

Machine Learning Approach for Crime Prediction: A Comparative Analysis

Gadisa Adamu Mitiku

gadisa.adamu@astu.edu.et

Adama Science and Technology University

Lelise Mitiku Dabale

Adama Science and Technology University

Research Article

Keywords: Crime prediction, Machine learning, Support Vector Machine, Long Short-Term Memory, Recurrent Neural Network

Posted Date: March 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9038165/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

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 urban and rural regions. 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 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 accuracy, precision, recall, F1-score, and confusion matrix metrics 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.

1. Background of 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).

In Ethiopia, several issues, such as increased rates of joblessness 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 to 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 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.

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.

2. Literature Review and Related Works

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 a 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 (Zhanget 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 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 beyond reactive measures by providing predictive insights based on historical crime data (Gahalot, Dhiman, & Chouhan, 2020).

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, the 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.

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. The study introduces a novel dataset that includes 6,574 crime incidents collected over seven years. The information in this dataset was gathered from several 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, outdatedness, 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 shows a summary of related works compared with ours.

Table 1: summary of related works

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.

3. Methodology

This section provides a brief history that complements the design, execution, and assessment of the suggested architecture by outlining the techniques, instruments, and processes utilized to analyze the research challenges. It contains techniques for sample selection, data collecting, and analysis 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 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.

Generally, in this section research design, including data collection, preprocessing, and the specific machine learning models used are discussed.

Data Collection: The data was obtained from the Ethiopian Federal Police Commission, encompassing records from three sub-cities between September 2011 and October 2016 E.C. This dataset includes approximately 8000 crime incidents, with features such as crime type, date, time, location, offender demographics (age, gender, marital status, education), and victim information. Crimes are classified into four categories: Violent Crime (VC), Property Crime (PC), State Crime (SC), and Victimless Crime (VCC).

Data Preprocessing

Data preprocessing involves several steps to prepare the raw data for analysis. Missing values were handled by mode imputation for categorical data and mean/median imputation for numerical data. Normalization techniques were applied to ensure that numerical features such as age and time were on the same scale. Categorical variables like crime type were encoded using one-hot encoding and label encoding, which allowed the machine learning models to interpret these features.

Feature Selection

Out of the 25 features available in the dataset, 17 were selected for the final model. Features such as offender and victim age, gender, location, and crime type were identified as the most relevant for predicting future crimes. Personal information like names and phone numbers was excluded to ensure privacy and data protection.

Model Selection

The machine learning and deep learning models tested in this study include

- **Support Vector Machine (SVM):** A robust model for multi-class classification, effective in handling high-dimensional data.
- **K-Nearest Neighbors (KNN):** A simple but effective model for identifying local crime patterns.
- **Recurrent Neural Networks (RNN):** Suited for time-series data, capturing temporal patterns.

- **Long Short-Term Memory (LSTM):** An advanced version of RNN that handles long-term dependencies, particularly useful for sequential crime data.
- **Multilayer Perceptron (MLP):** A neural network with multiple layers capable of modeling complex relationships between variables.

Hyperparameter Tuning

Grid search and cross-validation techniques were used to optimize the models. Parameters like the learning rate, batch size, and the number of epochs were adjusted to maximize performance and minimize overfitting.

Evaluation Metrics

The models were evaluated using accuracy, precision, recall, F1-score, and confusion matrix to assess their performance in predicting crime types. Each of these metrics provides insight into how well the models classify crimes, with accuracy measuring overall correctness, precision focusing on the proportion of true positives, and recall emphasizing the model's ability to identify true positive cases.

4. Proposed Solution

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 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 socioeconomic 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. Overall, this research 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 is shown below.

5. Results and Discussion

This section comprehensively analyzes the performance of various machine learning and deep learning models applied to the crime prediction task. The models evaluated in the study include Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Multilayer Perceptron (MLP). These models were trained and tested on a dataset from the Ethiopian Federal Police Commission (EFPC), which included 8000 records spanning from September 2011 to October 2016 E.C.

The results are discussed in terms of accuracy, precision, recall, F1-score, and the confusion matrix, which are standard performance metrics used to evaluate the efficacy of classification models. Each model's ability to predict different types of crimes, including Violent Crime (VC), Property Crime (PC), State Crime (SC), and Victimless Crime (VCC), was compared, with a focus on determining the most effective model for real-world crime prediction applications.

We evaluated 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%, a precision of 92.8%, a 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 Table 2 below.

Table 2
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%

In addition to accuracy, Precision, recall and F1 score 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.

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 learned 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.

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 throughout 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 learned 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 5: Training and validation loss graph of MLP model

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 MLP models is shown in the following figure.

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.

6. Conclusion

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.

7. Contributions of the work

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 Perceptron (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 the 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.

Declarations

Availability of data and materials: Access to the datasets used for this study is available from the corresponding author upon reasonable request.

Competing interests: No conflicts of interest.

Funding: Not applicable.

Acknowledgments: not applicable

Authors' contributions: Lelise Mitiku Dabale was involved in data collection, cleaning, data analysis, interpretation, and manuscript preparation. Gadisa Adamu Mitiku helped in data analysis and manuscript preparation. Both read and approved the manuscript.

Performance Improvement: We evaluated the effectiveness of five distinct algorithms. For classifying crimes, Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Multilayer Perceptron (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 the 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.

References

1. 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*.
2. 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.
3. Bachman R, Schutt RK. The practice of research in criminology and criminal justice. Sage; 2013.
4. Bardenet R, Brendel M, Kégl B, Sebag M. (2013). *Collaborative hyperparameter tuning*.
5. Paper presented at the international conference on machine learning.
6. Bengio Y, Goodfellow I, Courville A. Deep learning. Volume 1. MA, USA: MIT Press Cambridge; 2017.
7. Blackburn* AG, Matthews M. Crime and punishment in Ethiopia: A country profile. *Int J Comp Appl criminal justice*. 2011;35(2):167–81.

8. Bothos JM, Thomopoulos SC. (2016). *Factors influencing crime rates: An econometric analysis approach*. Paper presented at the Signal Processing, Sensor/Information Fusion, and Target Recognition XXV.
9. Buduma N, Buduma N, Papa J. Fundamentals of deep learning. O'Reilly Media, Inc.; 2022.
10. Burkart N, Huber MF. A survey on the explainability of supervised machine learning. *J Artif Intell Res*. 2021;70:245–317.
11. Butt UM, Letchmunan S, Hassan FH, Ali M, Baqir A, Sherazi H. H. R. (2020).
12. Spatio-temporal. crime hotspot detection and prediction: a systematic literature review.
13. IEEE access, 8, 166553–74.
14. Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Zdeborová L. Machine learning and the physical sciences. *Rev Mod Phys*. 2019;91(4):045002.
15. Celebi ME, Aydin K. Unsupervised learning algorithms. Springer; 2016. p. 9.
16. Cha S, Kang J. Corporate default prediction model using deep learning time series algorithm, RNN and LSTM. *J Intell Inform Syst*. 2018;24(4):1–32.
17. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. *Translational Vis Sci Technol*. 2020;9(2):14–14.
18. Chun SA, Paturu A, Yuan V, Pathak S, Atluri R, V., Adam R. N. (2019). *Crime prediction model using deep neural networks*. Paper presented at the Proceedings of the 20th Annual International Conference on Digital Government Research.
19. Cunneen C, Tauri JM. (2019). Indigenous peoples, criminology, and criminal justice.
20. *Annual Rev Criminol*, 2(1), 359–81.
21. DiPietro R, Hager GD. Deep learning: RNNs and LSTM. *Handbook of medical image computing and computer assisted intervention*. Elsevier; 2020. pp. 503–19.
22. Eckhardt CM, Madjarova SJ, Williams RJ, Ollivier M, Karlsson J, Pareek A, Nwachukwu BU. Unsupervised machine learning methods and emerging applications in healthcare. *Knee Surg Sports Traumatol Arthrosc*. 2023;31(2):376–81.
23. Feurer M, Hutter F. (2019). Hyperparameter optimization. *Automated machine learning: Methods, systems, challenges*, 3–33.
24. Gahalot A, Dhiman S, Chouhan L. (2020). *Crime prediction and analysis*. Paper presented at the 2nd International Conference on Data, Engineering, and Applications (IDEA).
25. Ghahramani Z. Unsupervised learning *summer school on machine learning*. Springer; 2003. pp. 72–112.
26. Goodfellow I. (2016). Deep learning: MIT Press.
27. Janka DG. A handbook on the criminal code of Ethiopia. Dejene Girma Janka; 2013.
28. Jenga K, Catal C, Kar G. Machine learning in crime prediction. *J Ambient Intell Humaniz Comput*. 2023;14(3):2887–913.

29. Jibat N, Nigussie B. Crime Situations and Reaction of Criminal Justice System in Oromia, Ethiopia. *Global J Hum social Sci.* 2014;14(1):73–9.
30. Kaelbling LP, Littman ML, Moore AW. (1996). Reinforcement learning: A survey.
31. *J Artif Intell Res*, 4, 237–85.
32. Kanimozhi N, Keerthana N, Pavithra G, Ranjitha G, Yuvarani S. (2021). CRIME type and occurrence prediction using a machine learning algorithm. *Int Conf ArtifIntell Smart Syst (ICAIS)*. 2021; 2021: 266–73.
33. 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).
34. Algorithms for verifying deep neural networks. *Found Trends® Optim*, 4(3–4), 244–404.
35. Longe O, Ngwa O, Wada F, Mbarika V, Kvasny L. Criminal uses of information & communication technologies in sub-Saharan Africa: trends, concerns and perspectives. *J Inform Technol Impact*. 2009;9(3):155–72.
36. Manengadan M, Nandan S, Subash N. (2021). Crime Data Analysis, Visualization, and Prediction Using LSTM. *databases*, 9(10), 17.
37. Meskela TE, Afework YK, Ayele NA, Teferi MW, Mengist. T. B. (2020).
38. Designing time series. crime prediction model using long short-term memory recurrent neural network. *Int J Recent Technol Eng*, 9(4), 402–5.
39. Mia MMA, Biswas SK, Urmi MC, Siddique A. An algorithm for training multilayer perceptron (MLP) for Image reconstruction using neural network without overfitting. *Int J Sci Technol Res*. 2015;4(02):271–5.
40. Micheli-Tzanakou E. Artificial neural networks: an overview. *Network: Comput Neural Syst*. 2011;22(1–4):208–30.
41. Michelucci U, Michelucci U. (2018). Hyperparameter Tuning. *Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks*, 271–322.
42. Minar MR, Naher J. (2018). Recent advances in deep learning: An overview. *arXiv preprint arXiv:1807.08169*.
43. Mishra C, Gupta D. (2017). Deep machine learning and neural networks: An overview.
44. IAES. *international J Artif Intell*, 6(2), 66.
45. Naeem S, Ali A, Anam S, Ahmed MM. (2023). An unsupervised machine learning algorithm: Comprehensive review. *Int J Comput Digit Syst*.
46. Nasteski V. An overview of the supervised machine learning methods. *Horizons b*. 2017;4(51–62):56.
47. Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. (2009). Sample selection bias and presence-only distribution models: implications for.
48. background and pseudo-absence data. *Ecological Applications*, 19(1), 181–197.

49. Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, Iyengar SS. A survey on deep learning: Algorithms, techniques, and applications. *ACM Comput Surv (CSUR)*. 2018;51(5):1–36.
50. Qiang W, Zhongli Z. (2011). *Reinforcement learning model, algorithms and its application*.
51. Paper presented at. the 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC).
52. Reier Forradellas RF, Nández Alonso SL, Jorge-Vazquez J, Rodriguez. M. L. (2020).
53. Applied machine. learning in social sciences: neural networks and crime prediction.
54. *Social Sci*, 10(1), 4.
55. Safat W, Asghar S, Gillani SA. Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. *IEEE Access*. 2021;9:70080–94.
56. Samek W, Montavon G, Lapuschkin S, Anders CJ, Müller K-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247–278.
57. Saraiva M, Matijošaitienė I, Mishra S, Amante A. Crime prediction and monitoring in porto, portugal, using machine learning, spatial and text analytics. *ISPRS Int J Geo-Information*. 2022;11(7):400.
58. Sewak M, Sahay SK, Rathore H. An overview of deep learning architecture of deep neural networks and autoencoders. *J Comput Theor Nanosci*. 2020;17(1):182–8.
59. Shakya AK, Pillai G, Chakrabarty S. Reinforcement learning algorithms: A brief survey. *Expert Syst Appl*. 2023;231:120495.
60. Shamsuddin NHM, Ali NA, Alwee R. (2017). *An overview of crime prediction methods*. Paper presented at the 2017 6th ICT International Student Project Conference (ICT-ISPC).
61. Shohan FT, Akash AU, Ibrahim M, Alam MS. (2022). Crime Prediction using Machine Learning with a Novel Crime Dataset. *arXiv preprint arXiv:2211.01551*.
62. Shrestha A, Mahmood A. (2019). Review of deep learning algorithms and architectures.
63. *IEEE access*, 7, 53040–65.
64. ToppiReddy HKR, Saini B, Mahajan G. Crime prediction & monitoring framework based on spatial analysis. *Procedia Comput Sci*. 2018;132:696–705.
65. 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*.
66. 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.
67. Vaquero Barnadas M. Machine learning applied to crime prediction. Universitat Politècnica de Catalunya; 2016.
68. Walker JT, Maddan S. Statistics in criminology and criminal justice. Jones & Bartlett Learning; 2019.
69. 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).

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

71. Zhai X, Ali AAS, Amira A, Bensaali F. MLP neural network-based gas classification system on Zynq SoC. *IEEE Access*. 2016;4:8138–46.

Figures

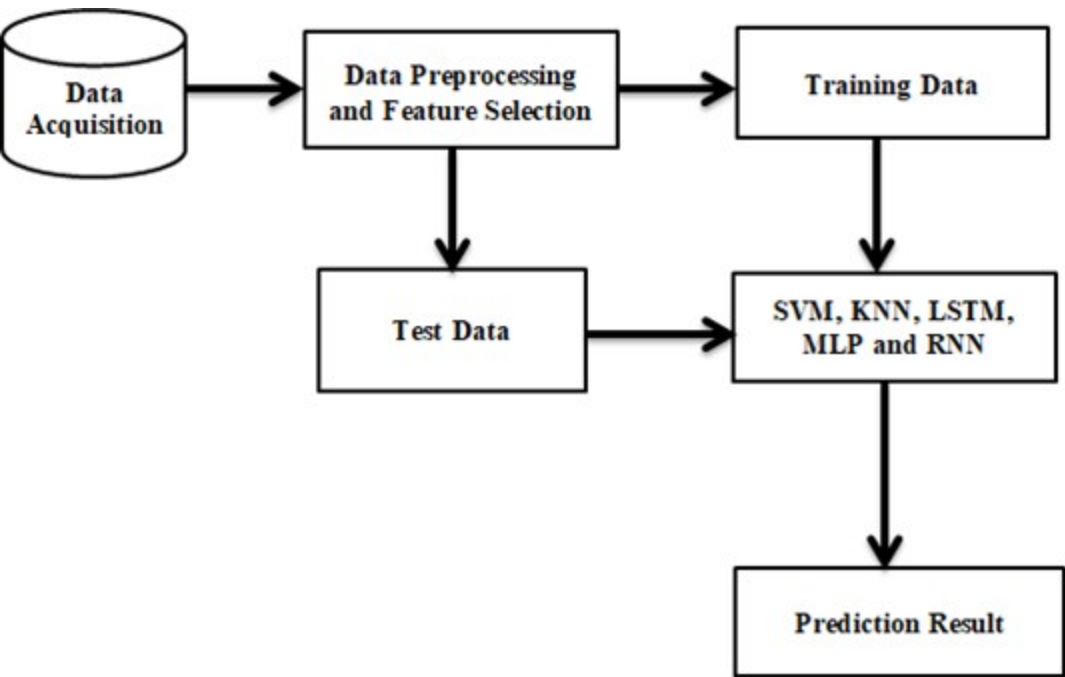


Figure 1

Proposed Solution

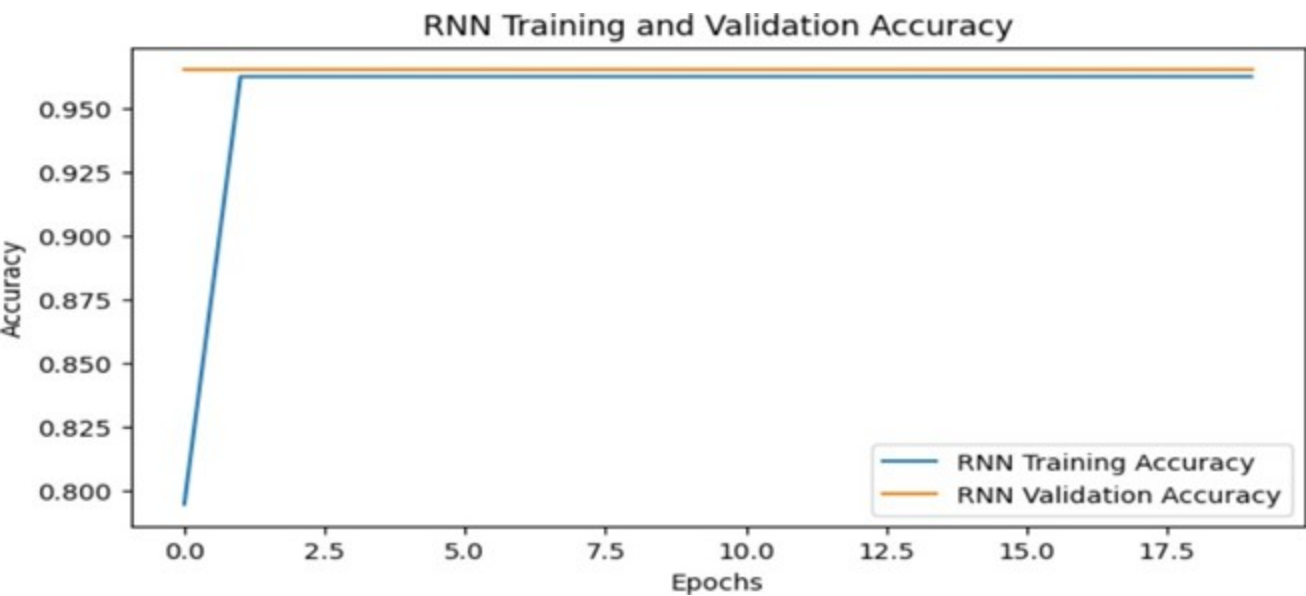


Figure 2

Training and validation Accuracy graph of RNN model

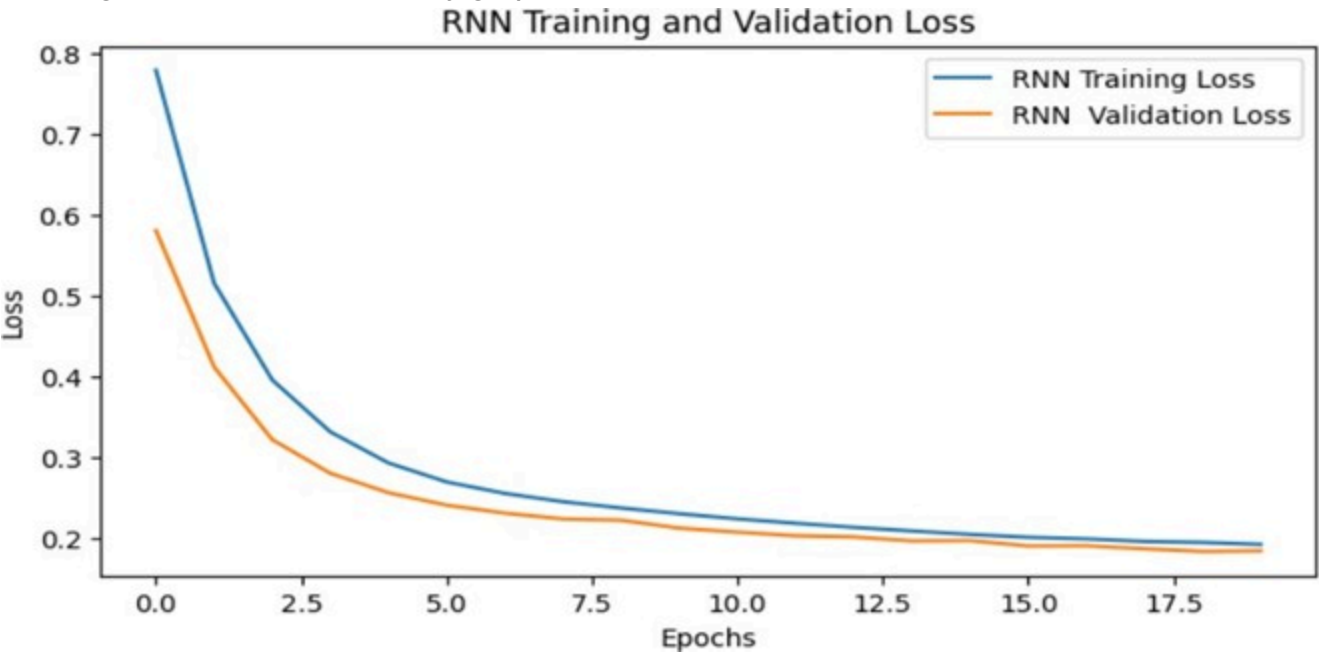


Figure 3

Training and validation loss graph of RNN model

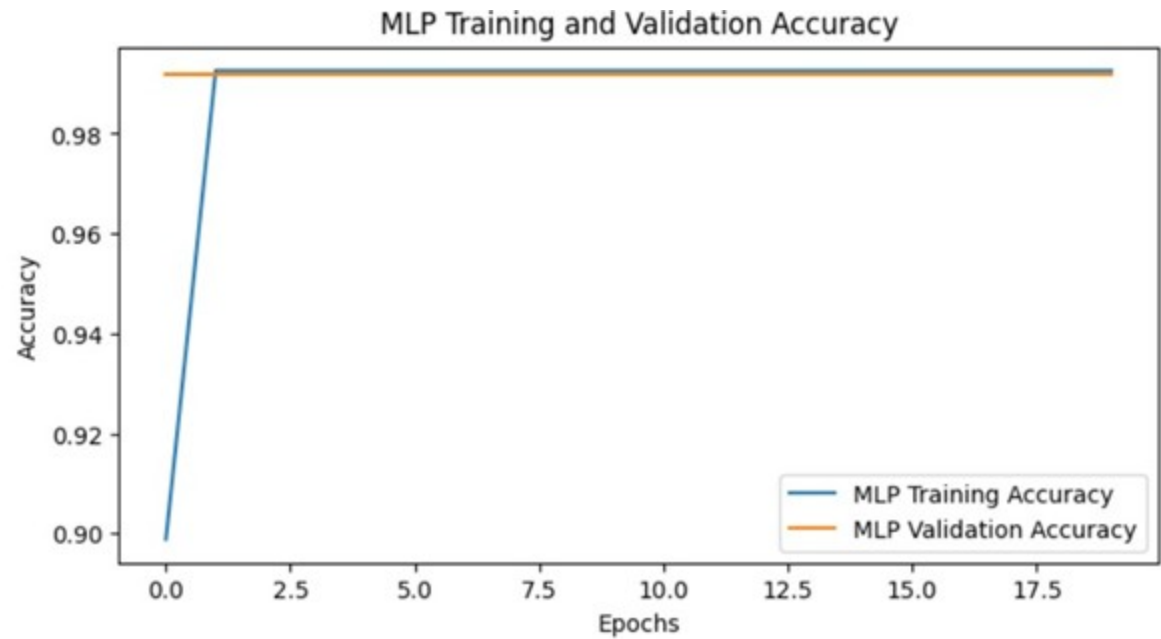


Figure 4

Training and validation Accuracy graph of MLPmodel

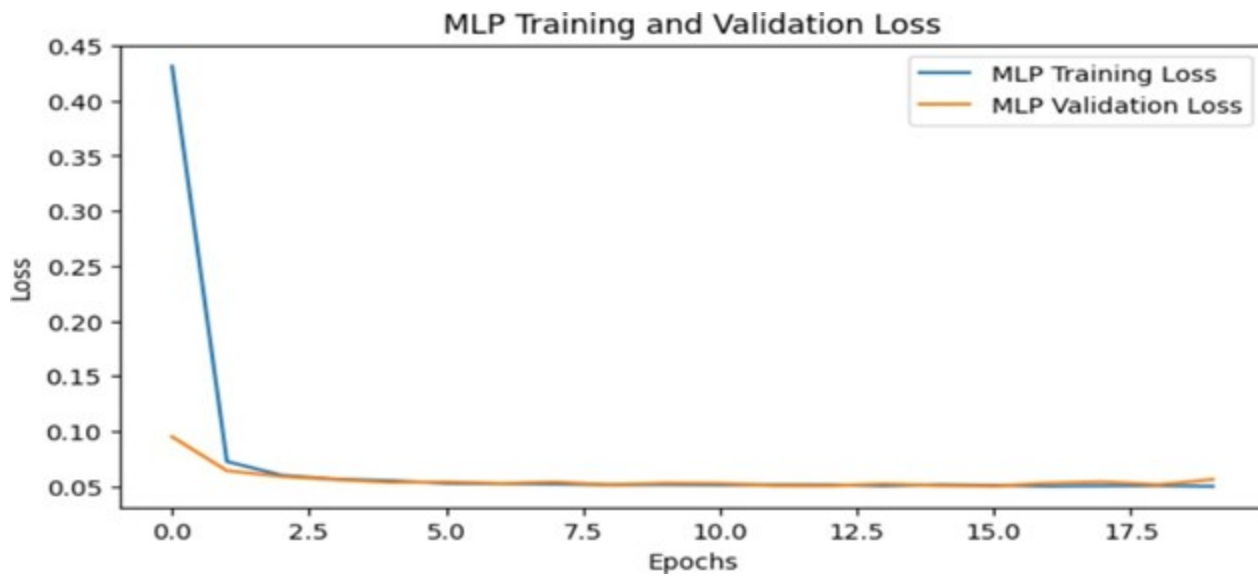


Figure 5

Training and validation loss graph of MLPmodel

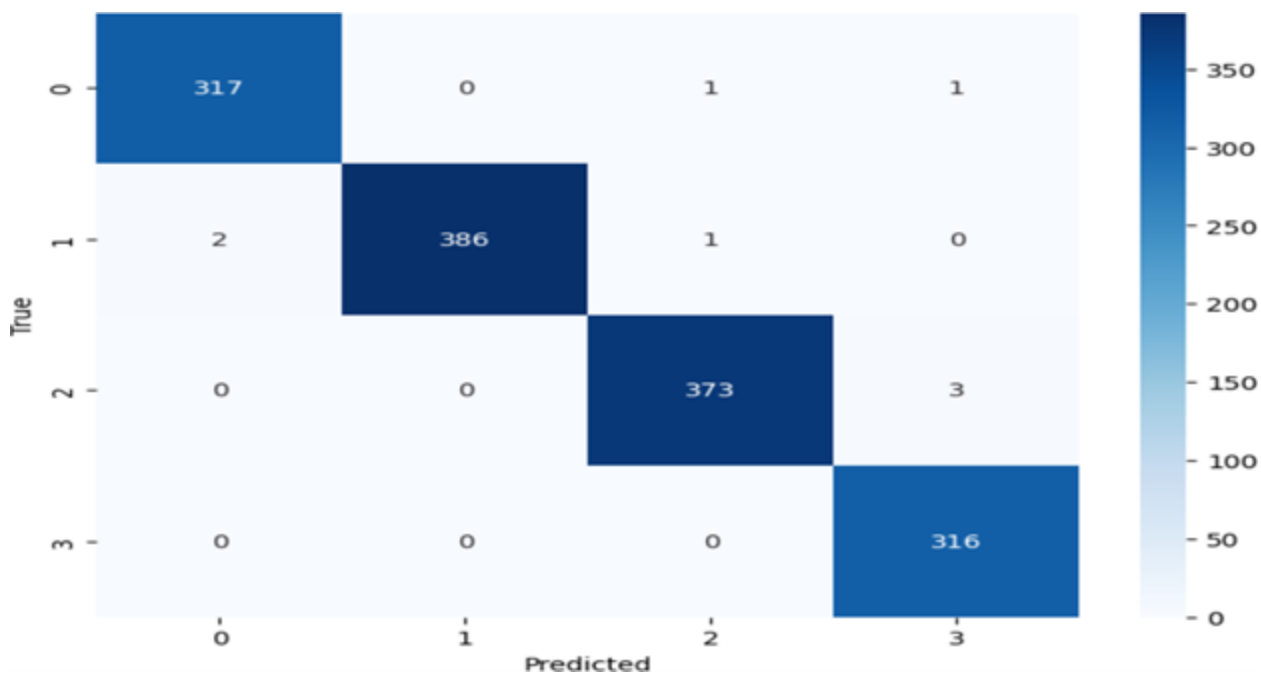


Figure 6

Confusion Matrix of MLP