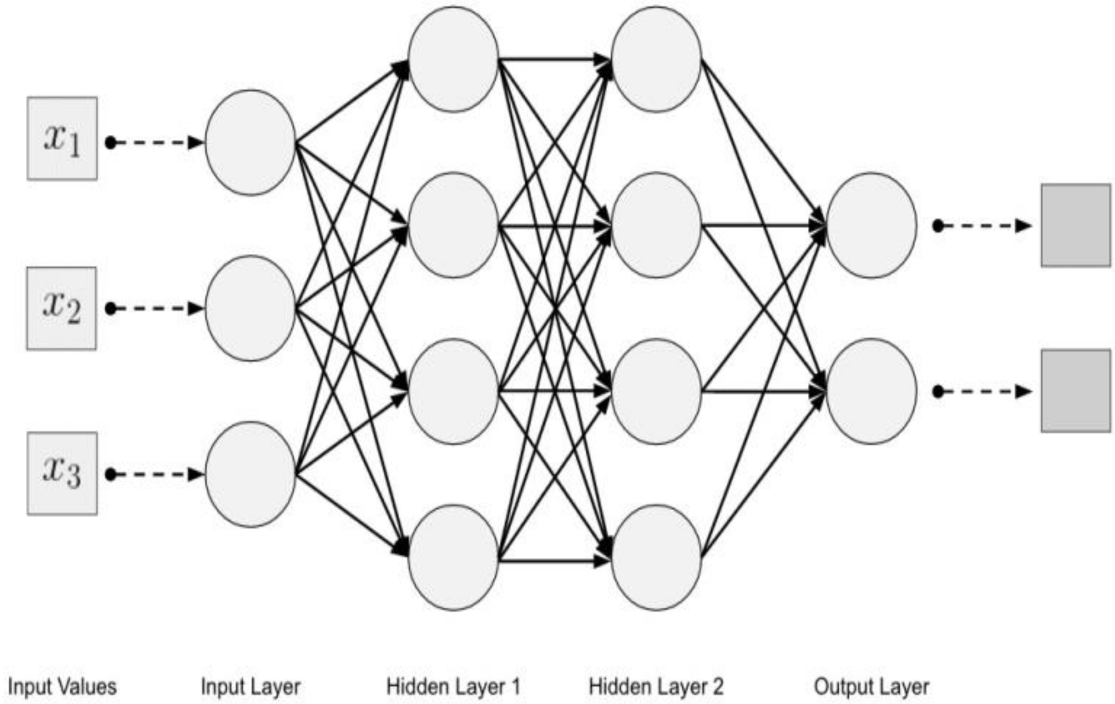

Figure 2.1: Multilayer Artificial Neural Network (Patterson & Gibson, 2017)

The above figure 2.1, shows feed forward artificial neural network; it is also known as multilayer perceptron. It has inputs from the external world. It consists input layer, hidden layers and output layer. There is also output which is out from the output layer. The function of the input layer to send signal for the hidden layers. The hidden layers will do computational analysis and send the result to the output layer.

ANN consists computing devices called neurons that are connected to each other in a complex communication network, through which the brain is able to carry out highly complex computations.

The behavior of neural networks is shaped by its network architecture. A network's architecture can be defined by: number of neurons, number of layers and types of connections between layers.

2.3.2. Feedforward Artificial Neural Network

Feed forward artificial neural network is a type of artificial neural network where connections between nodes do not create a cycle. They are arranged in layers form where the first layer is input layer and the last layer is output layer. In the middle we can use hidden layers which helps to improve the performance of the learning model. These hidden layers do not have connection to the external world. If a neural network consists of hidden layers then called multilayer perceptron or deep feed forward artificial neural network or simply deep neural network. Multilayer perceptron uses variety of learning techniques, the most popular is backpropagation method. Backpropagation uses gradient descent on the weights of the connections to minimize the error on the output of the network.

It is the foremost known and easy-to-understand (Patterson & Gibson, 2017). Each layer can consist a different number of neurons and each layer is fully connected to the next layer.

$$y_l = p_l(w_l p_{l-1}(w_{l-1} x_{l-1} + \ b_{l-1}) + b_t \qquad (2.1)$$

Where,
y_l , output signal

p_l , is a vector activation function

w_l , is weight from each neuron in a layer

x_l , is the input signal from l-1 and,

b_l, is an arbitrary offset vector (bias)

The performance of the prediction will be determined by the correct values of the weights and biases. The method of fine-tuning the weights and biases from the input data is known as training the Neural Network. Each iteration of the training process will have the following steps:

- Calculating the predicted output ŷ, known as feedforward

- Updating the weights and biases, known as backpropagation.

2.3.3. Recurrent Neural Network

Recurrent Neural Network is a type of neural network where the output from previous step are fed as input to the current step. In traditional neural networks, all of the inputs and outputs are independent of each other, but in cases like when it is required to predict the next data of a given dataset, the previous data are required and it will be mandatory to remember the previous data. Thus, RNN came into existence, which solved this issue with the help of a hidden layer. The basic and most important feature of RNN is Hidden state, which remembers some information about a sequence (Medsker & Jain, 2001). RNN is useful in time-series prediction, speech synthesis, natural language processing and other tasks. This RNN has drawback of exploding and vanishing gradient descent.

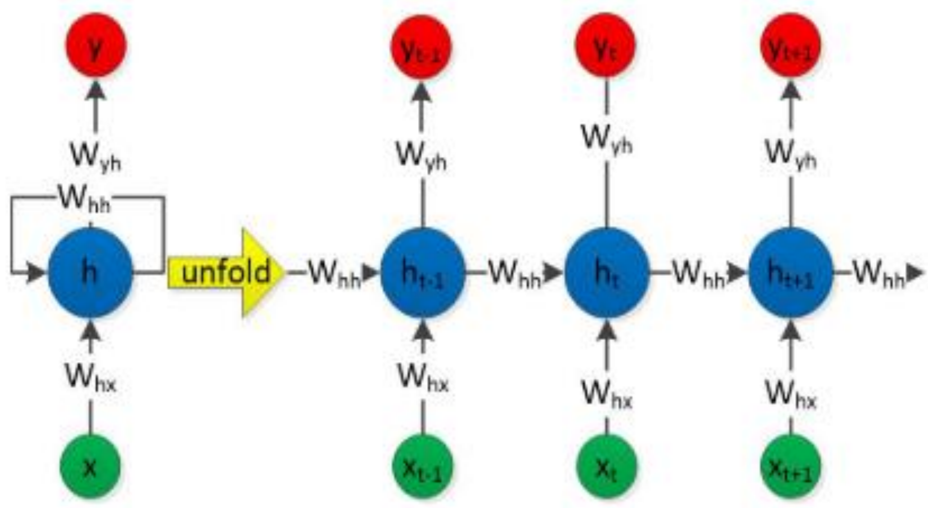

Figure 2.2: Recurrent neural network (Zheng et al., 2017) shows the architecture of RNN.

Where:

x: is input

h: hidden state

y: output

w: weight

The RNN cell takes two different inputs; the first one is the outputs of the last hidden state and the second one is observation at time t in the hidden state. Then the RNN will predict the next output at a time t+1.

RNN is a class of artificial neural networks where connections between nodes form a directed graph along with a temporal sequence. This allows it to exhibit temporal dynamic behavior.

RNN takes input from x (input variable) and from the previous hidden state. This behavior helps to remember the previous state and to predict the next outcome (temporal prediction) even if it suffers from exploding and vanishing gradient descent.

2.3.4. Long Short-Term Memory

LSTM is a special kind of RNN which is capable of learning long-term dependencies. It has feedback connections. It can also process the entire sequence of the data. The drawbacks of RNN has solved with LSTM. LSTM has long memory to remember previous activities and sequences. It helps in preventing errors that comes with backpropagation through time and layers (Brownlee, 2017). The popular way to train RNN is backpropagation through time. This makes the problem of the vanishing gradients, which causes the parameters to capture short-term dependencies while the information from earlier time steps decays. the reverse of vanishing gradient descent is exploding gradients which causes the error to grow drastically with each time step. By using the gates to selectively retain information that is relevant and forget information which is not relevant. LSTM overcome the problem of vanishing gradient descent. Lower sensitivity to the time gap makes LSTM networks better for analysis of sequential data than simple RNNs. LSTMs holds information separately from the normal flow of the RNN in a gated cell. Information can be kept in, read or written from a cell, like data in a computer's memory. The cell makes judgements via gates that open and close about, when to allow reads and writes, what to store, and removals. Yet, these gates are analog, applied with element-wise duplication by sigmoid, unlike the digital storage on computers, which are all in the range of 0-1. LSTMs excel in predicting, classifying, learning, and processing sequential data. It can be applied on video analysis, caption generation, weather forecasting, stock market prediction and the like (Greff et al., 2017).

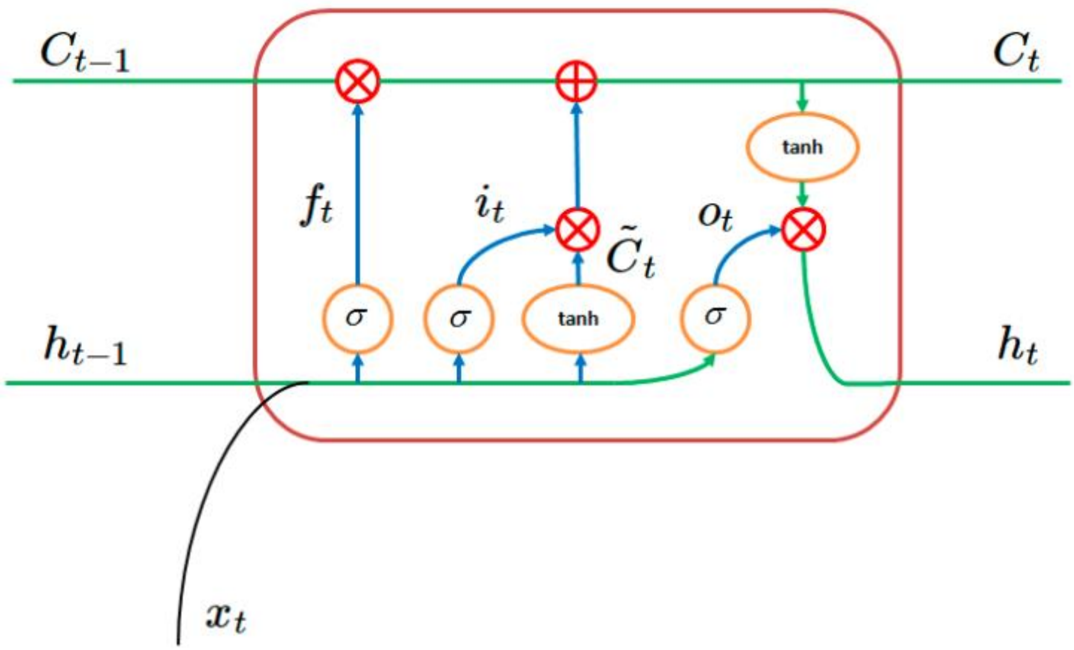

Figure 2.3: LSTM-RNN (Tran, 2016) shows the architecture of LSTM-RNN.

Where;

x_t: input at time t

h_t: hidden state at time t

c_t: cell state at time t

f_t: forget gate at time t

i_t: input gate at time t

o_t: output gate at time t

LSTM-RNN is a type of RNN which have LSTM cells as neurons in some of their layers. Much like Convolutional Layers assist a neural network to learn about image features, LSTM cells assist the neural network to learn about temporal features in data, it was a major struggle which other machine learning models traditionally was tried to achieve.

LSTM Neural Networks takes a whole matrix as its input. Generally, LSTM-RNN will have two inputs. These are; the previous LSTM cell's output (which gives it some information about the previous input) and its own input column.

The weights and biases to the input gate control the extent to which a new value flows into the cell. The weights and biases to the forget gate control the extent to which a value remains in the cell. The weights and biases to the output gate control the extent to which the value in the cell is used to compute the output activation of the LSTM block (Brownlee, 2017).

Cell state is usually called long-term memory. The looping arrows that allows information from previous to be stored in LSTM cell indicate the recursive nature of the cell. (Hussein et al., 2018).

$$c_t = f_t * c_{t-1} + i_t * \hat{c}_t \tag{2.2}$$

The forget gate modifies the cell state and adjust by the input modulation gate. It realizes that there might be change in the context after encounter its first loop.

$$f_t = \sigma(w_f[h_{t-1}, x_t]+b_{t)} \tag{2.3}$$

The input gates determine which information should enter to the cell state. The input gate has a range of [0, 1] and is a sigmoid function. The sigmoid function will only add memory and not to be able to remove/forget memory.

$$i_t = \sigma(w_i[h_{t-1}, x_t]+b_{i)} \tag{2.4}$$

$$\hat{c}_t = tanh(w_c[h_{t-1}, x_t]+b_{c)} \tag{2.5}$$

Output gate that is usually called focus vector highlights which information should go to the next hidden state.

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (2.6)$$

$$h_t = o_t * tanh(c_t) \quad (2.7)$$

2.3.5. Deep Learning approaches to predict crime

Stec & Klabjan (2018) have used deep learning approach for predicting crimes. The authors combined CNN and RNN. The RNN used for predicting temporal crime data, whereas the CNN used for predicting spatial attributes. Chicago and Portland crime data were used for evaluating the model. Crime area, public transportation, weather, and census data were used as features. The authors compare the performance of feed forward ANN, CNN, RNN, and the hybrid of CNN and RNN. As of their experiment the combined approach outperforms other algorithms which were evaluated in the experiment.

Predictive analysis of crimes has done using RNN (Krishnan et al., 2018). They have used Indian crime dataset. And they have also discussed that they have got better result than of the previous research studies. DNN using feature-level data fusion approach was used to predict crime occurrences (Kang & Kang, 2017). They have used multiple datasets to forecast crimes using different and relevant features. Demographic, housing, economical information, education, crime type, spatial, temporal and environmental data are the some of the features from their dataset.

Stalidis et al., (2018) evaluated the performance of 13 different deep learning algorithms. The researchers have used spatial and temporal data temporal data to build the model with respect to crime types. Similar to Stec & Klabjan (2018) they have also deployed LSTM for time-series forecasting and CNN for spatial visualization. Philadelphia, Minneapolis, Seattle, DC Metro, and San Francisco crime reports were used for model building. Three different experiments were conducted in the research. By considering temporal data first, and spatial data second, considering spatial data first and temporal data second, and lastly by making spatial and temporal data parallel.

Real time crime forecasting (B. Wang et al., 2017) using deep RNN for the temporal prediction and CNN for spatial. Los Angeles crime record was used including all types of crimes. As we have discussed above, researchers usually combine RNN, especially LSTM with other algorithms.

2.4. Techniques in crime analysis and prediction

Generally there are different techniques to predict and analyze crime. (Leul, 2003) have developed crime prediction using Oromia region crime record dataset. They compared two algorithms and predict crimes based on criminals' gender. The research did not consider location-based crime prediction. Research developed by Saeed et al. (2015) also compared two algorithms. The result showed that Naïve Bayes' is better than decision tree. The researchers have used UCI machine learning repository dataset having 2215 instances with 128 attributes. Whereas (Iqbal et al., 2013) have compared Naïve Bayes' and Decision tree classifier for prediction and the experiment have developed using USA crime dataset. The result shows Decision tree outperforms Naïve Bayes' classifier. This research has selected 12 attributes from 128 and have used 1994 instances.

Crime prediction using Bahir Dar city crime record was developed by (Endalew, 2017) and got less accuracy than the accuracy of researches that have done in other countries with the same algorithm. The researcher has used 7 attributes among the available 15 with 3000 records and described that some of the rest attributes had missing values. The researcher has selected to ignore these attributes rather filling the missing values. This research has described that the crime record attributes of Ethiopian is not suitable for crime analysis models even if it does not list what attributes should be added. This research did not consider specific location analysis of the crime. According to Leul, (2003), Tsion, (2019) crimes can be predicted with good performance, even if the attribute number differ.

The authors of Chen et al. (2004) have developed a general framework that can identifies relationships between techniques applied in criminal and intelligence analysis. The framework mainly consists 3 data mining tasks, these are: named entity extraction, detecting deceptive criminal identity, and criminal network analysis. The authors have used modified version of artificial intelligence entity extractor to extract named entity, string comparator for deceptive detection and concept space approach for criminal network analysis. The authors have analyzed static criminal network and recommend to use the dynamic network analysis.

A framework proposed to analyze crime data using decision tree and K-means algorithm (Al-janabi, 2011). This research tends to extract patterns, predict crimes, analyze criminal networks, and identify possible suspects. The authors have used classification based on crime type, and clustering based on crime and criminal attributes. A dataset to build and evaluate the model have used from open source repository. The dataset had 11 attributes, but the authors have used 9 of them by discarding crime name and location. Even though, location is a critical attribute for analyzing crime data (ToppiReddy et al., 2018).

Jabar et al. (2013) proposed a model that helps for crime analysis using genetic algorithm and association rule. The authors have used 24 types of attributes from crime, criminal and geo-crime datasets. They have used a dataset which is available on Sheriff's office. The data from 2003 G.C. to 2010 G.C. have used to build model and the data from 2010 G.C. to 2013 G.C. have used for testing the model.

Bang et al. (2017) have designed data mining framework that helps to extract crime patterns using unsupervised learning. The framework has been designed based on 2 crime types; these are assault and theft. The framework was designed based on data-driven approach. It has been evaluated using real world crime dataset of Seoul, in Republic of Korea.

Nasridinov et al., (2013) have developed crime prediction tool which employed classification-based technique. However, the research was conducted without the use of city map to assist location-based data. There are some researches which holds both demographic and location-based crime records (X. Wang et al., 2012). The research has considered location-based, demographic data, and social media analysis to predict crimes.

Mowafy et al. (2018) proposed a framework which integrates text mining technique with the crime analysis process. The proposed crime mining framework consists 6 components. These are: gathering unstructured crime data, applying a proper text mining technique, dissemination and feedback. The authors have described that applying crime analysis on unstructured data will reduce performance of the applied algorithm. Text mining should be applied on unstructured crime record.

ToppiReddy et al. (2018) have also deployed a framework that uses different visualization techniques to show the trend of crimes and to predict the crime using machine learning methods. Web mapping & visualization-based crime prediction tool are the major tools they have used for visualization. The tool has six modules, these are: visualization of crime data using google maps, visualization of exact location of crime with 3D view, visualization based on type of crime, visualization of crime hotspots, crime frequency report, and interactive crime frequency report using graph and bar chart. The framework is based on the spatial analysis, using K nearest neighbor and Naïve Bayes algorithm. This research has used 5 attributes from the available 11 attributes. The attributes are: crime type, location, date, latitude, and longitude. This research enables to locate where the crime will happen using longitude and latitude lines, it did not consider the demographic factors of the criminals.

Crime prediction model which predicts the to be committed crime type in specific location and time using Portland police bureau crime records built by (Nguyen et al., 2017). The researchers have also used demographical data like educational background, economical and ethnic background of the people in that area. There are different types of machine learning algorithms. Among these, the authors have used and compared Support Vector Machine (SVM), Random Forest, Gradient Boosting Machines, and Neural Networks. Effective crime prediction should predict crime type, location, and time and also expected to analyze demographic data. SVM and ANN performs well on the second dataset, which includes demography. Shermila et al. (2018) predicts the criminals' age, gender and their relationship with victims of the crime. The authors have used Multilinear Regression, K-Neighbors Classifier and Neural Networks with San Francisco Homicide dataset. The criminals' age predicted using Multilinear Regression, whereas criminals' gender and their relationship with victims predicted using K-Neighbors and Neural Networks.

The performance of K-Neighbor classifier, GaussianNB, MultinomialNB, BernoulliNB, SVM and Decision Tree was compared in a research of (Bharati & RA.K, 2018). K-

Neighbor and Decision Tree outperform the rest algorithms. Chicago's city crime record was used to build the model. The authors have used location data and in line with crime types.

2.5. Related works

The researchers of Aitelbour et al. (2018) used machine learning algorithms to predict the crime type based on spatial and temporal features of the crime record. They have compared Naïve Bayes, Decision Tree, Random Forest, Neural Network, and SVM using the same dataset. The Naïve Bayes and SVM was not fitting with their data format. Whereas, Decision Tree, Random Forest, Neural Network achieves 38%, 59%, and 81% respectively.

Shermila et al. (2018) have used Multilinear Regression to predict the criminals' age. The researchers intended to predict who will be the possible criminals and the relationship with victims. They have deployed K-Neighbors and Neural Networks to predict criminals' gender and their relationship with victims. Demographic data, time and location is used to predict crime types by (Nguyen et al., 2017). The authors used SVM, Random forest, Gradient boosting machine, and ANN to predict crime types.

LSTM-RNN is used to predict time-series data. A research of Krishnan et al. (2018) were developed to predict count of crimes to be happened in future months and years. They have employed LSTM on Indian crime records. A research conducted using Addis Ababa, Bole sub-city crime records Tsion (2019) used to predict crime type on specific time and location. The researchers have used LSTM method for developing the model. From different features of crime records the researchers used time, date, location and crime types to feed for the model. The research showed that crime has a time-series nature and can be predicted using time-series forecasting algorithms (Stec & Klabjan, 2018).

Crime hotspot prediction using recurrent neural network (Zhuang et al., 2017). The authors used spatio-temporal neural network to predict crimes and recommend to use demographic data in addition to time variables. Researchers has also used recurrent neural network with convolutional neural networks to predict hate speech from tweets (B, Robinson, & Tepper, 2018). They deployed GRU and CNN to detect hate speech in social media.

The approach to combine LSTM with other algorithm has conducted by (Gensler et al., 2016). The researchers used the combination of LSTM with Autoencoder to predict renewable energy power plants. German solar farm dataset has been used to measure the performance of the Auto-LSTM algorithm. Makris et al. (2019), Kaliakatsos-papakostas et al. (2017) combined feed forward artificial neural network and LSTM-RNN for music and rhythm composition. LSTM algorithm combined with CNN is also used to identify fake news on twitter (Ajao et al., 2018). The flexible nature of a neural network used to combine LSTM with CNN (Stec & Klabjan, 2018). The authors feed spatial features for the CNN and temporal features for the LSTM. Their outputs feed to the next hidden layer then to produce the result on output layer.

Table 2.1: Summary of related works

No.	Research Group	Objective	Algorithm Used	Remark
1.	Aitelbour et al., (2018)	Predicting crime based on specific time and location.	Naïve Bayes, Decision tree, Random forest, ANN, SVM was compared.	Time-series algorithms did not consider.
2.	Shermila et al., (2018)	Predicting possible suspects and identify the relationship between suspect and victims.	Multilinear regression, K-Neighbor, and Neural Networks.	Crime type is not predicted.
3.	Nguyen et al., (2017)	To predict crimes based on time, location and demography data.	SVM, Random forest, Gradient boosting machine, and ANN	Did not use time-series forecasting algorithms.
4.	Krishnan et al., (2018)	Predicting crime counts to be happened in the future.	LSTM	Crime types and location are not considered.
5.	Tsion, (2019)	To apply LSTM and analyze its feasibility on crime data.	LSTM	The research considered location and time data only.
6.	Zhuang et al., (2017)	Predict crimes based on spatial and temporal data.	Spatio-temporal neural network.	Other features of crime record are not considered.
7.	K. Wang et al., (2017)	Predicting quantities of crimes based on time and location.	LSTM, GRU-LSTM	Demography did not consider and focused on quantities of crimes.

8.	Kaliakatsos-papakostas et al., (2017), and Makris et al., (2019)	Predicting drum's sequence.	FFANN and LSTM	It was deployed for music composition.
9.	Stec & Klabjan, (2018)	To predict crimes based on spatial and temporal data.	RNN and CNN, also compared result with ANN	Demography data did not consider.

All of the previous works did not use hybrid of FFANN and LSTM for crime prediction.

CHAPTER THREE
RESEARCH METHODOLOGY

This chapter discusses about the research methods and methodology that we have followed. The description of data collection, data preparation or pre-processing tasks of the data employed in the study, data analysis, model development, model evaluation mechanism and the tools that we have used to do the research will be discussed in this chapter.

3.1. Data collection

In order to accomplish this research, we have used crime record dataset of Bahir Dar city. The dataset collected from the 7 police stations of the city. The data were available in manual record which later on converted to electronic format. We have gathered the dataset using document analysis technique. We have gathered and used primary data for this research. Even though there are crime records starting from a long time ago, we have used crime records starting from September 2005 up to April 2011 of E.C. because of time constraint. The time of the dataset is selected randomly. We have gathered 11,363 records with 15 attributes. The attributes are age of criminals', gender of criminals', marital status of criminals', job of criminals', educational status (educational background) of the criminals, crime type, kebele, zone of kebele, time, date, month, year on which the crime has committed, age, gender, and job of the victims.

3.2. Data Preparation

The collected data from real world dataset was recorded manually. We have encoded it and made in electronic format. The data was not ready in order to be processed using deep learning algorithms. These data contain incomplete, noisy and inconsistent data which can reduce the performance of the algorithm. The basic tasks of data preparing include attribute selection, data cleaning and transformation of data for modelling tools (Endalew, 2017).

From the available total attributes, we have selected 15 of them. For the sake of privacy, we did not take name, phone number, and address of both the criminals' and the victims, as well as birth place of the criminals.

Among the collected data 80 of the criminals' age column are empty. 3 of the criminals' gender column is empty. Among the criminal's marital status column 10 of them are empty. 27 of the recorded criminals' job record are empty. Among the criminals' educational status column 128 of them are empty. 4 of the criminal records hold empty crime type attributes. There are 36 records in which the kebele record are empty. Among the kebele-zone 5699 of them are empty. Kebele-zone used to describe sub-area under kebele. 2263 of the time columns are empty and 3002 records of the time column was also recorded in a general type of description like, in day or in night. 107 of the date columns are empty. 45 of the month columns are empty. 16 of the year columns are empty. The victims that are available on the dataset are individual persons, organization and the law of the country. There are 2184 crime record on which the victims are law. Whereas there are 789 crime record on which the crime has committed on organizations. The rest 8280 of the crime records are crimes that are committed on individuals. There are 110 empty records of victims' age. 1 record from the victims' gender is empty. Among the collected data 126 records of the victim job column are empty.

Ethiopia has 13 months and the crime records were also recorded using these 13 months formats. In order to make appropriate for the data analysis, we have changed the 13^{th} month data to the 12^{th} month.

For this research we used deep learning algorithms which needs clear training and testing data. Chen et al. (2004) described that missing data would limit the prediction accuracy.

Table 3.1: List and description of attributes

Attribute name	Description	Type
Criminal's age	The age of the criminal	Number
Criminal's gender	The gender of the criminal	Text
Criminal's marital status	The marital status of the criminal	Text

Criminal's job	The job of the criminal	Text
Criminal's educational status	The educational status of the criminal	Number and text
Crime type	The crime type committed	Text
Kebele	The kebele in which the crime has committed	Number
Kebele-zone	Zone of the kebele in which the crime has committed	Text
Time	The time on which the crime has committed	Number and text
Date	The date in which the crime has committed	Number
Month	The month in which the crime has committed	Number
Year	The year in which the crime has committed	Number
Victim's age	The age of the victim	Number
Victim's gender	The gender of the victim	Text
Victim's job	The job of the victim	Text

As the above table shown the types of the data, it consisted data in different format which later on need to convert to the same data format.

Table 3.2: List of attributes and missed values

Attribute name	**Missed value**
Criminal's age	80
Criminal's gender	3
Criminal's marital status	10
Criminal's job	27
Criminal's educational status	128

Crime type	4
Kebele	36
Kebele zone	5699
Time	2263
Date	107
Month	45
Year	16
Victim’s age	110
Victim’s gender	1
Victim’s job	126

3.3. Data preprocessing

We have 15 attributes in our dataset. In order to build a model a data should be preprocessed. According to Bray & Ferlay (2005) they have classified age groups into 18 groups. 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, and 85 and above. We have grouped age of criminals and victims based on this standardization. Our dataset consists criminals from age 11 to 85 and victims age from 0 to 95. Here 0 represents age, which is below 1 year.

As of the international standard classification of occupations (International Labour Organization, 2012), there are 10 major different types of occupations in which they consist sub-major, minor, and unit groups of occupations under them. These are major group 1- Managers, major group 2- professionals, major group 3- technicians and associate professionals, major group 4- clerical support workers, major group 5- services and sales workers, major group 6- skilled agriculture, forestry and fishery workers, major group 7- craft and related trades workers, major group 8- plant and machine operators and assemblers, major group 9- elementary occupations, and major group 0- armed forces occupations. This international standard classification of occupations did not consider students. Therefore, we have prepared additional group which holds students and housewife. Because of they do not have direct income from works. As of the Meti, (2016) researches economical background have direct relationship with committing crimes. There are 51 different types of occupations in our dataset for both the criminals and victims of the crime. We have used ISCO to group occupations of criminals and victims.

Our dataset consists different datatypes like number data, and text data, and combination of text and numbers. To make the model building efficient we have converted the text data to numerical format. After converting those to the similar format, our dataset contained only numerical records.

For the Age column we have grouped them into 18 groups within 5-years gaps (Bray & Ferlay, 2005). For gender column, we have two values male and female, we have assigned 0 for male and 1 for female. For the marital status column, we have four different values married, unmarried, divorced and widowed. We have assigned 0 for married, 1 for unmarried, 2 for widowed and 3 for divorced. For the job column, we have grouped them into 10 major occupation types. There are 10 major types of occupation. These values later on converted to between 0 and 1 using data normalizing mechanism.

For education status column we have used international standard classification of education 2011 (*The International Standard Classification of Education (ISCED) 2011*). There are 10 educational levels. From level 0 up to level 9. For crime type column, there are 82 different types of crimes recorded in the dataset. We have grouped them into 4 major crime types based on Meti, (2016). We have assigned 0 for VC, 1 for PC, 2 for SC, and 3 for VCC in which the data later on normalized and converted to values between 0 and 1.

Kebele has recorded using number data, therefore we did not need to convert it. There are 17 kebeles in Bahir Dar city. We dropped kebele-zone column since 5699 records of it are empty. Time was recorded using combination of text and number. This column consisted 2263 empty cells and 3002 of them are recorded in a general format like, day or night. Therefore, we have also dropped time column. Date, month, year were recorded with numerical value, so we did not change the data format for these columns.

3.3.1. Missed value handling

Our dataset contains categorical data. The recommended way to fill missing values for categorical data is using mode values (Fujikawa & Ho, 2002). We have used mode to fill missing data. We have dropped kebele-zone and time columns. They have 5699 and 2264 empty records respectively. The time column has additional 3002 records which have listed as day and night only. And this cannot describe the specific time on which the crime has committed. We have also dropped records which have 8 and more than 8 empty values. Among the 11363 rows 16 of them had 8 empty columns. We have dropped those rows and left with 11347 rows.

3.3.2. Attribute selection

In addition to the features that we have recorded in our dataset, there are name of criminals' and victims', phone number, address of the criminals' and victims' as well as birth place of criminals'. For the sake of privacy, we have rejected name, phone number and address. We have also dropped birth place, even if it does not represent race, it will create bias among criminals when the prediction model developed. We have recorded the rest 15 attributes in our dataset. Among the 15 features time and kebele-zone 2264 and 5699 empty rows respectively. Additional 3002 rows of the time attribute were defined as day time and night time. Generally, almost half record of the two features was empty. Filling these rows with missing value handling mechanisms will result bias. The recommended way is to drop the feature.

3.3.3. Redundant value handling

Among 11347 rows 1077 of them are redundant. Building the model with redundant data results overfitting on the model. We have dropped those rows and left with 10270 rows.

3.3.4. Data Normalization

One of the mandatory techniques in deep learning is feature scaling. They are two types of feature scaling; Standardization and Normalization. We used normalization approach which scale each feature of the dataset to a fixed range between 0 and 1. This is done by subtracting the minimum value from 'X' (actual value) and dividing by the maximum value minus the minimum values.

$$X' = \frac{X - Xmin}{Xmax - Xmin} \quad (3.1)$$

3.4. Data Visualization

This section discusses about the analysis of Bahir Dar city crime records from September 2005 up to April 2011 E.C. We have summarized the data using table and graph. After preprocessing the crime records, we have 10270 records with 13 features.

Table 3.3: Crime distribution and criminals' age

Criminals' Age	VC	PC	SC	VCC	Grand Total
10-14	6	6			12
15-19	813	966	92	83	1954
20-24	1274	1194	259	226	2953
25-25	1047	742	356	250	2395
30-34	489	303	149	149	1090
35-39	298	161	109	124	692
40-44	181	61	88	89	419
45-49	106	48	57	75	286
50-54	66	38	30	45	179
55-59	36	13	21	34	104
60-64	29	14	13	30	86
65-69	25	11	11	12	59
70-74	9	3	7	7	26
75-79	2	3	2	3	10
80-84	2			1	3
Above 85			1	1	2

Grand Total	4383	3563	1195	1129	10270

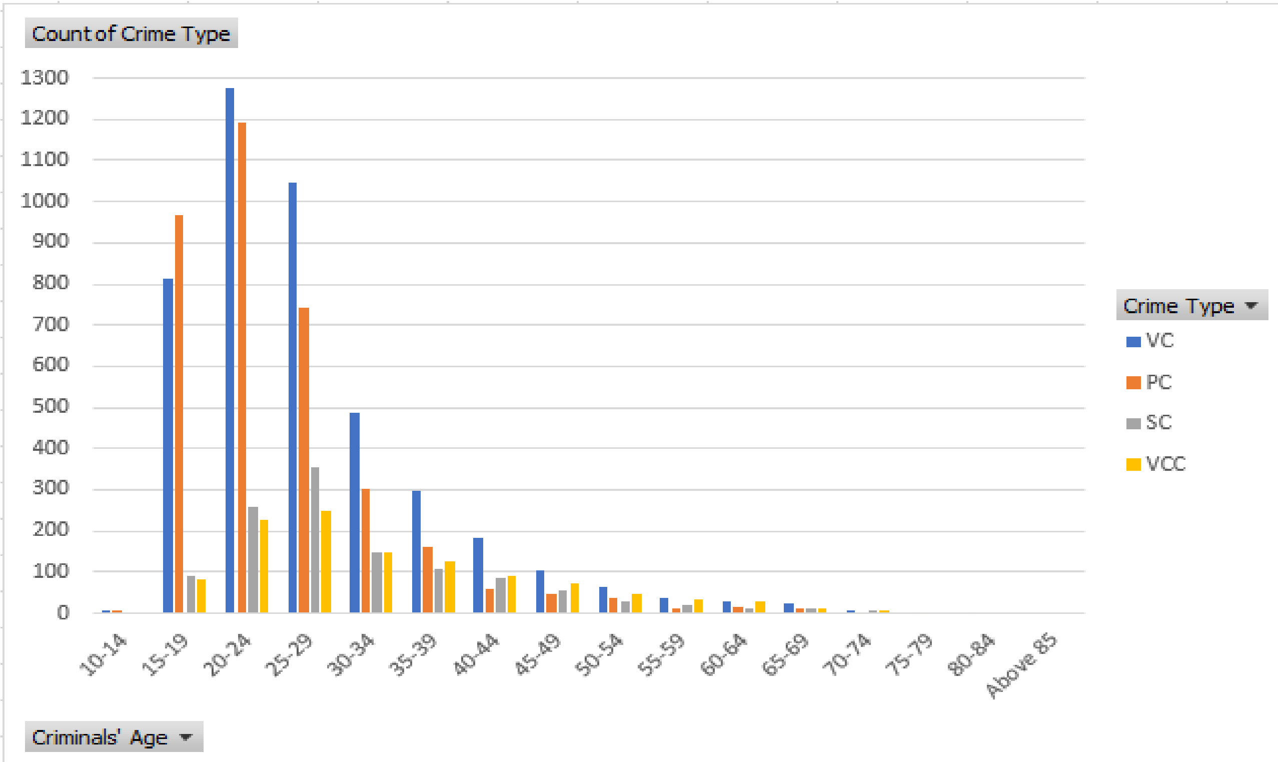

Figure 3.1: Crime distribution and criminals' age

As shown above in the table 3.3 and figure 3.1 from age 10 to 14 are less prone to commit crimes. Usually peoples from age 15 to 29 do crimes highly than the others whereas, the crime committing rate will decrease after 50.

Table 3.4: Crime distribution and criminals' gender

Criminals' Gender	VC	PC	SC	VCC	Grand Total
Male	4098	3398	1119	983	9598
Female	285	165	76	146	672
Grand Total	4383	3563	1195	1129	10270

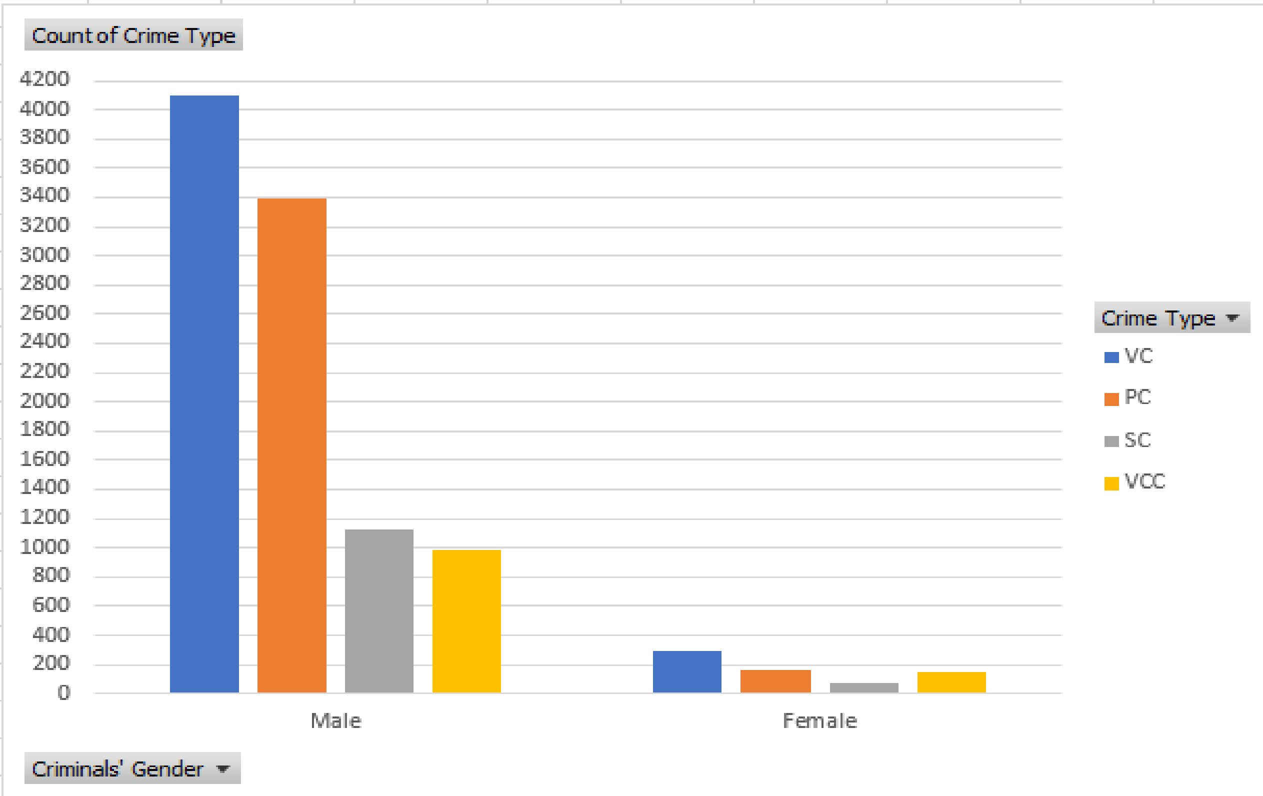

Figure 3.2: Crime distribution and criminals' gender

Table 3.4 and figure 3.2 shows the crime distribution and its relation with criminals' gender. Males do crimes highly than of females.

Table 3.5: Crime distribution and criminals' marital status

Criminals' Marital Status	VC	PC	SC	VCC	Grand Total
Married	1411	835	640	645	3531
Un-Married	2927	2715	547	480	6669
Divorced	44	13	7	3	67
Widowed	1		1	1	3
Grand Total	4383	3563	1195	1129	10270

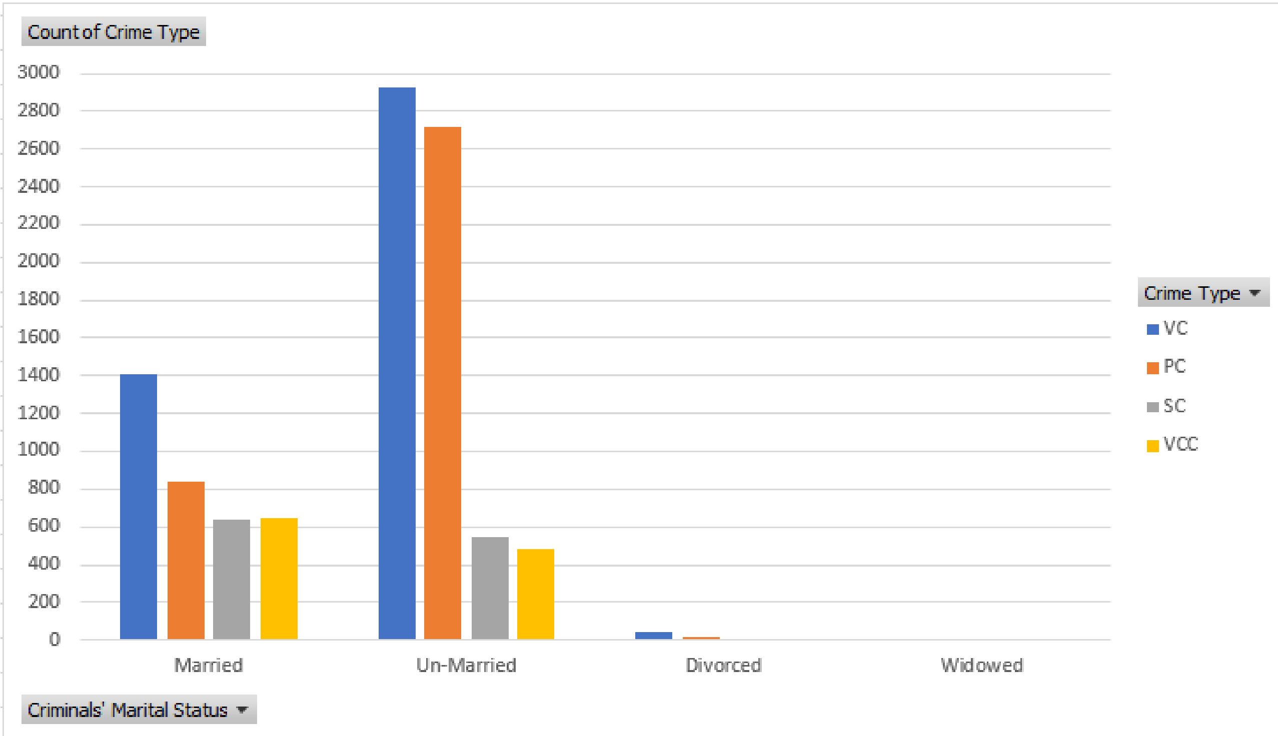

Figure 3.3: Crime distribution and criminals' marital status

The summary of table 3.5 and figure 3.3 shows that un-married persons commits crime highly than of married, widowed and divorced persons. Married persons are on the second rank next to un-married persons by committing crimes. Relatively divorced and widowed persons do crimes rarely, but not only the marital status determine to commit crimes but also other demographical attributes.

Table 3.6: Crime distribution among kebeles

Kebele	VC	PC	SC	VCC	Grand Total
1	118	67	24	27	236
2	112	82	20	14	228
3	50	36	13	10	109
4	165	187	80	18	450
5	157	98	22	16	293
6	262	149	41	16	468

7	132	64	23	17	236
8	159	143	41	56	399
9	57	54	16	17	144
10	223	175	40	44	482
11	860	621	273	244	1998
12	264	318	116	62	760
13	628	482	125	281	1516
14	565	502	140	104	1311
15	37	27	7	4	75
16	435	422	103	156	1116
17	159	136	111	43	449
Grand Total	4383	3563	1195	1129	10270

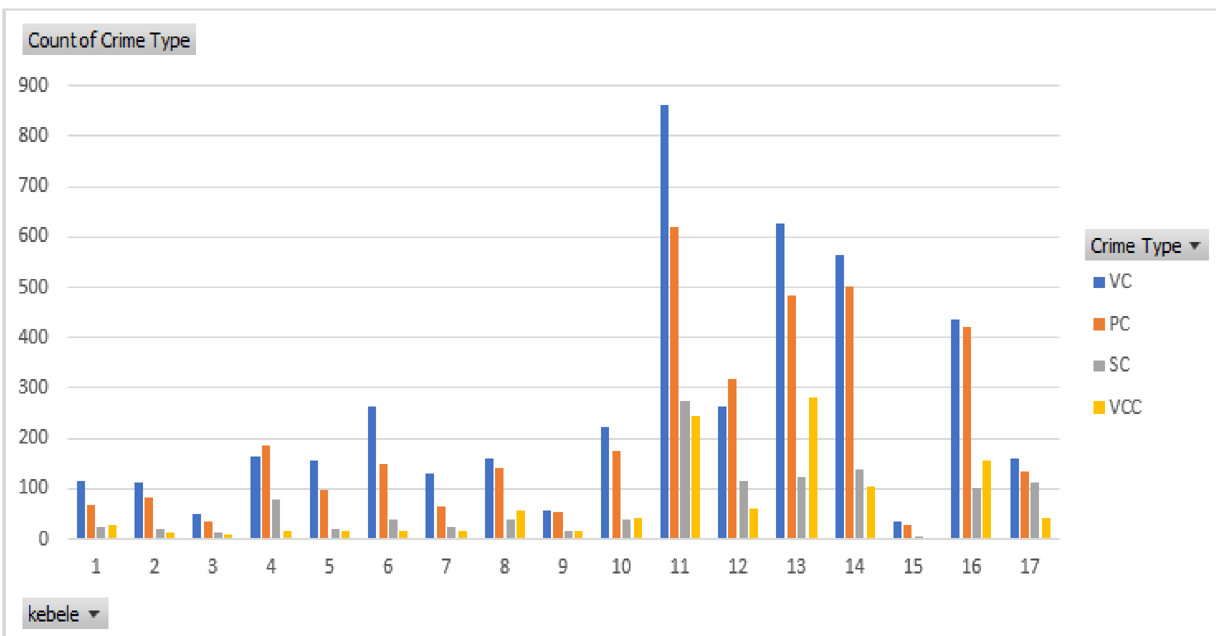

Figure 3.4: Crime distribution among kebeles

Table 3.6 and figure 3.4 shows crime distribution among 17 kebeles of Bahir Dar city.

Table 3.7: crime distribution among months

Month	VC	PC	SC	VCC	Grand Total
1	434	330	94	99	957
2	461	351	136	120	1068
3	352	298	104	92	846
4	392	279	107	111	889
5	343	279	80	89	791
6	406	300	100	106	912
7	387	314	94	83	878
8	312	312	95	88	807
9	317	248	105	81	751
10	318	249	78	86	731
11	331	297	92	98	818
12	330	306	110	76	822
Grand Total	4383	3563	1195	1129	10270

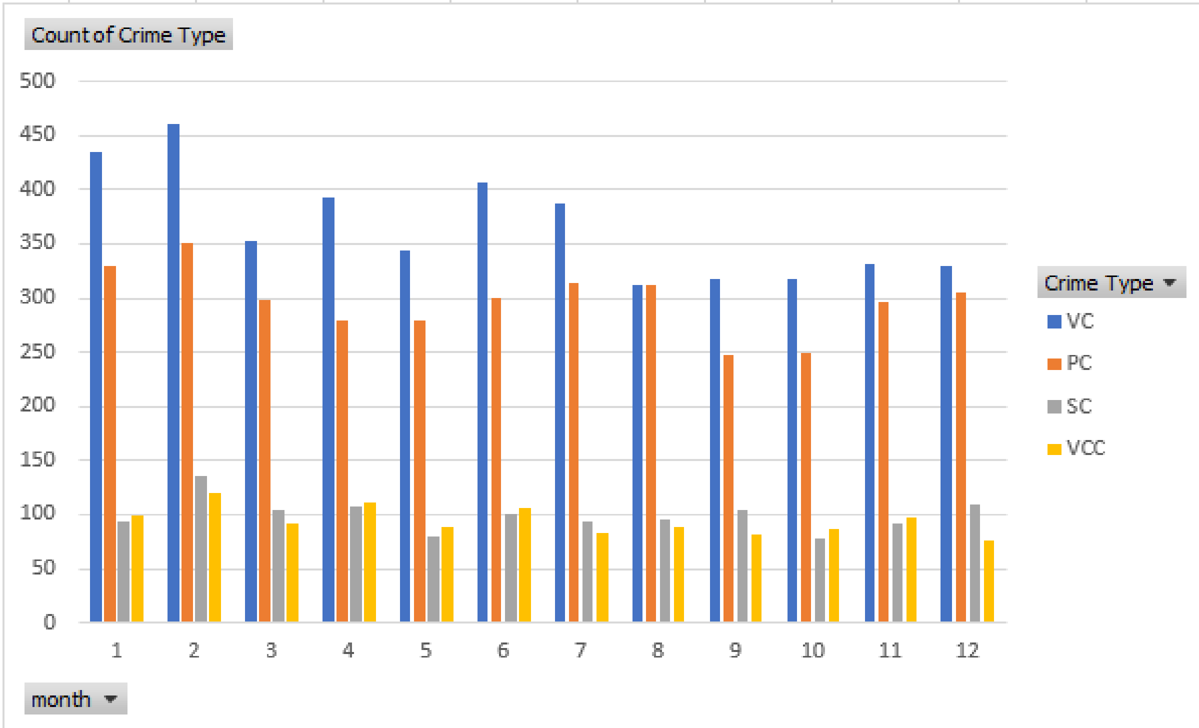

Figure 3.5: crime distribution among months

We have summarized crime distribution among months using table 3.7 and figure 3.5.

Table 3.8: Crime distribution among years

Year	VC	PC	SC	VCC	Grand Total
2005	511	535	137	195	1378
2006	673	447	128	166	1414
2007	633	412	155	164	1364
2008	805	582	291	226	1904
2009	679	532	238	141	1590
2010	617	664	143	166	1590
2011	465	391	103	71	1030
Grand Total	4383	3563	1195	1129	10270

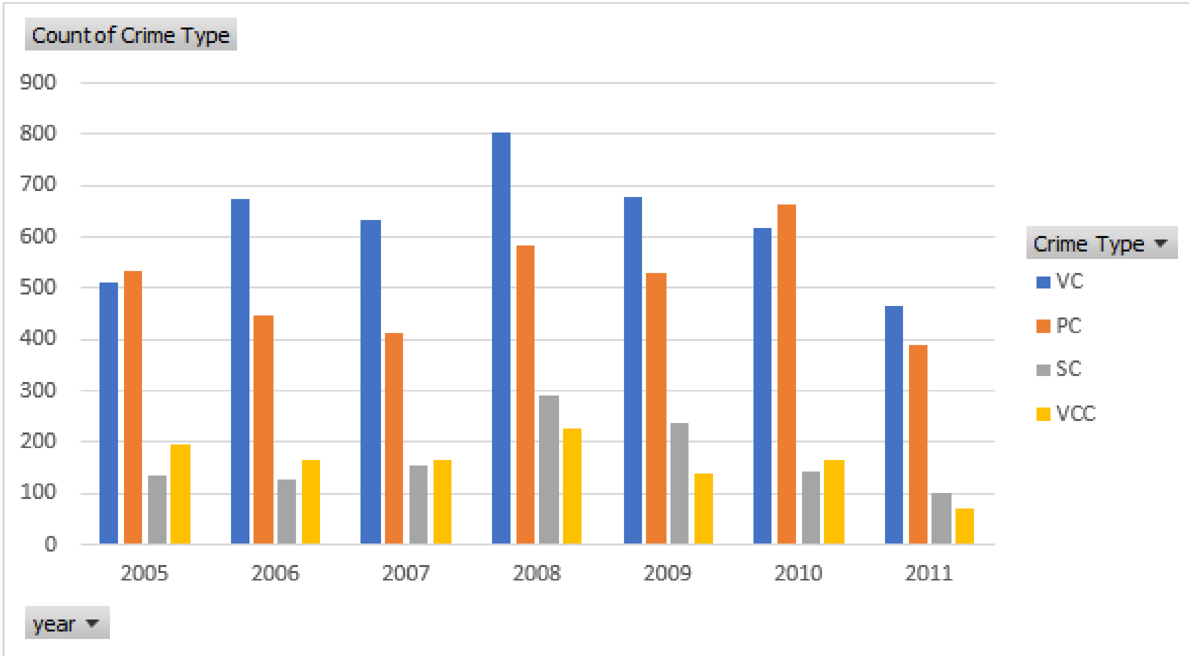

Figure 3.6: Crime distribution among years

Table 3.8 and figure 3.6 shows crime distribution among years from 2005 up to 2011 of E.C. the crime data of 2011 is from September up to April only due to the time constraint we had gathered the data up to the eighth month of 2011.

Table 3.9: Crime distribution and victims' gender

Victims' Gender	VC	PC	SC	VCC	Grand Total
Male	2937	2330	93	133	5493
Female	1152	817	51	34	2054
Grand Total	4089	3147	144	167	7547

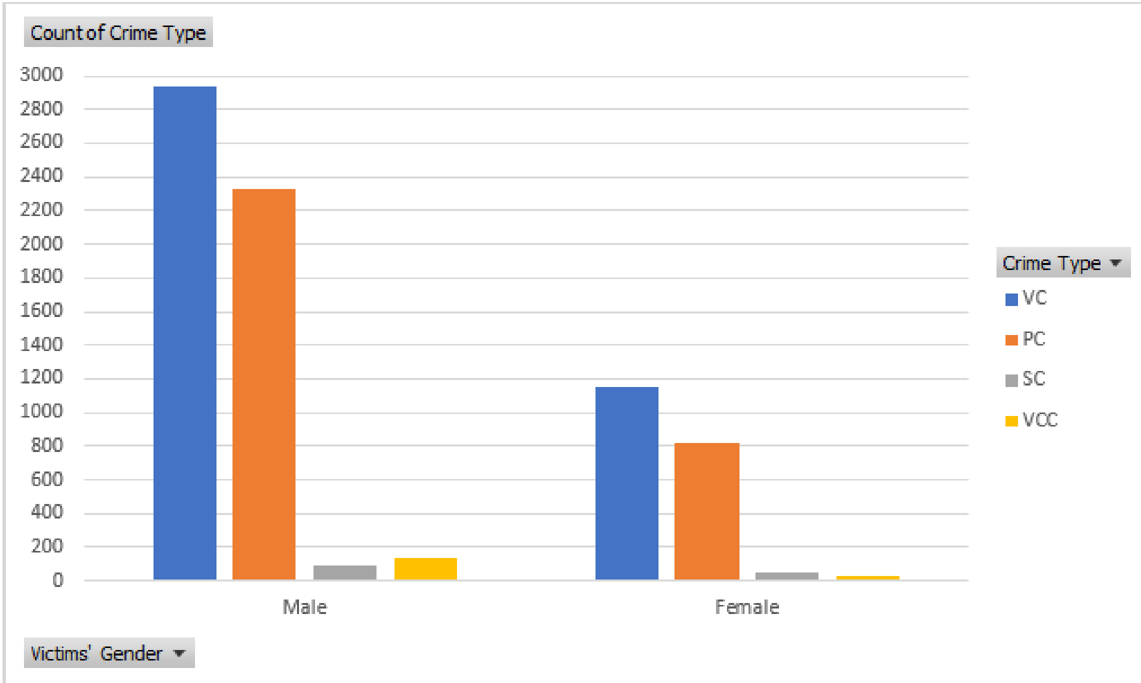

Figure 3.7: Crime distribution and victims' gender

Among the preprocessed 10270 records 7547 crimes were committed on individual persons. Whereas the rest 2723 crimes has committed on the law and organizations. Here by the law includes violating constitution and other rules. Some crimes like property crimes can be committed on organizations. As shown above on table 3.9 and figure 3.7 males are victims in a high rate than of females. When we compare this data with table 3.4, females are more vulnerable to be victims than they are to commit crimes.

Table 3.10: Crime distribution among victims' age

Criminals' Age	VC	PC	SC	VCC	Grand Total
0-4	41	39	2	4	86
5-9	17	1			18
10-14	60	2	2	1	63
15-19	491	193	8	3	689
20-24	955	657	31	17	1637

25-25	912	727	7	30	1700
30-34	449	436	20	19	911
35-39	346	352	17	34	752
40-44	230	248	19	12	507
45-49	195	160	19	7	381
50-54	117	131	4	14	281
55-59	91	81	3	11	187
60-64	57	53	4	10	123
65-69	59	37	4	4	104
70-74	45	23	3	1	73
75-79	15	4			22
80-84	2	1	1		3
Above 85	7	2	144		10
Grand Total	4087	3147		167	7547

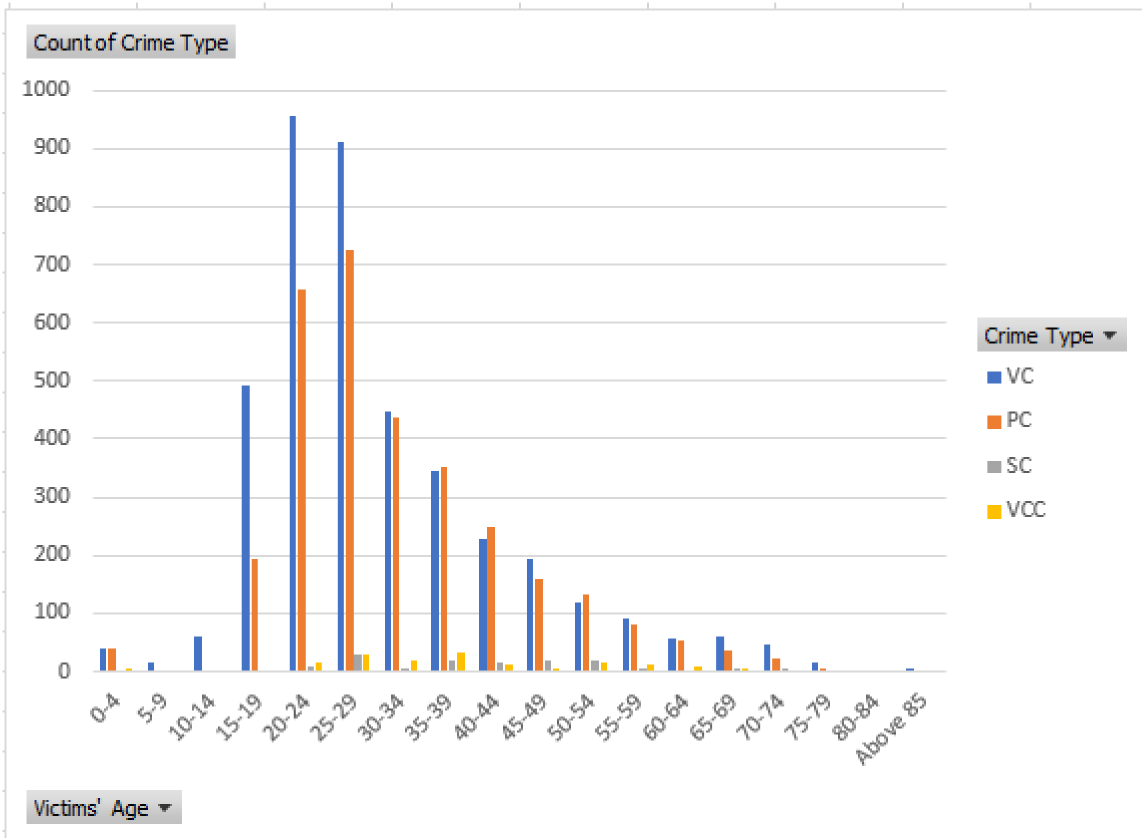

Figure 3.8: Crime distribution among Victims' age

Table 3.10 and figure3. 8 shows crime distribution and victims' age. Persons who are from age 15 to 49 are highly victims.

3.5. The proposed model

This section describes about the developed model. We have built the model by combining feed forward artificial neural network and LSTM-RNN algorithms. The reason behind selecting these machine learning algorithms are the nature of our dataset. The dataset consists categorical data and time-series data. As of Tsion, (2019) research LSTM is the recommended algorithm for time-series forecasting. LSTM is a type of RNN which solved the vanishing problem of RNN and have long memory. Researches like Shermila et al. (2018) shows using artificial neural network results good for categorical type of data. We have combined these two approaches for predicting crimes. Makris et al., (2019) have also used the hybrid of feed forward neural network and LSTM-RNN to predict the next movement of the drum by using rhythm and related features of music.

Combining feed forward artificial neural network and long short-term memory recurrent neural network (Makris et al., 2019) helps to get the benefits of both algorithms. LSTM is useful and efficient in learning sequences and time-series data. Tsion, (2019) has showed the applicability of time-series forecasting on crime and crime is time-based incident. Whereas FFANNs are effective in predicting categorical data. The Nguyen et al., (2017) proved ANN shows a promising result on demography based crime prediction. The hybrid of FFANN and LSTM-RNN was applied on rhythm composition (Makris et al., 2019), (Kaliakatsos-papakostas et al., 2017). We have used merge layer to predict crime types based on time, location and demography.

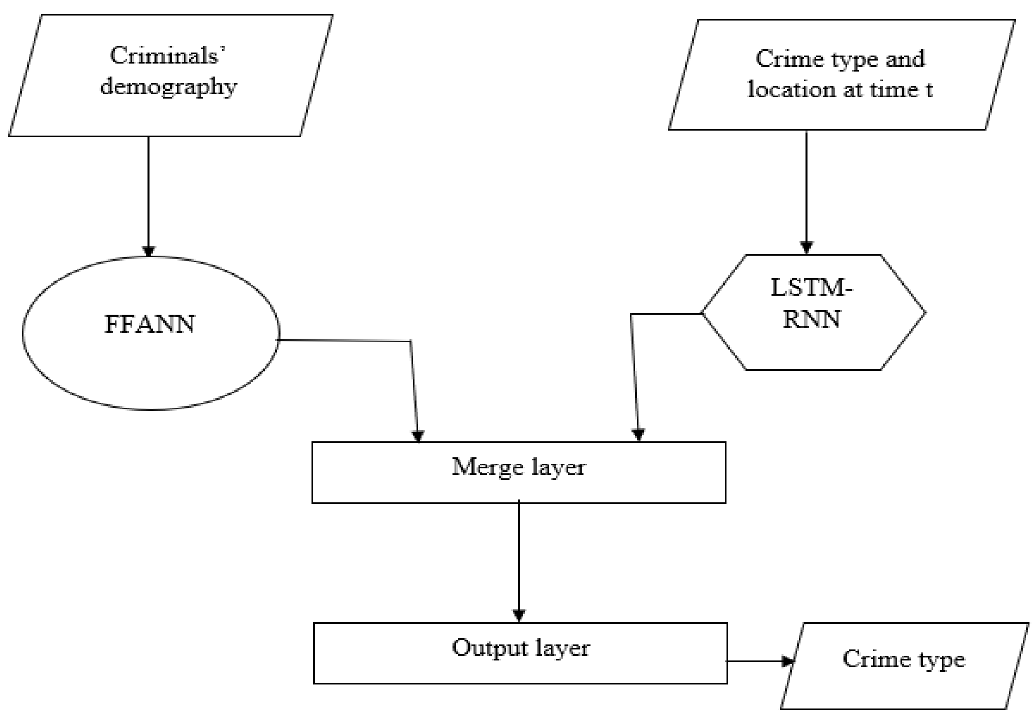

Figure 3.9: The proposed hybrid model

The above diagram figure 3.9 shows the proposed hybrid model. There is a component which is used to receive demographical data of criminals' and predict the crime type to be committed by that criminal using FFANN. Another component is used to receive time-based crime data and location so as predict crime types using LSTM-RNN. The outputs of the two neural networks will be merged on a component called merging layer. It is used to merge two different network layers on one layer. Then the result will go to output layer. Therefore, we will have crime type based on criminals' demography, location and time. Combining FFANN and LSTM-RNN is known as conditional neural sequence learner (CNSL) (Makris et al., 2019).

The FFANN for the hybrid model takes demographical data of criminals' as input and produce crime type to be committed as output. After initializing the network, it forward propagates, update weight by computing backpropagate error. It trains the network through this process to produce the outcome.

The LSTM-RNN for the hybrid model takes location and crime data at time t to produce the to be committed crime type at t+1 and specific location. It starts from the first defined time in the dataset (the first record since it is ordered in ascending order) and increase by one to the end of the dataset. During this order it applies the LSTM layer computation, by computing input gate, forget gate and output gate. After training the network with this approach it will generate the output.

The hybrid model takes the output from the FFANN and LSTM-RNN. Their results will be merged in merging layer and predict the possible crime type to be committed.

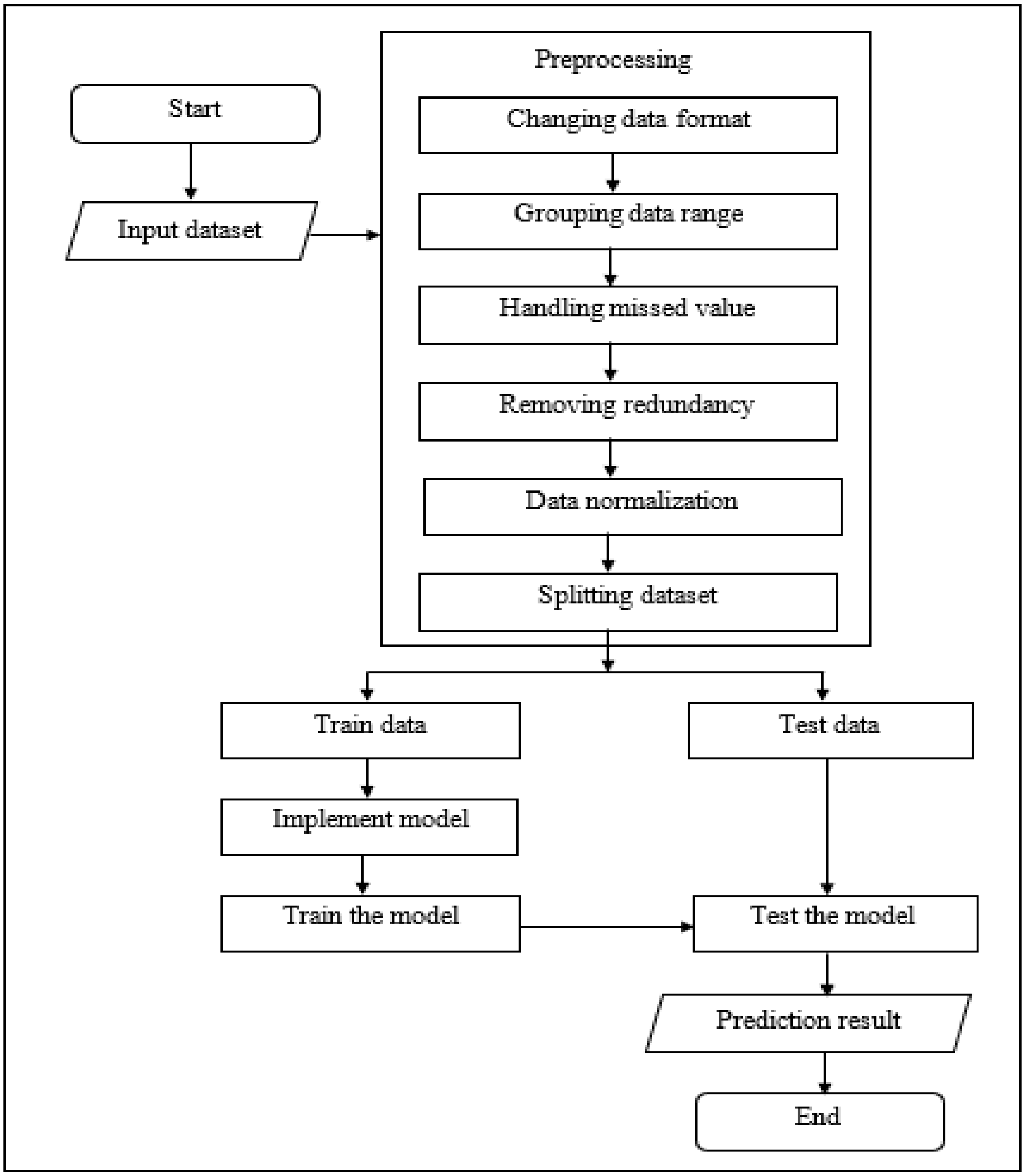

Figure 3.10: Process flow diagram

Figure 3.10 shows the designed prediction model. The model has components.

1. Input dataset: this is the first component which is responsible to accept the input dataset. The dataset contains a data before preprocessing.
2. Preprocessing: preprocessing is the second component which we used to clean the data that we got from the first component. This component has three sub components. They are:

2.1. Changing data format: our data was stored in different formats which contained text and number. This will be difficult for the prediction module to process such kind of data. We have converted the text data into numbers.

2.2. Grouping data range: the dataset contained a data within different range. We have grouped them based on different previous research works as mentioned above under data preprocessing section.

2.3. Handling missing value: this subcomponent helps to handle records which consists missing value. As we discussed above in the preprocessing section, we have used mode mechanism to fill missed value. We have dropped 2 columns and 16 rows because of they have consisted many empty cells.

2.4. Removing redundancy: our dataset had redundant data which can create overfitting on the model. We have removed all of the redundant data.

2.5. Data normalization: applying the data scaling mechanism. The data will be between 0 and 1.

2.6. Splitting dataset: for training the model and testing its performance the data need to be split into train and test data. We have used 80% of the data for training and 20% for testing the model.

3. Train data: this is used to train the model.
4. Test data: this data helps to evaluate the model which trained by using train data.
5. Implement model: we built the hybrid model of FFANN and LSTM-RNN even if we have compared the result with FFANN and LSTM-RNN.
6. Train the model: in this module the train data will be feed to the implemented model and the model will train based on the given data.
7. Test the model: in this module the test data will be feed to the model in order to measure the performance of the model.
8. Prediction result: this module shows the result of the prediction model. The to be committed crime type will be predicted.

3.6. Tools

In this section we discussed about the tools that we have used during developing our model.

3.6.1. Python

Python is an interpreted, object-oriented, general purpose, high-level programming language with dynamic semantics (Rossum, 2020). Its data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development (RAD). It is simple and easy to learn programming language. Python supports modules and packages, which encourages program modularity and code reuse (Huenerfauth et al., 2009). It has many good features among that:

- Easy to code
- Free and open source
- It follows object-oriented approach
- Portable
- High level language
- It has large and standard library

3.6.2. Anaconda

Anaconda is a tool which is used to developing machine learning, deep learning and artificial intelligence. It is a Python and R- programming distribution used for data analysis and scientific computing. It is an open source project developed by Continuum Analytics, and it can be run on Windows, Mac OS X and Linux. Anaconda consists many packages like: NumPy, SciPy, Matplotlib, Pandas, IPython, and Cython (Weston & Bjornson, 2016). It consists more than 1500 packages. Anaconda provides the tools needed to easily:

- Import data from files, databases, and data lakes
- Manage environments with conda
- Share, collaborate on, and reproduce projects
- Deploy projects into production with the single click of a button

We have used anaconda navigator to develop our model. It is a desktop and graphical user interface-based application.

3.6.3. Spyder

Spyder is a scientific python development environment. It is a free integrated development environment (IDE) which is included with Anaconda. It includes editing, interactive testing, debugging, and introspection features. It has also functionalities of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and good visualization capabilities of a scientific package and model results. Spyder supports Ipython, and popular python libraries like Numpy, Scipy, and matplotlib (Raybaut, 2017). Some of Spyder features are:

- MATLAB-like PYTHONPATH management dialog box (works with all consoles)
- Direct links to documentation (Python, Matplotlib, NumPy, Scipy, etc.)
- Direct link to Python (x, y) launcher
- Keyboard shortcuts
- Syntax coloring schemes (source editor, history log, help)
- Console: background color (black/white), automatic code completion, etc.
- Editor: syntax coloring (Python, C/C++, Fortran)

3.6.4. NumPy

NumPy is the fundamental and general-purpose package for scientific computing in Python. It stands for 'Numerical Python'. It is a Python library that provides a multidimensional array object, various derived objects (like arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, input/output, linear algebra, statistical operations, random simulation and the like (Oliphant, 2006). Some of the features of NumPy are:

- A powerful N-dimensional array object
- Sophisticated functions
- Tools for integrating other programming language
- Useful for linear algebra and random number capabilities

3.6.5. Pandas

Pandas is a software library written in Python programming language for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and time series. Pandas can be used to deal with statistical data, to import csv files, data alignment, handling missing values and generally for data preprocessing (Mckinney, 2010). Some of Pandas features are:

- DataFrame object for data manipulation with integrated indexing.
- Tools for reading and writing data or to import datasets
- Data alignment and integrated handling of missing data.
- Reshaping and pivoting of data sets.
- Data structure column insertion and deletion.
- Group by engine allowing split-apply-combine operations on data sets.
- Dataset merging and joining.
- Hierarchical axis indexing to work with high-dimensional data in a lower-dimensional data structure.

3.6.6. **Keras**

Keras is a high-level neural network, user-friendly API (Application Program Interface), which is written in Python and capable of running on top of TensorFlow or Theano. It is designed to be modular, fast and easy to use (Moolayil, 2019), (Borr, 2018). Basic features of Keras are:

- It contains neural network building blocks like layers, activation functions and optimizers.
- It helps to develop deep learning models in quickly and easily.
- It supports layers like dropout, batch normalization and pooling which needs to develop convolutional and recurrent neural networks.

3.6.7. TensorFlow

TensorFlow is one of the most popular and widely used deep learning frameworks. It was developed and open sourced by Google and supports deployment across CPUs, GPUs, and mobile and edge devices too (Moolayil, 2019), (Borr, 2018). Some of the features of TensorFlow are:

- Efficiently works with mathematical expressions involving multi-dimensional arrays
- Good support of deep neural networks and machine learning concepts
- GPU/CPU computing where the same code can be executed on both architectures
- High scalability of computation across machines and huge data sets

CHAPTER FOUR
EXPERIMENT AND RESULT DISCUSSION

In this chapter we discussed about the experimentation and evaluation processes in line with the result of the models.

4.1. Model implementation procedures

As we have discussed above, after the data preprocessing step, the model is built by combining two deep learning algorithms. There is merge layer, dropout layer, activation function, loss function, optimizers and evaluation metrics which we used to implement and evaluate the model.

We collected crime record dataset from Bahir Dar city police stations. Preprocessing is done to make the data ready for the prediction model. Crime record datasets are divided into two parts. Training dataset used to train the model and to increase the performance of the system through different parameters, whereas the testing dataset to evaluate the system. About 10270 records with 13 features was collected. we have used 80/20 format for training and testing respectively. Among the 13 features we have merged date, month, and year into 1 column and left with 11 features. 8 of the features are used for the prediction model both during the training and testing phase. The rest 3 feature shows victims' data. We have used them during the analysis and data summarization.

Training phase: we have developed 3 different models. The first one is FFANN model. It consists of a sequence of dense layer, activation, and dropout layers. The input for the FFANN is criminals' demography and location, whereas the output is crime type to be committed. The neural network will learn the pattern from the given dataset. Dropout is also applied to prevent overfitting by altering the network architecture at training time. It ensures that no single node in the networks is responsible to learn a pattern.

The second one is LSTM-RNN model. It consists of a sequence of dense layer, LSTM layer, activation, and dropout layers. The input for the LSTM-RNN is demography and location at a time t, whereas the output is crime type at a time t+1.

The third experiment is the hybrid of FFANN and LSTM-RNN. It consists of a sequence of dense layer, LSTM layer, activation, merge layer, and dropout layers. The input for the hybrid model is the output crime type from the above two models, whereas the output is crime type at a time t+1 and specific location. For building and evaluating this model we feed demography of criminals for the FFANN and we feed crime type and location at a time t for the LSTM-RNN. The output of the two network layers will be merged on a merging layer to predict the crime type.

Validation phase: different techniques such as dropout at early stages and batch normalization between network layers are applied to better characterize or learn features and have higher accuracy. In this phase, our aim is to increase the performance of our model by increasing the accuracy or by decreasing the loss. Combination of parameters (weight and bias) that provide higher accuracy are used to classify testing datasets.

Testing phase: crime record different from training datasets have been given to the learned model to evaluate how well the system responds to new datasets.

After the prediction model built it should be trained using training data. The training set have to contain the target attribute. Data attribute that contains the target variable and training parameters to control learning parameters are also essential in the training process. We have used fit() function from the keras to train the model. It contains training data, training target, Epochs, Batch_size, and validation_data.

Batch_size: defines the number of samples to work through before updating the internal model parameters. Batch is like a for-loop iterating over one or more samples and make predictions. The predictions are compared to the expected output variables and error is calculated at the end of the batch. We used 64 for batch size. We have tried to the models with different batch values and we got better result with batch size of 64.

Epochs: the number of epochs determines the time that the learning algorithm will work through the entire training. It's an iteration over the entire training data and training target provided. The model is trained until the epoch of index `epochs` is reached. 10 epochs are used to train and validate the dataset. After the 10 epochs the loss value was not decreasing or it means the accuracy was not increasing. So, it only increased the running time which can affect the performance of the model. Therefore, we have used 10 epochs for building and evaluating the prediction models.

Validation_data: it's a data that is holdback from training our model and is used to give an estimate of model. It's used to give the unbiased estimation of the model when comparing or selecting between final models. It evaluates loss and model metrics at the end of the epoch. It has two parameters; these are test data and test target.

4.1.1. Merge layer

Merge layer is a layer which is found in keras library. It is used two merge two layers. There are different mechanisms to merge layers. We have used 'Concatenate' which is useful to merge two layers as input in one layer. It is also sequential.

4.1.2. Activation function

Among the three layers of neural networks input layer do not have any computation, it means it does not use activation layer. Hidden layers do not have contact with the external world. It computes all sort of computation on features entered through the input layer and transfer as output through the output layer. Usually activation function can be used in hidden and output layer.

Activation function helps to decide, whether a neuron should be activated or not. It will be determined by calculating weighted sum and further adding bias with it. The importance of the activation function is to introduce non-linearity into the output of a neuron. Neural networks have neurons that work in correspondence of weight, bias and their respective activation function. In a neural network, we need to update the weights and biases of the neurons on the basis of the error at the output. This process is known as back-propagation.

Activation functions make the back-propagation possible since the gradients are supplied along with the error to update the weights and biases.

Sigmoidal activation function: - is an activation function which is plotted as **'S'** shaped graph. It has a non-linear nature. Notice that X values lies between -2 to 2, Y values are very steep. This means, small changes in x would also bring about large changes in the value of Y. The activation function values will be 0 to 1.

$$A = \frac{1}{(1+e^{-x})} \tag{4.1}$$

We have used sigmoid in hidden layers. Sigmoid activation function used as the getting function for the input gate, output gate and forget gate. The vanishing gradient problem of RNN solved by LSTM network architecture, so we do not need to use additional activator for that problem.

4.1.3. Dropout

Dropout refers to ignoring units or neurons, during the training phase of certain set of neurons which is chosen at random will be ignored or dropped out. This regularization method is supported by keras which is provided by dropout core layer. The main reason for using dropout is to prevent overfitting. Usually 20% of dropout regularization is used and gives a good result in preventing overfitting and retaining model accuracy. We have also used 20% of dropout.

4.1.4. Optimizer

The aim of deep learning is to reduce the difference between the predicted output and the actual output which is also called as a Cost function or Loss function. In order to minimize the cost function, we need to find the optimized value for weights by achieving many iterations with different weights, which is called gradient descent. Generally, gradient descent is an iterative machine learning optimization algorithm to reduce the cost function. Optimizers are algorithms or methods that is used to change the attributes of the neural network such as weights and learning rate in order to reduce the losses.

Adam: - is a replacement optimization algorithm which can be used instead of the classical stochastic gradient descent procedure to update network weights iterative based in training data in deep learning models. It combines the best properties of the Adaptive gradient (AdaGrad) and Root Mean squared Propagation (RMSProp) algorithms to provide an optimization algorithm which can handle sparse gradients on noisy data by maintaining a per-parameter learning rate. The learning rate is calculated depending on the past gradients that have been computed for each parameter.

$$\Theta_{t+1} = \Theta_t - \frac{n}{\sqrt{G_t + \epsilon}} * g_t \tag{4.2}$$

Where:

θ: is parameter to be updated

η: is the initial learning rate

ϵ: some small value that is used to avoid division of 0

G_t: sum of squares of past gradients

g_t: gradient estimate at time t

4.1.5. Loss

Loss is a method used to evaluates how well the algorithm models the data. Among different types of loss function, we have used mean squared error loss function for this study.

Mean squared error: - it measures the average of squared difference between predictions and actual observations. It’s only concerned with the average magnitude of error irrespective of their direction.

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{4.3}$$

Where:

y_i : prediction value and

$\hat{y}_i$: actual value

4.1.6. Metrics

We have used accuracy, MAE, and MSE to measure the performance of the model. It is a popular metrics for classification algorithms.

Accuracy: - Accuracy is number of correct predictions made by the model over all kinds of predictions made. It can be calculated in different ways. We have calculated by using the following approach.

Accuracy= 1-MSE (4.4)

Mean absolute error (MAE): - is one of the popular metrics for continuous data. It measures the average magnitude of the errors in the set of predictions, without considering their direction. It is the average over the test sample of the absolute differences between the prediction and actual observation where all individual differences have equal weight.

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \tag{4.5}$$

Mean squared error (MSE): - is calculated by taking the average of the square of the difference between the original and predicted values of the data.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{4.6}$$

4.2. Result

In this section, we discussed about results that we got from the prediction models. Experimental evaluation of the proposed model for predicting crime types is described in detail. Experimental evaluation approves the realization of the proposed model architecture. Results got from feedforward artificial neural network, long short-term memory, and the hybrid of the two algorithms are compared. Loss value during training and validation is also compared.

4.2.1. Feedforward neural network prediction

We have implemented the feed forward neural network and measure its performance. The model was built using demographical attributes of criminals and location. During implementing this model, the time sequence of the data did not consider because of the nature of FFANN.

```
Train on 8216 samples, validate on 2054 samples
Epoch 1/10
8216/8216 [==============================] - 1s 169us/step - loss: 0.1155 - val_loss: 0.0997
Epoch 2/10
8216/8216 [==============================] - 0s 36us/step - loss: 0.1140 - val_loss: 0.0994
Epoch 3/10
8216/8216 [==============================] - 0s 27us/step - loss: 0.1133 - val_loss: 0.0991
Epoch 4/10
8216/8216 [==============================] - 0s 30us/step - loss: 0.1135 - val_loss: 0.0988
Epoch 5/10
8216/8216 [==============================] - 0s 28us/step - loss: 0.1117 - val_loss: 0.0986
Epoch 6/10
8216/8216 [==============================] - 0s 26us/step - loss: 0.1115 - val_loss: 0.0985
Epoch 7/10
8216/8216 [==============================] - 0s 29us/step - loss: 0.1113 - val_loss: 0.0983
Epoch 8/10
8216/8216 [==============================] - 0s 27us/step - loss: 0.1115 - val_loss: 0.0982
Epoch 9/10
8216/8216 [==============================] - 0s 28us/step - loss: 0.1105 - val_loss: 0.0981
Epoch 10/10
8216/8216 [==============================] - 0s 29us/step - loss: 0.1097 - val_loss: 0.0979
```

Figure 4.1: Training and Validation loss of FFANN

By applying FFANN on the crime record dataset we got 90.20% accuracy, with 10.97% and 9.79% loss during training and validation phase respectively. As we have discussed above in chapter 4, we have used demography of criminals' and location of the crime for building and evaluating the FFANN model.

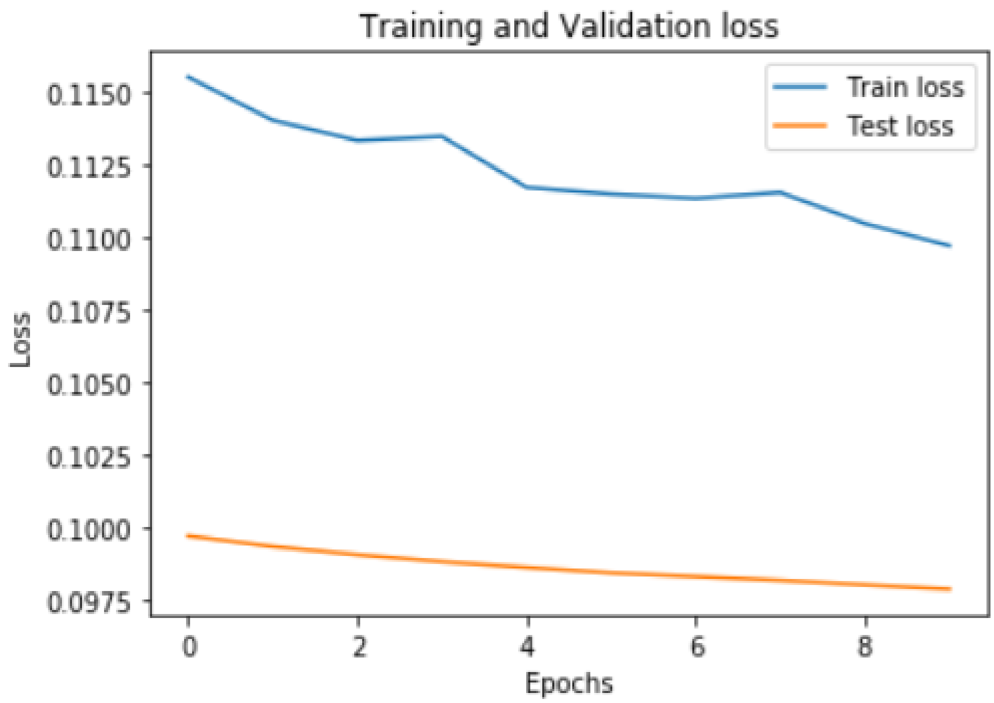

Figure 4.2: Training and validation loss graph of FFANN

4.2.2. LSTM-RNN prediction

LSTM-RNN is implemented using location and demographic data of criminals. We got 94.80% accuracy with 9.95% and 5.19% loss during training and validation respectively. The following diagram shows the training and validation loss using 10 epochs.

```
Train on 8216 samples, validate on 2054 samples
Epoch 1/10
8216/8216 [==============================] - 4s 510us/step - loss: 0.2115 - val_loss: 0.0985
Epoch 2/10
8216/8216 [==============================] - 1s 109us/step - loss: 0.1083 - val_loss: 0.0824
Epoch 3/10
8216/8216 [==============================] - 1s 111us/step - loss: 0.1046 - val_loss: 0.0786
Epoch 4/10
8216/8216 [==============================] - 1s 108us/step - loss: 0.1036 - val_loss: 0.0744
Epoch 5/10
8216/8216 [==============================] - 1s 107us/step - loss: 0.1026 - val_loss: 0.0700
Epoch 6/10
8216/8216 [==============================] - 1s 107us/step - loss: 0.1016 - val_loss: 0.0658
Epoch 7/10
8216/8216 [==============================] - 1s 103us/step - loss: 0.1008 - val_loss: 0.0616
Epoch 8/10
8216/8216 [==============================] - 1s 101us/step - loss: 0.1002 - val_loss: 0.0578
Epoch 9/10
8216/8216 [==============================] - 1s 103us/step - loss: 0.0998 - val_loss: 0.0544
Epoch 10/10
8216/8216 [==============================] - 1s 107us/step - loss: 0.0995 - val_loss: 0.0519
```

Figure 4.3: Training and Validation loss of LSTM-RNN

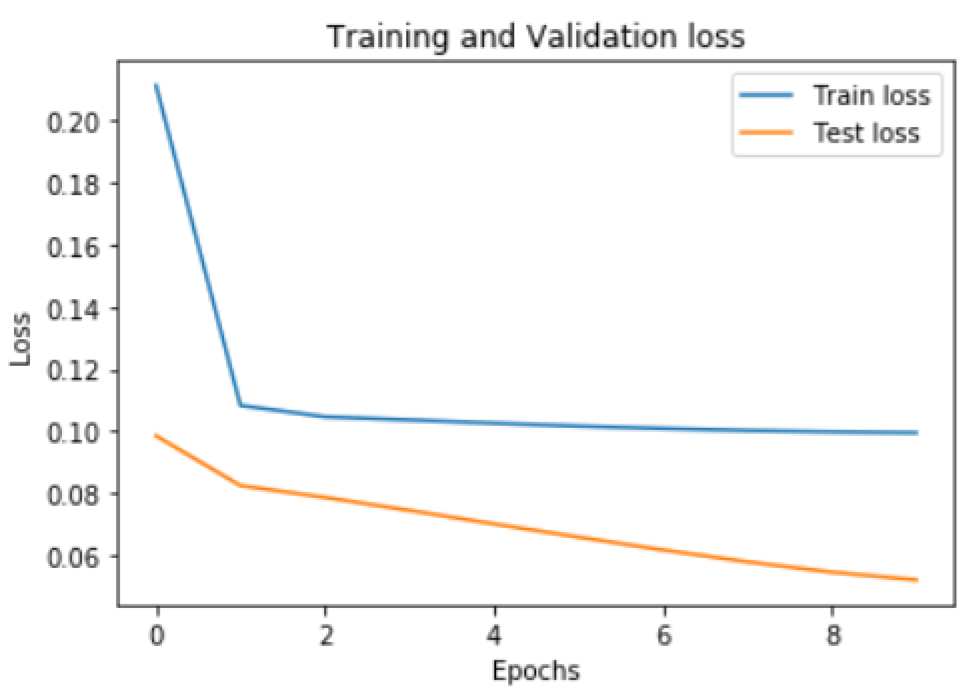

Figure 4.4: Training and validation loss graph of LSTM-RNN

4.2.3. The proposed model prediction

The hybrid of FFANN and LSTM was build using demographical, date and location attributes and got 95.07% accuracy with 4.97% and 4.93% loss during training and validation phase respectively.

```
Train on 8216 samples, validate on 2054 samples
Epoch 1/10
8216/8216 [==============================] - 4s 468us/step - loss: 0.8294 - val_loss: 0.7217
Epoch 2/10
8216/8216 [==============================] - 1s 100us/step - loss: 0.2754 - val_loss: 0.2924
Epoch 3/10
8216/8216 [==============================] - 1s 102us/step - loss: 0.1025 - val_loss: 0.1310
Epoch 4/10
8216/8216 [==============================] - 1s 102us/step - loss: 0.0621 - val_loss: 0.0798
Epoch 5/10
8216/8216 [==============================] - 1s 102us/step - loss: 0.0563 - val_loss: 0.0664
Epoch 6/10
8216/8216 [==============================] - 1s 103us/step - loss: 0.0551 - val_loss: 0.0621
Epoch 7/10
8216/8216 [==============================] - 1s 102us/step - loss: 0.0540 - val_loss: 0.0581
Epoch 8/10
8216/8216 [==============================] - 1s 102us/step - loss: 0.0528 - val_loss: 0.0569
Epoch 9/10
8216/8216 [==============================] - 1s 100us/step - loss: 0.0513 - val_loss: 0.0519
Epoch 10/10
8216/8216 [==============================] - 1s 117us/step - loss: 0.0497 - val_loss: 0.0493
```

Figure 4.5: Training and Validation loss of the hybrid model

As we have discussed above, we have used demographic data of criminals' in FFANN and location and crime type at time t in LSTM-RNN and we have merged their result on a merging layer called 'Concatenate'.

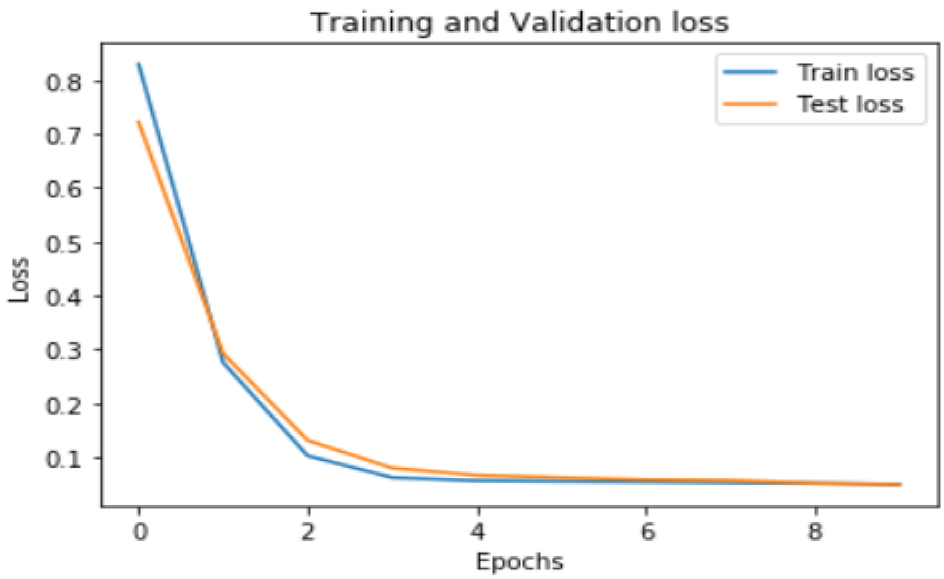

Figure 4.6:Training and validation loss graph of the hybrid model

We have compared the models using accuracy and MSE during training and validation phases. The hybrid model shows high accuracy with minimum error (MSE) than of the others. The following graphs shows the crime type prediction of the hybrid model.

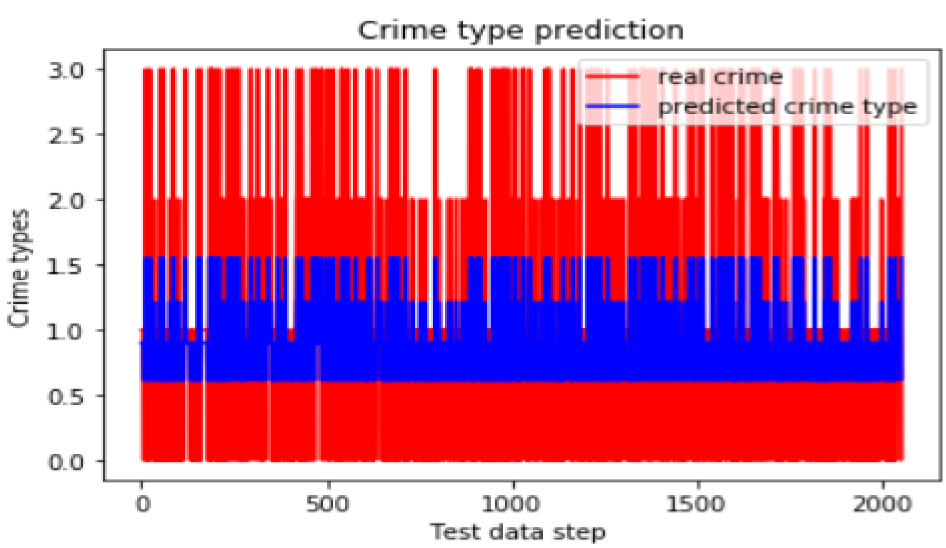

Figure 4.7: Crime type prediction

The above diagram shows crime type prediction for 2054 records that is, for the 20% of the total dataset. The crime record for the test set is from January 2010 up to April 2011 E.C. The graph divides the Y-axis into 7 parts even if we had 4 crime types. The prediction result lies between 0 and 2, more likely the result is near to 0 and 1. These results represent violent and property crimes respectively. Based on the prediction violent crime and property crimes will be happen highly than of state and victimless crimes. There are small amount of state crime and victimless crimes in our dataset. This has some influence in the prediction model. Deep learning needs many data to be trained, if the dataset is too small it will be difficult for the model to identify patterns. Even if the dataset in this research is more than 10,000, state crimes and victimless crimes are rarely committed in the city so do in the dataset. The model predicts crimes that are frequently committed in the city, those are violence crimes and property crimes.

Table 4.1: Result comparison

	FFANN	**LSTM-RNN**	**Hybrid model**
MAE	0.249	0.181	0.212
MSE	0.097	0.051	0.049
Accuracy	90.20%	94.80%	95.07%

We have compared the algorithms using MAE, MSE, and accuracy. In both MSE and accuracy the hybrid model shows the best performance than of the two algorithms. When we compared the result using MAE, LSTM-RNN is with a minimum MAE then the hybrid model. It shows LSTM-RNN has a least error than of the others. Generally, the hybrid model outperforms the FFANN and LSTM-RNN by having high accuracy and minimum MSE.

CHAPTER FIVE
CONCLUSION AND RECOMMENDATION

In this section conclusion, contribution and future works are discussed.

5.1. Conclusion

This research has used a hybrid approach of deep learning algorithms. After having systematic review of literatures and related works, we have identified predicting crimes is given low emphasis and are unresolved problems yet especially in Ethiopia. We have also identified that different scholars focus on either location and time data or demography data even though both are important.

The objective of this research is to apply the hybrid of feedforward artificial neural network and long short-term memory recurrent neural network so as to predict the to be committed crime types. The crime type is predicted based on criminals' demography, time and location. The data was collected from Bahir Dar city administration.

An exhaustive study is performed on the given dataset and algorithms. The research approach consists five phases. The first one is data gathering, the crime data from Bahir Dar city police stations is collected. The second is data pre-processing, it is about making the data ready for the prediction model. Thirdly, the prediction model is implemented to accurately predict crimes. Fourth, the model is trained using training dataset. Finally, the model is evaluated using test dataset. We have built FFANN, LSTM-RNN and the hybrid of the two algorithms. We have compared the accuracy of the algorithms and they scored 90.20%, 94.80%, and 95.07% for FFANN, LSTM-RNN and hybrid model respectively.

5.2. Contribution

As a contribution to the scientific world, we have prepared a dataset with 10,270 records and 15 features. This will be useful for researchers who wants to do their research using crime record dataset especially for those who need Ethiopia crime dataset. One of the limitations was the availability of dataset in electronic format. We have combined two deep learning algorithms; feedforward artificial neural network and long short-term memory recurrent neural networks.

5.3. Recommendation and Future work

The hybrid model proposed in this research work is used in prediction of crimes. We would say the proposed model is fit to the current requirement though there are some issues that need additional work. In this section, therefore, we insight two key points that remain a challenge and of course were limitation of this research work. The first one is it will be better to calculate the information gain and correlations of attributes rather than selecting them by the truth on the ground. Then the prediction will have high accuracy based on the feature result from information gain and correlation. Second, we have used a feed forward approach in both algorithms even if they have backward learning mechanisms, we recommend that to use bi-directional neural network approaches since their learning rate is high. However, it will be much efficient and effective if the crime is predicted with a promising result, we believe that an attempt in the future study should consider it to achieve better result in predicting crimes. Hence, this research work presented different contributions that can be further improved or implemented on additional efforts.

REFERENCES

Aitelbour, H., Ounacer, S., Elghomari, Y., Jihal, H., & Azzouazi, M. (2018). A crime prediction model based on spatial and temporal data. *Periodicals of Engineering and Natural Sciences*, *6*(2), 360–364.

Ajao, O., Bhowmik, D., & Zargari, S. (2018). *Fake News Identification on Twitter with Hybrid CNN and RNN Models*. 226–230. https://doi.org/https://doi.org/10.1145/3217804.3217917

Al-janabi, K. B. S. (2011). A Proposed Framework for Analyzing Crime Data Set Using Decision Tree and Simple K-Means Mining Algorithms. *Journal of Kufa for Mathematics and Computer*, *1*(3), 8–24.

B, Z. Z., Robinson, D., & Tepper, J. (2018). *Detecting Hate Speech on Twitter Using a Convolution-GRU Based Deep Neural Network* (Vol. 1). https://doi.org/10.1007/978-3-319-93417-4

Bang, S. H., Lee, M., Ji, B., Kim, J., Park, H., & Cho, H. (2017). *Data Mining Framework for Identifying Crime Patterns Using Unsupervised Learning : Focusing on Assault and Theft*. (December 2016).

Bharati, A., & K, S. R. A. (2018). *Crime Prediction and Analysis Using Machine Learning*. 1037–1042.

Bharati, A., & RA.K, S. (2018). *Crime Prediction and Analysis Using Machine Learning*. 1037–1042.

Borr, R. (2018). *Introduction to Deep Learning with TensoFlow and Keras libraries with some examples in biomedical research at the Hospital Clinic of Barcelona*.

Bray, F., & Ferlay, J. (2005). *Age standardization*. 112–115.

Brownlee, J. (2017). *Long Short-Term Memory Networks With Python* (1.0).

Chen, H., Chung, W., Jie, Xu, J., Wang, G., Yi, Q., & Chau, M. (2004). Crime Data Mining : A General Framework. *IEEE*.

Cruz-benito, J. (2016). *Systematic Literature Review & Mapping*. https://doi.org/10.5281/zenodo.165773

Endalew, A. (2017). Crime Analysis For The Case of Bahir Dar City Administration Using Data Mining Approach. Bahir Dar University.

Fujikawa, Y., & Ho, T. (2002). Cluster-based Algorithms for Filling Missing Values. *In 6th Pacific-Asia Conferene on Knowledge Discovery and Data Mining,*. Taiwan.

Gensler, A., Henze, J., Sick, B., & Raabe, N. (2016). Deep Learning for Solar Power Forecasting – An Approach Using Autoencoder and LSTM Neural Networks. *IEEE International Conference on Systems, Man, and Cybernetics*.

Gottlieb, S., & Arenberg, S. (2000). *Crime Analysis : From Concept to Reality*. California, office of criminal justice planning.

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM : A Search Space Odyssey. *Transactions on Neural Networks and Learning Systems*, 1–12. https://doi.org/10.1109/TNNLS.2016.2582924

Hagen E. Frank. (2010). *Crime Types and Criminals* (J. Westby, L. Dutro, E. Oettinger, K. Wiley, T. Herlinger, V. Reed-Castro, … J. Reed-Banado, Eds.). London, United Kingdom: SAGE publications Ltd.

Hassen, S. (2014). A CRITICAL ANALYSIS OF INDIGENOUS AND MODERN POLICING IN ETHIOPIA. South Africa.

Hilbert, M. (2016). Big Data for Development : A Review of Promises and Challenges. *Development Policy Review*, *34*(1), 135–174. Retrieved from https://doi.org/10.1111/dpr.12142

Huenerfauth, M., State, P., Google, G. V. R., & Caltech, R. P. M. (2009). *Introduction to Python*.

Hussein, R., Palangi, H., Ward, R. K., & Wang, Z. J. (2018). Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals. *Clinical Neurophysiology*, *130*(1), 25–37. https://doi.org/10.1016/j.clinph.2018.10.010

Iqbal, R., Azrifah, M., Murad, A., & Mustapha, A. (2013). An Experimental Study of Classification Algorithms for Crime Prediction. *Indian Journal of Science and Technology*, *6*(3).

Jabar, E. K., Hashem, S. H., & Hessian, E. M. (2013). Propose Data Mining AR-GA Model to Advance Crime analysis. *IOSR Journal of Computer Engineering*, *14*(5), 38–45.

Jadhav, S. B., Rajpal, N., Mulla, H., & Patel, N. (2017). Cops Friend : A Crime Pattern

Analysis using Machine Learning to Reduce the Crime Rate. *International Journal for Scientific Research & Development*, *5*(04), 272–276.

Kalantari, M., Ghavagh, A. R., Toomanian, A., & Dero, Q. Y. (2016). A New Methodological Framework for Crime Spatial Analysis Using Local Entropy Map. *Modern Applied Science*, *10*(9), 179. https://doi.org/10.5539/mas.v10n9p179

Kaliakatsos-papakostas, M., Karydis, I., & Kermanidis, K. L. (2017). Combining LSTM and Feed Forward Neural Networks for Conditional Rhythm Composition. *Conference on Communications in Computer and Information Science*. https://doi.org/10.1007/978-3-319-65172-9

Kang, H., & Kang, H. (2017). Prediction of Crime Occurrence From Multi- Modal Data Using Deep Learning. *Https://Doi.Org/10.1371/Journal.Pone.0176244*, 1–19.

Keyvanpour, M., Javideh, M., & Reza, M. (2010). Detecting and investigating crime by means of data mining : a general crime matching framework. *Procedia Computer Science*, 872–880. https://doi.org/10.1016/j.procs.2010.12.143

Krishnan, A., Sarguru, A., & Sheela, A. C. S. (2018). PREDICITVE ANALYSIS OF CRIME DATA USING DEEP LEARNING. *International Journal of Pure and Applied Mathematics*, *118*(20), 4023–4031.

Leul, W. (2003). The Application of Data Mining in Crime Prevention : The Case of Oromia Police Commission. Addis Ababa University.

Lin, Y., & Chen, T. (2017). *Using Machine Learning to Assist Crime Prevention*. 1029–1030. https://doi.org/10.1109/IIAI-AAI.2017.46

Lin, Y., Yen, M., & Yu, L. (2018). *Grid-Based Crime Prediction Using Geographical Features*. https://doi.org/10.3390/ijgi7080298

Makris, D., Kaliakatsos-Papakostas, M., Karydis, I., & Kermanidis, K. L. (2019). Conditional neural sequence learners for generating drums' rhythms. *Neural Computing and Applications*, *31*(6), 1793–1804. https://doi.org/10.1007/s00521-018-3708-6

Mcclendon, L., & Meghanathan, N. (2015). *U SING M ACHINE L EARNING A LGORITHMS TO A NALYZE C RIME D ATA*. *2*(1), 1–12.

Mckinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceeding of the 9th Python in Science Confrence*, 56–61.

Medsker, L. R., & Jain, L. C. (2001). *Recurrent Neural Networks Design and Applications*.

Meti, K. (2016). An Assessment of Socio-Economic Factors on Crime: A Case Study of Kaliti Correctional Administration, Ethiopia. *Unpublished Master's Thesis, Addis Ababa University*.

Moolayil, J. (2019). *Learn Keras for Deep Neural Networks*. https://doi.org/10.1007/978-1-4842-4240-7

Mowafy, M., Rezk, A., & El-bakry, H. M. (2018). General Crime Mining Framework for Unstructured Crime Data Prediction. *International Journal of Computer Application*, *4*(8). https://doi.org/10.26808/rs.ca.i8v4.02

Nasridinov, A., Ihm, S., & Park, Y. (2013). *A Decision Tree-Based Classification Model for Crime Prediction*. https://doi.org/10.1007/978-94-007-6996-0

Natarajan, M. (2016). Crime in developing countries: The contribution of crime science. *Crime Science*, *5*(1). https://doi.org/10.1186/s40163-016-0056-7

Nguyen, T. T., Hatua, A., & Sung, A. H. (2017). Building a Learning Machine Classifier with Inadequate Data for Crime Prediction. *Journal of Advances in Information Technology*, *8*(2), 141–147. https://doi.org/10.12720/jait.8.2.141-147

Oliphant, T. E. (2006). *Guide to NumPy*.

Organization, I. L. (2012). *International Standard Classification of Occupation*. Geneva, Switzerland: International Labour Organization.

Patterson, J., & Gibson, A. (2017). *Deep Learning A practitioner's approach* (1st editio). Retrieved from http://oreilly.com/catalog/errata.csp?isbn=9781491914250

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Raybaut, P. (2017). *Spyder Documentation*.

Rossum, G. V. and python development team. (2020). *Python Tutorial Release3.8.1*. 1.

Roth, R. E., Ross, K. S., Finch, B. G., Luo, W., & Maceachren, A. M. (2009). *A user-centered approach for designing and developing spatiotemporal crime analysis tools 1 . Introduction to Crime Analysis and GeoVISTA CrimeViz*. (Norman 1988).

Saeed, U., Sarim, M., Usmani, A., Mukhtar, A., Shaikh, A. B., & Raffat, S. K. (2015). Application of Machine learning Algorithms in Crime Classification and Classification Rule Mining. *Research Journal of Recent Sciences*, *4*(3), 106–114.

Sathyadevan, S., & Devan, M. S. (2014). Crime Analysis and Prediction Using Data Mining. *First International Conference on Networks and Soft Computing*, 406–412.

Shermila A., M., Bellarmine, A. B., & Santiago, N. (2018). Crime Data Analysis and Prediction of Perpetrator Identity using Machine Learning Approach. *2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, 107–114.

Sreedevi, M., Reddy, A. H. V., Venakata, C., & Krishna, S. (2018). *Review on Crime Analysis and Prediction Using Data Mining Techniques*. 3360–3369. https://doi.org/10.15680/IJIRSET.2018.0704019

Stalidis, P., Semertzidis, T., Daras, P., & Member, S. (2018). Examining Deep Learning Architectures for Crime Classification and Prediction. *ArXiv:1812.00602v1 [Cs.LG]*, 1–12.

Stec, A., & Klabjan, D. (2018). Forecasting Crime with Deep Learning. *ArXiv:1806.01486 [Stat.ML]*, 1–20.

The International Standard Classification of Education (ISCED) 2011 (Vol. 5). (2012). https://doi.org/10.1007/BF02207511

Thotakura, S. (2011). Crime: A Conceptual Understanding. *Indian Journal of Applied Research*, *4*(3), 196–198. https://doi.org/10.15373/2249555X/MAR2014/58

ToppiReddy, H., Saini, B., & Mahajan, G. (2018). Crime Prediction & Monitoring Framework Based on Spatial Analysis. *Procedia Computer Science*, *132*(Iccids), 696–705. https://doi.org/10.1016/j.procs.2018.05.075

Tran, T. (2016). Creating A Text Generator Using Recurrent Neural Network. Retrieved November 2, 2019, from https://chunml.github.io/ChunML.github.io/project/Creating-Text-Generator-Using-Recurrent-Neural-Network/

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, *14*, 207–222.

Tsion, E. (2019). Feasibility Analysis of Long Short-Term MemoryRecurrent Neural Network in Time Series Crime Prediction: A Case of Bole Sub City Police Departement. *Unpublished Master's Thesis Adama Science and Technology University*. Retrieved from Adama Science and Technology University

Wang, B., Zhang, D., Zhang, D., Brantingham, P. J., & Bertozzi, A. L. (2017). Deep Learning for Real Time Crime Forecasting. *ArXiv:1707.03340v1 [Math.NA]*, 1–4.

Wang, K., Zhu, P., Zhu, H., & Cui, P. (2017). An Interweaved Time Series Locally Connected Recurrent Neural Network Model on Crime Forecasting. *Springer International Publishing*, *2*, 466–474.

Wang, X., Brown, D. E., & Gerber, M. S. (2012). Spatio-Temporal Modeling of Criminal Incidents Using Geographic , Demographic , and Twitter-derived Information. *IEEE*, 36–41.

Weston, S., & Bjornson, R. (2016). Introduction to Anaconda. *Yale Center for Research Computing Yale University*.

Yamuna, S., & Bhuvaneswari, N. S. (2012). □ Datamining Techniques to Analyze and Predict Crimes Output : *International Journal of Engineering and Science*, 243–247.

Zheng, J., Xu, C., Zhang, Z., & Li, X. (2017). *Electric Load Forecasting in Smart Grid Using Long-Short-Term-Memory based Recurrent Neural Network*. 1–6. https://doi.org/10.1109/CISS.2017.7926112

Zhuang, Y., Almeida, M., Morabito, M., & Ding, W. (2017). Crime Hot Spot Forecasting : A Recurrent Model with Spatial and Temporal Information. *IEEE International Conference on Big Knowledge*. https://doi.org/10.1109/ICBK.2017.3