Enhanced Crime Prediction: Leveraging CNN-LSTM Fusion for Improved Accuracy and Temporal Pattern Recognition

K.Ramalakshmi[1], Satheeshkumar Palamnisamy[2,*], L.Krishna Kumari[3], Khalaf Ibrahim Osamah [4], Karthikeyan Ganesan[5], Sameer Algburi[6], Habib Hamam[7]

[1]Department of Electronics and Communication Engineering, P.S.R Engineering College, Sivakasi, Tamilnadu, India, ramalakshmi@psr.edu.in
[2] Department of ECE, BMS Institute of Technology and Management, Bengaluru, India, satheeshp@bmsit.in
[3]Department of Electronics and Communication Engineering, Ramco Institute of Technology, Rajapalayam, Tamilnadu, India, krishnakumari@ritrjpm.ac.in
[4]Department of Solar, Al-Nahrain Research Center for Renewable Energy, Al-Nahrain University, Jadriya, Baghdad, Iraq, usama81818@nahrainuniv.edu.iq
[5]Department of Electronics and Communication Engineering, P.S.R. Engineering College (Autonomous), Sevalpatti, Sivakasi, Tamilnadu, India, karthikeyan@psr.edu.in .
[6]Al-Kitab University, College of Engineering Techniques, sameer.algburi@uoalkitab.edu.iq
[7]Uni de Moncton, NB, 1EA 3E9, Canada, Habib.Hamam@umoncton.ca

Abstract:

Urbanization is a contributing factor to numerous social problems. Crime is one of these issues that exists in every city on the planet. A significant amount of data is gathered by police databases, which may be examined to lower the rate of crime. The investigation of criminal activity and the estimation of several crimes ruins one of the most exciting complications for investigators. It is not unusual for people in developing nations like India to hear about crimes occurring frequently. With the quick development of cities, we have to be continually conscious of our surroundings. To effectively manage such expanded and intelligent crimes, it is significant to investigate the recent crime trends and make organized and inclusive countermeasures to encounter the new movements of crime. Preventing crimes before they occur is always preferable to dealing with them after they have already occurred. Because of the recent dramatic advancements in machine learning technology, research on data analysis and crime prediction systems is crucial to lowering the rate of crime. The hybrid prediction method observes crime rates in order to avoid the unfortunate. A hybrid model based on deep learning approaches that combines a convolutional neural network (CNN) model and a long-short-term memory (LSTM) model to improve the accuracy of crime rate prediction is proposed. The CNN layers are added first in this CNN-LSTM method, and the LSTM layers are then added. On the procedure of the CNN model for feature extraction, and subsequent to the LSTM method to understand the features through time steps that have a high density layer for output,. The CNN and LSTM models provide a complete guide for crime rate analysis. The performance of the CNN and LSTM models is provided with 97.8% accuracy. When compared to the traditional methods, the proposed method yields high accuracy.

Keywords: Hybrid model; Crime rate; Long Short-Term Memory; Convolutional Neural Network

I. Introduction

Day by day, excellence of lifespan has enhanced as the inclination of progressive industrialization, information, and specialization fast-tracked due to the unceasing and swift economic evolution and the progress of science and technology. Nevertheless, several issues like unemployment, poverty, and crime never go away. Amid these social glitches, crime issues have become added intelligent, progressive, differentiated, and widened, creating it hard to envisage and survive with crimes beforehand they happen. Traditional data mining can recognize the arrangements and tendencies from organized information but through the progression in expertise, inventive data mining can excerpt the data from organized and formless data. Data mining give assistance to criminal detectives for exactly and competently examine the huge amount of information. Prediction is the most important process that has involved substantial attention given the potential suggestion of positive prediction for all applications. The two types of predictions namely prediction of unavailable data or pending trends, or else a class label for a few data is predicted. Then classification takes place. If the classification process is designed depended on a training set, class label of an object is predicted based on object attribute values and class attribute values. The prediction is used to estimate absent numerical values in time-correlated data. Its main idea is used for a huge volume of earlier value for reflecting possible upcoming data value. The clustering process is similar to the classification process whereas clustering is the organization of data in classes. Unlike classification, the class labels are unknown in clustering and the algorithm is to discover the labels of adequate classes. The clustering is recognized as an unsupervised categorization, due to its classification is not revealed by the given class labels. Several clustering methods are developed based on the principle of exploiting the resemblance between same class bodies such as intra-class resemblance and diminishing resemblance among different class entities such as inter-class resemblance. Section I presents introduction, Section II illustrates the existing works related to this research, the preceeding section III analyses the materials and methods of the work proposed, section IV illustrates the results and discussion and finally section V concludes and describe the future directions of the work proposed.

II.Related works:

Sivanagaleela et al. 2019 examined factors related to crime from earlier years. This system focuses more on the location of crimes than it does on identifying the perpetrators. To classify and group the crime data for all crimes, like murder, kidnapping, cheating, theft, burglary, robbery, crime against women, and other crimes, fuzzy C-Means algorithm and naive Bayes classification are employed here. A regression model was presented by Yadav et al. (2017) to predict the crime rate in different states from 2001 to 2014. Many machine learning processes are utilized by Safat et al., 2021 to analyze time series data through LSTM and autoregressive unified moving average (ARIMA) models. These processes included support vector machine, logistic regression, k-nearest neighbors, decision tree, Naïve Bayes, random forest, eXtreme Gradient Boosting and multilayer perceptron. On both data sets, the root mean square error and mean absolute error of the LSTM time series analysis were performed to a largely acceptable degree. In comparison to other methods, this one offers enhanced predictive accuracy for future trends, early crime detection, and hotspots with higher crime rates. These insights can be utilized to inform police tactics and practices. Kumar et al. (2020) used the KNN prediction method to observe the crime rate. It will make educated guesses about the kind of crime, when, where, and when it might occur. The behaviors in crime over a region will be provided by this data, which may be useful for criminal investigations. The assemble-stacking based crime estimate approach, which is grounded on SVM set of rules, is a real reliable method presented by Kshatri et al., 2021 for recognizing the suitable forecasts of crime by employing learning-based approaches. Exhausting the SVM procedure, domain-specific formations can be accomplished in association to supplementary machine learning approaches like Random Forest as well as Naïve Bayes. The ensemble model that has the premier correlation coefficient plus average and absolute errors sometimes performs better than the other models. Classification accuracy of the proposed method was 99.5%. To consistently estimate crime tendencies in every region and automatically identify high-risk crime areas in urban zones, Catlett et al. (2018) presented a methodology centered on spatial investigation and auto-regressive representations. The process yields a spatiotemporal crime prediction method, which is made up of an assemblage of regions with high crime rates and a group of interconnected crime prognosticators. Each of these components exemplifies a predictive model for estimating the number of crimes that would occur in a given region. The experimental evaluation demonstrates that the suggested technique realizes greater accuracy in spatial and temporal crime predicting over rolling time horizons, using real-world information composed in the region of Chicago. To predict crime in the upcoming years, Agarwal et al. (2018) investigated statistical approaches such as Functional Coefficient Regression, Weighted Moving Average and Arithmetic-Geometric Progression. For this, the criminality statistics for the Indian states from 2001 to 2013 were utilized. Information on crimes committed between 2001 and 2011 has been used to forecast crimes in 2012 and 2013. These estimated values have been contrasted with information on actual crimes that occurred in 2012 and 2013. The accuracy of the suggested approaches is indicated by the variation among the actual records and the estimated values of those years. To autonomously and successfully identify the high-risk regions in the city, Han et al. (2020) recommended a daily crime prediction framework that integrated the LSTM and Spatial-Temporal Graph Convolutional Network (ST-GCN). The model's dataset is carried by topological maps of urban areas, and it principally consists of two parts to derive the factors of theft crimes collectively called a spatial-temporal features extraction part and temporal feature extraction part. A method for predicting crimes utilizing weather and Twitter data was presented by Sandagiri et al. in 2020. The two modules of the suggested method are the crime prediction and detection modules. The crime detection module identifies the Twitter posts about the offence. The recognition module is developed with a Bidirectional Encoder Representations from Transformers based methodology. Next, an ANN is used to implement the extrapolation module. The efficacy of the BERT and ANN centred forecast method is demonstrated by the empirical analysis of our suggested strategy. According to Joshi et al.'s 2019 presentation, law enforcement continues to trail behind criminals convincingly. Improving the efficiency of investigations and analyzing the rising number of crimes are two of the police's top priorities. This study develops an intuitive user interface for mapping criminal activity using QGIS, as well as for visualizing, analyzing, and forecasting patterns and trends in criminal activity using clustering algorithms like K-Means, agglomerative algorithms, and predictive algorithms like SVM and Random Forest. According to Pratibha et al. [2020], it is important to keep an accurate database of all crimes committed since it may be useful in the future. Law enforcement agencies can prevent crimes before they happen by having the ability to predict potential crimes in the future. Tarlekar et al.'s [2021] analysed the prediction of crime using an efficient method for categorizing the different kinds of crimes, explaining their motives, and forecasting future crimes. Official police report and data that were scraped from reputable websites are both included in the dataset. The hotspot areas are determined by the system through the analysis of crime reports. Analysts of crime information can assist law prosecution in finding criminals more rapidly. Chauhan et al.'s [2017] analysis of unstructured data revealed previously undiscovered, valuable information. Utilizing analytical and predictive methods to detect criminal activity is known as predictive policing, and research has shown that this approach is largely successful in doing so. Using a machine learning algorithm, Asad et al. [2020] were able to predict computer malware infection rates based on the features of the malware. Both gradient-boosting algorithms and supervised machine-learning algorithms are being used. A rough time and risk map of the incident of specific types of crimes, for example violent thefts, armed hold-ups, and home invasions, was used by Baloian et al. in 2017. The police can then make use of this information to increase patrols in the appropriate areas, which will lower the rate of crime. Ghankutkar et al., [2019] provided a report that analyses a real time crime news. The recommended system is a web-based application that affords a description on crime-related news by analyzing instaneous crime information in the arrangement of online articles. Rather than the other way around, Raza et al.'s [2021] research concentrated on predicting the areas based on crimes. Additionally, this will help estimate the amount of crimes that will occur in various cities within a given period. Additionally, by being aware of the projected trend, detectives and law enforcement will be able to take preventative measures, which will lower the frequency of these incidents. Here, the districts (also known as regions) are predicted using the Random Forest algorithm based on various crime rates.

III. Materials and Methods

CNN AND LSTM Model

CNN and LSTM methods were used to simulate and predict crime rates which are explained in the following Figure 1

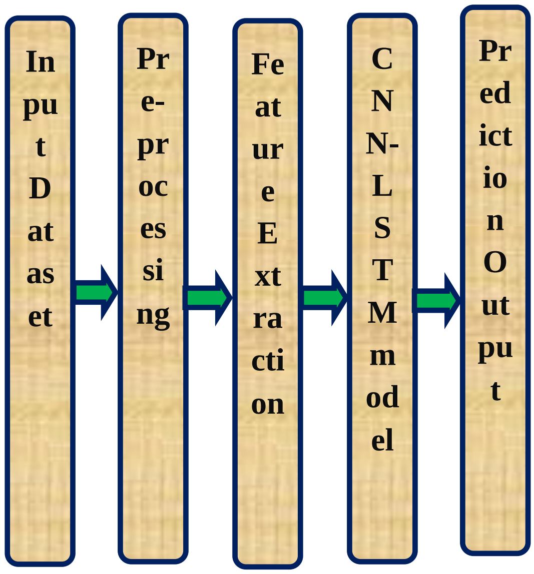

Figure 1 Prediction using CNN-LSTM

At first, the data are extracted and stored as a data frame from the input dataset which is named (crime data from dataset.csv) Then we pre-process the data by selecting the first 33 columns only and hence the unnamed columns are dropped or neglected. Then we get the state name and crime case name from the user to extract the data that is related to those state name and crime case names. We select the total crime cases row along the extracted data. Hence now we have the total crime cases for that particular state dataset. Now we feed the extracted crime case values to the CNN+LSTM train that model. After training, we predicted the future for the next 10 years by using the trained CNN model.

Data Set Details

Data sets with various states are gathered from the "Crime in India" website. After every year, the State Crime Records Bureaux gathers information to compile the description from the District Crime Records Bureaux and forward it to NCRB with the source. Megacities data is also gathered independently.

CNN and LSTM Workflow

The data on daily crime incidence were collected and segmented into a training group for a model development and testing group for assessment. The CNN as well as LSTM models are pooled and not influenced by exogeneous meteorological variables were espoused to apt the crime rate prevalence through exhausting about the details regarding of the training group. Using the test group data, the four fitted models' forecasting accuracy was confirmed. The following Figure 2 explains the workflow of the suggested model to identify the excellent approach for envisaging the criminal rate by integrating LSTM models incorporating exogenous meteorological variables.

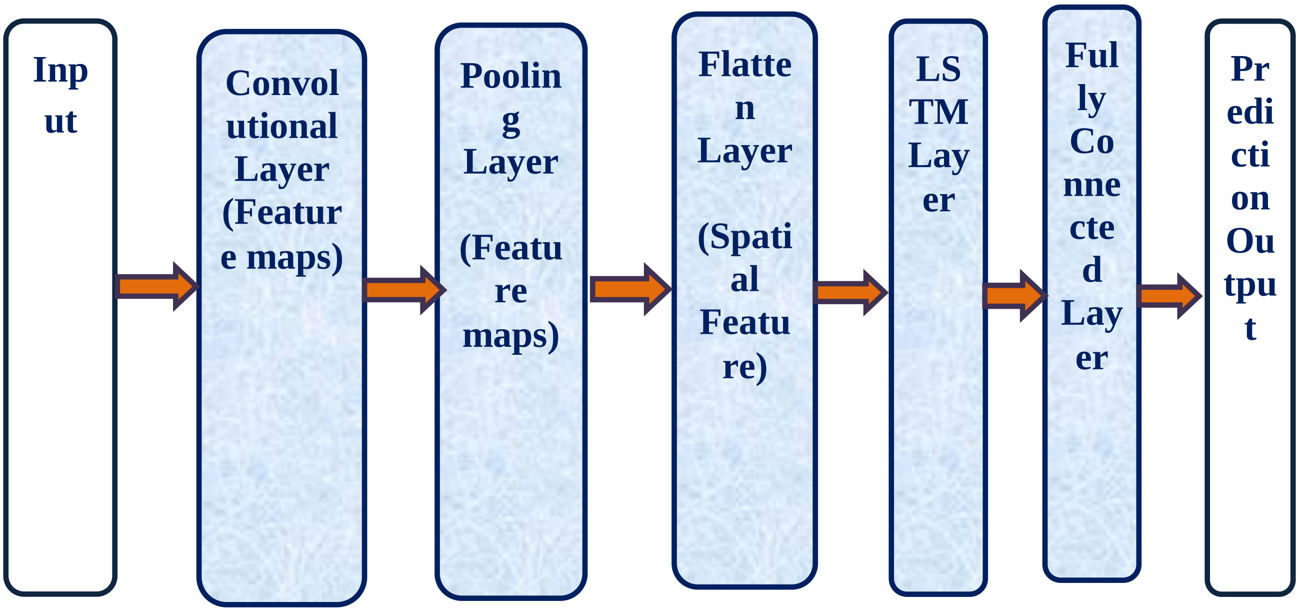

Figure 2 Prediction using CNN and LSTM Workflow

Convolutional Neural Network (CNN)

CNNs are superior to DNNs in many ways. The primary one resembles more resembles the human visual handling scheme, being optimized greatly in the configuration of 2D and 3D images processing, and being proficient in acquiring and abstracting 2D features demonstrating effectiveness in the learning process. The max-pooling layer is proficient at assimilating shape disparities. Similarly, it is also made up of sparse connections that have weights attached to them. CNNs have a many less constraints than a fully linked network of the identical size. The optimization algorithm is a powerful tool for almost all CNNs, and it is less vulnerable to the diminishing gradient issue. When the optimization algorithm instructs the entire web to directly reduce loss function, CNN can generate well-tuned weights. The system framework of CNNs is shown in Figure 3. It is composed of two primary components: a classifier and feature extractors. In the feature extraction layers, the layer from the output of each layer is taken directly before it takes as input. In the same way, it feeds its output into the layer that comes after it. Convolution, max-pooling, and classification are the three different layers that produce the CNN architecture. In exist, two different kinds of layers in the middle levels and network's low. Max-pooling and convolutional layers are what they are. Max-pooling operations are performed on odd-numbered layers and convolutions are executed on even-numbered layers. Feature mapping is the name given to the clustering of the convolution and max-pooling layers output nodes hooked on a 2D plane. The development of every plane of the layer is achieved through the amalgamation of one or added planes of preceding layers. A plane's nodes are connected to a tiny portion of every connected plane in the layer above. Every convolutional layer node has its features obtained from the input images using convolution processes happening in the input nodes.

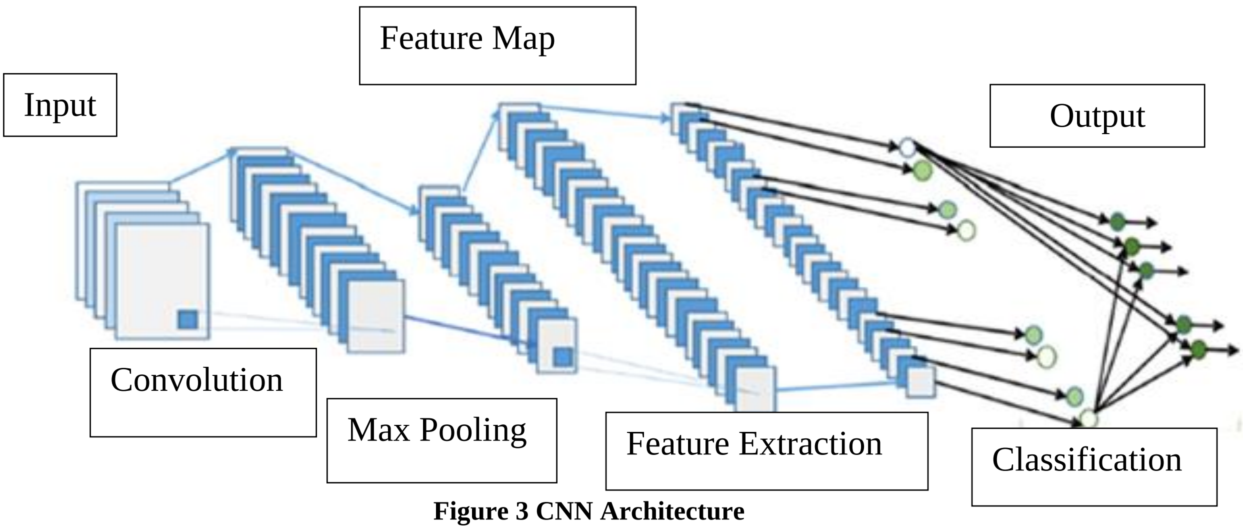

Figure 3 CNN Architecture

Figure 3 shows the structural framework of the CNN comprises of input layer, multiple alternating convolution and max-pooling layers, one fully connected layer, and one classification layer. The features are propagated from lower-level layers derive higher-level features. Depending on the size of the kernel, the dimensions of features are being reduced for the convolutional and max-pooling operations in turn when the highest layer or level, features are propagated. However, to ensure classification accuracy improved features of the input images are represented, the number of feature maps is typically improved. The fully connected network is also known as the classification layer. The result of the former layer of the CNN is taken as the input for this fully connected network. Because of the better performance, the classification layer has been implemented using feed forward network. Concerning the final neural network's weight matrix dimension is, the extracted features are taken as inputs in the classification layer. Alternatively, in terms of network or learning parameters, the fully connected layers are costly. Average pooling and global average pooling are some of the new techniques emerged in recent years. On the other hand, these techniques can be used to fully connected networks. It uses a soft-max layer for the calculation of the score of the respective class in the uppermost classification layer. The classifier affords result for the corresponding classes, depend upon the highest score. The succeeding sector explores the mathematical details of different layers of CNNs.

Convolutional Layer

Within this layer, learnable kernels are used to convolve the feature maps from earlier layers. The output kernel passes over a linear or non-linear activation function to create the result feature maps. Multiple input feature maps can be amalgamated using any one of the output characteristics maps. Generally speaking, it can be stated as:

$$x_j^l = f\left(\sum_{i\epsilon M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \text{----> (1)}$$

Where x_j^l denotes the current layer output, x_i^{l-1} denotes the outcome of the earlier layer, k_{ij}^l denotes the current layer kernel, and b_j^l denotes the current layer biases. M_j signifies a input maps selection. There is an additive bias b assigned to every result map. The input maps will be convolved using different kernels, to create the matching output maps. Ultimately, a non-linear or linear activation function is smeared to the output maps.

Sub-sampling Layer

The downsampled process is carried out on the input maps by the Sub-sampling layer, also referred to as the pooling layer. Within this layer, the constant terms are the quantity of input and output feature maps. Namely, exactly N output maps will exist if it has N input maps. Reliant on the size of the down-sampling mask, the down-sampling operation will result in an abridgment the output maps each dimension. For example, all the images' output dimensions will be half of their respective input dimensions if a 2 × 2 downsampling kernel is applied. This procedure can be written as

$$X_j^l = down\left(X_j^{l-1}\right) \qquad \text{-----> (2)}$$

where a subsampling function is indicated by down (.). In this layer, two types of operations are typically carried out: max-pooling or average pooling. If the average pooling method is used, the function typically chooses the average value by adding up all N × N patches of the feature maps from the former layer. Alternatively, in the event of max-pooling, identified the maximum value from the N×N patches of the feature maps. The result map's scopes are subsequently decreased by a factor of n. In certain exceptional cases, a scalar multiplies each output map. Other subsampling layers, for example the fractional max-pooling layer and the subsampling with convolution layer, have been proposed.

Classification Layer

One of the fully connected layers utilizes the features that were fetched from a convolutional layer in the earlier steps to determine the score of respective class. The features of the last layer are characterized as vectors that have scalar quantities and are provided to the fully connected layers. We employ the fully connected feed-forward neural layers as a soft-max classification layer. Regarding the amount of merged layers in the network model, there are no fast and hard rules. However, two to four layers are typically found in various architectures for example LeNet, Alex Net, and VGG Net. Other methods have been suggested in recent years due to the high computational cost of fully connected layers. These also adds the global average pooling layer and the average pooling layer, which greatly aid in lowering the total number of network parameters. The fully linked layer informs in the backward propagation through the CNNs, adhering to the general procedure of fully connected neural networks. By applying the complete convolutional process to the features spanning the convolutional layer and its immediately preceding layer, we update the filters of the convolutional layer.

Parameters of the network and Memory requirement for CNN

To gauge a deep learning model's complexity, one crucial metric is its number of computational parameters. One way to calculate the output feature map size is as follows:

$$M = \frac{(N-F)}{S} + 1 \qquad \text{-----> (3)}$$

here N is the amount of input features, F is the amount of filters, M is the amount of output feature map dimensions, and S is the number of stride lengths. Typically, padding is used during convolution processes to protect the identically sized I/O feature maps. The size of the kernel determines how much padding is used.

$$P = (F - 1)/2, \text{ -----> (4)}$$

In this case, P represents the amount of padding, and F the kernels' dimensions. The models are compared based on several criteria. However, majority of time, both the entire quantity of memory and the amount of network factors are carefully considered. Using the following formula, the number of parameters ($Parm_l$) of the l^{th} layer is determined:

$$Parm_l = (F \times F \times FM_{l-1}) \times FM_l. \quad \text{-----> (5)}$$

After adding bias with the weights, the above equation will be

$$Parm_l = (F \times (F + 1) \times FM_{l-1}) \times FM_l \text{ ---> (6)}$$

here FM_l denotes the quantity of result feature maps, and FM_{l-1} denotes the quantity of input feature maps.

LSTM Model

One classification of artificial neural network is a recurrent neural network. To keep the network graph in a stable internal state, they augment more heaviness to the network and produce phases. Weightiness which are back and forward spread over layers can be preserved. A series of repeating neural network modules is the structure of every recurrent neural network. This reiterating section in traditional RNNs will have a simple arrangement, similar to sole tanh layer. LSTMs also exhibit these type of chain structure, even though we structure the repeating module differently. There are four layers in a neural network, each of which interacts uniquely. Neural network layer is not a single. Each line in the figure represents the complete vector, linking ouput of one node to the other nodes. The yellow boxes are denoted by the trained neural network layers, and pointwise operations, like vector addition, are denoted by the pink circles. Lines that fuse specify concatenation, but lines that bifurcate indicate that their content is derivative and goes to different places. By keeping an error rate higher over many time steps, the network can keep learning. Long-term dependencies can thus be learned by the network. LSTM networks utilize gates and memory cell to report the vanishing gradient disputes. Rather than biology, circuitry is the main source of inspiration for these. Three gates namely input, output, and forget as well as one memory cell are present in every neuron. By preventing or permitting the flow of information, these gates serve to protect it. The input gate controls the volume of data that is stored in the cell from the preceding layer. On the other hand, the output layer assumes responsibility for deciding how plentiful info about this cell's status is passed on to the subsequent layer. The forget gate regulates how long a value stays in the cells as a result of future operations, which helps forget some previous values.

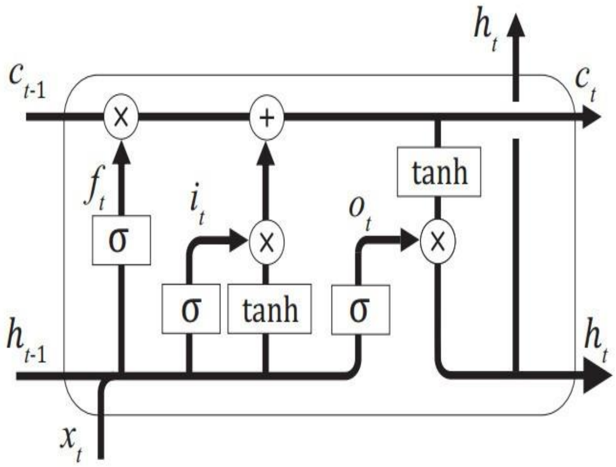

Figure 4: The structure of the LSTM cell

Multiple input variable problems can be modelled by LSTM with near-perfect ease. This is especially helpful in time series forecasting, as it might present challenges to adapt classical linear methods to multivariate estimating difficulties. One concealed LSTM layer then a typical feedforward result layer make up the original LSTM model. This model is extended by the Stacked LSTM. Each of its several hidden LSTM layers has several memory cells. The several LSTM layers from the LSTM model is identified as a stacked LSTM construction. Instead of directing a single value output to the lower LSTM layer, the upper LSTM layer drives a sequence of values. Because stacked LSTM is a consistent process for tough sequence prediction problems, the model's accuracy increases. We fit the data in this literature using a three-layer arranged LSTM.

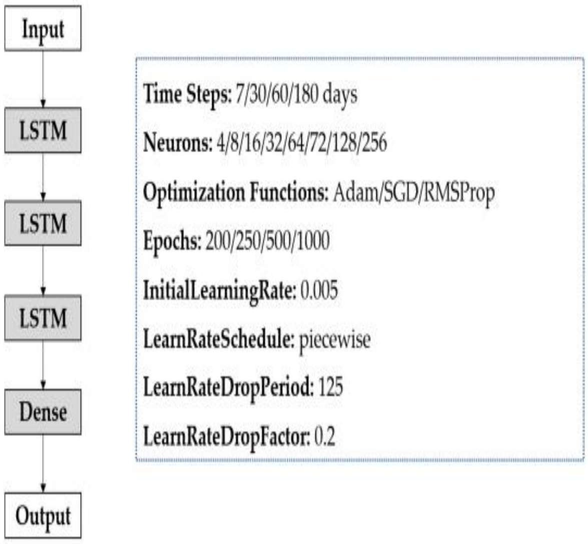

Figure 5: Stacked LSTM architecture

The LSTM model's training and prediction can be separated into the following three stages. Since the LSTM approaches were profound to the gauge of the input information, the information was first rescaled and regulated to the sort of 0 to 1. Second, the time steps for the univariate and multivariate LSTM were fixed to 7/30/60/180. This implies that we predicted the incidence of the following day using the data from the preceding 7/30/60/180 days. At last, a three-layer arranged LSTM structure was created. There is one hidden layer in each LSTM layer, which was configured for the LSTM approach and has neurons with choices of 4/8/16/32/64/72/128/256. Adam, SGD and RMSProp are the alternative optimization functions. These learning procedures were conducted over 200, 250, 500, and 1000 epochs. We gave the typical instructions to descent the learning rate each 125 epochs are multiplied by 0.2, with a preliminary learning rate of 0.005. Based on the aforementioned findings, we selected the best model by minimizing the test set's RMSE.

IV. Result and Discussion

The crime rate was predicted using the CNN and LSTM models. The simulation and validation of the outcomes are done with the Python programming language. After calculation, the performance metrics Accuracy, MSE, and RMSE are contrasted with those of other machine learning algorithms. The raw data set includes all available information on India's crime rate from 2001 to 2013, including details on crimes like robbery, theft, rape, and murder. Figure 6 displays the data set cases below.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9840 entries, 0 to 9839
Data columns (total 33 columns):
 #   Column                                                Non-Null Count  Dtype
---  ------                                                --------------  -----
 0   STATE/UT                                              9840 non-null   object
 1   DISTRICT                                              9840 non-null   object
 2   YEAR                                                  9840 non-null   int64
 3   MURDER                                                9840 non-null   int64
 4   ATTEMPT TO MURDER                                     9840 non-null   int64
 5   CULPABLE HOMICIDE NOT AMOUNTING TO MURDER             9840 non-null   int64
 6   RAPE                                                  9840 non-null   int64
 7   CUSTODIAL RAPE                                        9840 non-null   int64
 8   OTHER RAPE                                            9840 non-null   int64
 9   KIDNAPPING & ABDUCTION                                9840 non-null   int64
 10  KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS           9840 non-null   int64
 11  KIDNAPPING AND ABDUCTION OF OTHERS                    9840 non-null   int64
 12  DACOITY                                               9840 non-null   int64
 13  PREPARATION AND ASSEMBLY FOR DACOITY                  9840 non-null   int64
 14  ROBBERY                                               9840 non-null   int64
 15  BURGLARY                                              9840 non-null   int64
 16  THEFT                                                 9840 non-null   int64
 17  AUTO THEFT                                            9840 non-null   int64
 18  OTHER THEFT                                           9840 non-null   int64
 19  RIOTS                                                 9840 non-null   int64
 20  CRIMINAL BREACH OF TRUST                              9840 non-null   int64
 21  CHEATING                                              9840 non-null   int64
 22  COUNTERFIETING                                        9840 non-null   int64
 23  ARSON                                                 9840 non-null   int64
 24  HURT/GREVIOUS HURT                                    9840 non-null   int64
 25  DOWRY DEATHS                                          9840 non-null   int64
 26  ASSAULT ON WOMEN WITH INTENT TO OUTRAGE HER MODESTY   9840 non-null   int64
 27  INSULT TO MODESTY OF WOMEN                            9840 non-null   int64
 28  CRUELTY BY HUSBAND OR HIS RELATIVES                   9840 non-null   int64
 29  IMPORTATION OF GIRLS FROM FOREIGN COUNTRIES           9840 non-null   int64
 30  CAUSING DEATH BY NEGLIGENCE                           9840 non-null   int64
 31  OTHER IPC CRIMES                                      9840 non-null   int64
 32  TOTAL IPC CRIMES                                      9840 non-null   int64
dtypes: int64(31), object(2)
memory usage: 2.5+ MB
Ln: 5  Col: 0
```

Figure 6: Input Data Formation

A prompt input (state) from the user trains the LSTM model undergoes testing using this method. The skilled LSTM and CNN models receive prompt data from the user, which they use to predict future crime rates. The input parameters for the crime prediction are displayed in the figure 6.

```
Enter the name of the state in India, that you want to predict the cases in future : TAMIL NADU
```

Figure 7: State Wise Prediction

```
please enter which crime you want to predict for TAMIL NADU : MURDER
```

Figure 8: Crime Wise Prediction

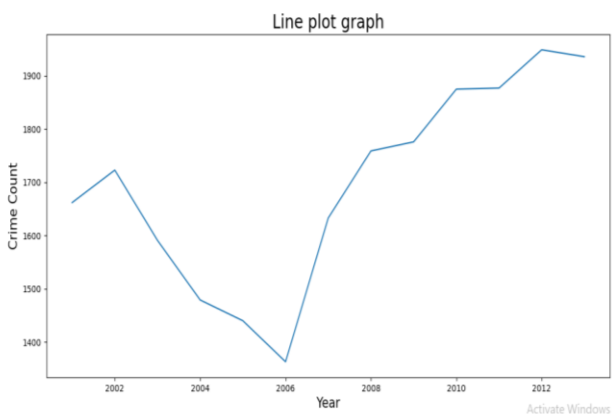

Figure. 9: Crime Count vs Year (Existing data)

The output of the line plot of the crime count versus year for the available data from 2002 to 2013 is shown in Figure 9 above. The plot explains the previous year's data set details (2002 to 2013).

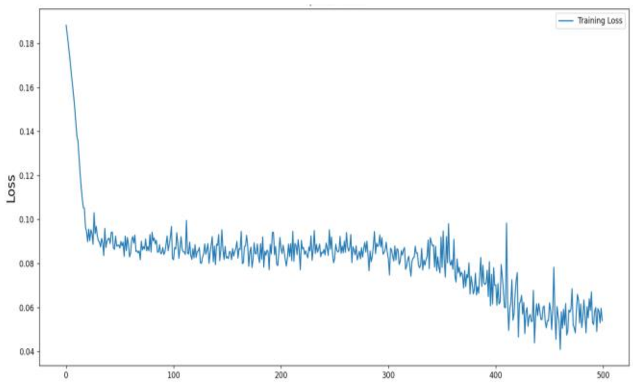

Figure 10 CNN Training result

The CNN model's training loss for the relevant data set is shown in Figure 10. For different iterations, the model's loss function obtained a minimum error value.

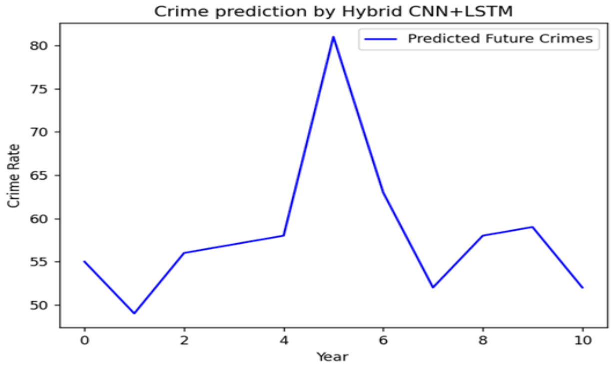

Figure 11: Prediction result using Hybrid CNN+LSTM Model

Figure 11 uses a hybrid CNN and LSTM model to predict data from 2014 to 2025. The CNN model with individual processing only uses linear data for prediction. CNN is used in this work to extract features.

Table 1 Hyper parameters of CNN and LSTM	
Optimization Function	Adam
Momentum	0.9
LSTM Dropout rate	0.25
Initial learning rate	0.00001
Schedule of learning rate	Piecewise
LSTM Activation function	Sigmoid
Loss function	Categorical Cross-Entropy

Hyper parameters utilized in the network is displayed in table 1. The LSTM training result above can be used to calculate the loss and accuracy. For the validation data set, the suggested LSTM models exhibit loss values of 0.0348 and an accuracy of 97.8%.

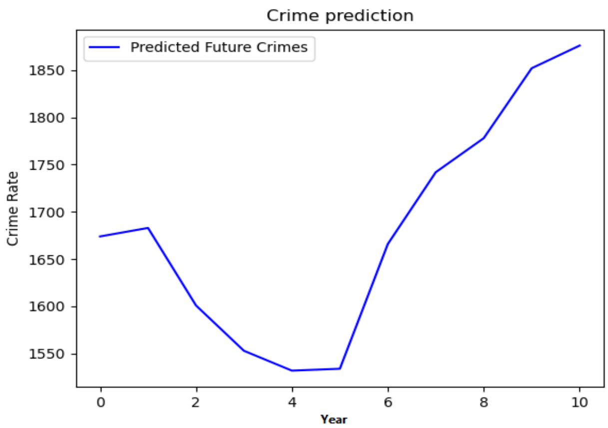

Figure 12: Future 10 years Predicted value by LSTM

Figure 12 uses the LSTM approach to envisage the crime rates for the next ten years.

It is an example of CNN-LSTM trained for the anticipated crime rate in the upcoming year. To obtain a higher accuracy, the LSTM processes the CNN's residual error.

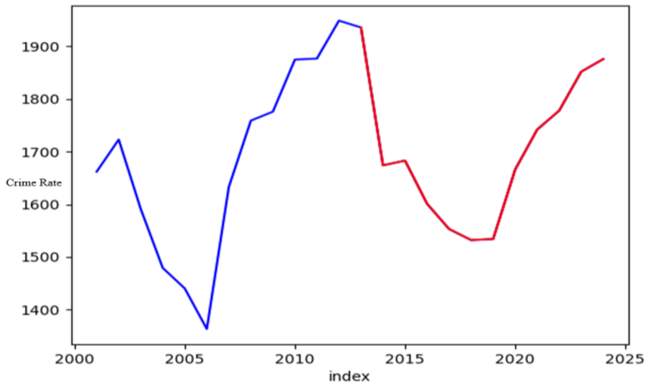

Figure13: CNN-LSTM Predicted Result

The predicted value for the future is :

	Prediction
2014	1674.0
2015	1683.0
2016	1601.0
2017	1553.0
2018	1532.0
2019	1534.0
2020	1666.0
2021	1742.0
2022	1778.0
2023	1852.0
2024	1876.0

Figure 14: Future Predicted Results

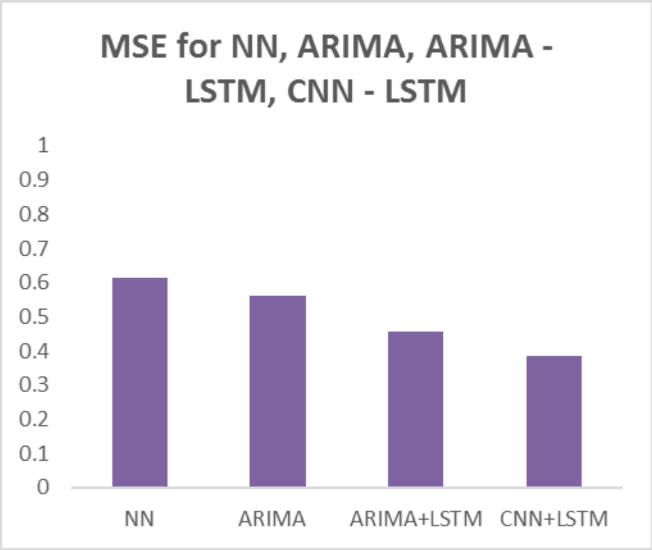

Figure 15: MSE for NN, ARIMA, ARIMA - LSTM, CNN – LSTM

The average of the squared forecast error values yields the mean squared error (MSE). By constraining the forecast error significances to be positive, squaring them too increases the significance of the greater errors. In comparison to NN, ARIMA, and LSTM, the CNN and LSTM models have a lower mean squared error, as seen in Figure 15.

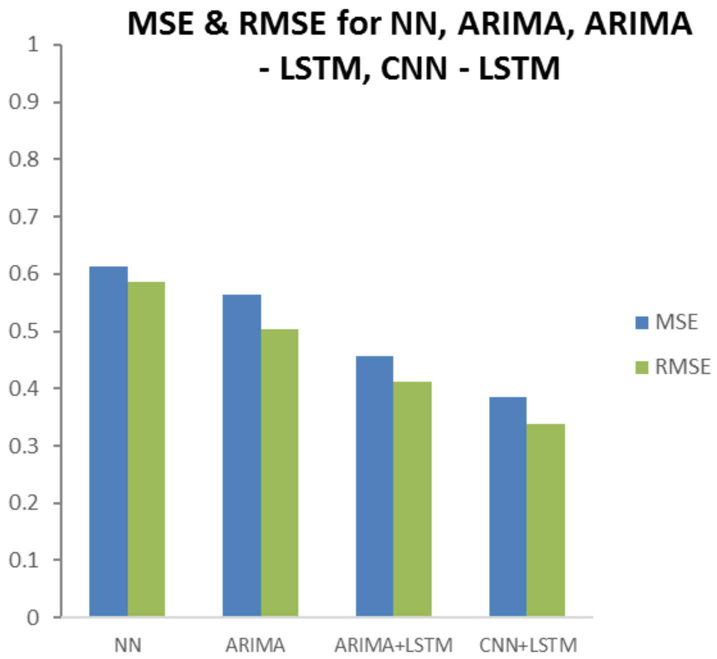

Figure 16: Comparative Analysis of MSE & RMSE for NN, ARIMA, ARIMA

LSTM, CNN-LSTM

The estimates accuracy is computed from root mean squared error (RMSE) and MSE. The RMSE and MSE calculate an estimator's average deviation from the true parameter. The estimate and the absolute criteria are on the same scale. Whereas RMSE is in the estimate scale, MSE is in the squared deviation scale. The performance analysis of MSE and RMSE about the other algorithms is shown in Figure 16 above. Classification accuracy is the ratio of the accurate estimations to the total input trials. The percentage of accurate predictions made using the test data is called as accuracy. To compute divide the total number of predictions by the number of accurate predictions. Figure 17, which compares CNN and LSTM models to other machine learning algorithms, demonstrates the high accuracy of these models.

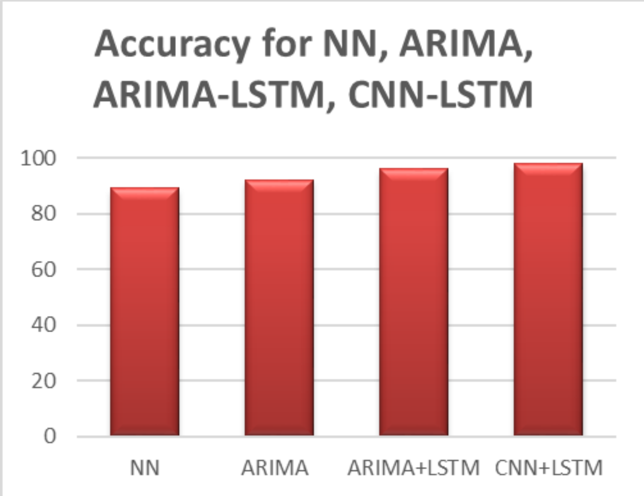

Figure 17: Comparative Analysis of Accuracy for NN, ARIMA, ARIMA-LSTM, CNN-LSTM

Table 2: Comparison of performances of Proposed and Existing Methods

Performance Metrics	Proposed and Existing Methods			
	NN	**ARIMA**	**ARIMA-LSTM**	**CNN-LSTM**
Accuracy (%)	89	92	96	97.8
MSE	0.612	0.563	0.456	0.384
RMSE	0.586	0.504	0.411	0.339

Table 2 depicts the comparison outcomes related to the enactment of other machine learning approaches. The comparison of results shows that CNN-LSTM has high accuracy compared to other existing methods.

Among the methods, CNN-LSTM stands out with the highest accuracy of 97.8%, surpassing ARIMA, ARIMA-LSTM, and NN. CNN-LSTM also exhibits superior predictive performance, boasting the lowest MSE (0.384) and RMSE (0.339) values. ARIMA-LSTM follows closely behind CNN-LSTM, with the second-highest accuracy of 96%, and the second-lowest MSE (0.456) and RMSE (0.411). ARIMA performs moderately, with an accuracy of 92%, while ARIMA-LSTM and CNN-LSTM outshine it in both MSE and RMSE metrics. NN has the lowest accuracy at 89% and the highest MSE and RMSE among the methods, suggesting it is the least accurate for this task. The results emphasize the effectiveness of combining convolutional and recurrent neural networks (CNN-LSTM) for the given prediction task. Lower MSE and RMSE values of CNN-LSTM indicate its ability to minimize prediction errors and provide more accurate forecasts. The proposed NN method lags behind in accuracy and precision compared to the time-series models and hybrid models (ARIMA, ARIMA-LSTM, CNN-LSTM). Choosing the appropriate model should consider both accuracy and precision metrics, and CNN-LSTM appears to be the most promising method based on the provided results. It's crucial to deliberate the precise characteristics of the data and potential trade-offs among accuracy and computational complexity when selecting the best model for practical applications.

V. Conclusion

CNN and LSTM, two deep neural network models, were consumed in this article to forecast the crime rate. The modeling complexity notwithstanding, the outcomes were promising. The findings showed that in terms of the trend and variation of the crime rate, the predictions made by the CNN and LSTM models outperformed those made by the traditional model. The strength of this suggested model is its ability to outdo the prior methods in terms of accuracy. Furthermore, using multiple LSTM layers reduces the training time of the CNN and LSTM models. It has been demonstrated that the suggested network is an advanced technique for modeling, analyzing, and forecasting crime rates and yielded an accuracy of 97.8%. In the future, optimization algorithms will be used to adjust the hyperparameters of the CNN and LSTM models to accurately classify crime rates.

Conflict of Interest

The authors state no conflict of interest

COMPETING INTERESTS

Not Applicable

FUNDING INFORMATION

Natural Sciences and Engineering Research Council of Canada (NSERC) and New Brunswick Innovation Foundation (NBIF) for the financial support of the global project.

Author Contributions:

CONCEPTUALIZATION, K.RAMALAKSHMI

DATA CURATION, L.KRISHNA KUMARI AND SAMEER ALGBURI.;

FORMAL ANALYSIS, SATHEESHKUMAR PALANISAMY AND HABIB HAMAM.;

FUNDING ACQUISITION, HABIB HAMAM.;

INVESTIGATION, K.RAMALAKSHMI

PROJECT ADMINISTRATION, SATHEESHKUMAR PALANISAMY AND KARTHIKEYAN GANESAN.;

RESOURCES, L.KRISHNA KUMARI, OSAMAH IBRAHIM KHALAF. AND KARTHIKEYAN GANESAN.;

SOFTWARE, SAMEER ALGBURI AND HABIB HAMAM;

VALIDATION, K.RAMALAKSHMI, OSAMAH IBRAHIM KHALAF. AND SATHEESHKUMAR PALANISAMY;

VISUALIZATION, SATHEESHKUMAR PALANISAMY;

WRITING – REVIEW & EDITING, OSAMAH IBRAHIM KHALAF.

DATA AVAILABILITY STATEMENT

Not Available

RESEARCH INVOLVING HUMAN AND /OR ANIMALS

Not Available

INFORMED CONSENT

Not Available

5. Acknowledgement

The authors thank Natural Sciences and Engineering Research Council of Canada (NSERC) and New Brunswick Innovation Foundation (NBIF) for the financial support of the global project. These granting agencies did not contribute in the design of the study and collection, analysis, and interpretation of data.

REFERENCES

1. A. Kumar, A. Verma, G. Shinde, Y. Sukhdeve and N. Lal, (2020) "Crime Prediction Using K-Nearest Neighboring Algorithm," International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-4, doi: 10.1109/ic-ETITE47903.2020.155.
2. B. Sivanagaleela and S. Rajesh, (2019) "Crime Analysis and Prediction Using Fuzzy C-Means Algorithm," 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 595-599, doi: 10.1109/ICOEI.2019.8862691.
3. C. Catlett, E. Cesario, D. Talia and A. Vinci, (2018) "A Data-Driven Approach for [1 Spatio-Temporal Crime Predictions in Smart Cities," IEEE International Conference on Smart Computing (SMARTCOMP), 2018, pp. 17-24, doi: 10.1109/SMARTCOMP.2018.00069.
4. N. Joshi, S. Shaikh, A. Tripathy, S. Sen, J. Janson, S. Karnik, B. Verghese, (2019) "Crime Anatomization Using QGIS," IEEE 5th International Conference for Convergence in Technology (I2CT), 2019, pp. 1-4, doi: 10.1109/I2CT45611.2019.9033910.
5. Pratibha, A. Gahalot, Uprant, S. Dhiman and L. Chouhan, (2020) "Crime Prediction and Analysis," 2nd International Conference on Data, Engineering and Applications (IDEA), 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170731.
6. S. Agarwal, L. Yadav and M. K. Thakur, (2018) "Crime Prediction Based on Statistical Models," Eleventh International Conference on Contemporary Computing (IC3), 2018, pp. 1-3, doi: 10.1109/IC3.2018.8530548.
7. S. P. C. W. Sandagiri, B. T. G. S. Kumara and B. Kuhaneswaran, (2020) "ANN Based Crime Detection and Prediction using Twitter Posts and Weather Data," International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 2020, pp. 1-5, doi: 10.1109/ICDABI51230.2020.9325660.
8. S. S. Kshatri, D. Singh, B. Narain, S. Bhatia, M. T. Quasim and G. R. Sinha, (2021) "An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach," in IEEE Access, vol. 9, pp. 67488-67500, 2021, doi: 10.1109/ACCESS.2021.3075140.
9. S. Tarlekar, R. Bhosle, E. D'souza and S. Sheikh, (2021)"Geographical Crime Rate Prediction System," IEEE India Council International Subsections Conference (INDISCON), 2021, pp. 1-6, doi: 10.1109/INDISCON53343.2021.9582218.
10. S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma and N. Yadav, (2017) "Crime pattern detection, analysis & prediction," International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017, pp. 225-230, doi: 10.1109/ICECA.2017.8203676.
11. W. Safat, S. Asghar and S. A. Gillani, (2021) "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," in IEEE Access, vol. 9, pp. 70080-70094, 2021, doi: 10.1109/ACCESS.2021.3078117.

12. X. Han, X. Hu, H. Wu, B. Shen and J. Wu, (2020) “Risk Prediction of Theft Crimes in Urban Communities: An Integrated Model of LSTM and ST-GCN,” in IEEE Access, vol. 8, pp. 217222-217230, 2020, doi: 10.1109/ACCESS.2020.3041924.
13. S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory” Neural Computation, vol 9, no. 8, pp 1735-1780, 1997, doi:10.1162/neco.1997.9.8.1735.
14. Palanisamy, S., Rubini, S.S., Khalaf, O.I. et al. Multi-objective hybrid split-ring resonator and electromagnetic bandgap structure-based fractal antennas using hybrid metaheuristic framework for wireless applications. Sci Rep 14, 3288 (2024). https://doi.org/10.1038/s41598-024-53443-z
15. Iqbal Jebril & P. Dhanaraj & Ghaida Muttashar Abdulsahib & SatheeshKumar Palanisamy & T.Prabhu & Osamah Ibrahim Khalaf, 2022. "Analysis of Electrically Couple SRR EBG Structure for Sub 6 GHz Wireless Applications," Advances in Decision Sciences, Asia University, Taiwan, vol. 26(Special), pages 102-123, December.
16. Palanisamy, S, Thangaraju, B. Design and analysis of clover leaf-shaped fractal antenna integrated with stepped impedance resonator for wireless applications. Int J Commun Syst. 2022; 35(11):e5184. doi:10.1002/dac.5184.
17. E. Suganya, T. Prabhu, Satheeshkumar Palanisamy, Praveen Kumar Malik, Naveen Bilandi, Anita Gehlot, "An Isolation Improvement for Closely Spaced MIMO Antenna Using λ/4 Distance for WLAN Applications", International Journal of Antennas and Propagation, vol. 2023, Article ID 4839134, 13 pages, 2023. https://doi.org/10.1155/2023/4839134.
18. Palanisamy, Satheeshkumar, BalakumaranThangaraju, Osamah Ibrahim Khalaf, YouseefAlotaibi, SalehAlghamdi, and FawazAlassery. 2021. "A Novel Approach of Design and Analysis of a Hexagonal Fractal Antenna Array (HFAA) for Next-Generation Wireless Communication" Energies 14, no. 19: 6204. https://doi.org/10.3390/en14196204.
19. Palanisamy S, Thangaraju B, Khalaf OI, Alotaibi Y, Alghamdi S. Design and Synthesis of Multi-Mode Bandpass Filter for Wireless Applications. Electronics. 2021; 10(22):2853. https://doi.org/10.3390/electronics10222853.
20. SatheeshKumar & Balakumaran T. (2021). Modeling and simulation of dual layered U-slot multiband microstrip patch antenna for wireless applications. Nanoscale Reports, 4(1), 15 – 18. https://doi.org/10.26524/nr.4.3.
21. Ghaida Muttashar Abdulsahib, Dhana Sekaran Selvaraj, A. Manikandan, SatheeshKumar Palanisamy, Mueen Uddin, Osamah Ibrahim Khalaf, Maha Abdelhaq, Raed Alsaqour, Reverse polarity optical Orthogonal frequency Division Multiplexing for High-Speed visible light communications system, Egyptian Informatics Journal, Volume 24, Issue 4, 2023, 100407, ISSN 1110-8665, https://doi.org/10.1016/j.eij.2023.100407.
22. Palanisamy, S., Nivethitha, T., Alhameed, M.R., Udhayakumar, A., Hussien, N.A. (2023). Urban Wastewater Treatment for High Yielding in Agriculture Through Smart Irrigation System. In: Swaroop, A., Kansal, V., Fortino, G., Hassanien, A.E. (eds) Proceedings of Fourth Doctoral Symposium on Computational Intelligence . DoSCI 2023. Lecture Notes in Networks and Systems, vol 726. Springer, Singapore. https://doi.org/10.1007/978-981-99-3716-5_52
23. Shivappriya, S.N., Palanisamy, S., Mohammed Shareef, A., Ali Zearah, S. (2023). Significance of Literacy in Minimizing Infant Mortality and Maternal Anemia in India: A State-Wise Analysis. In: Swaroop, A., Kansal, V., Fortino, G., Hassanien, A.E. (eds) Proceedings of Fourth Doctoral Symposium on Computational Intelligence . DoSCI 2023. Lecture Notes in Networks and Systems, vol 726. Springer, Singapore. https://doi.org/10.1007/978-981-99-3716-5_73
24. Xue, X.; Shanmugam, R.; Palanisamy, S.; Khalaf, O.I.; Selvaraj, D.; Abdulsahib, G.M.A Hybrid Cross Layer with Harris-Hawk-Optimization-Based Efficient Routing for Wireless Sensor Networks. Symmetry 2023, 15, 438.https://doi.org/10.3390/sym15020438.
25. A. P. Mukti, L. Lusiana, D. Titisari, and S. Palanisamy, “Performance Analysis of Twelve Lead ECG Based on Delivery Distance Using Bluetooth Communication”, j.electron.electromedical.eng.med.inform, vol. 5, no. 1, pp. 46-52, Jan. 2023.
26. AngurajKandasamy, SaravanakumarRengarasu, Praveen KittiBurri, SatheeshkumarPalanisamy, K. Kavin Kumar, Aruna Devi Baladhandapani, Samson AlemayehuMamo, "Defected Circular-Cross Stub Copper Metal Printed Pentaband Antenna", Advances in Materials Science and Engineering, vol. 2022, Article ID 6009092, 10 pages, 2022. https://doi.org/10.1155/2022/6009092.
27. Sam P.J.C., Surendar U., Ekpe U.M., Saravanan M., Satheesh Kumar P. (2022) A Low-Profile Compact EBG Integrated Circular Monopole Antenna for Wearable Medical Application. In: Malik P.K., Lu J., Madhav B.T.P., Kalkhambkar G., Amit S. (eds) Smart Antennas. EAI/Springer Innovations in Communication and Computing.Springer, Cham.https://doi.org/10.1007/978-3-030-76636-8_23.
28. Satheesh Kumar, P., Jeevitha, Manikandan (2021). Diagnosing COVID-19 Virus in the Cardiovascular System Using ANN. In: Oliva, D., Hassan, S.A., Mohamed, A. (eds) Artificial Intelligence for COVID-19. Studies in Systems, Decision and Control, vol 358.Springer, Cham.https://doi.org/10.1007/978-3-030-69744-0_5.
29. S, D.; Palanisamy, S.; Hajjej, F.; Khalaf, O.I.; Abdulsahib, G.M.; S, R. Discrete Fourier Transform with Denoise Model Based Least Square Wiener Channel Estimator for Channel Estimation in MIMO-OFDM. Entropy 2022, 24, 1601. https:// doi.org/10.3390/e24111601.
30. Nivethitha, T., Palanisamy, S. K., MohanaPrakash, K., &Jeevitha, K. (2021).Comparative study of ANN and fuzzy classifier for forecasting electrical activity of heart to diagnose Covid-19. Materials today. Proceedings, 45, 2293–2305. https://doi.org/10.1016/j.matpr.2020.10.400.
31. Palanisamy, S. (2022).Predictive Analytics with Data Visualization. Journal of Ubiquitous Computing and Communication Technologies, 4(2), 75-96. doi:10.36548/jucct.2022.2.003.
32. Prabhu, T., et al. "An Intensive Study of Dual Patch Antennas with Improved Isolation for 5G Mobile Communication Systems." In Future Trends in 5G and 6G: Challenges, Architecture, and Applications; CRC Press: Boca Raton, FL, USA, 2021.

33. Kumar, P.S.; Chitra, P.; Sneha, S. Design of Improved Quadruple-Mode Bandpass Filter Using Cavity Resonator for 5G Mid-Band Applications. In Future Trends in 5G and 6G: Challenges, Architecture, and Applications; CRC Press: Boca Raton, FL, USA, 2021, pp. 219–234.
34. P. Satheesh Kumar and S. Valarmathy, "Development of a novel algorithm for SVMBDT fingerprint classifier based on clustering approach," IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012), 2012, pp. 256-261.
35. K. Suganyadevi, V. Nandhalal, S. Palanisamy and S. Dhanasekaran, "Data Security and Safety Services using Modified Timed Efficient Stream Loss-Tolerant Authentication in Diverse Models of VANET," 2022 International Conference on Edge Computing and Applications (ICECAA), 2022, pp. 417-422, doi: 10.1109/ICECAA55415.2022.9936128.
36. P. S. Kumar, S. Boopathy, S. Dhanasekaran and K. R. G. Anand, "Optimization of Multi-Band Antenna for Wireless Communication Systems using Genetic Algorithm," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-6, doi: 10.1109/ICAECA52838.2021.9675686.
37. C. Chauhan and S. Sehgal, "A review: Crime analysis using data mining techniques and algorithms," *2017 International Conference on Computing, Communication and Automation (ICCCA),* 2017, pp. 21-25, doi: 10.1109/CCAA.2017.8229823.
38. S. Ghankutkar, N. Sarkar, P. Gajbhiye, S. Yadav, D. Kalbande and N. Bakereywala, "Modelling Machine Learning for Analysing Crime News," *2019 International Conference on Advances in Computing, Communication and Control (ICAC3),* 2019, pp. 1-5, doi: 10.1109/ICAC347590.2019.9036769.
39. D. M. Raza and D. B. Victor, "Data mining and Region Prediction Based on Crime Using Random Forest," *2021 Int. Conf. on Artificial Intelligence and Smart Systems* (*ICAIS*), 2021, pp. 980-987, doi: 10.1109/ICAIS50930.2021.9395989.
40. A. B. Asad, R. Mansur, S. Zawad, N. Evan and M. I. Hossain, "Analysis of Malware Prediction Based on Infection Rate Using Machine Learning Techniques," *2020 IEEE Region 10 Symposium (TENSYMP),* Jun 2020, pp. 706-709, doi: 10.1109/TENSYMP50017.2020.9230624.
41. N. Baloian, E. Bassaletti, M. Fernandez, O. Figueroa, P. Fuentes, R. Manasevich, M. Orchard, S. Penafiel, J. A. Pino, M. Vergara, "Crime prediction using patterns and context," *Proceedings of the 2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD),* 2017, pp. 2-9, doi: 10.1109/CSCWD.2017.8066662.