Machine Learning Approach for Crime Prediction: A comparative Analysis

Lelise Mitiku Dabale

A Thesis Submitted to the Department of Computer Science and Engineering

School of Electrical Engineering and Computing

Presented in Partial Fulfillment of the Requirement for the Degree of Master of Science in computer science and Engineering

Office of Graduate Studies

Adama Science and Technology University

October, 2024

Adama, Ethiopia

Machine Learning Approach for Crime Prediction: A comparative Analysis

Lelise Mitiku Dabale

Sintayew Hirphasa (PhD)

A Thesis Submitted to the Department of Computer Science and Engineering

School of Electrical Engineering and Computing

Presented in Partial Fulfillment of the Requirement for the Degree of Master of Science in computer science and Engineering

Office of Graduate Studies
Adama Science and Technology University

October, 2024

Adama, Ethiopia

DECLARATION

Hereby declare that this Master Thesis entitled "**Machine Learning Approach for Crime Prediction: A comparative Analysis**" is my original work. That is, it has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials that are used for this thesis have been duly acknowledged through citation.

Lelise Mitiku Dabale

______________________________	______________________	______________
Name of student	Signature	Date

ADVISOR RECOMMENDATION

I/we, the advisor(s) of this thesis, hereby certify that I/we have read the revised version of the thesis entitled "**Machine Learning Approach for Crime Prediction: A comparative Analysis**" prepared under my/our guidance by **Lelise Mitiku Dabale** submitted in partial fulfillment of the requirements for the degree of Masters of Science in computer science and Engineering. Therefore, I/we recommend the submission of the revised version of the thesis to the department following the applicable procedures.

Sintayew Hirphasa (PhD)	____________________	____________________
Advisor/ Supervisor	Signature	Date

APPROVAL PAGE

I/we, the advisors of the thesis entitled "**Machine Learning Approach for Crime Prediction: A comparative Analysis**" and developed by **Lelise Mitiku Dabale** hereby certify that the recommendation and suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis.

Sintayew Hirphasa (PhD) ______	______	______
Advisor	Signature	Date

We, the undersigned, members of the Board of Examiners of the thesis by **Lelise Mitiku Dabale** have read and evaluated the thesis entitled "**Machine Learning Approach for Crime Prediction: A comparative Analysis"** and examined the candidate during the open defense. This is, therefore, to certify that the thesis is accepted for partial fulfillment of the requirement of the degree of Master of Science in computer science and Engineering.

______	______	______
Chairperson	Signature	Date
______	______	______
Internal Examiner	Signature	Date
______	______	______
External Examiner	Signature	Date

Final approval and acceptance of the thesis are contingent upon the submission of its final copy to the Office of Postgraduate Studies (OPGS) through the Department Graduate Council (DGC) and School Graduate Committee (SGC).

______	______	______
Department Head	Signature	Date
______	______	______
School Dean	Signature	Date
______	______	______
Office of Postgraduate Studies	Signature	Date

ACKNOWLEDGMENTS

I want to start by saying thank you God and the Mother of God for their guidance and blessings throughout this journey. I am extremely thankful of my thesis advisor, Dr. Sintayew Hirphasa, for his guidance, support, and encouragement. His knowledge and consistent assistance were critical to the completion my paper.

I sincerely thank the Ethiopian Federal Police Commission for their cooperation in providing the crime data necessary for this research. Their support was crucial in enabling the successful completion of this study. In addition, I'm thankful to Mr. Chalsan Damana and Mr. Gadisa Adamu for their valuable contributions and assistance during challenging times. They were greatly appreciated for their help and advice.

Lastly, I extend my sincere thanks to my family for their love and support. I would especially like to thank my sister for her continuous support and encouragement. Their belief in me kept me motivated and determined to achieve my goals.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DL	Deep Learning
DNN	Deep Neural Network
E.C	Ethiopian Calendar
EFPC	Ethiopia Federal Police Commission
EDL	Education Level
FN	False Negative
FP	False Positive
KNN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MLP	Multilayer Perceptron
PC	Property Crime
RL	Reinforcement Learning
RNN	Recurrent Neural Network
SC	State Crime
SVM	Support Vector Machine
VC	Violent Crime
VCC	Victimless Crime

ABSTRACT

Crime is a global issue that affects countries at all stages of development. It impacts the local economy, quality of life, and societal well-being, leading to various social problems. In Africa, including Ethiopia, crime rates vary widely across regions, both urban and rural. One of the main challenges faced by law enforcement agencies, such as the Ethiopian Federal Police Commission, is the lack of sophisticated analytical tools for predicting and preventing illegal activities. This limitation often leads to the use of traditional and time-consuming crime analysis methods, which are often ineffective in addressing the rising crime rates. Previous research has used various algorithms to predict crime, but there is still room for improvement in terms of accuracy and practical application. This research focuses on crimes categorized under criminal law, using a dataset acquired from the Ethiopian Federal Police Commission. The aim is to develop an automatic crime type prediction model by combining machine learning and deep learning techniques. We chose five models, each bringing unique strengths that contribute to accurate crime type predictions based on the available data. The first step involved preprocessing the dataset by cleaning the data, normalizing numerical features, handling missing values, encoding categorical variables, and selecting features. The performance of the models was assessed by dividing the data into training and testing sets. Next, we built and trained the predictive models. Each model was trained on the training set using selected features, and techniques like cross-validation were employed to tune hyperparameters and prevent overfitting. Model performance was evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix to gauge how well each model predicts crime types. Among the models, MLP achieved the highest accuracy at 98.64%, proving the most effective for crime type prediction.

KEYWORDS: Crime prediction, Machine learning, Support Vector Machine, K-Nearest Neighbors, Multilayer Perceptron, Long Short-Term Memory, Recurrent Neural Network

CHAPTER ONE
INTRODUCTION

1.1 Background of the study

Crime is a significant issue that affects countries worldwide, regardless of their level of development. It has serious consequences on the economy, reducing productivity and investment, and it negatively affects the quality of life and the general well-being of societies, often leading to various social issues (Alsaqabi, Aldhubayi, & Albahli, 2019). In Africa, and specifically in Ethiopia, crime rates show significant variation between urban and rural areas. Several factors contribute to these criminal activities, including the growing unemployment rate, high population density, and instability in security (Alsaqabi et al., 2019). The impact of crime extends beyond the direct harm to the victims; it weakens social ties, disrupts community life, and hampers economic progress.

To address these issues, Ethiopia has made efforts to improve security by passing laws aimed at strengthening law enforcement agencies and promoting safer neighborhoods. However, despite these steps, there are still ongoing challenges, such as political instability, organized crime networks, and corruption that continue to make the crime situation worse (Blackburn* & Matthews, 2011). Criminal activities can occur at any time and in any place. Often, criminals commit offenses in areas where they feel comfortable, and once they succeed, they tend to repeat the crime under similar circumstances (Bothos & Thomopoulos, 2016). This suggests that crime incidents often follow predictable patterns or behaviors, which can be studied and analyzed.

To maintain public safety and security, it is crucial to predict and prevent crime. In recent years, technological advancements, particularly in the fields of machine learning and predictive analytics, have provided new ways to forecast and reduce criminal activities (Carleo et al., 2019). These technologies allow law enforcement agencies to shift from traditional methods to more data-driven approaches, making it easier to anticipate crime trends. However, even with these advancements, the job of crime analysts remains tough, as they often have to manually analyze crime patterns.

Identifying crime patterns manually can be slow and complicated. This is where data-driven methods and predictive software come into play, allowing authorities to analyze large-scale trends and background crime rates.

Crime prediction involves the application of sophisticated analytical tools to support and improve existing crime detection systems (Choi, Coyner, Kalpathy-Cramer, Chiang, & Campbell, 2020). With the increasing availability of crime-related data and the constant improvements in technology, researchers now have more opportunities to explore how machine learning can be used for crime detection and prevention.

The rapid rise in criminal activities makes it even more important to develop and implement technology that can help law enforcement agencies solve cases more quickly and efficiently (Choi et al., 2020). In Ethiopia, like in many other countries, there is a growing need for more effective ways to predict and prevent crime. Ethiopia's unique socio-economic and political context adds layers of complexity to the problem (Chun et al., 2019). However, by utilizing technology and adopting data-driven approaches, law enforcement agencies in Ethiopia can enhance their ability to predict criminal activities and reduce crime rates, ultimately creating a safer and more secure environment for its citizens (Meskela, Afework, Ayele, Teferi, & Mengist, 2020).

1.2 Motivation

Crime is a complicated problem that significantly impacts societies, economies, and public safety worldwide. In Ethiopia, a number of issues, such as increased rates of joblessness and impact the dynamics of criminal activity, varying population densities, and security instability (Bachman & Schutt, 2013). The nation faces issues like unstable politics, organized crime, and corruption despite legislative measures to resolve security concerns. One of the main issues facing law enforcement agencies, such as the Federal Police Commission in Ethiopia, is the lack of sophisticated analytical tools for predicting and preventing illegal activities (Cunneen & Tauri, 2019). This limitation often leads traditional and time- consuming crime analysis methods, which can be ineffective in addressing the rising incidence of crime (Biks, 2024). The belief that deep learning and advanced machine learning techniques offer a potential route to improving crime prediction methodologies is what drives this research. By leveraging predictive analytics, law enforcement agencies can enhance their ability to allocate resources more effectively and transition towards a more proactive approach to crime prevention (Walker & Maddan, 2019).

Ultimately, this research aims to address existing gaps in crime prediction mechanisms and Help create the total public safety and security in Ethiopia. By implementing state-of-the-art technical applications within the law enforcement framework, the study seeks to demonstrate how the intelligent use of machine learning and deep learning can significantly improve the Commission's ability to predict and prevent crime, thereby creating a safer environment for all Ethiopians.

1.3 Statement of the Problem

According to Blackburn and Matthews (2011), crime rates are increasing in both seriousness and variety. Ethiopia's stability and security are upheld by the Federal Police Commission (EFPC). But as (Longe, Ngwa, Wada, Mbarika, & Kvasny, 2009) point out, the Commission has a difficult time predicting and preventing illegal activity. The Commission's capacity to proactively handle evolving criminal trends is restricted by the lack of advanced analytical instruments and new methods for analyzing crime data (Blackburn* & Matthews, 2011). As the number of crimes grows and security becomes unstable, authorities struggle to accurately predict crime, resulting in a rise in unlawful activity across. Current methods for identifying crime hotspots are largely manual and based on location, which makes them time-consuming (Janka, 2013).

Using machine learning approaches, some researchers have attempted to address the problem of crime prediction. However, the studies have identified critical issues. The datasets they used were biased and insufficient. They did not get all of the information they needed from police reports. So information did not fully represent all possible crime situations or types. This could lead to missing important factors and reduce the reliability of the data, affecting the accuracy of their results. Biased datasets usually contain information about only specific crime records, which can lead to inaccurate predictions. (Phillips et al., 2009) explains that using large and unbiased datasets improves accuracy because they cover a wider range of scenarios and reduce the chance of bias, making the findings more reliable.

Kanimozhi et al. (2021) researched crime type and occurrence prediction using machine learning but relied on online datasets. These differ from real-world conditions, affecting the accuracy of the results. Considering the actual environment is crucial for developing an effective model (A. Newman, 2021).

P. Yerpude (2020) emphasizes that key attributes for crime prediction include demographic information, socio-economic factors, and time and location details. These attributes help identify crime patterns, which improve predictive model accuracy and encourage the use of efficient crime prevention measures. Safat et al. (2021) didn't incorporate socioeconomic and demographic characteristics in their use of machine and deep learning methods. Additionally, some researchers used outdated data in their research on crime prediction and monitoring, missing important changes in crime patterns. These concerns are the focus of this thesis by developing deep learning and machine learning models. By applying advanced analytical tools, we aim to improve crime prediction strategies and enhance the Commission's ability to predict and prevent crime.

1.4 Research Questions

- RQ1. What key hyperparameters greatly influence the effectiveness of crime prediction models?
- RQ2.Which attributes play a crucial role in enhancing the accuracy of crime prediction models?
- RQ3.Which algorithms deliver the most accurate and efficient outcomes for crime prediction?

1.5 Objective of the Study

1.5.1 General Objective

The primary goal of this study is to create a model that accurately predicts crimes by comparing various machine learning algorithms.

1. 5.2 Specific Objectives

- ✓ To identify key attributes that significantly improves the accuracy of crime prediction models.
- ✓ To determine and examine the key hyperparameters that significantly impacts the performance of crime prediction models.
- ✓ To evaluate and compare different algorithms to identify highly efficient one for crime prediction.

1.6 Significance of the Study

The goal of this paper is to improve crime prevention by utilizing historical crime data analysis and predictive algorithms. By applying machine learning and deep learning techniques, law enforcement can shift from being reactive to proactive in predicting criminal activity, which leads to more accurate prediction.

The research provides valuable insights into crime patterns, helping law enforcement anticipate criminal behavior, predict future crime types, and allocate resources more effectively, such as police forces and patrols. Identifying crime hotspots enables preventive measures, which ultimately reduces crime rates. A safer city improves the quality of life, making it more attractive for living, working, and investing, which boosts economic growth. This research is a crucial phase in the right direction strengthening public safety and security by improving crime prevention methods through advanced technology.

1.7 Scope of the Study

In this research, we developed a model using a dataset that we obtained from the Ethiopia Federal Police Commission. Data from three sub-cities are included in this dataset, covering the period from September 2011 to October 2016 E.C. In the Ethiopian context, there are three major types of crimes: traffic crimes, crimes under social law, and crimes addressed by criminal law (Jibat & Nigussie, 2014) . For this research, we consider only crimes categorized under criminal law.

1.8 Limitations of the Study

This study does not take into account seasonal variations, such as the impact of holidays and public events, but rather excludes environmental data. The study excludes environmental data and takes in to account seasonal fluctuations, such as the impact of holidays and public events. In the absence of this information, it is impossible for the model to comprehend the factors that impact crime rates throughout the year or in different settings.by utilizing seasonal and environmental information, it is possible to enhance the accuracy of predictions and gain insight in to how crime is affected by various factors.

1.9 Thesis Organization

The second chapter offers an overview of crime prediction, including how machine learning and deep learning algorithms are used in this field. It also discusses the factors influencing crime prediction and reviews related research conducted by various researchers. Chapter three is dedicated to the research methodology. This chapter covers data preparation, preprocessing, feature selection, model training, performance measuring metrics, and the hardware and software tools used in the paper. The Fourth chapter discusses the proposed solution and its implementation. It details the tasks completed during the implementation process, including data preprocessing and feature selection, algorithm training, and evaluations. Chapter five describes outcomes of the performance evaluation of the models. Chapter six concludes the thesis, offering conclusions, recommendations, contributions, and suggestions for future work.

CHAPTER TWO

LITERATURE REVIEW AND RELATED WORK

2.1 Crime Prediction

According to (Shamsuddin, Ali, & Alwee, 2017) The process of predicting future crime rates within a given geographic area or demography involves the methodical examination of past crime statistics and behavioral tendencies. It will function as proactive strategy, enabling law enforcement agencies to allocate resources strategically and intervene preventively, aiming to reduce crime rates and enhance public safety (Almanie, Mirza, & Lor, 2015).Traditionally, crime prediction relied on statistical models and historical data analysis to identify trends and patterns in criminal behavior(Zhang et al., 2022). However, these regular methods often lacked precision and failed to adapt dynamically to evolving criminal tactics and societal changes, leading to limitations in their effectiveness (ToppiReddy, Saini, & Mahajan, 2018). A subset of artificial intelligence is machine learning, will emerge as a promising paradigm in crime prediction due to its capacity to analyze vast datasets, identify complex patterns, and generate predictive models. The use of this technology strengthens law enforcement to move beyondreactive measures by providing predictive insights based on historical crime data (Gahalot, Dhiman, & Chouhan, 2020).

2.2 Machine Learning in Crime Prediction

A machine learning algorithm is a computational process that uses input data to achieve a desired task without being literally programmed to produce a particular outcome. These algorithms are in a sense in that they automatically alter or adapt their architecture through repetition (i.e., experience) so that they become better at achieving the desired task (Jenga, Catal, & Kar, 2023). The algorithm then optimally configures itself, according to Jordan & Mitchell (2015), so that it can generate the intended result from new, unseen data in addition to producing the desired result when given the training inputs. The "learning" component of machine learning is this training(Vaquero Barnadas, 2016). It is not necessary for the training to be restricted to an initial adaption over a set amount of time. As with humans, a good algorithm can practice “lifelong” learning as it processes new data and learns from its mistakes. There are many ways that a computational algorithm able to change in response to instruction (Reier Forradellas, Náñez Alonso, Jorge-Vazquez, & Rodriguez, 2020). The input data can be selected and weighted to provide the most decisive outcomes. The algorithm's mathematical parameters may vary and can be changed by iterative improvement.

It might be connection of possible computational pathways that it arranges for optimal results. It can determine probability distributions from the input data and use them to predict outcomes (Wu, Wang, Cao, & Jia, 2020). Algorithms for machine learning are often categorized into supervised machine learning, unsupervised machine learning and reinforcement learning.

2.2.1 Supervised Machine Learning

The different algorithms used in supervised learning produce a function that connects inputs to intended outputs. Prediction tasks are characterized by a conventional formulation of supervised learning because the objective is to predict or classify a particular result of interest. By examining multiple input-output examples of the function, the learner must be able to approximate the behavior of a function that translates a vector into one of several classes (Nasteski, 2017). A basic machine-learning model splits the learning process into two stages: testing and training. Samples from the training data are used as input in the training process, and the learner uses these features to develop the learning model. The learning model makes advantage of the execution engine during the testing process (Burkart & Huber, 2021).

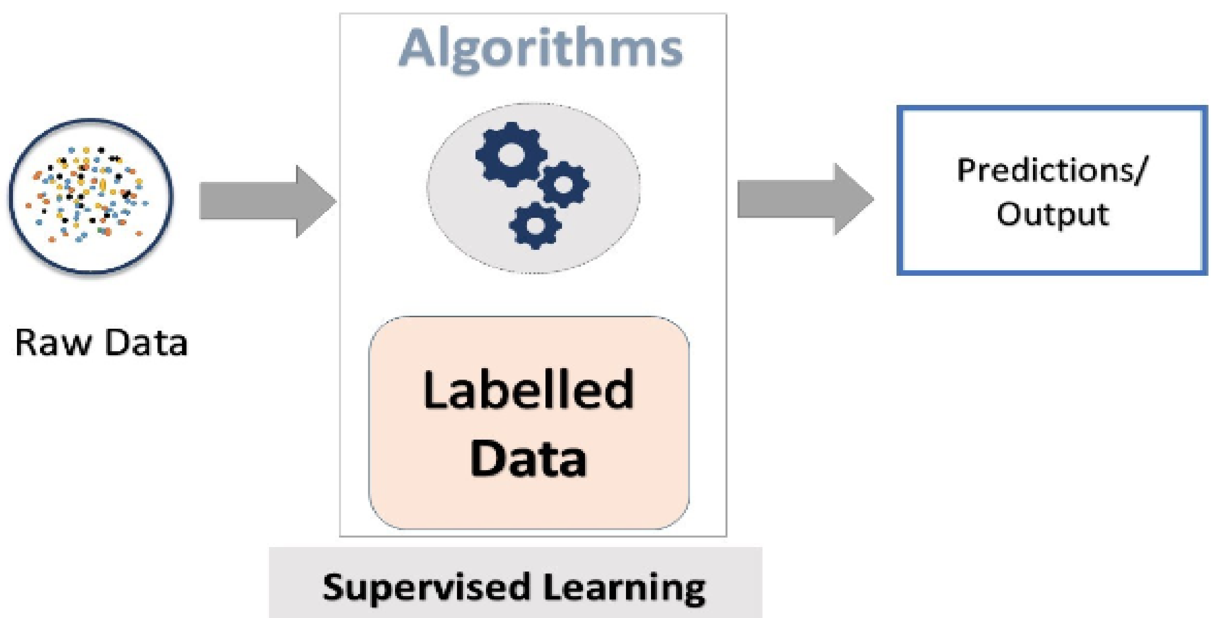

Figure 1: Supervised learning process (Radhoush, Whitaker, & Nehrir, 2023)

2.2.2 Unsupervised Machine Learning

A machine can learn using unsupervised learning if it is not under any supervision. The machine is trained with a collection of unlabeled, unclassified, and uncategorized data, and the algorithm must act on that data independently (Eckhardt et al., 2023). Restructuring the input data into new features or a collection of objects with related patterns is the aim of unsupervised learning (Naeem, Ali, Anam, & Ahmed, 2023). Few features are learnt from the data by the unsupervised learning algorithms. It recognizes the class of the data when it is introduced by using the previously learnt features. Clustering and feature reduction are its primary applications (Celebi & Aydin, 2016). In exploratory data analysis, the most popular unsupervised learning technique is clustering, which is used to uncover hidden patterns or groupings in the data. Dimensionality reduction, which aims to minimize the number of input variables in a dataset while retaining as much of the pertinent information as feasible, is another use for unsupervised learning (Ghahramani, 2003). Additional uses include association rule learning, which is used to uncover intriguing correlations between variables in big datasets, like those frequently found in market basket analysis, and anomaly detection, in which the algorithm learns what constitutes "normal" in the data and can detect outliers or anomalies.

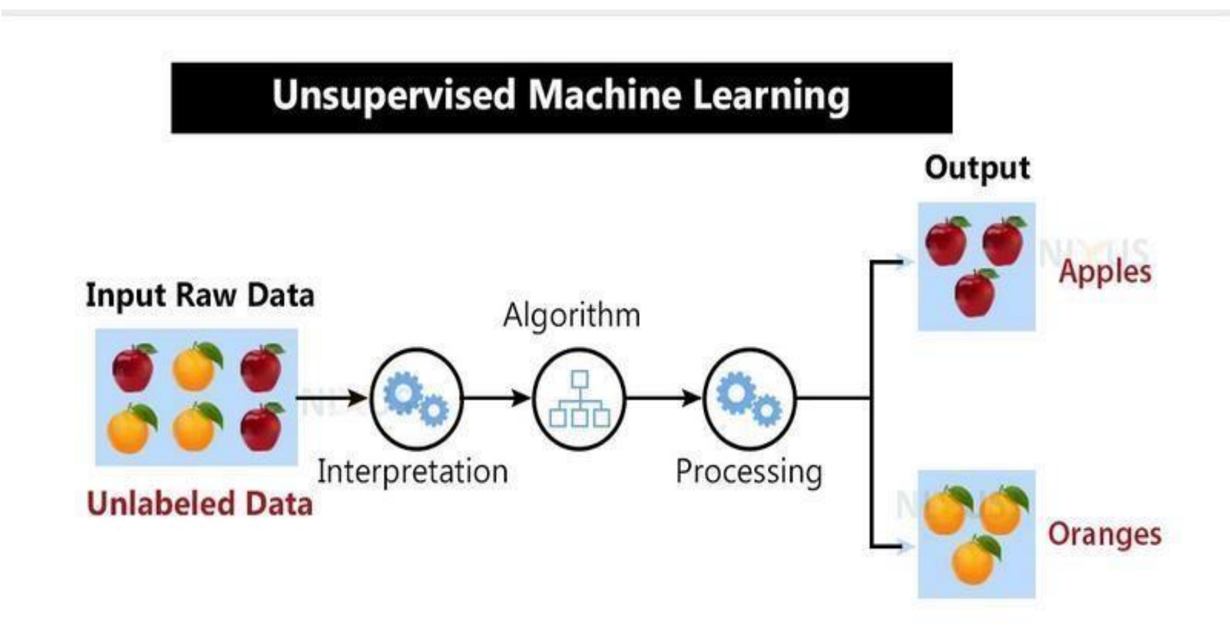

Figure 2: Unsupervised learning process (Ghahramani, 2003)

2.2.3. Reinforcement Learning

A learning agent that uses reinforcement learning receives rewards for correct actions and penalties for incorrect ones. Reinforcement learning is a feedback-based learning approach. With these inputs, the agent automatically gains knowledge and enhances its functionality (Qiang & Zhongli, 2011). The agent engages in interaction and exploration with the environment in reinforcement learning (Shakya, Pillai, & Chakrabarty, 2023).An agent's objective is to maximize reward points since doing so enhances performance (Kaelbling et al., 1996). Robotics is one of the main fields in which reinforcement learning (RL) is applied. Using RL algorithms, robots may be trained to accomplish intricate tasks including grabbing objects, navigating through surroundings, and even playing games. (Kaelbling, Littman, & Moore, 1996) Autonomous cars also employ reinforcement learning to acquire safe and effective.

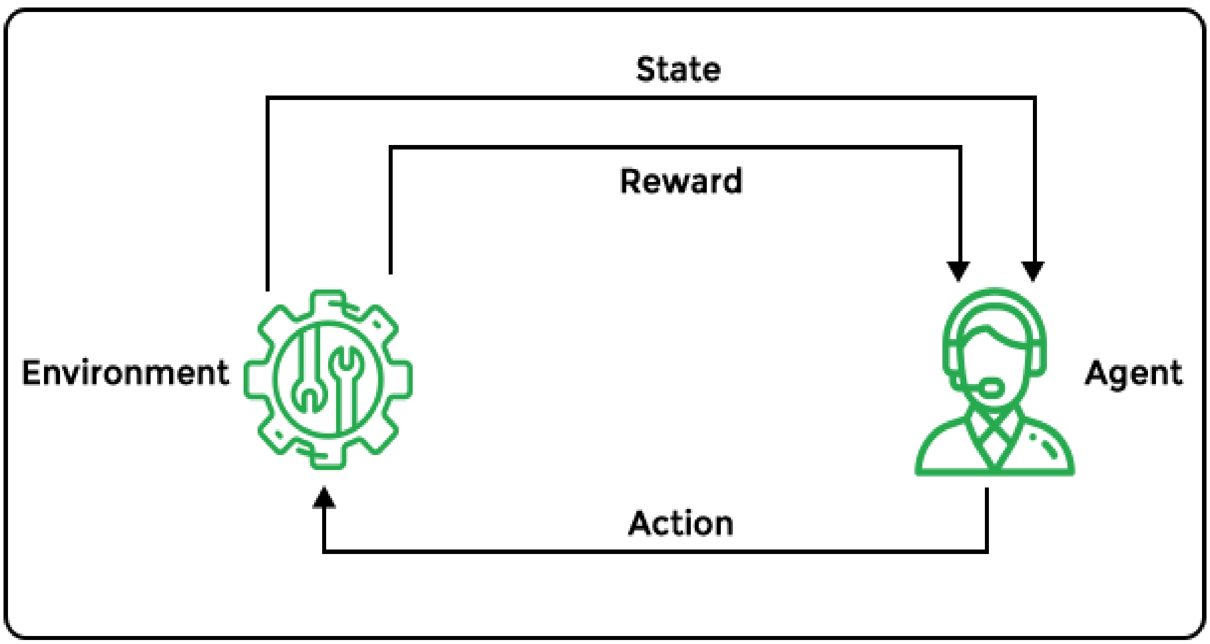

Figure 3: Reinforcement learning process (Kaelbling et al., 1996)

2.3 Deep Learning Techniques

A subfield of machine learning known as deep learning (DL) excels at handling unstructured data considerably better. According to (LeCun, Bengio, & Hinton, 2015), deep learning approaches surpass machine learning techniques. It facilitates step-by-step feature learning by computer models using data at several levels. According to (Goodfellow, 2016), deep learning gained popularity as more data became available and hardware advanced. It can handle a lot of features when working with unstructured data, which allows it to gain increased power and flexibility. Soon, a broad range of applications for human assistance will use deep learning algorithms. (LeCun et al., 2015)They are inclined to allow a computer to experiment with its presumptions. The majority of deep learning algorithms simulate the neuronal connections seen in the human brain in order to integrate artificial intelligence into a computer system. This enhances operating precision and speed for numerous crucial operations. These days, computers are programmed to carry out certain jobs (Goodfellow, 2016). Before the deep learning algorithm can be examined, input is needed. Similar to human vision, images can be analyzed in a variety of ways. According to (Bengio, Goodfellow, & Courville, 2017), numerical data, voice, digital signals, and sensor outputs are all regarded as inputs for deep learning algorithms. Nevertheless, the deep learning algorithm needs a suitable preprocessing step to remove the differences in format (Wang, Chen, Xu, & Jin, 2015).

This leads to an increase in the algorithm's precision. Noise is eliminated and the region of interest is marked during the preprocessing stage. Subsequently, the study is uninterruptedly focused on a specific location (Bengio et al., 2017). Data is sent through multiple layers in a deep learning algorithm, each of which can extract features and send them to the next layer gradually (Buduma, Buduma, & Papa, 2022). The primary distinction between deep learning and classical machine learning, according to (Shrestha & Mahmood, 2019), is that whereas DL uses a high number of hidden layers to learn many features, deep learning automatically learns features for more difficult tasks. Over the past few years, it has gained recognition (Pouyanfar et al., 2018).

2.3.1 Deep Neural Network

Many (deep) layers of extremely optimized algorithms and architectures make up a deep neural network. Numerous deep neural network algorithms exist, such as the multilayer perceptron, which uses multiple hidden neurons to learn data deeply, and the convolutional neural network, which has more applications in images, and the recurrent neural network, (Mishra & Gupta, 2017)which is frequently used with sequential and time series data (Liu et al., 2021). Without the requirement for manually created feature extractors, representations from data sets are learnt using a multilayer perceptron model. In contrast to shallow learning, which has fewer unit layers, deep learning, as its name suggests, has more or deeper processing layers (Sewak, Sahay, & Rathore, 2020). Transitioning from shallow to deep learning architectures has made it possible to map (Samek, Montavon, Lapuschkin, Anders, & Müller, 2021).

2.3.2 Recurrent Neural Network

A Recurrent Neural Network (RNN) is a type of neural network where the output from one step is fed back as input into the next step (Cha & Kang, 2018). Unlike traditional neural networks, where each input and output is handled independently, RNNs are designed to handle situations where past information is important. For example, when trying to predict the next word in a sentence, the previous words are essential (DiPietro & Hager, 2020). In traditional networks, this wouldn’t be possible because they don’t "remember" previous inputs. To solve this problem, RNNs were created, and they use something called a "Hidden Layer" to retain information from earlier steps. This hidden layer allows the network to remember and use the context of previous inputs when making predictions (Minar & Naher, 2018).

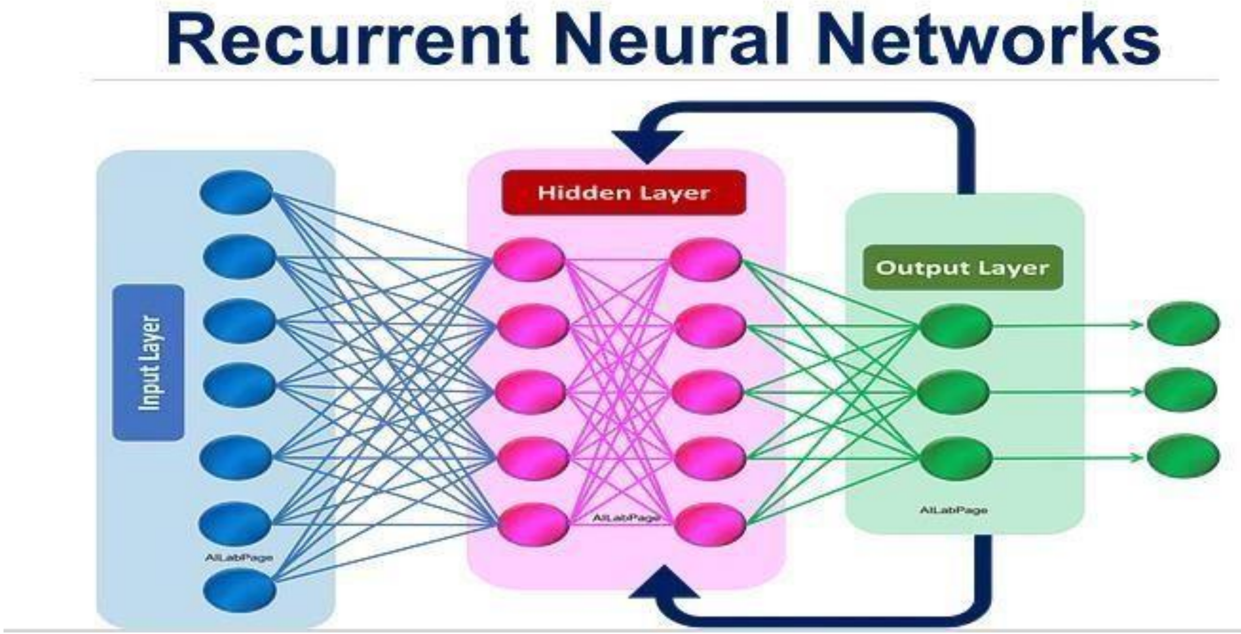

Figure 4: Structure of Recurrent Neural Network (Xiao & Zhou, 2020)

2.3.3. Application of RNN

Recurrent Neural Networks (RNNs) are widely used when making predictions about future events. In the financial industry, RNNs help forecast stock prices and predict the overall movement of the stock market, whether it will go up or down. For instance, a study in (Minar & Naher, 2018) used RNNs in autonomous vehicles to predict the vehicle's path, helping to avoid accidents by anticipating possible collisions.

Beyond finance and self-driving cars, RNNs are also heavily used in tasks like sentiment analysis, machine translation, text analysis, and generating captions for images. For example, if a company wants to understand how audiences felt after watching a movie based on their reviews, RNNs can analyze the text to detect the emotions behind the reviews. This automation is helpful for companies that don’t have enough time to manually read, categorize, and analyze thousands of reviews. An RNN can perform this task quickly and with great accuracy (Cha & Kang, 2018).

2.3.4. Long Short-Term Memory

Long Short-Term Memory (LSTM) is an advanced version of a recurrent neural network (RNN) designed for tasks that involve sequences and time series. It is particularly effective at capturing long-term patterns in data, which is important for complex problems like language translation and speech recognition (Cha & Kang, 2018). LSTM has been used to improve systems such as Google's speech recognition, Google Translate, and Amazon's Alexa. By 2017, Facebook was using LSTM to process over 4 billion translations per day. Unlike traditional RNNs, which often struggled with performance issues like the exploding or vanishing gradient problem, LSTM overcame these challenges, making it more reliable for training deep neural networks (Manengadan, Nandanan, & Subash, 2021). LSTM networks, equipped with memory cells and a multi-layered structure, were created to help neural networks retain long-term dependencies in data (Sherstinsky, 2020; Li & Zhao, 2019). Its architecture includes a memory cell, an input gate, an output gate, and a forget gate, allowing it to handle tasks like speech recognition, language modeling, and many others. The memory cell in an LSTM holds information over time, while gates manage how data centers and exits the cell. This structure allows LSTMs to keep information for longer periods, making them ideal for tasks that require remembering long-term dependencies. The figure below shows the flow of an LSTM, where Ct−1 represents the old cell state, Ct is the current cell state, ht−1is the output of the previous cell, and ht is the output of the current cell. The input gate is represented by *it*, the forget gate by *ft*, and Ot is the output from the sigmoid gate layer (Butt et al., 2020). LSTMs were developed to address the issue of short memory in traditional RNNs. They use gates to control how information flows through the network. A simple diagram below illustrates how an LSTM functions from the input layer to the output layer and how each state contributes to the process(Butt et al., 2020).

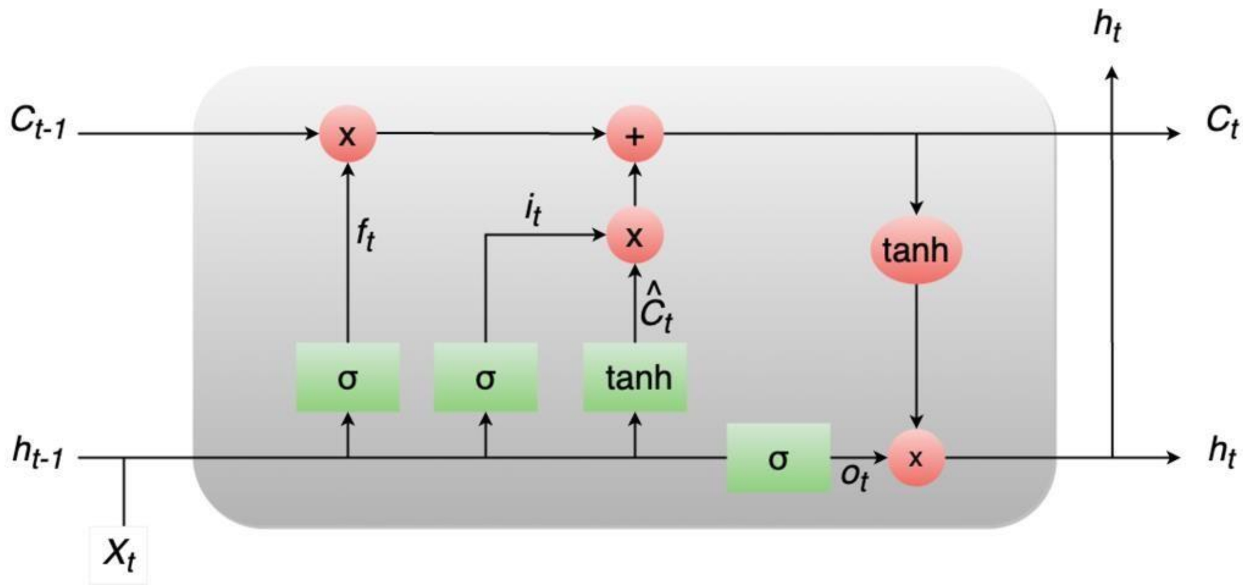

Figure 5: Flow diagram of LSTM cell (Hiransha et al., 2018)

The cell state, represented by Ct−1 (previous cell state) and Ct (current cell state), is responsible for transferring important information from one step of the network to the next. This allows the model to carry forward useful data through time. The forget layer, ft, decides which As information travels over the network, it should be retained or deleted. Its role is essential for determining what data is important enough to store and pass on, helping the LSTM focus on relevant information while ignoring what is not needed.

Traditional methods of crime prediction have been revolutionized by machine learning applications, which use sophisticated algorithms to examine previous crime data. Information technology, statistics, probability, artificial intelligence, psychology, neuroscience, and many more fields all present a sizable area of study called machine learning. (Manengadan et al., 2021). By creating a model that accurately represents a chosen dataset, machine learning can address problems with ease.The applications of ML in crime prediction span various domains within law enforcement. These techniques assist in identifying crime hotspots by geospatial analysis, enabling strategic resource allocation and targeted intervention efforts. Additionally, temporal analysis and trend prediction provide insights into the cyclical nature of criminal behavior, facilitating proactive measures during expected spikes in criminal activities. Furthermore, ML aids in risk assessment and suspect profiling, supporting law enforcement agencies in identifying high-risk individuals or locations prone to criminal involvement (Lipton, Kale, Elkan, & Wetzel, 2015).

2.3.5 Multilayer Perceptron

Artificial intelligence includes artificial neural networks. Multilayer perceptron are one kind of neural network (MLP). Simple, networked neurons or nodes make up an MLP, which performs a nonlinear mapping of input data to output data. These nodes are coupled by weights, and their output signals are adjusted by a nonlinear activation function and rely on the total of their inputs (Mia, Biswas, Urmi, & Siddique, 2015). By mixing numerous basic nonlinear activation functions, an MLP can handle extremely complicated nonlinear functions. The MLP can only represent linear functions if the activation function is linear. Classifying MLPs involves dividing the outputs into distinct groups. To obtain strong performance on new, unseen data is the primary objective of training an MLP (Micheli-Tzanakou, 2011). It has been demonstrated that MLPs work well for applications including classification, function approximation, and prediction. They are especially helpful in scenarios (such as nonlinear systems) when a comprehensive theoretical model is impractical.

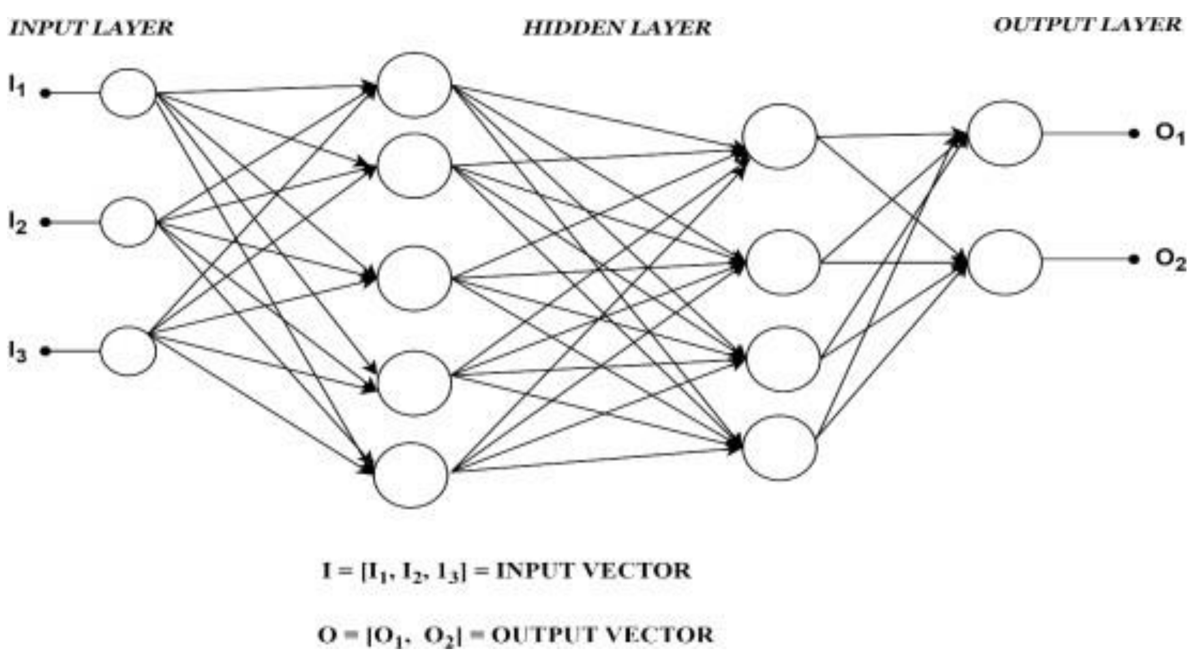

Figure 6: Multilayer Perceptron with Two Hidden Layers

The inputs of each perceptron in an MLP neural network are connected to the inputs of another perceptron from the layer before it, per (Karlik & Olgac, 2011). The weighted total of the inputs and the specified threshold will determine whether the perceptron fires. Three layers should ideally make up an ANN/MLP. There are three layers: the concealed layer, the output layer, and the input layer. Nodes in the input layer stand for the quantity of system variables. Data moves from the output layers to the hidden layers (Micheli-

Tzanakou, 2011). Weight variables connected to each connector regulate the flow. The system's categorization judgment is represented by the output layer nodes.

To predict and classify each input, the output node values are compared to specific thresholds (Karunathilaka, 2020). Artificial neural networks are inspired by how biological neurons in the brain function. In the human brain, a neuron sends a signal (called an action potential) when the total input signals it receives surpass a certain threshold, represented as Ɵ. This threshold isn't fixed for every neuron and varies slightly, making it difficult to predict exactly how a neuron will behave. However, neurons can adjust their firing thresholds over time, allowing the brain to learn and adapt to new information. Similarly, in artificial neural networks, the learning process involves adjusting the thresholds, or weights, between nodes.

In a basic artificial neural network known as a perceptron, this idea is applied, but it can only solve simple, linear problems. However, the Multilayer Perceptron (MLP) model, a more advanced version, can handle complex, non-linear problems (Zhai, Ali, Amira, & Bensaali, 2016). MLPs are feed-forward networks, indicating that data flows in a single direction, from input to output. These networks typically comprise several tiers of nodes, and the connections between them only go forward, without looping back. MLPs are commonly trained using a method called back propagation, which modifies the network's weights to reduce mistakes during training. This design is widely used today for tasks involving both classification (grouping data into categories) and regression (predicting numerical outcomes) (Informatik & Joachims, 1998; Osowski et al., 2004).

2.4 Factors Influencing Crime Prediction

Several factors have a big impact on crime prediction, including socio-economic conditions, demographic details, location, and time. Socio-economic factors, such as income levels, unemployment, and poverty, often lead to higher crime rates (Towers, Chen, Malik, & Ebert, 2018). People who are struggling financially may become depressed and turn to crime. High unemployment can leave individuals with fewer options, leading some to commit theft, robbery, or commit other crimes. Poverty and inequality also increase social tension, making economic conditions an important factor in crime prediction (Shamsuddin et al., 2017).

Demographic factors, like population density, age, and gender, also contribute to crime. High population areas may experience more criminality as a result of the growing population, providing more opportunities for criminal activity (Bothos & Thomopoulos, 2016). Younger demographics, particularly those in the 16–30 age range, are frequently more prone to commit crimes, especially in regions with high rates of unemployment among young people. In terms of numbers, men are more likely than women to commit certain types of crimes, therefore gender disparities are also important (Towers, Chen, Malik, & Ebert, 2016).

Location and time are key elements in understanding crime patterns. Crimes often happen in specific areas, called "hotspots," where criminal activity is repeated. Certain times of the day or night also tend to have more crimes. By analyzing historical crime data, law enforcement can predict when and where crimes are likely to occur(Towers et al., 2016). Community engagement and trust in law enforcement are other important factors. Because neighbours watch out for one another, cooperative communities with close links typically have lower crime rates. Conversely, crime tends to increase in regions where law enforcement is scarce or there is mistrust of the police since criminals believe they have a lower chance of getting caught (Detotto & Otranto, 2010). Advancements in technology, such as data analytics and machine learning, have improved the ability to predict crime. These tools evaluate enormous volumes of data to identify patterns that might not be obvious at first, assisting law enforcement in making more precise predictions. Additionally, changes in the economy, government policies, and legal reforms can also influence crime rates by improving social conditions or law enforcement strategies. Understanding and incorporating these factors into crime prediction models is essential to accurately forecast and prevent crime(Towers et al., 2016).

2.5 Empirical Review

Many literature sources, including journals, articles, and internet resources, will be thoroughly to be able to have a broad understanding of the research issue on crime prediction. Numerous studies have been undertaken aiming to implement systems for crime prediction, as outlined below.

Researcher Tesfaye et al. (2023) concentrate on utilizing machine learning to predict crime. They employed Decision Tree and Naïve Bayes algorithms for their analysis, where the Decision Tree obtained an 85% accuracy rate, reflecting its effectiveness in this context. They used a dataset of 2,100 records. However, this dataset is considered insufficient for identifying comprehensive crime patterns. The limited size and potential lack of coverage may not fully represent the diversity of crime types and locations, which could result in an incomplete knowledge of the trends in crime.

In their study titled "Crime Prediction using Machine Learning," Detotto & Otranto (2010) and Shohan et al. (2022) make significant advances by constructing a comprehensive crime dataset for the study. They address the lack of structured and organized information about crime in the nation. Study introduces a novel dataset that includes 6,574 crime incidents collected over seven years. The information in this dataset was gathered from a number of sources, including the Bangladesh National Census report and daily newspaper archives. The researchers evaluated five supervised machine learning algorithms on this dataset and 79% accuracy was attained by applying the Extra Trees algorithm.

The research by Safat, Asghar, and Gillani (2021) focuses on predicting and forecasting crime trends in urban areas using deep learning and machine learning methods on a dataset of 4,770 records from Los Angeles and Chicago. The study utilizes only time and location attributes to build its models, employing various algorithms such as Logistic Regression, SVM, XGBoost, KNN, Multilayer Perceptron, and Random Forest. Among these, XGBoost achieves the highest accuracy at 88% for crime prediction. The Exploratory data analysis includes detailed visualizations of crime rates, types, and high-crime areas, providing valuable insights for law enforcement organizations to improve the distribution of resources and preventative measures.

The Naïve Bayes technique is employed in the study by Kanimozhi, Keerthana, Pavithra, Ranjitha, and Yuvarani (2021) to examine crime trends with an emphasis on time and location-based crime prediction. They display the most prevalent crime types, along with their locations and timings, using crime data from the online platform Kaggle. Even though Kaggle is a well-known resource, official police records and internet data may not always match. Internet data might not be entirely accurate in reflecting, out datedness, or missing details. This may affect how accurate forecasts are. Although the study's high accuracy rate of 93.07% is impressive, it’s important to keep it in mind when assessing the results, online data may not exactly match official records.

The paper by Saraiva, Matijošaitienė, Mishra, and Amante (2022) explores crime prediction and monitoring using models like Logistic Regression, Decision Tree, Random Forest, and SVM. They used a dataset of around 1,680 records from 2016 to 2018, covering offenses such as auto theft, wallet theft, physical assaults, domestic violence, threats, drunk driving, gun and drug trafficking, and driving without a license. With 83% accuracy, the Random Forest model performed well. The study uses this crime data analysis to forecast current affairs. The researchers contribute by creating a framework that uses historical data to predict future crimes, helping law enforcement agencies better allocate their resources.

Worku Shobe (2021) studies crime prediction using four machine learning classification techniques: Decision trees, Naïve Bayes, Support Vector Machine (SVM), and Linear Regression. The dataset, consisting of 3,200 instances, is sourced from the Addis Ababa Police Commission, with some data collected online. The study pinpoints important variables that impact crime and claims to have reached 94% accuracy using the SVM algorithm. By examining these factors, the research provides valuable insights to support law enforcement efforts and improve resource allocation.

Table 1: Summary Empirical Review

Authors	**Does it include Demographic information?**	**Does it include Socio Economic information?**	**Does it include Time and Location information?**	**Is Data set size > 3500?**	**Is accuracy > 90%?**	**Hyper parameters**
(Tesfaye et al., 2023)	✓	✓	✓	✗	✗	✗
(Shohan, Akash, Ibrahim, & Alam, 2022)	✓	✓	✓	✓	✗	✗
(Safat, Asghar, & Gillani, 2021)	✗	✗	✓	✓	✗	✓
(Kanimozhi, Keerthana, Pavithra, Ranjitha, & Yuvarani, 2021)	✗	✗	✓	✓	✓	✗
(Saraiva, Matijošaitienė, Mishra, & Amante, 2022)	✓	✓	✓	✗	✗	✗
(Worku et al.,2021)	✓	✓	✗	✗	✓	✗
Our Study	✓	✓	✓	✓	✓	✓

As demonstrated in related work, crime prediction efforts often face challenges such as biased and insufficient datasets, relying solely on historical data without considering recent events leads to outdated predictions that do not reflect current trends. Additionally, past research does not include key attributes crucial for accurate crime prediction. Addressing these issues, our work focuses on collecting more diverse datasets that include socio-economic, demographic, time, and location factors. Using real-time data can significantly improve the accuracy and relevance of crime predictions by incorporating the most current information.

CHAPTER THREE

METHODOLOGY

3.1 Overview

This chapter provides a brief history that complements the design, execution, and assessment of the suggested architecture by outlining the techniques, instruments, and processes utilized to analyses the research challenges. It contains techniques for sample selection, data collecting, and analysis in order to meet the main goal of the study. In further detail, the research method, approach, data collection techniques, research finding measures or metrics, testing and assessment process, and algorithms used are covered in the research methodology chapter. We discussed the approaches or models that were applied in the recommended approach to forecast crime in order to look at explanatory research concerns. Next, we evaluated the performance of several machine learning and deep learning models, such as SVM, KNN, RNN, LSTM, and MLP.

3.2 Overall Methodology of the Research

This approach systematically examines how research is conducted and explores crime prediction using machine learning. We reviewed existing methods to gain insights into the literature, identify problems, prepare data, and test our models. The steps we followed to meet our objectives are as follows:

- Identifying the Research Area: We began by reviewing various sources, including research articles, journals, books, and related works.
- Problem Identification: Within the designated research area, we were able to identify the precise issue.
- Data Collection: Relevant data was gathered to address the identified problem.
- Data Preprocessing: The collected data was cleaned and prepared for analysis.
- Feature Selection and Model Design: We determined the necessary parameters to build and design the proposed solution.
- Analysis and Evaluation: The proposed algorithms were analyzed and tested to assess the solution's effectiveness.
- Results Analysis and Reporting: We compared our approach's performance to that of other approaches and presented the results.

Figure 7 shows the overall methodology employed in this thesis. The process starts with evaluating relevant publications to identify gaps, followed by problem formulation, data collection, preprocessing, model development, and testing. Lastly, a discussion of the findings is had, followed by recommendations and conclusions.

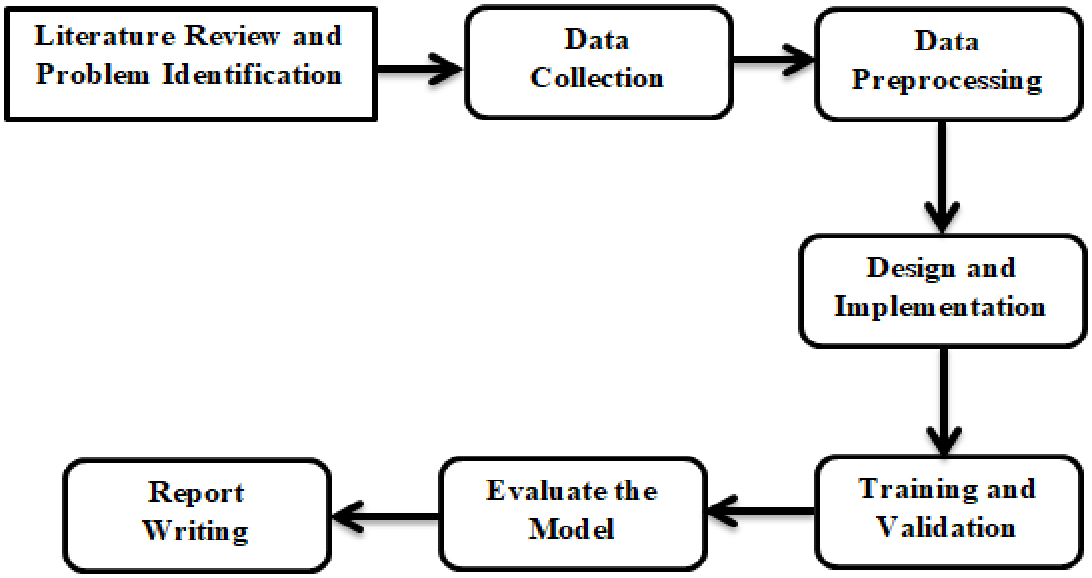

Figure 7: Research methodology workflow diagram

3.3 Data Collection

For this research, data was carefully collected, totaling about 8,000 datasets from the Ethiopia Federal Police Commission. The socioeconomic, demographic, temporal, and spatial information found in the database was chosen because it is crucial for examining and predicting crime trends. The Ethiopia Federal Police Commission granted us official authorization so that we could make sure the data collection procedure complied with moral and legal requirements. Three sub-cities' police stations provided the information. Although crime records are available from earlier periods, we focused on records from September 2011 to October 2016 E.C. due to time constraints. There are 8,000 records with 25 attributes in the data that was gathered, such as Crime Code, Crime Level, Date, Time, Woreda, Location, Crime Type, Victim and Offender details, including Age, Sex, Job, Religion, and Marital Status.

3.4 Data Preparation

The selected algorithms were not instantly capable of processing the gathered data. To guarantee the quality and reliability of the model's predictions, data preparation is an essential stage. Cleaning up the data, dealing with missing values, and converting different data types into ones that might be used for modeling were important challenges (Endalew, 2017). A number of columns had missing values: the criminal's marital status had 12 missing entries, the educational status had 101 missing entries, and 40 records lacked Woreda information. With 540 crime records involving the law, 233 involving organizations, and 7,227 involving individual victims, the victims in the dataset included both individuals and the law. Proper data preparation helps guarantee that machine learning and deep learning models receive clear and accurate training and testing

3.5 Data preprocessing and Feature Selection

An essential step in getting data ready for analysis is data preparation. Before analysis can start, errors, missing values, and inconsistencies are frequently present in real-world datasets. Since the data originates from various sources, accuracy and dependability must be ensured by cleaning and organizing it (Han, 2016). Data preparation is the process of converting, cleaning, and organizing unprocessed data into a format that can be used. Because it directly affects the accuracy of the outcomes, this stage must be done well. Converting data formats, normalizing values, and encoding categorical variables are common data preparation procedures.

For the crime type column in our dataset, we categorized crimes into four major groups: Violent Crime (VC), Property Crime (PC), State Crime (SC), and Victimless Crime (VCC), based on Meti (2016). These categories received the following numerical values from us: 0 for VC, 1 for PC, 2 for SC, and 3 for VCC. Standardization techniques were then used to the data. Another crucial step in the procedure is feature selection. It helps improve the performance and efficiency of predictive models by focusing on the most relevant features while eliminating irrelevant or redundant ones. Not all attributes contribute equally to the prediction process, so selecting the right features enhances model accuracy, reduces computational costs, and speeds up the analysis. From the 25 attributes in our dataset, we selected 17 that were most relevant for analysis, excluding personal details such as names, phone numbers, addresses, and birthplaces of both criminals and victims to ensure privacy. The selected features include DATE, which indicates when the crime occurred and helps analyze patterns over time, and TIME, This provides the time of day in order to detect trends of crime related to time. WOREDA provides geographical context for analyzing crime distribution across different areas, while PLACE/LOCATION pinpoints where crimes occur, highlighting high-crime zones.

CRIME TYPE categorizes different types of crime, providing understanding of the characteristics of criminal activity. VICTIM AGE and VICTIM SEX provide data for analyzing age and gender-related crime patterns, while VICTIM JOB helps understand circumstances and potential risk factors related to the victim's occupation.

Similarly, attributes related to offenders, such as OFFENDER AGE and OFFENDER SEX, offer demographic insights into crime patterns among offenders. VICTIM RELIGION and VICTIM MARITAL STATUS add further demographic context, while OFFENDER EDL (Education Level), OFFENDER MARITAL STATUS, and OFFENDER JOB provide insights into the background and potential motives of offenders. By focusing on these key attributes, the feature selection process ensures that the predictive model uses the most relevant data, significantly enhancing its accuracy and effectiveness.

3.6 Dataset Description

Data is needed for machine learning in order to learn and make judgments. In the event that the gathered data are in a format that is not machine-readable, they must be transformed. The data kinds of attributes in the crime dataset are explained in the table below.

Table 2: Dataset Description

No	Attribute Names	Description of attributes	Data-type
1	Crime Code	A special code for identification that is linked to every offence that is recorded.	Number
2	Crime Level	The severity or classification of the crime.	Text
3	Date	The date when the crime was reported or occurred.	Number
4	Time	The time when the crime was reported or occurred.	Number
5	Woreda	A geographical area or district within a region or city.	Number
6	Place/ Location	The specific location where the crime occurred.	Text
7	Crime Type	The category or nature of the crime.	Text
8	Victim Age	The victim's age at the time of the offence.	Number
9	Victim Sex	The gender of the victim.	Text
10	Victim Job	The occupation or job of the victim.	Text
11	Victim Religion	The religious affiliation of the victim.	Text
12	Victim Marital Status	The marital status of the victim.	Text

13	Offender Age	How old the criminal was.	Number
14	Offender_Sex	The gender of the offender.	Text
15	Offender_EDL	Education Level of the offender.	Text
16	Offender Job	The occupation or job of the offender.	Text
17	Offender Marital Status	The marital status of the offender.	Text

3.7 Data Visualization

This section discusses the analysis and visualization of the crime dataset. After preprocessing the crime records, we have 7800 records with 17 features. Understanding the distribution and connections between these variables is made easier with the aid of visualization.

3.7.1 Distribution of Crimes by Offender gender

The distribution of crimes according to the gender of the offenders is examined in this subsection. Understanding the gender dynamics in criminal activities can provide insights into potential social and psychological factors influencing crime rates.

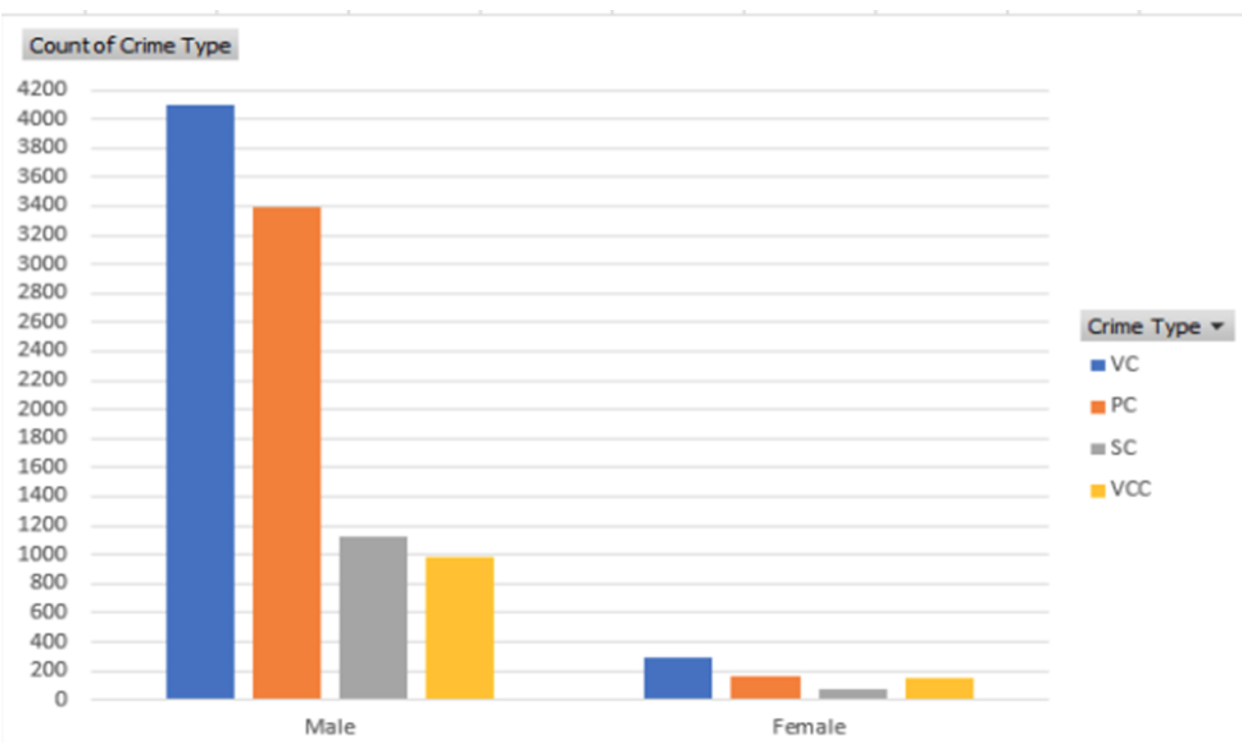

Figure 8: Crime distribution and Offender gender

According to the figure, men commit crimes at a higher rate than women. The trend highlights a significant gender disparity in crime rates, suggesting that males are more prone to engaging in criminal activities. Understanding this difference can help in developing targeted crime prevention strategies.

3.7.2 Distribution of Crimes by Victim's Age

We created a histogram of the ages of the offenders in order to comprehend the age distribution of criminals.

This visualization reveals which age groups are most frequently victimized.

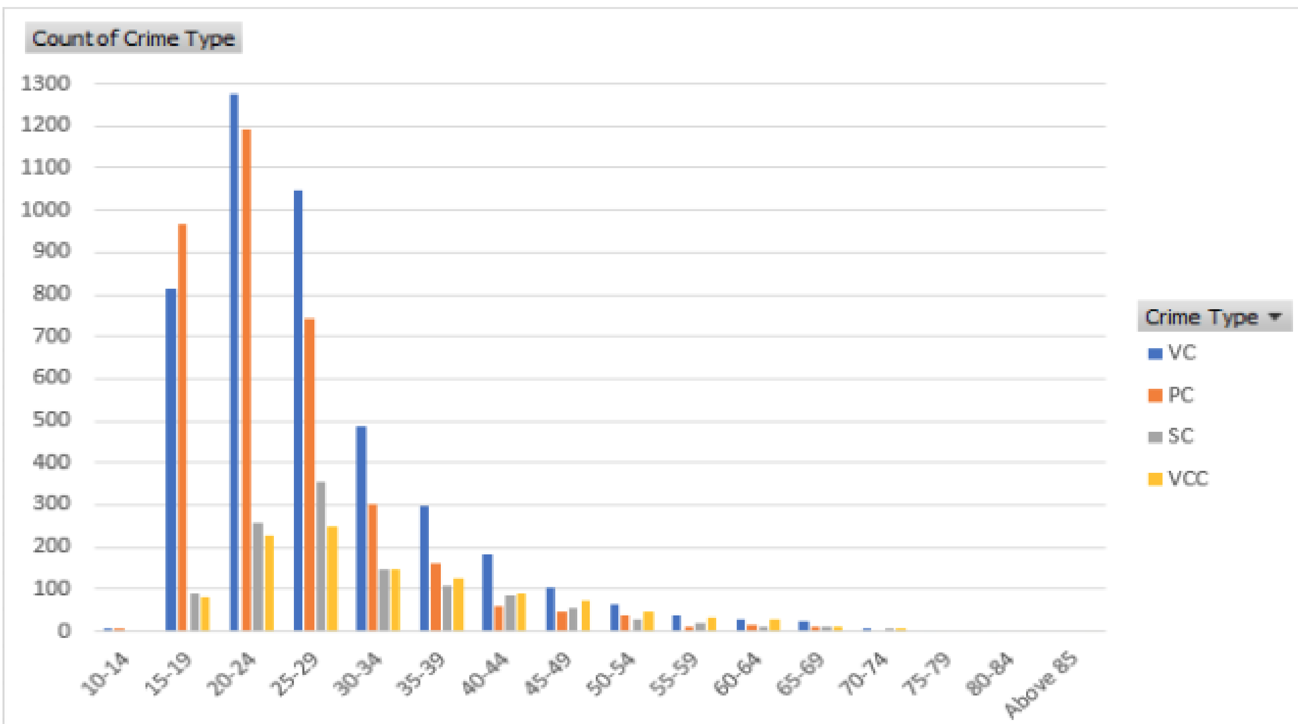

Figure 9: Crime distribution and offenders age

Ages 10 to 14 have lower crime rates, as Figure illustrates. People between the ages of 15 and 29 tend to commit crimes more frequently than people of other ages. But for those above 50, the crime rate drops considerably.

3.7.3 Crime distribution and criminals' marital status

The distribution of crimes according to the perpetrators' marital status is examined in this subsection. An analysis of the relationship between criminal behavior and marital status can shed light on the socioeconomic variables that affect crime.

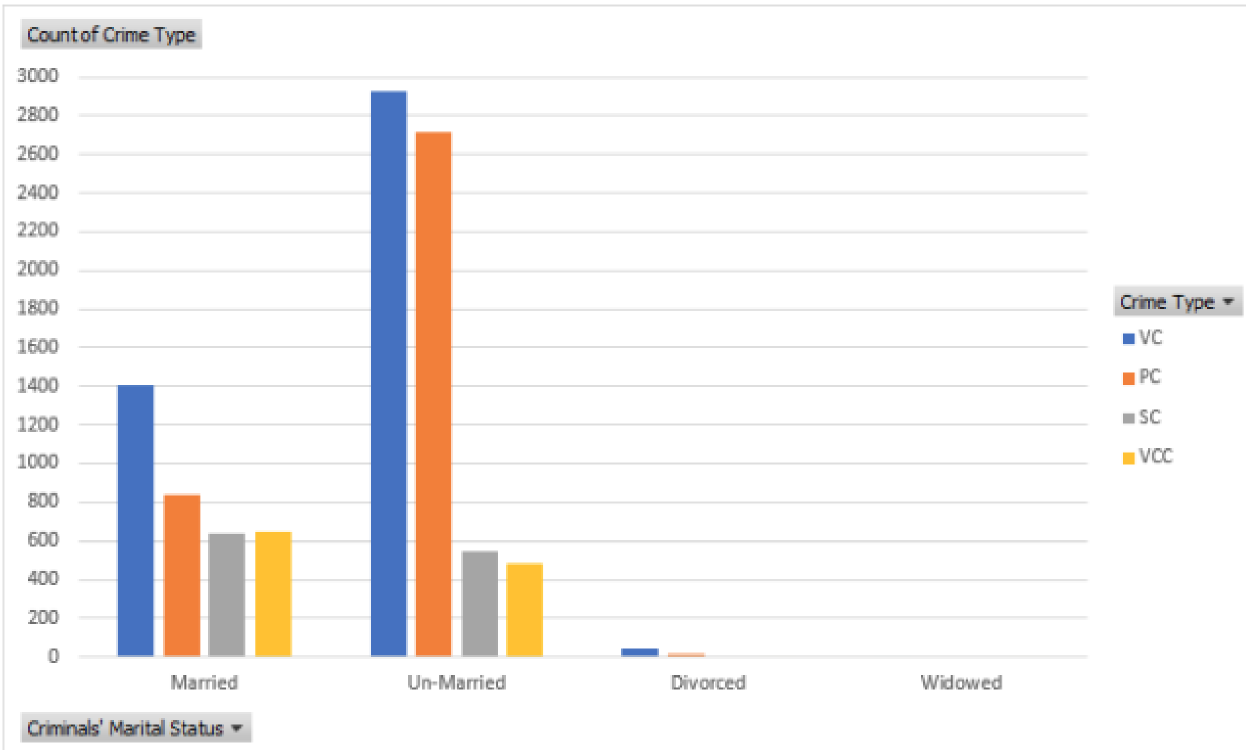

Figure 10: Crime distribution and criminals' marital status

The figure shows that unmarried individuals commit crimes more frequently than married, widowed, and divorced individuals. Married individuals rank second in committing crimes,following unmarried individuals. Divorced and widowed individuals commit crimes relatively rarely. It's crucial to remember, though, that criminal behavior is also influenced by variables other than marital status, such as other demographic characteristics.

3.7.4 Crime Type Distribution

Understanding the distribution of crime types in the dataset is crucial for evaluating the balance and reliability of the model's performance. The dataset contains four types of crimes: Violent Crime (0) with 1,706 instances, Property Crime (1) with 1,816 instances, State Crime (2) with 1,722 instances and Victimless Crime (3) with 1,754 instances. This distribution is fairly balanced, as there is no significant overrepresentation or underrepresentation of any crime type. A balanced dataset like this ensures that the model can learn effectively from each type of crime without being biased towards any particular category.

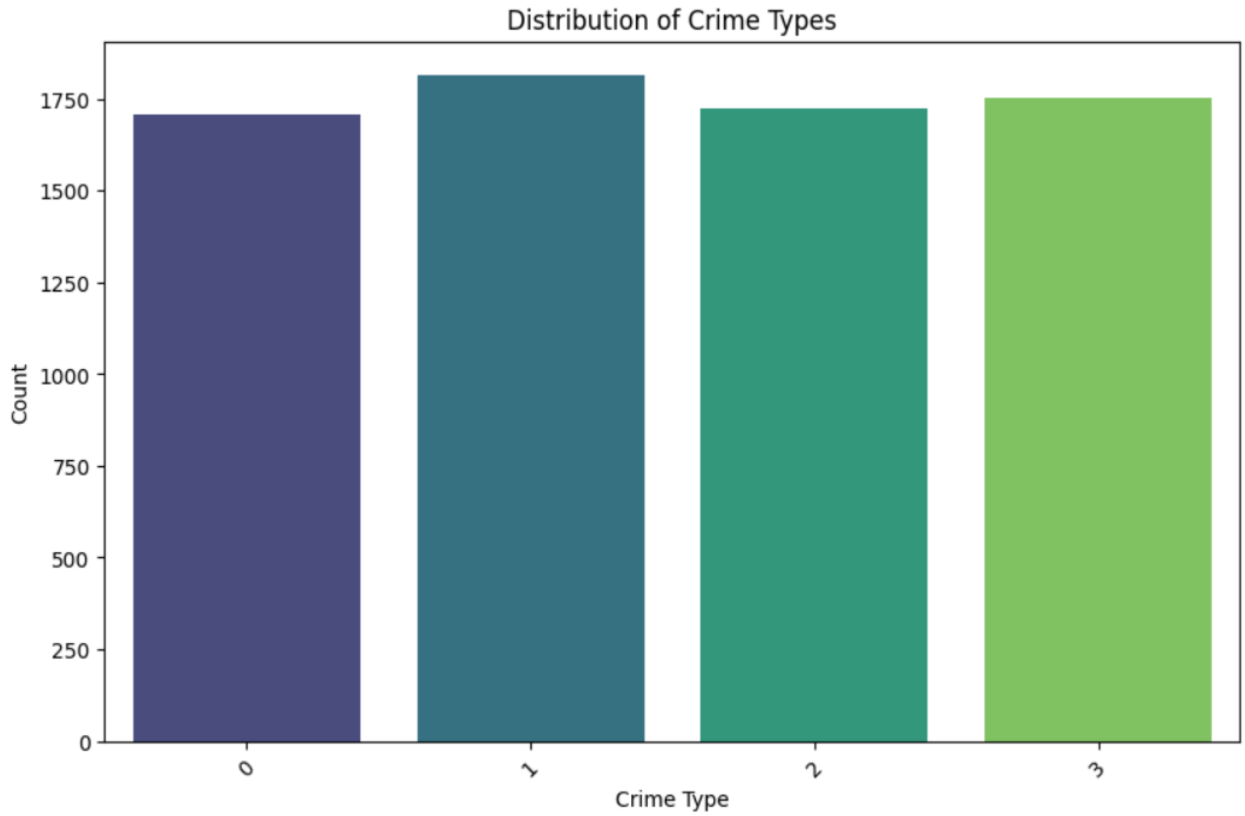

Figure 11: Distribution of crime type

3.8 Modeling

Using algorithms such as Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Multilayer Perceptron (MLP), this section addresses the creation of crime prediction models. The selection of these models was contingent upon the characteristics of the dataset, encompassing both numerical and categorical data.

Before training the models, the dataset was preprocessed. This involved encoding categorical variables with one-hot encoding and normalizing numerical features to ensure the models performed optimally. SVM was chosen because it is effective for multi-class classification tasks, which is useful for predicting different types of crimes. It handles high-dimensional data well, making it suitable for datasets with many features, such as those with various demographic details. KNN was selected for its simplicity and effectiveness in capturing local patterns in crime data, which can help identify hotspots or areas with similar crime patterns. RNN and LSTM were used because they are designed for sequential data. These models were set up with input layers, recurrent layers, dropout layers, and dense output layers.

They were trained with specific epochs and batch sizes, evaluated on various metrics, and compared to find the best model for predicting crime types.

MLPs were used to examine intricate data patterns, such as the connections between socioeconomic variables and the frequency of crimes. This approach helps in understanding how different factors influence crime rates.

3.9 Performance Measures of Algorithms

A number of measures, including accuracy, precision, recall, F1-score, and confusion matrix, were used to assess the effectiveness of the crime prediction models.

Accuracy: calculates the percentage of accurately predicted cases (both true positives and true negatives) out of all predictions produced in order to assess the overall correctness of the model.

$$\text{Accuracy} = \frac{\text{True Positives + True Negative}}{\text{Total Predictions}}$$

Precision: demonstrates the amount of the model's positive predictions that come true. It illustrates how well the model can evade erroneous positives.

$$\text{Precision} = \frac{T\text{rue positives}}{T\text{rue positives + false positives}}$$

F1 Score: calculates the harmonic mean of precision and recall to provide a single metric that balances the two, particularly in cases when the distribution of classes is not uniform.

$$\text{F-Measure} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Recall (Sensitivity): evaluates how well the model can recognize real positive cases. It is the proportion of all actual positives to true positives.

$$\text{Recall} = \frac{T\text{rue Positives}}{T\text{rue Positives +False Negatives}}$$

These metrics were used to evaluate the effectiveness of each algorithm and determine which model more reliably and accurately predicts various types of crimes.

3.10 Hyperparameter Tuning

Hyperparameter tuning is a crucial step in machine learning for improving model performance (Michelucci & Michelucci, 2018). In contrast to model parameters, hyperparameters regulate the way the model learns from data and are established before the learning process starts. Learning rate, batch size, number of layers, and units in neural networks are examples of common hyperparameters (Bardenet, Brendel, Kégl, & Sebag, 2013). Correctly adjusting these hyperparameters can greatly increase the models precision, speed, and generalization. Several important hyperparameters for the selected models SVM, KNN, LSTM, RNN, and MLP were adjusted in this study. Grid search and trial and error methods were used to optimize the hyperparameters to find the values that maximized each model's performance in crime prediction (Feurer & Hutter, 2019). Three crucial hyperparameters for SVM tuning were gamma, C (the regularization parameter), and the kernel type. The highest performing kernel for identifying complicated linkages in crime data was found to be the RBF (Radial Basis Function) kernel. Gamma regulates the impact of individual data points, whereas the C parameter manages the trade-off between preventing overfitting and correctly classifying training data mistakes. By adjusting these parameters, the SVM model's decision limits can be strengthened and its ability to classify different sorts of crimes is improved.

The distance metric, weight function, and K value (number of neighbours) were the three main hyperparameters for KNN that were tweaked. In order to strike a compromise between model complexity (higher K) and noise sensitivity (lower K), the ideal value for K was chosen. The optimal distance metric was determined to be the Euclidean distance, and the weight function was set to distance to give closer neighbours more weight in the classification process. The model's capacity to categorize crime incidents based on similarities to nearby data points is improved by fine-tuning these hyperparameters.

Numerous critical hyperparameters, such as the number of units, batch size, learning rate, and dropout rate, were optimized in LSTM models. The amount of data stored in each memory cell is determined by the number of units; higher numbers enable the learning of more intricate patterns. Dropout was designed to randomly remove neurons during training in order to minimize overfitting. A learning rate of 0.01 was discovered to strike a balance between speed and stability during training, and the batch size was optimized for faster convergence. Because of the Adam optimizer's effectiveness with big datasets, it was utilized. The LSTM model's capacity to manage sequential data and make precise predictions about the types of crimes is enhanced by these hyperparameters used together.

Similar to the LSTM model, the RNN's unit count, dropout rate, batch size, and learning rate were adjusted. The dropout rate helps in preventing overfitting, while the number of units in the hidden layers regulates the network's memory capacity. The choice of learning rate and batch size was made to guarantee rapid convergence without compromising accuracy. The Adam optimizer was selected because to its effectiveness in managing gradients in time-series data. The RNN performs better at forecasting crime patterns over time when certain hyperparameters are tuned.

A variety of hyperparameters, including the number of layers/neurons, batch size, learning rate, and activation functions, have to be adjusted for the MLP model. Several hidden layers with ReLU activation for non-linearity and Softmax activation in the output layer for multi-class classification made up the optimal structure. It was discovered that a modest learning rate and small batch size struck a balance between preventing overfitting and enabling quick learning. Once more, the Adam optimizer was selected due to its deep network adaptability. The MLP model outperformed other models in crime prediction challenges by optimizing these hyperparameters.

3.11 Development Tools

Tools for modeling and taking this study into practice include both software and hardware. Below is a brief description of the study's hardware and software.

3.11.1 Hardware

A typical desktop personal computer was utilized for this study's data preprocessing, model creation, and execution. Among the hardware specs are:

- **Processor**: Intel Core i7 processor for fast computation and model training.
- **Memory (RAM)**: 16GB RAM to handle large datasets and complex models
- **USB flash drive: Used** to provide portable storage of big data

3.12.2 Software

The following software tools were used for data preprocessing, model development, and implementation:

- **Python**: programming language for creating code and putting machine learning models into practice. Python's extensive library ecosystem includes NumPy, pandas, scikit-learn, and TensorFlow provided essential tools for data manipulation, model development, and deep learning.

- **Jupyter Notebook**: Interactive development environment used for writing and executing Python code. Jupyter Notebook allowed for easy experimentation and visualization of data and models.
- **Anaconda**: Python distribution that includes various libraries and tools for data science and machine learning. Anaconda simplified package management and environment setup for the project.
- **TensorFlow and Keras**: Deep learning frameworks used for developing and training RNN and LSTM models. TensorFlow provided the core functionalities for building deep learning models, while Keras offered a high-level interface for easier model development.
- **Scikit-learn**: Python machine learning package used to implement KNN and SVM models. Scikit-learn offer a comprehensive range of tools for data preprocessing, model evaluation, and model selection.
- **Matplotlib and Seaborn**: Libraries for Python that are used to visualise data. Plots and charts were made using Matplotlib and Seaborn in order to analyse the dataset and assess the efficacy of the model.

CHAPTER FOUR

PROPOSED SOLUTION

4.1 Chapter Overview

We offer our address to the noted issue in this chapter. This study's primary objective is to predict different sorts of crimes utilizing deep learning and machine learning models, including KNN, SVM, LSTM, MLP, and RNN. We use the accuracy, precision, recall, F1-score, and confusion matrix to evaluate how well these models perform. As with any research project, the first step is to provide a remedy for the identified problem. Our suggested remedy is predicated on an enhanced algorithm created to effectively forecast the types of crimes.

4.2 Proposed Model Architecture

This work aims to develop an automated crime type prediction model by combining deep learning and machine learning methods. The models that were selected are RNN, MLP, LSTM, KNN, SVM, and RNN. Based on the dataset, each model offers distinct advantages that help produce precise crime type predictions.

The suggested architecture offers a broad overview of the solution design and describes the implementation procedures. You can find thorough explanations of every algorithm in the section on research methodology. As data is an important component of machine learning, preparing the dataset by encoding categorical variables and normalizing numerical characteristics is the first stage. After performing further preprocessing operations like dimension reduction and noise removal, the data is divided into training and testing sets in order to assess model performance. After preprocessing, the next step is to build and train the predictive models. Each model (KNN, SVM, RNN, MLP, and LSTM) is trained on the training set using selected features, with techniques like cross-validation employed to tune hyperparameters and prevent overfitting. Model performance is assessed using metrics such as accuracy, precision, recall, F1-score, and confusion matrix to gauge how well each model predicts crime types. KNN is selected for its simplicity and effectiveness in identifying local patterns in crime data, helping to pinpoint hotspots or areas with similar crime patterns. SVM are chosen for their ability to handle multi-class classification tasks and high-dimensional data, making them suitable for datasets with various demographic details. RNN and LSTM are utilized for their capability to analyze sequential data. RNNs are effective in recognizing patterns in sequences, while LSTMs handle long-term dependencies, such as trends over time. MLP is included to capture complex patterns and relationships in the data, such as the link between socio-economic factors and crime occurrences.

Lastly, the best performance for the dataset is identified by comparing the outcomes of each model. This requires considering the advantages and disadvantages of various performance measures and choosing the model that provides the optimum balance between interpretability and accuracy.

The project attempts to develop a strong system for predicting different sorts of crimes by utilizing KNN, SVM, RNN, MLP, and LSTM. Each model offers a distinct benefit to improve the overall predictive capabilities and handle the complexity of crime data. The proposed architecture of the crime type prediction model using these machine learning and deep learning techniques are shown below.

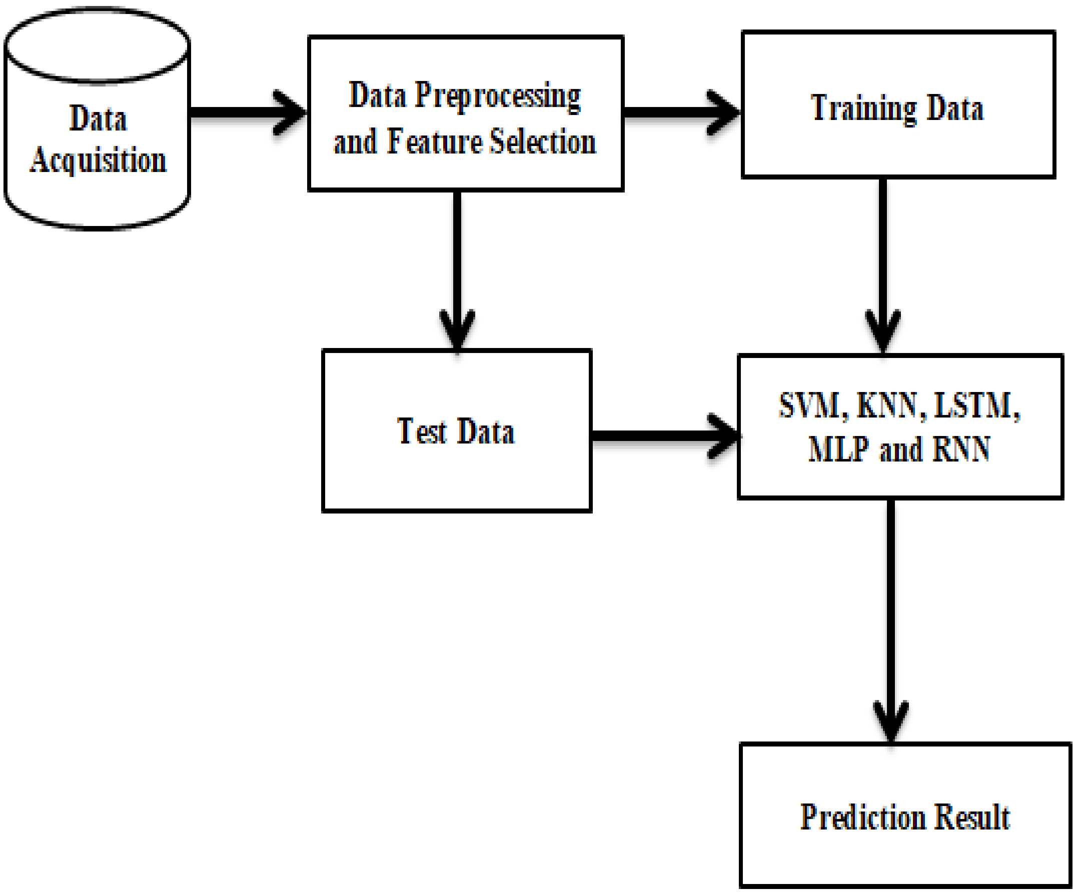

Figure 12: Proposed Solution

4.3 Proposed Data Preprocessing

The performance of the crime type prediction model may be impacted by the extraneous noise that is frequently present in real data that is gathered from diverse sources. To convert unstructured data into a format that is clear and machine-readable, effective data preparation is essential. The specific procedures and techniques implemented in the preprocessing pipeline are described in this section. Taking care of missing values, a frequent problem in datasets, is part of data cleansing. Missing values were replaced using methods such as mode imputation for categorical characteristics and mean or median imputation for numerical features. Data transformation includes normalization and encoding of categorical variables. Normalizing numerical features ensures a consistent scale, which simplifies training and improves performance. Normalization techniques such Standardization were used. Categorical features were converted into a machine-readable format using encoding techniques like one-hot encoding, which creates binary columns for each category, and label encoding; this gives each group its own unique number.

To enhance model performance, feature engineering entails identifying and extracting the most useful features. A comma-separated value (CSV) file was created and saved with the processed data organized. Python and other data analysis tools can quickly parse CSV files, which make them perfect for use in model construction.

Dividing the dataset into training and testing sets is the last stage in the preprocessing pipeline, which assesses the model's performance on untested data. Preparing raw data for machine learning algorithms making sure it's clean, standardized, and in a format that works for training models is the major objective of data preprocessing. The crime type prediction model performs better and is more accurate when effective preprocessing is used. The figure below depicts the architecture of the pipeline for preparing data.

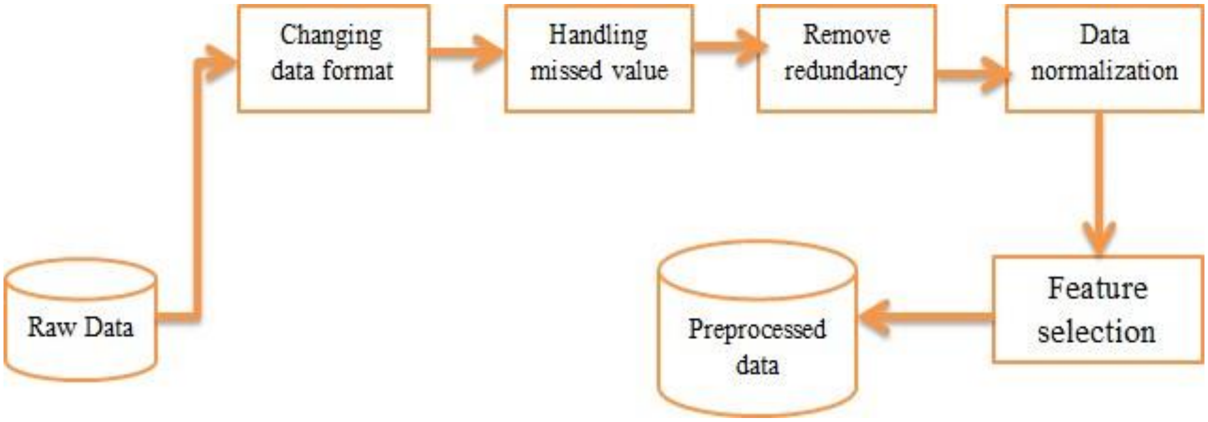

Figure 13: Data preprocessing steps

4.4 Proposed Solution

The developed model, which combines a number of machine learning and deep learning models, such as KNN, SVM, RNN, MLP, and LSTM architectures, is described in this part. These models were chosen because they improve crime type prediction accuracy by capturing intricate patterns in crime data. Before training the suggested models, we specify the required parameters, including batch size, number of epochs, activation function, optimizer, loss function, and metrics. These parameters can have a big impact on the models' performance and are essential for the training process. The number of epochs dictates how many times the model iterates through the whole training dataset, whereas batch size specifies how much data the model examines before modifying the weights. Activation functions give the model nonlinearity and facilitate capturing. Based on the estimated gradients, the optimizer chooses the algorithm that will be used to update the model's weights. The difference between the expected and actual output is computed by the loss function, which is used to determine how effectively the model works. Metrics are measurable indicators of the performance of the model that can be used to compare models or adjust their hyperparameters.

4.5 Implementation of the Proposed Solution

We spoke about the suggested model's application for crime prediction in this section. We also spoke about setting up the Python working environment and implementing Jupyter Notebooks.

Environment Setup Python 3.9.0: was selected as the language to be used while implementing the suggested technique.

Jupyter Notebook: is the most popular and recommended integrated development environment (IDE) for academics to update Python code, visualize specific diagrams, train, and test charts, and track process progress. The suggested procedure is carried out using a Jupyter Notebook with a 6.5.2 version installed via Python installation. Python libraries such as:

Pandas: Pandas is an effective library for working with and analyzing data.

NumPy: A core Python package for numerical computation is called NumPy.

Matplotlib: Using the Python charting package Matplotlib, you can create interactive, animated, and static representations.

SciPy: A NumPy-based package called SciPy offers more technical and scientific computing features.

Scikit-learn: offers a wide range of tools for data mining, classification, regression, clustering, dimensionality reduction, and model selection. It is a machine-learning library. It provides a standard interface for putting different machine learning and deep learning algorithms into practice and analyzing them.

TensorFlow: TensorFlow is a prominent open-source deep learning package created by Google. It offers a complete ecosystem for developing and training machine learning models, particularly deep neural networks.

Keras: High-level neural network API Keras can be used with Microsoft Cognitive Toolkit (CNTK), TensorFlow, or Theano. It provides a user-friendly and straightforward interface for constructing and training deep learning models.

L2 Regularization (Ridge): L2 regularization modifies the loss function by adding a penalty proportional to the squared size of the model's weights.

4.6 Implementation of proposed models

In this part, we clarify how the proposed machine learning and deep learning models KNN, SVM, RNN, MLP, and LSTM were implemented to predict crime types. Before implementing each model, the first step is to set up the necessary environment. We start by importing the required Python libraries, which are essential for training the models. These libraries include popular ones like pandas, numpy, scikit-learn, tensorflow, and keras. After importing the libraries, we load the crime dataset, stored as a CSV file, from the local disk into the Jupyter Notebook. This dataset will be used to train and test the models. Once the data is loaded and preprocessed, we proceed with the implementation and training of each model, evaluating their performance in predicting crime types.

4.6.1 RNN Model Implementation

The Recurrent Neural Network is designed to handle sequential data by recognizing patterns and dependencies that unfold over time. Unlike other models, RNNs are unique because they pass information from one step to the next, allowing them to remember and use previous inputs as the sequence progresses. This helps the model better understand the sequence of data. The RNN architecture includes one or more hidden layers with recurrent units, such as Simple RNN units, that store information from earlier steps in the sequence. For training the RNN, we used an optimization algorithm called Adam, which helps the model learn more efficiently. We experimented with two different batch sizes, 16 and 32, to see which would work best. For classification, we used the SoftMax activation function, which helps the model assign probabilities to different classes. Since we are dealing with a multi-class classification problem, we applied the categorical cross-entropy loss function to measure how well the model is performing.

We trained the model for a range of 10 to 100 epochs to see how the performance improved over time. After testing different epochs, we found that training for 20 epochs gave stable results, so we settled on that number. We also used a learning rate of 0.01, which controls how much the model adjusts its weights with each update.

The input data was processed step-by-step, allowing the RNN to capture temporal patterns in the sequence. We split the dataset into training and testing sets with two different ratios: 80% for training and 20% for testing, and 70% for training and 30% for testing. For the final implementation, the model chosen from these divides with the highest accuracy was used.

4.6.2 LSTM Model Implementation

The Long Short-Term Memory network is a type of Recurrent Neural Network (RNN) that is designed to work well with sequential data, like time-series information. Unlike traditional RNNs, LSTMs are better at capturing long-term dependencies in data because they have a special architecture. This architecture includes memory cells that help the model retain important information for longer periods, which is useful for avoiding the vanishing gradient problem that RNNs often face.

For our implementation, we trained the LSTM model using the Adam optimizer, which is commonly used for training deep learning models. We experimented with batch sizes of 16 and 32, which refer to the number of samples processed before updating the model's weights. To handle the classification of multiple crime types, we used the SoftMax activation function, which is suitable for multi-class problems, and the categorical cross-entropy loss function, which measures how well the model's predictions match the actual data.

The model was trained over a range of 10 to 100 epochs, with 20 epochs proving to be enough since the model's performance became stable after that. We used a learning rate of 0.01, which controls how much the model's weights are adjusted during training, and this helped the model converge effectively. The LSTM model processes the data step-by-step, learning both short-term and long-term patterns from the sequences.

We split the dataset into training and testing sets, using both 80%/20% and 70%/30% ratios. The model that achieved the highest accuracy from these splits was selected for the final implementation. This approach ensured that we found the best version of the model for predicting crime types accurately.

4.6.3 MLP Model Implementation

One kind of neural network consisting of multiple layers of neurons is called a multilayer perceptron. The input layer, hidden layers, and an output layer are some of these layers. The MLP is particularly useful for jobs like classification since it is made to learn intricate patterns by comprehending nonlinear correlations between the input data and the output.

The Adam optimizer, which helps modify the model's weights during training to enhance performance, was applied to our MLP model. To determine which batch size worked best for our model, we experimented with 16, 32, and other values. To assist the model learn more intricate patterns, we introduced non-linearity in the hidden layers using the ReLU activation function. We utilized the SoftMax function in the output layer since it works well for multi-class classification jobs. We measured the model's performance during training using the categorical cross-entropy loss function because our challenge entailed predicting numerous classes.

After testing, we decided to use 20 epochs instead of the full 100 that we had trained the model for, as it did not yield significant improvements in outcomes. By setting the learning rate at 0.01 the model was able to learn gradually without experiencing significant leaps. Additionally, we tested with various dataset splits, utilizing, in one case, 70% of the data for training and 30% for testing, and 80% of the data for training and 20% for testing in another. For the final implementation, we ultimately chose the model from these divides with the highest accuracy.

4.6.4 SVM Model Implementation

Encouragement Strong machine learning algorithms like vector machines are frequently applied to prediction problems. The primary goal of support vector machines (SVM) is to identify the optimal border, or hyper plane, between the various data classes. The preparation of the dataset is the first step in the implementation of an SVM model. Typically, this entails dividing the data into two portions: one for model training and the other for performance evaluation. Although other splits like 70% training and 30% testing are also common, we typically use 80% of the data for training and 20% for testing. Defining the SVM model comes next after the data has been separated.

Choosing a kernel function, which converts the data into a higher-dimensional space and facilitates class separation, is a crucial stage in this process. Radial basis function (RBF), polynomial, and linear are common choices for kernels. To further balance the trade-off between having a low training error and a low testing error, we also modify additional parameters, such as the regularization value.

The training data is used to instruct the SVM algorithm once the model has been defined. The algorithm searches for the hyper plane that maximizes the margin, or the distance, between the classes in order to determine which one best divides them. Lastly, we test the model on the testing data to determine its performance in identifying newly added data points after training. The SVM model's performance is then contrasted between various training/testing splits in order to determine which version produces the best accuracy.

4.6.5 KNN Model Implementation

K-Nearest Neighbours is a machine learning technique that is easy to understand and may be applied to tasks involving regression and classification. The fundamental notion of KNN is that it uses the classification of its neighboring data points to determine how to classify a given data point. The first step in putting the KNN model into practice is to separate the data into training and testing sets. Typically, 80% of the data is used for training and 20% is used for testing; however, a 70-30 split may also be employed.

After the data is ready, we choose the number of neighbours (k) to be taken into account while defining the KNN model. Setting k to 3 or 5 is a common beginning point, and this value can be changed based on the model's performance. We use the training data to train the model after determining k. With KNN, there are no complicated computations involved in training all that is required is storing the training data.

When the model predicts, it measures the separation between each new data point and every other point in the training set, often using the Euclidean distance. After choosing the k closest neighbors, a majority vote among those neighbors determines the class of the new data point. Lastly, we use the testing data to make predictions to test the model, and we assess the model's accuracy to determine how well it performed. Additionally, we may experiment with various choices of k to determine which model yields the most accuracy.

CHAPTER FIVE

RESULTS AND DISCUSION

The results of the research experiments are presented in this chapter. Crime categories are predicted using two machine learning and three deep learning algorithms: SVM, KNN, MLP, LSTM, and RNN. The efficacy of any algorithm in predicting crimes is ascertained by analyzing and contrasting its output. Furthermore, the findings' ramifications and importance in the field of crime prediction are examined.

5.1 Model Performance Evaluation

The effectiveness of the machine learning models used for predicting crime is assessed in this section. We evaluate the models with different hyperparameters using a variety of measures, such as accuracy, precision, recall, F1 score, and confusion matrix. These measurements give an accurate view of each model's performance. The presented findings represent the mean of ten-fold cross-validation, guaranteeing an accurate evaluation of performance. The SVM model obtains an 89% accuracy rate, 90% precision, 89.08% recall, and 89.07% F1 score. This suggests that while the SVM model is highly accurate and exact, it is not as good at finding every positive case. The KNN model shows an accuracy of 91.89%, precision of 92.8%, recall of 92.02%, and an F1 score of 92.03%. It performs similarly to the SVM model, yielding an excellent F1 score due to its high precision and slightly poorer recall. With an accuracy of 96.71%, precision of 96.1%, recall of 96.07%, and an F1 score of 96.74%, the RNN model operates quite well. These outcomes show how successful the RNN model is, as evidenced by its balanced and excellent ratings on all criteria. With an accuracy of 97.92%, precision of 97.93%, recall of 97.92%, and an F1 score of 97.93%, the LSTM model performs even better. The LSTM model performs marginally better than the RNN model, especially in precision and recall, demonstrating its prowess with sequential data. The MLP model reaches the highest performance, with an accuracy of 98.64%, precision of 98.68%, recall of 98.64%, and an F1 score of 98.64%. The MLP model excels in all metrics, making it the most accurate and reliable model for crime prediction. In conclusion, all models work well, but the MLP model outperforms the others all around, having the highest accuracy and the most evenly distributed precision and recall. The RNN and LSTM models are excellent substitutes and both exhibit remarkable performance. Even though they are a little less efficient, the SVM and KNN models nevertheless offer insightful data. These results show that deep learning models MLP and LSTM in particular are useful for predicting crimes.

Additionally, optimum hyperparameters for each algorithm were used, and cross-validation was performed before the final testing. An overview of these algorithms' performance based on F1 score, accuracy, precision, recall, and other metrics is given in the Table below.

Table 3: Performance of the Algorithms

Algorithm	Accuracy	Precision	Recall	F1 Score
SVM	89%	90%	89.08%	89.07%
KNN	91.89%	92.8%	92.02%	92.03%
RNN	96.71%	96.1%	96.07%	96.74%
LSTM	97.92%	97.93%	97.92%	97.93%
MLP	98.64%	98.68%	98.64%	98.64%

5.1.1 Training and Validation Loss

Training and validation loss are critical indicators of a model's learning process and ability to generalize to new data. They provide insight into how well the model matches the training set and performs on unknown data, respectively.

For the deep learning models, including RNN, LSTM, and MLP, the training and validation loss were tracked during the training process. The following sections detail the observed loss trends for each model.

5.1.2 Training and Validation Loss and Accuracy of RNN Model

Recurrent Neural Network (RNN) model performance was assessed during training by closely monitoring the training and validation loss and accuracy metrics. The training began with an initial accuracy of 71.15% and a loss of 0.9163. In contrast, the validation accuracy started much higher at 96.52%, with a validation loss of 0.5803.

The training accuracy increased throughout the course of the 20 epochs, hitting 96.51% by the last epoch, while the training loss continuously dropped to 0.1811. The computer appears to have successfully learnt the patterns in the training data based on this decreasing trend in loss. The training process demonstrated that the model was successfully optimizing its parameters with each iteration.

Throughout the training process, validation accuracy stayed high and steady, holding a constant value of roughly 96.52% from the second epoch on. Similarly, there was a decline in the validation loss with time, with values falling from 0.5803 in the first epoch to 0.1839 in the last. The validation accuracy remained largely constant, and the validation loss decreased gradually, indicating that the model performed well in generalizing to new data and did not exhibit substantial overfitting.

All things considered, these patterns show that the RNN model successfully strikes a balance between retaining its predictive ability on the validation set and learning from the training data. The model appears to be well-suited for the crime prediction job, as evidenced by its consistent performance across training and validation criteria, which is also indicative of robust and reliable performance. This graph displays the RNN model's training and validation losses.

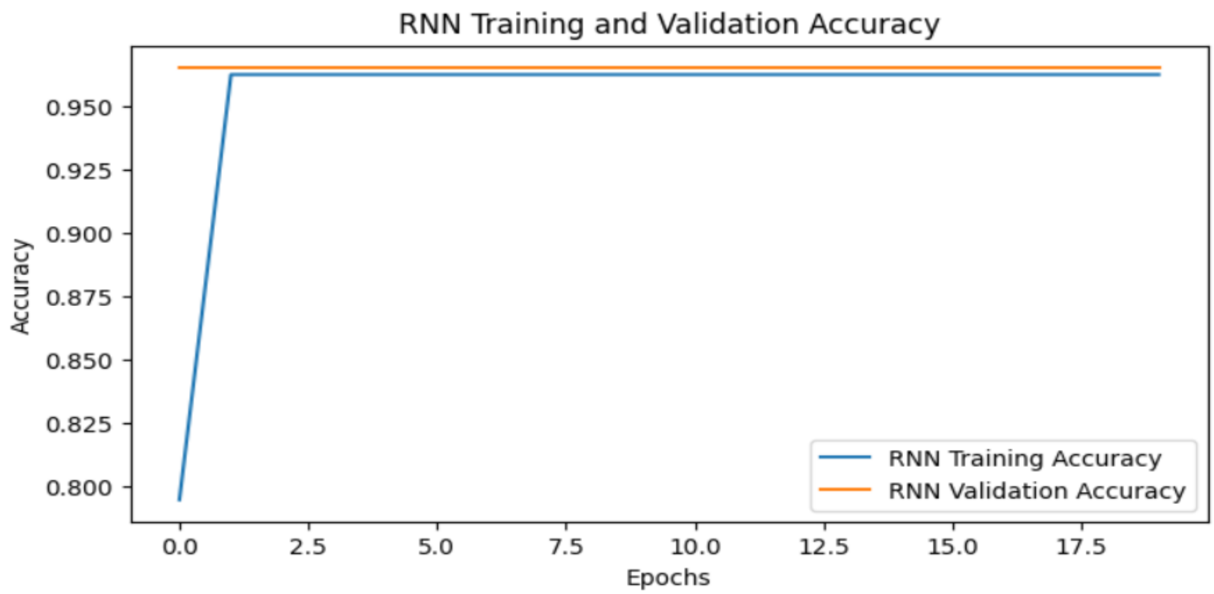

Figure 14 : Training and validation Accuracy graph of RNN model

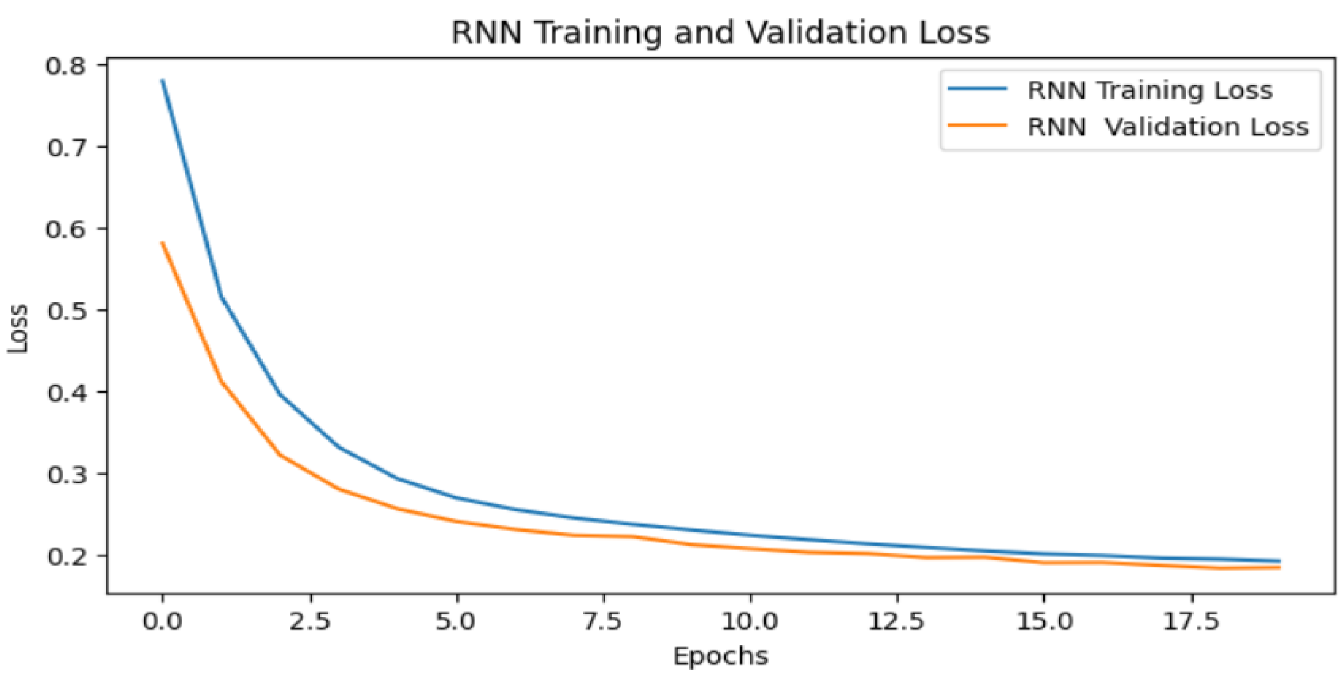

Figure 15: Training and validation loss graph of RNN model

5.1.4 Training and Validation Loss and Accuracy of LSTM Model

The training process of the LSTM model was conducted over 20 epochs. The model achieved 42.47% training accuracy and 1.2889 training loss during the first epoch, and 50.89% validation accuracy and 0.8450 validation loss during the second epoch. As the training progressed, both the training and validation metrics improved significantly. By the 10th epoch, the training accuracy reached 97.79%, and the training loss decreased to 0.1479. The validation accuracy remained stable at 97.68%, with a validation loss of 0.1480, indicating that the model was learning effectively without overfitting. By the final epoch (Epoch 20), the model achieved a training accuracy of 97.95% and a reduced training loss of 0.1169. The model's strong performance on untested data was demonstrated by the validation accuracy, which continuously remained high at 97.68% and the validation loss, which was minimized to 0.1289. The LSTM model appears to have successfully learned the patterns in the data, as evidenced by the steady improvement in both training and validation metrics. This, together with its high accuracy and low loss values, make the model well-suited for the given task. This graph displays the LSTM model's training and validation losses.

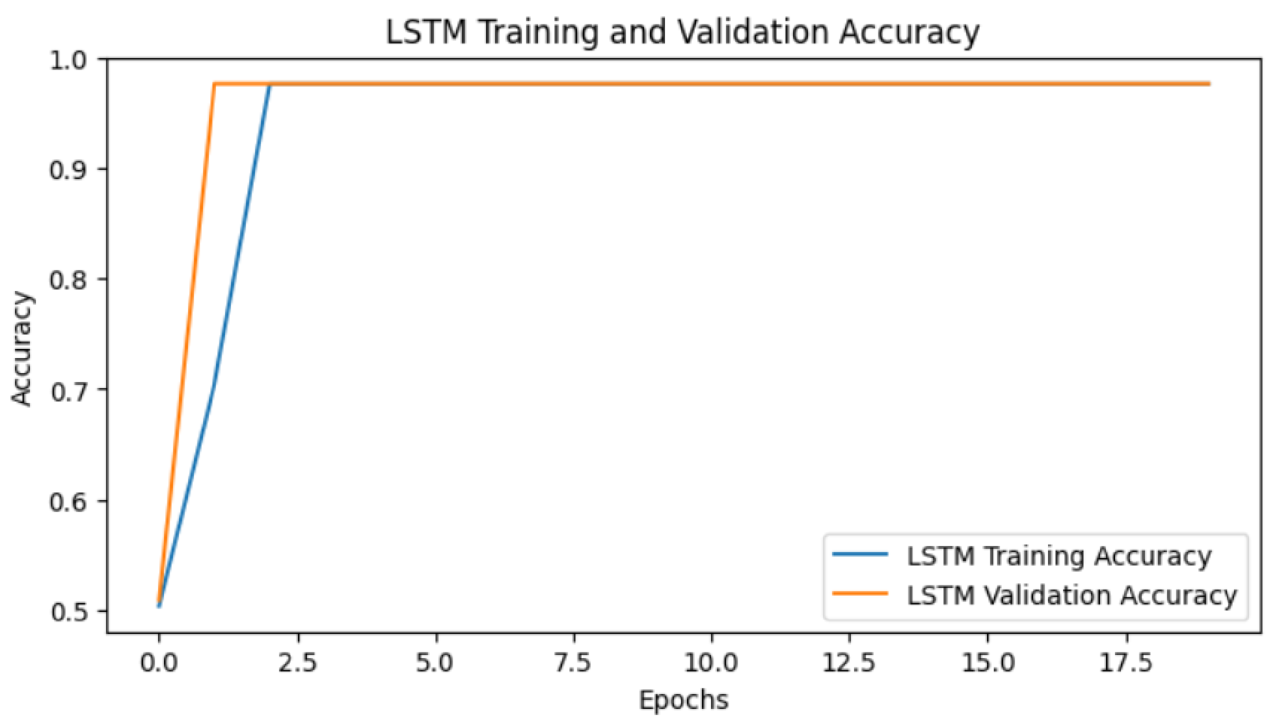

Figure 16: Training and validation Accuracy graph of LSTM model

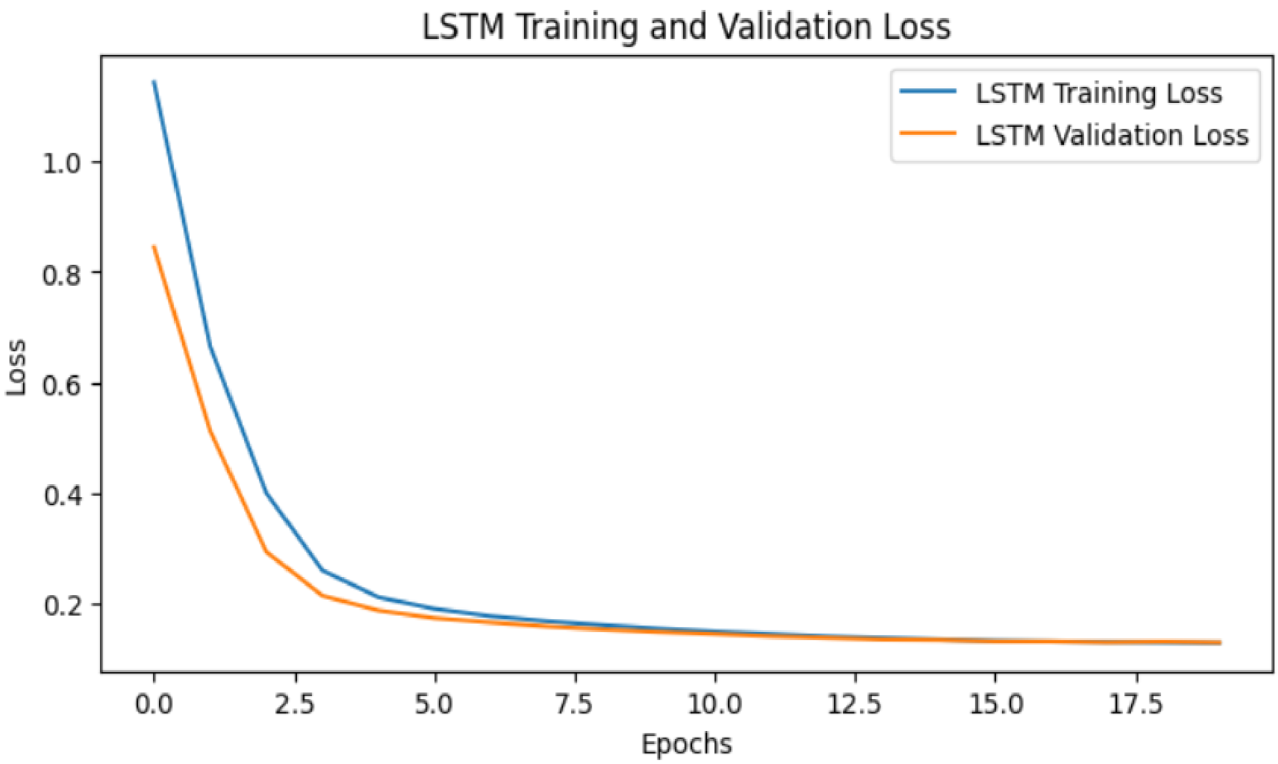

Figure 17: Training and validation loss graph of LSTM model

5.1.5 Training and Validation loss and accuracy of MLP Model

Starting with an initial training accuracy of 88.16% and a training loss of 0.8376, the MLP model was trained over a period of 20 epochs. The model had a good start, as evidenced by the first epoch's validation accuracy of 96.16% and validation loss of 0.2483. There was a noticeable improvement in both accuracy and loss values during the training phase. The training accuracy increased to 98.26% and the loss decreased to 0.0902 by the fifth period. Simultaneously, the validation loss dropped to 0.0755 and the validation accuracy was high at 98.57%. During the training process, the model maintained a high training accuracy of 98.32% and a training loss of 0.0838 until the final epoch (Epoch 20). With a validation loss of 0.0719 and a validation accuracy that held consistent at 98.57%, the model clearly learnt without overfitting and successfully generalized to the validation data. Based on the training and validation sets, the MLP model showed good accuracy and low loss values, indicating consistent and reliable performance. This graph displays the LSTM model's training and validation losses.

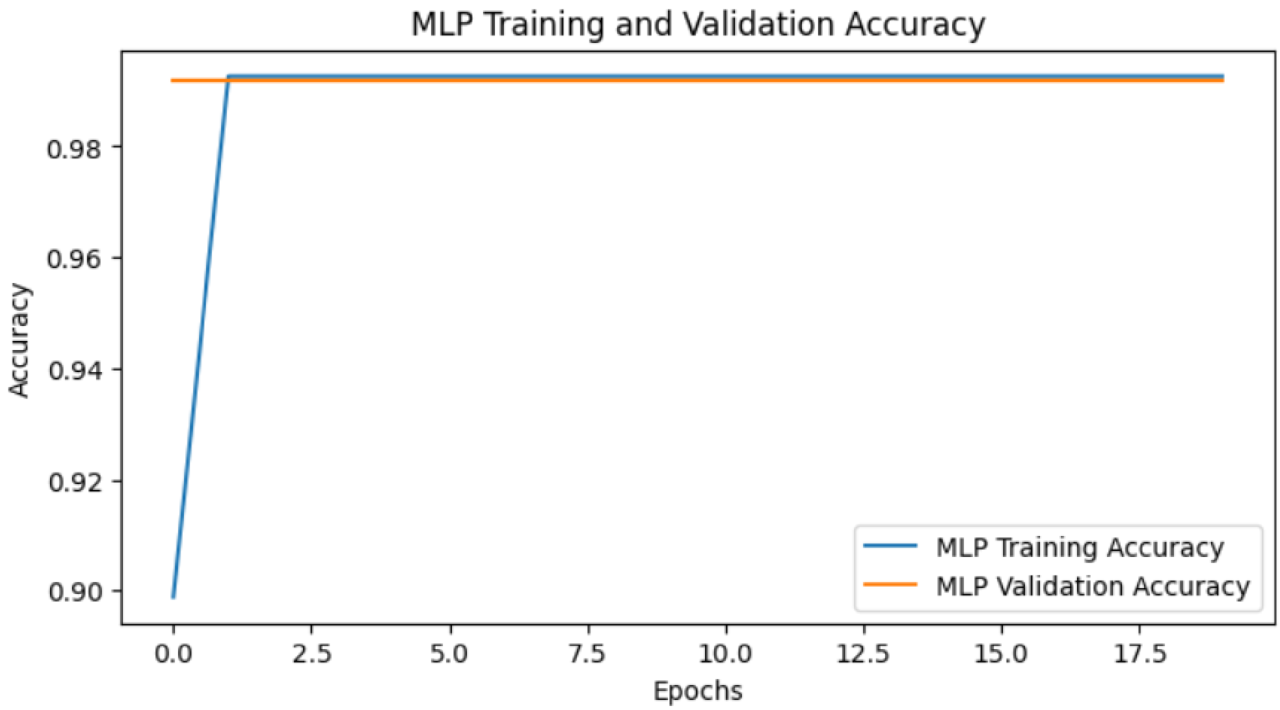

Figure 18: Training and validation Accuracy graph of MLP model

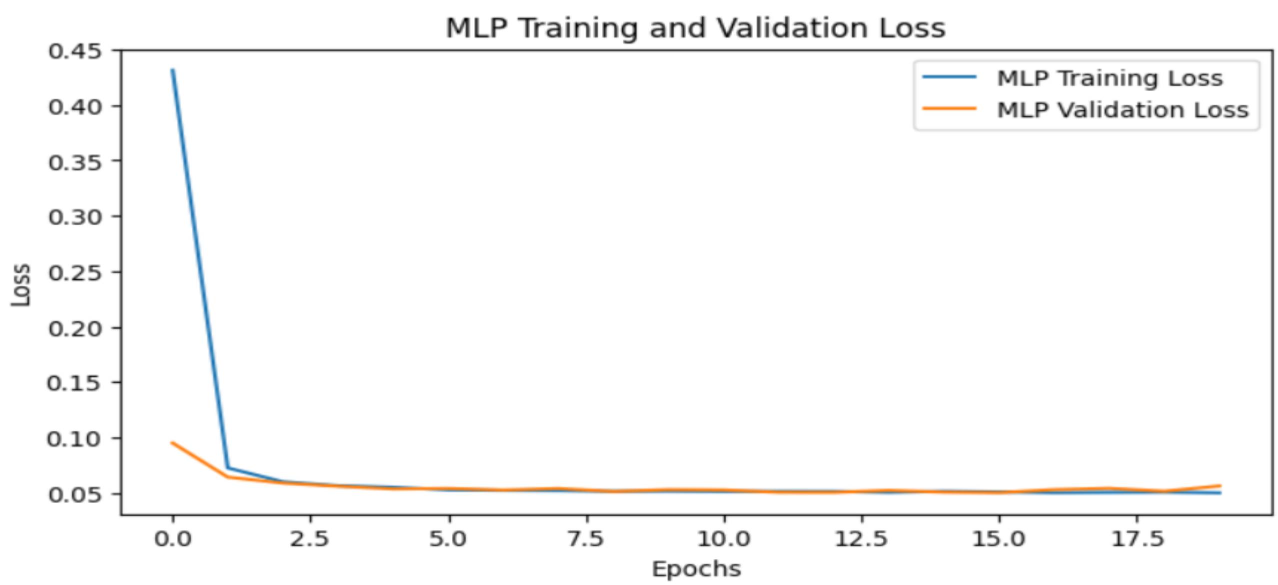

Figure 19: Training and validation loss graph of MLP model

5.2 Confusion Matrix

We employed a confusion matrix in addition to the performance metrics previously mentioned to assess the models. The confusion matrix looks at the following to determine how well our models work: False Positives (FP) are negative instances that are mistakenly predicted as positive, and False Negatives (FN) are positive instances that are mistakenly predicted as negative. True Positives (TP) are correctly predicted positive instances, and True Negatives (TN) are correctly predicted negative instances. The confusion matrix for RNN models is shown in the following figure.

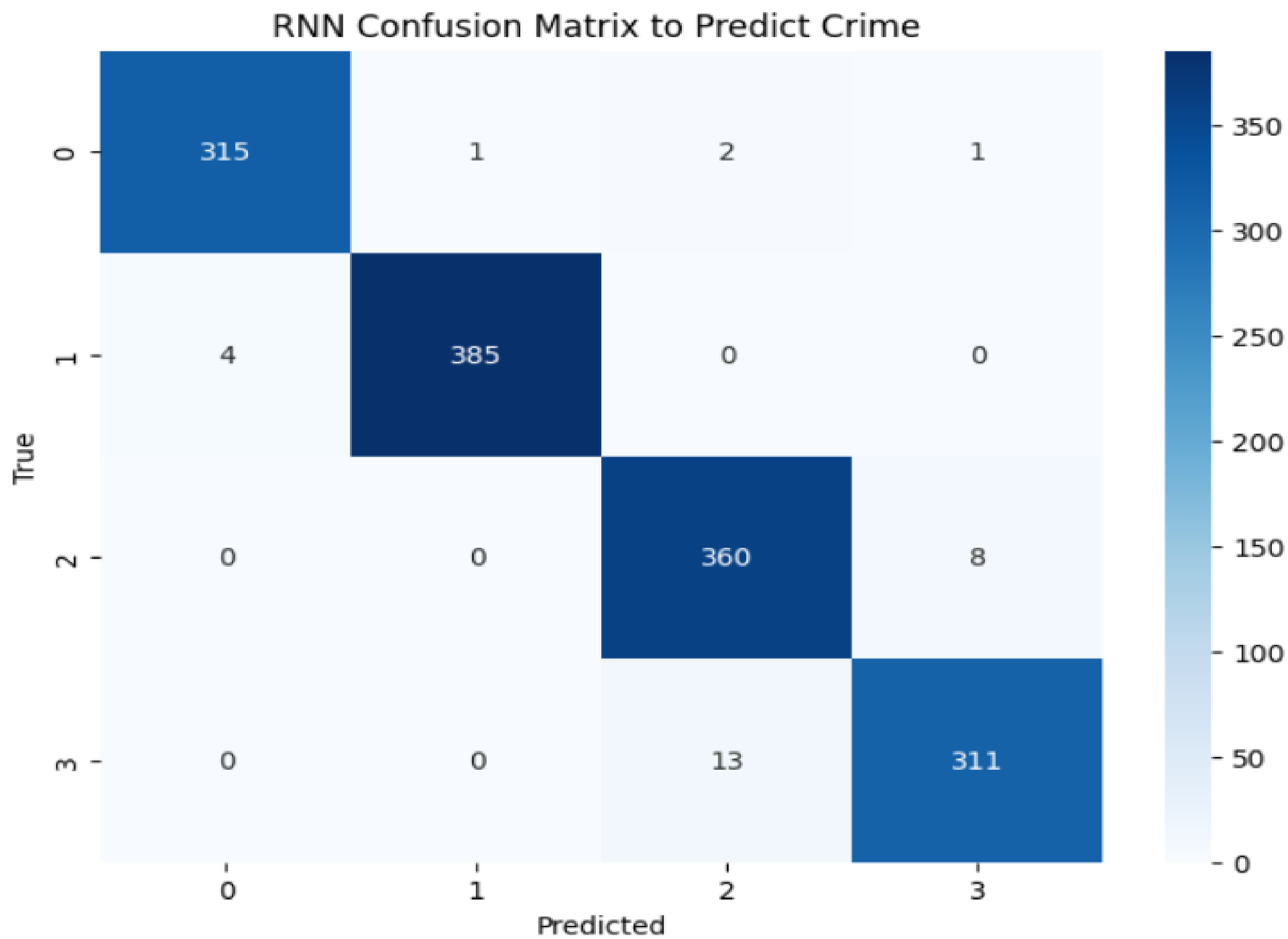

Figure 20: Confusion Matrix of RNN Model

From above figure 28 confusion matrix for the RNN model evaluating crime prediction with four classes (Violent Crime, Property Crime, State Crime, and Victimless Crime) demonstrates the model's performance across different crime types. For Violent Crime (Class 0), the model correctly predicted 315 instances, with only minimal misclassifications: 1 instance as Property Crime, 2 instances as State Crime, and 1 instance as Victimless Crime. This indicates high accuracy in identifying Violent Crimes. For Property Crime (Class 1), the model correctly identified 385 instances, with 4 instances incorrectly predicted as Violent Crime and none misclassified as State or Victimless Crime. This highlights the model's strong performance in recognizing Property Crimes. Regarding State Crime (Class 2), the model correctly predicted 360 instances, with no misclassifications to Violent or Property Crimes. 8 cases, however, were incorrectly identified as victimless crimes, suggesting that there is still room for improvement in the differentiation between these two categories of crimes.

Lastly, for Victimless Crime (Class 3), the model accurately predicted 311 instances. There were no misclassifications to Violent or Property Crimes, but 13 instances were incorrectly predicted as State Crime. Despite these misclassifications, the model generally performed well in predicting Victimless Crimes. Overall, the RNN model demonstrated high accuracy and effectiveness in distinguishing between the four crime types, with only a small number of misclassifications. The majority of instances were correctly classified

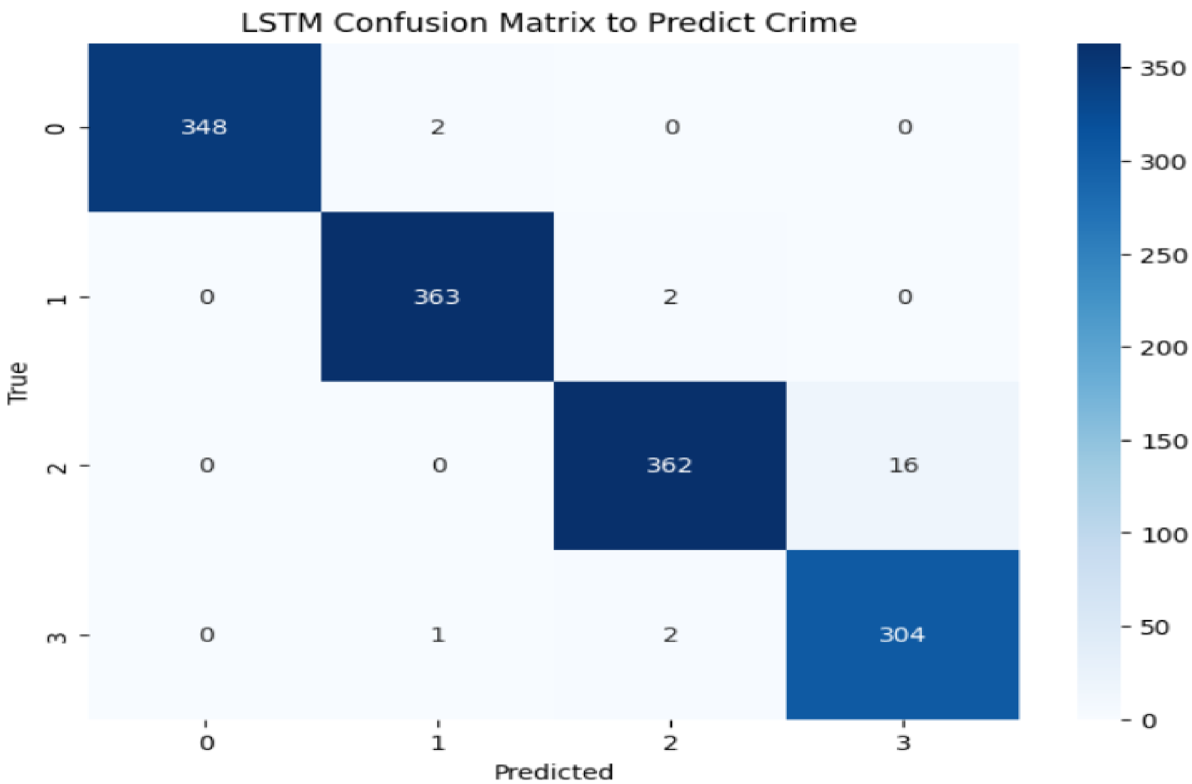

Figure 21: Confusion Matrix of LSTM Model

As we see on above figure the confusion matrix for the LSTM model on crime prediction includes four classes: Violent crime (0), Property crime (1), State crime (2), and Victimless crime (3). The outcomes demonstrate how extremely effectively the model functions. For violent crime (0), there were 348 true positives, with only a few misclassifications: 2 as property crime, no instances as state crime as victimless crime. Property crime (1) had 363 true positives, with slightly higher misclassifications: no instances as violent crime, 2 as state crime, and 0 as victimless crime. State crime (2) showed high accuracy with 362 true positives and no miss prediction as property crime, 16 instances as victimless crime. No instances of state crime were misclassified as violent and 1 instances as property crimes and 2 instances as state crime. Victimless crime (3) had perfect accuracy with 304 true positives and no misclassifications.

Overall, the model demonstrates high effectiveness in distinguishing between different types of crimes. The performance is particularly impressive for victimless crimes, with no misclassifications, and for violent and state crimes, where misclassifications are minimal. Property crimes had a few more misclassifications but still maintained high accuracy. These findings show that the LSTM model is strong and trustworthy for predicting the type of crime, accurately identifying positive and negative examples in a variety of crime categories.

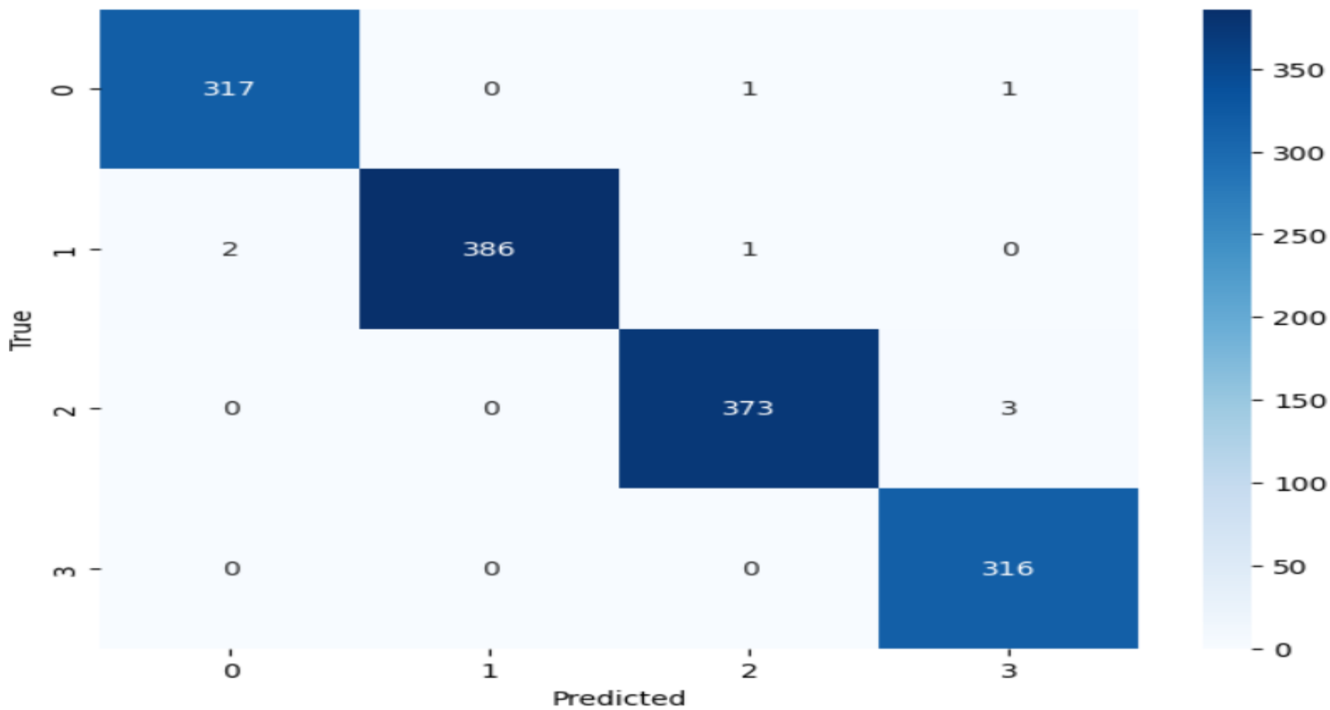

Figure 22: Confusion Matrix of MLP

Four classes are included in the confusion matrix for the MLP model on crime prediction: property crime (1), victimless crime (3), state crime (2), and violent crime (0). The algorithm accurately predicted 317 cases of violent crime (number 0), misclassifying only one as a state crime and one as a victimless crime. No crimes classified as property were misclassified. 386 true positives for property crime (1) were found, with 2 cases incorrectly classified as violent crimes and 1 incident as a state crime. No cases were incorrectly classified as victimless crimes. The algorithm accurately predicted 373 cases of state crime (2), misclassifying 3 as victimless crimes and none as violent or property crimes. Finally, victimless crime (3) had 316 true positives, with no misclassifications into any other crime categories. The MLP model predicts crime types better than the other four models, as shown in the performance evaluation part of this chapter. This is evident from MLP's exceptional accuracy and minimal misclassifications across all crime categories. The MLP model's strength lies in its ability to thoroughly understand and adapt to the dynamic nature of crime types. While the other deep neural network models, RNN and LSTM, achieved similar results, MLP stands out due to its superior performance.

The following factors contribute to MLP's success:

- **High Accuracy**: MLP achieves higher accuracy in predicting various crime types compared tothe other models.
- **Minimal Misclassifications**: The model shows fewer misclassifications, particularly invictimless crimes, where it exhibits perfect accuracy.
- **Robust Performance**: In every category, MLP performs better on a constant basis and is able to identify both good and negative examples. These aspects highlight MLP's ability to provide reliable and precise crime type predictions,making it the best choice among the evaluated models.

5.3 Research Question Discussion

Overall, our study addresses the research questions we aimed to answer by the end of the study,such as:

RQ1: What key hyperparameters greatly influence the effectiveness of crime prediction models?

Optimizing hyperparameters is crucial for improving the performance of crime prediction models. In our study, we found that using the Adam optimizer, a batch size of 32, a learning rate of 0.01, Softmax activation, categorical cross-entropy loss, and 20 epochs significantly enhanced model accuracy and speed of convergence.

Our models, such as LSTM and MLP, performed significantly better than traditional machine learning models, like Decision Trees and Naïve Bayes, which only achieved 85% accuracy in the study by Tesfaye et al. (2023), thanks to our focus on hyperparameter tuning, especially with the Adam optimizer and specific batch sizes. This demonstrates how selecting hyperparameters carefully is essential to getting better outcomes.

A chance to properly optimize model performance was lost in other recent investigations, such as those by Saraiva et al. (2022), which frequently did not thoroughly investigate how hyperparameters affect their outcomes. Our research shows that this strategy can considerably increase machine learning and deep learning models' accuracy and efficacy for crime prediction by carefully choosing and fine-tuning important hyperparameters.

RQ2: Which attributes play a crucial role in enhancing the accuracy of crime prediction models?

Improving the efficacy of crime prediction algorithms requires first determining which features are most crucial. Crime prediction models are based on temporal (e.g., time, date) and spatial (e.g., location) elements, as demonstrated by recent studies (Safat, Asghar, and Gillani, 2021; Kanimozhi et al., 2021). The fact that these characteristics are routinely used in different studies indicates how much of an impact they have on prediction accuracy. Safat et al., for example, employed time and position as their main qualities and used XGBoost to obtain an accuracy of 88%, demonstrating the significance of these features. However, they did not consider demographic or socioeconomic attributes, which our study incorporated, such as age, gender, and job status. By including these additional factors, our study achieved a higher accuracy of 98.86% with the MLP model, emphasizing the need for a broader set of attributes beyond just temporal and spatial data.

In contrast, Worku Shobe (2021) identified critical crime-related factors using a dataset from the Addis Ababa Police Commission but omitted location data, resulting in an accuracy of 94% with the SVM algorithm. This gap shows that location is a key missing element, which is crucial for capturing spatial crime trends. Our research further strengthened the model's effectiveness by integrating these attributes, along with demographic and socioeconomic data, enhancing overall model accuracy.

Our findings align with recent works, such as the study by Detotto & Otranto (2010) and Shohan et al. (2022), who used a broader dataset incorporating various crime types, and demographic information, achieving 79% accuracy with Extra Trees. The comparison reveals that models enriched with data including demographic, temporal, and spatial factors outperform those limited to basic features, thereby improving crime prediction outcomes.

RQ3: Which algorithms deliver the most accurate and efficient outcomes for crime prediction

Our study's comparison with other recent studies provides important new information about how well different crime prediction algorithms perform. The algorithms that were assessed included MLP, LSTM, KNN, RNN, and SVM. The MLP model outperformed the others, according to our analysis, with an accuracy of 98.64%. This result is higher than the greatest accuracy recorded by Naïve Bayes at 93.07% and by Safat et al. (2021), where XGBoost attained 88%. Additionally, our approach demonstrated a 5.57% improvement in accuracy compared to previous studies, highlighting how hyperparameter tuning and the selection of advanced models significantly enhanced performance. When compared to other more conventional models like Decision Trees and SVM, the MLP model's performance demonstrates its greater ability to identify intricate patterns in the data.

This result is in line with the cutting-edge research conducted by Saraiva et al. (2022), in which Random Forest only managed 83% accuracy on a smaller sample. Our study's LSTM model also performed well, especially when it came to processing sequential data, which makes it a good option for crime prediction tasks involving time-series data.

It is clear from comparing our results to those of Worku Shobe (2021), who claimed 94% accuracy using SVM, that deep learning models like LSTM and MLP provide improved accuracy because of their capacity to represent intricate relationships found in the data. Specifically, our LSTM model showed a 4.86% improvement in accuracy, indicating the benefit of using these deep learning approaches for more complex, sequential data. In comparison to conventional machine learning techniques, this suggests that they are more potent instruments for predicting crimes.

All things considered, the comparative research highlights the higher performance of sophisticated deep learning models in crime prediction tasks when combined with extensive data attributes and well-tuned hyperparameters. This fits in with the current literature's trend of using more complex algorithms for increased prediction accuracy, which boosts law enforcement's capacity to deal with crimes before they become serious.

CHAPTER SIX

CONCLUSION AND FUTURE WORKS

Crime prediction is a challenging but essential task for law enforcement agencies around the world. On this investigation, we focused on predicting crime types using several machine learning and deep learning algorithms. Our goal was to identify which algorithms could most accurately predict crime types using historical data from the Ethiopian Federal Police Commission.

We began by collecting and preprocessing crime data to ensure it was ready for analysis. This involved selecting the most important features that could help in predicting crime types. Once the data was prepared, we applied different models, including SVM, KNN, LSTM, RNN, and MLP. These models were trained using 80% of the data, while the remaining 20% was used for testing their performance. The results showed that all the models performed well, with varying levels of accuracy. The accuracy scores for each model were: SVM at 89%, KNN at 91.89%, RNN at 96.71%, LSTM at 97.92%, and MLP at 98.64%. Among these, the MLP model performed the best, achieving the highest accuracy.

Overall, this study demonstrates that machine learning and deep learning algorithms can effectively predict crime types based on historical data. These findings could assist law enforcement agencies in improving their ability to anticipate and respond to crime, potentially leading to better crime prevention strategies.

6.2 Contribution

The main contributions of this study are:

Performance Improvement: We evaluated the effectiveness of five distinct algorithms. For classifying crimes, Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Multilayer Perceptrons (MLP), Long Short-Term Memory (LSTM), and Recurrent Neural Networks (RNN) are utilized. Our study focused on improving the accuracy and effectiveness of crime prediction models. Through our analysis, we compared performance of these algorithms based on different performance evaluation metrics in predicting crime types. Finally, we found that MLP outperformed the other models, highlighting its effectiveness for crime prediction tasks.

Data Collection: We compiled an extensive dataset of 7,800 records with 17 attributes. This dataset, focusing on demographic, socioeconomic, location, time, and crime type information, is a valuable resource for researchers studying crime patterns in Ethiopia. It fills a crucial gap in the availability of crime data for the region and supports future research in crime prediction and law enforcement strategies.

6.3 Recommendations and Future Work

The crime prediction model developed in this research shows potential but can be further improved. Future models should integrate environmental data and take seasonal variations into account, including the effects of public events, festivals, and holidays. These factors significantly influence crime patterns and can provide deeper insights into how crimes vary under different conditions. By adding this information, the model's forecast accuracy would increase and it would be able to identify more dynamic crime trends.

For future work, researchers ought to look into how the season, the weather, and other outside variables could impact crime rates. Comprehending these factors can result in a more all-encompassing model that more accurately represents the dynamics of crime in the real world.

REFERENCES

Almanie, T., Mirza, R., & Lor, E. (2015). Crime prediction based on crime types and using spatial and temporal criminal hotspots. *arXiv preprint arXiv:1508.02050*.

Alsaqabi, A., Aldhubayi, F., & Albahli, S. (2019). *Using machine learning for prediction of factors affecting crimes in Saudi Arabia.* Paper presented at the Proceedings of the 2019 International Conference on Big Data Engineering.

Bachman, R., & Schutt, R. K. (2013). *The practice of research in criminology and criminal justice*: Sage.

Bardenet, R., Brendel, M., Kégl, B., & Sebag, M. (2013). *Collaborative hyperparameter tuning.* Paper presented at the International conference on machine learning.

Bengio, Y., Goodfellow, I., & Courville, A. (2017). *Deep learning* (Vol. 1): MIT press Cambridge, MA, USA.

Blackburn*, A. G., & Matthews, M. (2011). Crime and punishment in Ethiopia: A country profile. *International journal of comparative and applied criminal justice, 35*(2), 167-181.

Bothos, J. M., & Thomopoulos, S. C. (2016). *Factors influencing crime rates: An econometric analysis approach.* Paper presented at the Signal Processing, Sensor/Information Fusion, and Target Recognition XXV.

Buduma, N., Buduma, N., & Papa, J. (2022). *Fundamentals of deep learning*: " O'Reilly Media, Inc.".

Burkart, N., & Huber, M. F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research, 70*, 245-317.

Butt, U. M., Letchmunan, S., Hassan, F. H., Ali, M., Baqir, A., & Sherazi, H. H. R. (2020). Spatio-temporal crime hotspot detection and prediction: a systematic literature review. *IEEE access, 8*, 166553-166574.

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., . . . Zdeborová, L. (2019). Machine learning and the physical sciences. *Reviews of Modern Physics, 91*(4), 045002.

Celebi, M. E., & Aydin, K. (2016). *Unsupervised learning algorithms* (Vol. 9): Springer.

Cha, S., & Kang, J. (2018). Corporate default prediction model using deep learning time series algorithm, RNN and LSTM. *Journal of Intelligence and Information Systems, 24*(4), 1-32.

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Translational vision science & technology, 9*(2), 14-14.

Chun, S. A., Avinash Paturu, V., Yuan, S., Pathak, R., Atluri, V., & R. Adam, N. (2019). *Crime prediction model using deep neural networks.* Paper presented at the Proceedings of the 20th Annual International Conference on digital government research.

Cunneen, C., & Tauri, J. M. (2019). Indigenous peoples, criminology, and criminal justice. *Annual Review of Criminology, 2*(1), 359-381.

DiPietro, R., & Hager, G. D. (2020). Deep learning: RNNs and LSTM *Handbook of medical image computing and computer assisted intervention* (pp. 503-519): Elsevier.

Eckhardt, C. M., Madjarova, S. J., Williams, R. J., Ollivier, M., Karlsson, J., Pareek, A., & Nwachukwu, B. U. (2023). Unsupervised machine learning methods and emerging applications in healthcare. *Knee Surgery, Sports Traumatology, Arthroscopy, 31*(2), 376-381.

Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. *Automated machine learning: Methods, systems, challenges*, 3-33.

Gahalot, A., Dhiman, S., & Chouhan, L. (2020). *Crime prediction and analysis.* Paper presented at the 2nd International Conference on Data, Engineering and Applications (IDEA).

Ghahramani, Z. (2003). Unsupervised learning *Summer school on machine learning* (pp. 72-112): Springer.

Goodfellow, I. (2016). Deep learning: MIT press.

Janka, D. G. (2013). *A handbook on the criminal code of Ethiopia*: Dejene Girma Janka.

Jenga, K., Catal, C., & Kar, G. (2023). Machine learning in crime prediction. *Journal of Ambient Intelligence and Humanized Computing, 14*(3), 2887-2913.

Jibat, N., & Nigussie, B. (2014). Crime Situations and Reaction of Criminal Justice System in Oromia, Ethiopia. *Global journal of human social sciences, 14*(1), 73-79.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research, 4*, 237-285.

Kanimozhi, N., Keerthana, N., Pavithra, G., Ranjitha, G., & Yuvarani, S. (2021). CRIME type and occurrence prediction using machine learning algorithm. Int Conf Artif Intell Smart Syst (ICAIS). 2021; 2021: 266–73.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature, 521*(7553), 436-444.

Liu, C., Arnon, T., Lazarus, C., Strong, C., Barrett, C., & Kochenderfer, M. J. (2021). Algorithms for verifying deep neural networks. *Foundations and Trends® in Optimization, 4*(3-4), 244-404.

Longe, O., Ngwa, O., Wada, F., Mbarika, V., & Kvasny, L. (2009). Criminal uses of information & communication technologies in sub-Saharan Africa: trends, concerns and perspectives. *Journal of Information Technology Impact, 9*(3), 155-172.

Manengadan, M., Nandanan, S., & Subash, N. (2021). Crime Data Analysis, Visualization and Prediction Using LSTM. *databases, 9*(10), 17.

Meskela, T. E., Afework, Y. K., Ayele, N. A., Teferi, M. W., & Mengist, T. B. (2020). Designing time series crime prediction model using long short-term memory recurrent neural network. *International Journal of Recent Technology and Engineering, 9*(4), 402-405.

Mia, M. M. A., Biswas, S. K., Urmi, M. C., & Siddique, A. (2015). An algorithm for training multilayer perceptron (MLP) for Image reconstruction using neural network without overfitting. *International Journal of Scientific & Technology Research, 4*(02), 271-275.

Micheli-Tzanakou, E. (2011). Artificial neural networks: an overview. *Network: Computation in Neural Systems, 22*(1-4), 208-230.

Michelucci, U., & Michelucci, U. (2018). Hyperparameter Tuning. *Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks*, 271-322.

Minar, M. R., & Naher, J. (2018). Recent advances in deep learning: An overview. *arXiv preprint arXiv:1807.08169*.

Mishra, C., & Gupta, D. (2017). Deep machine learning and neural networks: An overview. *IAES international journal of artificial intelligence, 6*(2), 66.

Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). An unsupervised machine learning algorithms: Comprehensive review. *International Journal of Computing and Digital Systems*.

Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons. b, 4*(51-62), 56.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological applications, 19*(1), 181-197.

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., . . . Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM computing surveys (CSUR), 51*(5), 1-36.

Qiang, W., & Zhongli, Z. (2011). *Reinforcement learning model, algorithms and its application.* Paper presented at the 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC).

Reier Forradellas, R. F., Náñez Alonso, S. L., Jorge-Vazquez, J., & Rodriguez, M. L. (2020). Applied machine learning in social sciences: neural networks and crime prediction. *Social Sciences, 10*(1), 4.

Safat, W., Asghar, S., & Gillani, S. A. (2021). Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. *IEEE access, 9*, 70080-70094.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE, 109*(3), 247-278.

Saraiva, M., Matijošaitienė, I., Mishra, S., & Amante, A. (2022). Crime prediction and monitoring in porto, portugal, using machine learning, spatial and text analytics. *ISPRS International Journal of Geo-Information, 11*(7), 400.

Sewak, M., Sahay, S. K., & Rathore, H. (2020). An overview of deep learning architecture of deep neural networks and autoencoders. *Journal of Computational and Theoretical Nanoscience, 17*(1), 182-188.

Shakya, A. K., Pillai, G., & Chakrabarty, S. (2023). Reinforcement learning algorithms: A brief survey. *Expert Systems with Applications, 231*, 120495.

Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017). *An overview on crime prediction methods.* Paper presented at the 2017 6th ICT international student project conference (ICT-ISPC).

Shohan, F. T., Akash, A. U., Ibrahim, M., & Alam, M. S. (2022). Crime Prediction using Machine Learning with a Novel Crime Dataset. *arXiv preprint arXiv:2211.01551*.

Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE access, 7*, 53040-53065.

ToppiReddy, H. K. R., Saini, B., & Mahajan, G. (2018). Crime prediction & monitoring framework based on spatial analysis. *Procedia computer science, 132*, 696-705.

Towers, S., Chen, S., Malik, A., & Ebert, D. (2016). Factors influencing temporal patterns in crime in a large American city; a predictive analytics perspective. *Available at SSRN 2833583*.

Towers, S., Chen, S., Malik, A., & Ebert, D. (2018). Factors influencing temporal patterns in crime in a large American city: A predictive analytics perspective. *PLoS one, 13*(10), e0205151.

Vaquero Barnadas, M. (2016). *Machine learning applied to crime prediction.* Universitat Politècnica de Catalunya.

Walker, J. T., & Maddan, S. (2019). *Statistics in criminology and criminal justice*: Jones & Bartlett Learning.

Wang, H., Chen, S., Xu, F., & Jin, Y.-Q. (2015). *Application of deep-learning algorithms to MSTAR data.* Paper presented at the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).

Wu, S., Wang, C., Cao, H., & Jia, X. (2020). *Crime prediction using data mining and machine learning.* Paper presented at the The 8th International Conference on Computer Engineering and Networks (CENet2018).

Zhai, X., Ali, A. A. S., Amira, A., & Bensaali, F. (2016). MLP neural network based gas classification system on Zynq SoC. *IEEE access*, *4*, 8138-8146.