SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I2P109

Volume 11 Issue 2, 80-91, February 2024

Original Article

Prediction of Crime Rate in Diverse Environs Using Hybrid Classifier

S. Santhosh[1], N. Sugitha[2]

[1]*Department of Information Technology, Noorul Islam Centre for Higher Education, Tamilnadu, India.*
[1]*Scientist-E/Joint Director, Communications and Software Technology Group,*
Centre for Development of Advanced Computing (C-DAC), Kerala, India.
[2]*Department of Electronics and Communication Engineering, Saveetha Engineering College, Tamil Nadu, India.*

[1]*Corresponding Author : santhosh.vattappara@gmail.com*

Received: 01 November 2023 Revised: 05 December 2023 Accepted: 02 January 2024 Published: 06 February 2024

Abstract - *Crime is the fear and terror among the populace worldwide. Crime is an inherent component of the hazards we encounter daily. In recent times, the mass media has extensively covered numerous criminal incidents, including theft, rape and sexual offenses, robbery, murder, and kidnappings. Various works have been produced to understand the factors that lead to an individual committing a criminal act, the potential dangers involved, and strategies to prevent it. The crime computation technique aims to forecast crime rates, enabling police officers to avoid such incidents effectively. Based on this, a novel prediction approach utilizing a hybrid classifier is suggested. An evaluation of the proposed method was conducted using several criteria. The performance of this recently constructed hybrid prediction model is evaluated by comparing it with established models such as Genetic Algorithm, Particle Swarm Optimization, and Firefly Algorithm. Various performance measures, including error rate, sensitivity, specificity, precision, and execution time, are used for the comparison. Based on the results, this hybrid model is the most optimal crime prediction model compared to the other current models. The suggested approach is executed using the JAVA programming language.*

Keywords - *Artificial Neural Network, Crime prediction, Hybrid classifier, Feature selection, Support Vector Machine, Preprocessing, Prediction.*

1. Introduction

Historically, resolving crimes has been the privilege of the illegal equity and law requirement subject matter experts. With the increased utilization of automated systems to track fraud and trace convicts, PC information professionals have begun loaning thumbs in assisting law execution officials, including investigators, to quicken the road toward addressing corruption. Criminology is a tool used to identify misconduct and criminal characteristics. Criminology techniques can be used to assess both the likelihood of wrongdoing events and criminals. The police department, analyst organizations, and wrongdoing divisions can recognize the actual characteristics of an illegal with the aid of criminology [1].

Everywhere in society, there was corruption and ferocity. Well-prepared cultures have used many different crime prevention methods [2]. Understanding the crime characteristics was crucial in this framework to implement effective regulations against these criminal behaviours. Numerous factors have been linked in studies to crime, including psychological traits [3], ecological conditions [4], spatial patterns [6], and social and economic indicators [8].

However, it was simple to find debatable explanations for the origins of crimes [9]. A few issues that can result in inaccurate results include procedural problems with data collection and selection [10], inaccuracies related to data reporting [11], and incorrect statistical hypotheses [9].

Conceptualizations of criminal behaviour and crime have evolved throughout history. The “classic criminal theories” focus on individuals’ developmental experiences, biological factors, and societal influences in shaping their propensity for criminal behaviour. Alternatively, the “environmental theories” attribute corruption to the actions of criminals, their victims, and specific laws and administrations at particular eras and locales [12].

In this instance, the illicit behavior was the object of intrigue instead of the actual wrongdoers. Crime is a pervasive social problem that has a detrimental impact on the quality of life and financial development of a society. This factor is a crucial driver in determining whether individuals travel to different places and identifying spots to elude while drifting [13].

Law enforcement agencies are urging enhanced geographic statistics systems and innovative tactics to boost criminal research besides superior defense to communities in response to increasing crime rates [1].

Crimes can be anticipated due to the proactive behaviour of criminals within their familiar environments. Upon successfully perpetrating the transgression, they endeavor to repeat it in comparable circumstances. The probability of a crime occurring is contingent upon many factors, including the criminal's IQ and the location's security [7].

Criminals typically choose similar locations and times for their subsequent crimes. Recurrence is more likely to predict crime; however, this may not always be the case [5]. Even while crimes can happen anywhere, it was usual for criminals to focus on the corruption opportunities they encountered in the communities they were most familiar with [2].

By providing access to the data generated by the Support Vector Neural Network (SVN) technique and categorizing the characteristics, whereabouts, and timing of the event, we expect to enhance public consciousness regarding hazardous zones during specific periods. Hence, our proposed approach has the potential to reduce fatalities and enable individuals to steer clear of particular locations during specific periods.

Moreover, this knowledge can impact individuals' decisions regarding their residential environments [9]. Conversely, law enforcement authorities can utilize this technique to enhance criminal forecasting and anticipating [13]. The distribution of police resources would also benefit. This will make it possible for the police to deal with crime most of the time while still properly using their resources. By publicly disclosing all of these recommendations, we aim to enhance the safety of inhabitants and tourists.

As a result of modern-day communications and travel, criminal activity has become a topic of universal fear and loathing. Criminals are usually busy and working within their usual routines; thus, crimes could be anticipated. When accomplished, they try to replicate the crime in a similar setting.

Robbery, sexual assault-related sex offenses, theft, killing, as well as kidnapping are only some of the crimes that have been widely highlighted in the media recently. The crime computation method was developed to assist authorities in preventing crime by providing accurate predictions of future crime levels.

The aims of this research are specified as follows:

(1) A novel SVN is created to decrease classification complications while maximizing the accuracy of the model at the same time. This combines SVM as well as ANN techniques. Since this is done, things will get more superficial.
(2) The advancement of Artificial Flora (AF) has enhanced the reliability of categorization. As a result, the size reduction is decreased, and an efficient forecasting system is established.
(3) The suggested method's accuracy, sensitivities, and precision are evaluated, among other metrics.

2. Materials and Methods

There are four steps in the suggested methodology: information collection, preprocessing, feature extraction, and predictions, Figure 1.

Information is initially gathered through the Internet. Finally, in the preprocessing phase, we eliminate duplicates and fill in the blanks. Subsequently, the artificial flora optimization algorithm is used for feature selection. The final step is to feed the hybrid classifier the previously chosen characteristics. The 'SVM and ANN techniques are integrated into a hybrid classifier to create a new approach. This suggested methodology includes four stages.

a) Data assortment
b) Data preprocessing
c) Feature collection using the Artificial Flora algorithm
d) Hybrid SVM-ANN categorization

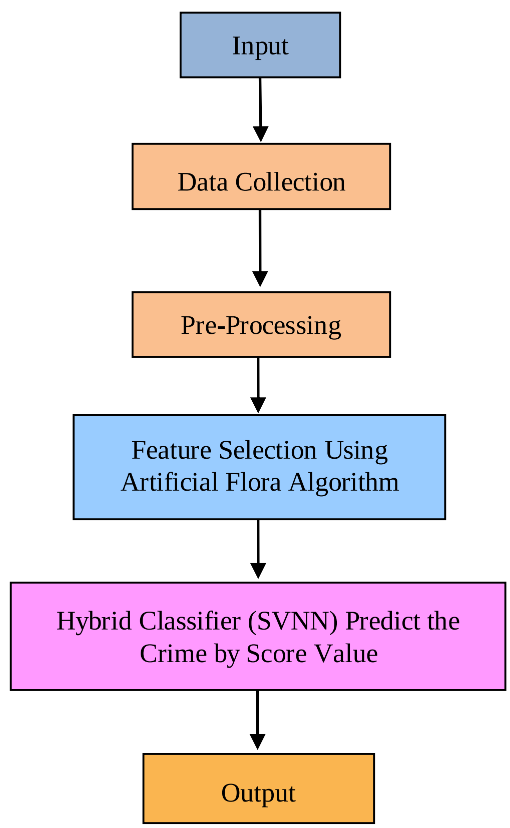

Fig. 1 Proposed method

2.1. Data Collection

The Internet is mined for crime statistics during processing, yielding valuable knowledge for implementing an efficient crime-prediction system. The gathered information is provided as a parameter.

2.2. Preprocessing

Collecting data over the Internet often results in a plethora of extraneous information, which slows down the system as it needs to be processed. The data regarding criminal acts were collected, followed by the preprocessing phase, which eliminated extraneous information. The information is cleaned up by removing all of the noise. That's why the forecast is more precise now.

2.3. Feature Selection Using Artificial Flora

This was a way to pick the best features for a smaller model, which improved the predictive power [1, 2]. Throughout this study, we use fake plants to examine the connection between qualities and their significance. Since the algorithms of artificial flora analyze each component separately, it can be used to investigate causal relationships between attributes.

The astute approach for Artificial Flora (AF)' seems to be another method described in this paper. This program emulates the phenomenon of plant movement and propagation in the environment. While plants are immobile, they can disperse seeds within a limited area, enabling their offspring to choose the most suitable environment.

Initially, plants dispersed their seeds for a specific purpose, and the initial plants that successfully separated and modified their new environment were those that had originated from separate, specialized species. The seeds' ability to persevere is correlated with their inherent credibility. If a seed does not adapt to its environment, it will die like its relatives did [3-5]. If a seed is successful, it will grow into a new plant, producing more seeds. Artificial plant computation utilizes plants' unique behavior to re-energize the structure through plant motion.

The four main components of a fake plant architecture are the mother plant, the offspring, the plant's geographic area, and the spread distance. The earliest plants hint at those that are ready to propagate through seed. This was the era when the initial seeds of vegetation that could not self-produce were produced through relatives. Manufacturing takes place in the vicinity of the plant. This spacing suggests the distance a seedling can germinate and propagate [6, 7].

Transformative, contagious, and individualized behavior are the three basic types of human action. The AF algorithm's feature selection procedure is broken down into the following steps:

i) Initialization
ii) Evolution behaviour
iii) Spreading behaviour
iv) Select behaviour
v) Termination

Replica plant architecture consists of four primary components: the parent plant, the plant's progeny, the plant's location, and the distance between them. The term "first plants" refers to plants capable of dispersing seeds. The cousins were the earliest progenitors of plants that could not distribute seeds freely during that time. The three primary types of behaviour are transformative, contagious, and individual.

The procedure for feature extraction by AF algorithm is outlined as follows:

2.3.1. Initialization

The artificial flora algorithm replicates natural plants by planting them in a controlled environment. The distribution of plants is determined by the matrix $A_{m,n,}$ where m denotes the dimension and n is the quantity of plants.

$$A_{m,n} = \text{rand}\ (0,1)\ x\ s\ x\ 2 - s \tag{1}$$

Where, the variable s represents the maximum limited area, whereas the array and (0, 1) represent a collection of uniformly distributed random numbers between 0 and 1.

2.3.2. Evolution Behavior

The initial dispersal of plant seeds occurs through aerial means, facilitated by proliferation and separation [14]. The spread separation is established by the distance of proliferation from the maternal plant and plant grandparents.

$$S_n = S_{1n}\text{rand}\ (o, 1)\ x\ r_1 + S_{2n}\text{rand}\ (o, 1)\ x\ r_2 \tag{2}$$

Where S_{1n} represents the distance at which the grandparent plant propagates, S_{2n} represents the distance at which the parent plant propagates, r1 and r2 are the coefficients of learning, and rand (0,1) is a randomly distributed number between 0 and 1. The novel grandparent dissemination distance was,

$$s'_{1n} = d_{2n} \tag{3}$$

The relative propagation distance between the original plant and its progeny remained constant but underwent alterations.

$$s'_{2n} = \sqrt{\frac{\sum_{m=1}^{J}\left(A_{m,n} - A'_{m,n}\right)^2}{J}} \tag{4}$$

2.3.3. Spreading Behaviour

The formation of the offspring occurs through the centrifugal process in the following manner:

$$A'_{1,n\times i} = S_{m,n\times i} + A_{m,n} \tag{5}$$

Where A' $_{m,n\times i}$ represents the location of the offspring plant, S m,n×i is a randomly generated number following a Gaussian dispersal with a mean of 0 and a variance of S $_{n.\,i}$ is the quantity of seeds that can be produced by one plant, and Am,n represents the location of the original plant.

Fitness Calculation

Once the solution commenced, the fitness of each solution was assessed. The exercise metric that yielded the highest predicted precision was determined to be the maximum. Equation (13) presented the exercise function.

$$\text{Fitness} = \text{Max (Accuracy)} \tag{6}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

Where TP signifies true positives, TN denotes true negatives, FN symbolizes false negatives, and FP signifies false positives, correspondingly.

2.3.4. Select Behavior

The vitality of plant progeny was determined using the survival probability equation.

$$a = \left| \sqrt{\frac{U(A'_{m,n\times i})}{F_{\max}}} \right| \times B_g^{(n\times i-1)} \tag{8}$$

Where, Bg is the selective probability and Bg $^{(n\times i-1)}$ lies between 0 and 1. It is evident that their fitness level is lower than those of offspring plants, which are farthest from the original plant [15]. Bg regulates the examination ability of the method and must be higher for the issue to obtain the local optimal solution quickly. Fmax represents the highest fitness level among the current generation's plant life, while U (A' $_{m,n\times i}$) denotes the fitness of the n^{th} solution.

The roulette wheel selection procedure will determine the viability of the progeny plant. This method is alternatively referred to as the proportionate selection method. The primary goal is to "accept in probability," indicating the presence of several choices, each assigned individual ratings. Nevertheless, the exam score's reliability is questionable. The decision is determined by the likelihood of being approved [16].

2.3.5. Termination

Choose plants from the current generation of plants as potential candidates for new distinct plants and repeat the above process until the preferred level of accuracy is attained or the determined quantity of iterations is attained. Therefore, the most advantageous feature is chosen in this procedure. The selected attribute enhances the precision of the crime forecasting algorithm.

2.4. Classifier as SVNN

A priori processing of the crime data, including feature selection, is required before the classification stage can begin. Appropriate features of each dataset ought to be used to categorize crimes [16]. For this purpose, we employ the SVN. This technique is needed to improve the margin during training by decreasing the values connected to the characteristics of the classification algorithm, hence decreasing the challenge of the classifiers without diminishing the quality of the training set [8, 9]. When comparing ANN and SVM, the hybrid method that emerged was SVN. The rate of crime is accurately predicted using SVN in the suggested technique.

2.5. Support Vector Machine (SVM)

The supervised learning algorithm-based learning model is comparable to Support Vector Machines (SVM). SVM analyses data and identifies the models employed for classification and regression analysis, Figure 3.

The SVM classifiers accomplish binary organization [1] by dividing a set of training vectors into two distinct classes $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$, where denotes vectors within a feature space of dimension d and is a group label.

The SVM study was generated by transferring the input vectors to an innovative feature space with a higher dimension, described as $\phi: R^d \rightarrow H^f$. Where, d<f. An optimal separating hyperplane was also constructed in creative feature space using a kernel function $K(x_i, x_j)$ that is the product of input vectors x_i and x_j where,

$$K(x_i, x_j) = \Phi(x_i).\Phi(x_j)$$

2.6. Artificial Neural Network (ANN)

The complete weighted input is received at node j, along with an output enable function (typically the sigmoid function) that transforms the weighted input of neuron n into its output function, Figure 4 and Equation 9.

$$S_j = \sum_{j=1}^{n} x_i w_{ij} \quad \text{and} \quad O_j = \frac{1}{1+e^{s_j}} \tag{9}$$

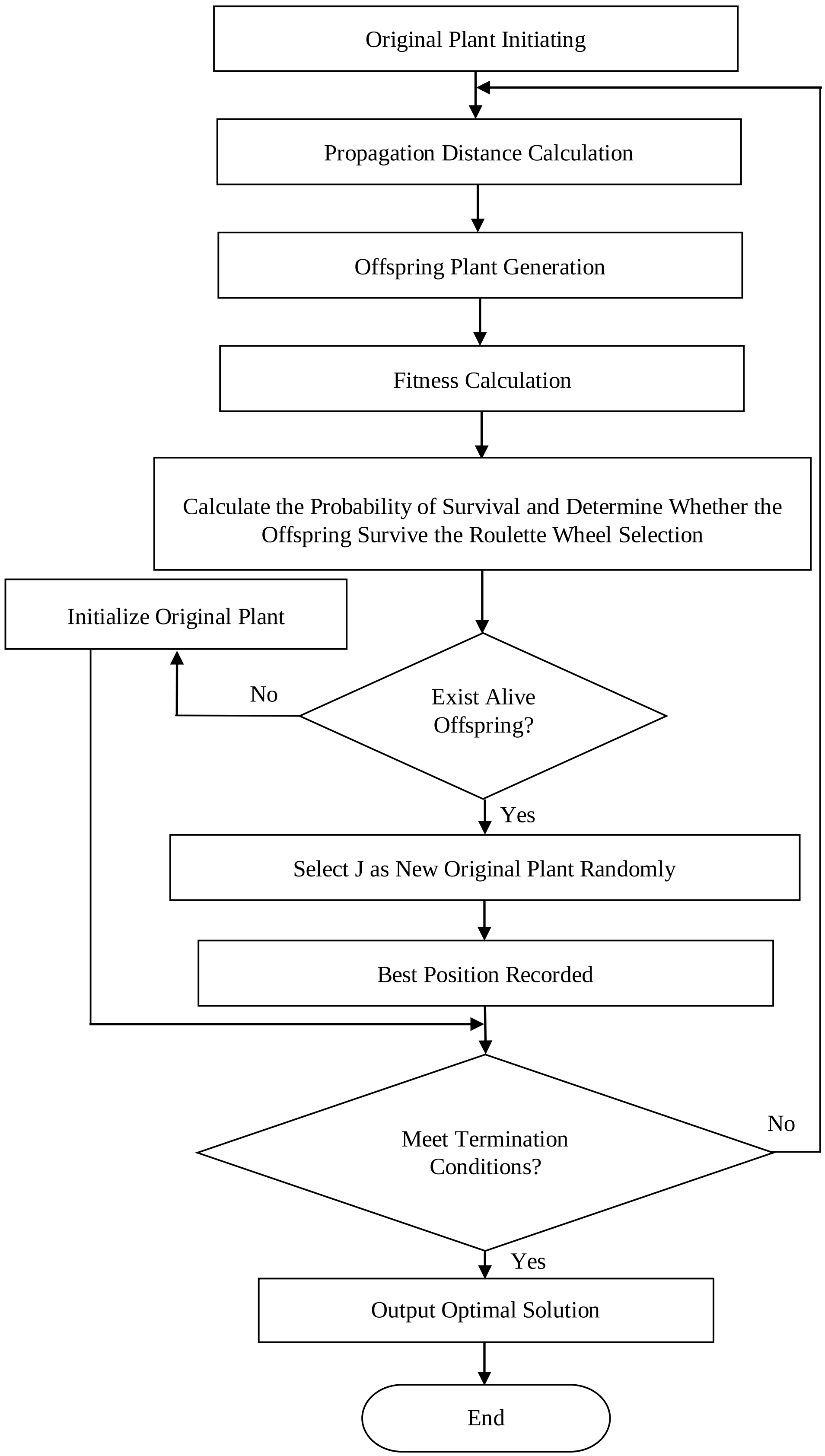

Fig. 2 Artificial Flora algorithm

Preparing mode and client mode are the two activity phases of the neuron and, by extension, the ANN. In the preparatory phase, the data set comprising authentic inputs and outputs could serve as an illustrative example to instruct the computer on predicting the outputs.

Regulated learning commences with arbitrary loads and modifies those loads for the assignment by employing backpropagation and slope drop search calculations [4]. One applies the differentiation between the qualities gained and the objective output values while studying driving errors.

The error task is contingent on the weight that must be modified to restrict the error. For a given training set $\{(x_1,t_1), (x_2,t_2),\ldots.(x_k,t_k)\}$, the k-ordered pairs that comprise the input and output formats, n-inputs and m-outputs, respectively, are defined by the equation for the output of each neuron:

$$E_j = \frac{1}{2}\sum_{j=1}^{k}\left(O_j - t_j\right)^2 \tag{10}$$

While from the training set, the network O_j represents the output produced when the input pattern and t_1 describes the target value. During the drive mode, each weight was modified by adding the amount to the earlier value.

$$\Delta w_{ij} = -\gamma \frac{\partial E}{\partial w_{ij}} \tag{11}$$

While, γ representing the constant that furnishes the learning rule, the greater the learning rate, the more rapidly it integrates. However, the search path for the optimal solution is obstructed, rendering the merger unattainable.

The NN model may then take another set of data with uncertain output values and forecast the related outputs autonomously once a group of good weights has been determined.

2.7. Support Vector Neural Network (SVNN)

There is no probability justification for classifying below and above the classifying hyperplane [10, 11] because the support vector classifier functions by inserting data points. In all of ANN, this represents the most crucial problem. The research problem is solved using ANN; however, it fails to explain why or how.

Due to this, trust in the system is diminished. In this research, we employ a hybrid approach to solve these problems. To make SVN classifiers, ANN and SVM classifiers are combined.

Choose the feature just after the process has been completed. Select the features that provide the best SVN record. Combining SVM and ANN produces a classification selection. The hybrid method combines ANN and SVM classifiers to form SVM-ANN classifiers.

In this case, the ANN classifier is initially shown the chosen features. Hidden layer ANN output is fed into SVM to drive the same goal. Figure 5 depicts the SVM-formed hybrids ANN classifier once training is complete.

Step 1 : ANN has three layers: input, hidden, and output. Numerous neurons are present in every layer and are represented as:

$(X_1, X_2.....X_n)$ - Input neurons
$(Y_1, Y_2....Y_n)$ - Hidden neurons.

In this step, the output layer was substituted with an SVM classifier. The weight linking between the input layer Xn and hidden layer Yn is described as W_{ij}.

Step 2 : Each node in the inquiry layer was initially augmented by a weight value common to both the hidden and query layers. The Equation (12) displayed the output of the hidden layer.

$$Y_j = B_j + \sum_{i=1}^{a} X_i W_{ij} \tag{12}$$

Where,
Bj - Bias value
W_{ij} - Weight value

Step 3 : Further, The output Y_j of the hidden layer was processed using the activation function. The activation function was specified in Equation (13).

$$F(Y_j) = \frac{1}{1+e^{-Y_j}} \tag{13}$$

Step 4 : We queried the SVM, which had output data hidden from view. SVM describes the structure using metadata to characterize the practices. The structure was saved after the cycle of preparations. The built framework was put through its paces in an analysis phase, Equation (14).

$$decision = \begin{cases} T_h \geq score\, ;\, Positive \\ T_h < score\, ;\, Negative \end{cases} \tag{14}$$

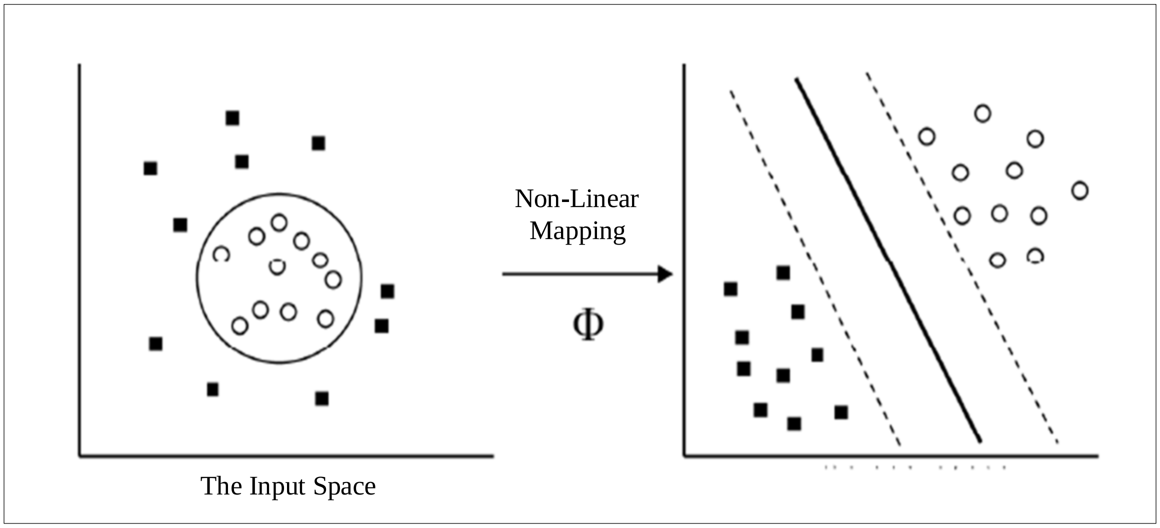

Fig. 3 SVM model generation

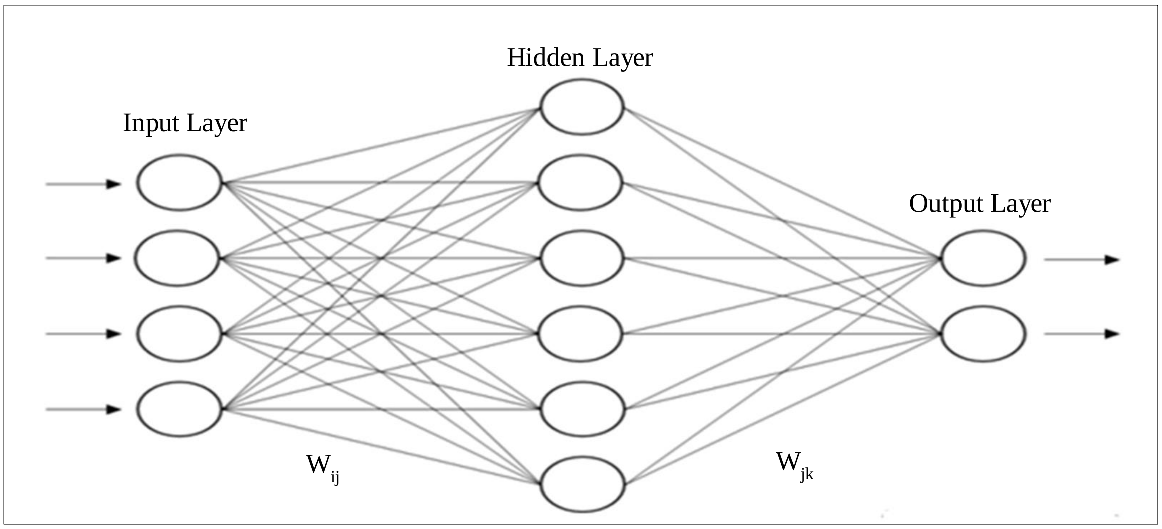

Fig. 4 Structure of an Artificial Neural Network

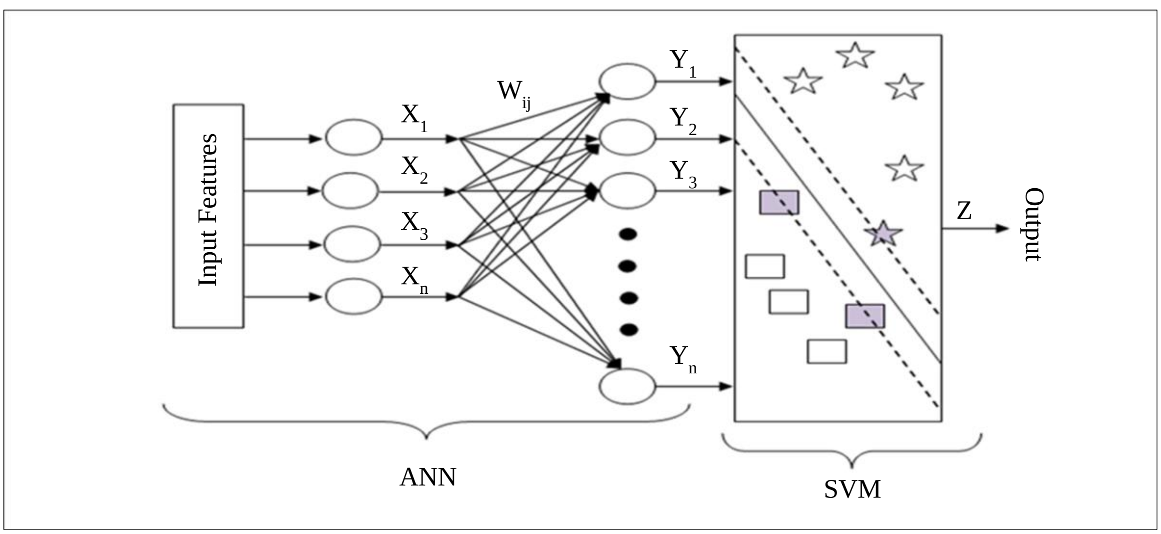

Fig. 5 Overall structure of optimal SVMANN

3. Results and Discussion

This section of the study aims to present and debate the program's results on calculated crime levels. We utilized JAVA to put the planned method into action. The proposed procedure uses a Windows computer with an Intel Core i5 CPU with 1.6 GHz speed and 4 GB RAM.

3.1. Dataset Explanation

Every piece of crime statistics is a compilation of statistics and other details about criminal activity in Los Angeles between 2010 and the present. This data is more reliable because it comes from actual crime study following and can be used to predict criminal behaviour. Violence, rape, as well as other significant violent offenses, as well as other statistics, are included in the Philadelphia Police Department's Crime Incidents (CI) database.

Datasets of criminal activity were gathered mainly through the Bloomington Police Department, abbreviated "CD." Countries belonging to the African, Indian, Caucasian, and Pacific Islander groups are just as different as the behaviors of their inhabitants. The Annual Crime Dataset (ACD) is a regularly updated database with information about crimes committed in the Austin, Texas area and the surrounding state of Texas. This data is subsequently utilized to investigate the potential involvement of specific individuals in those criminal activities.

3.2. Result Explanation

The acquired sample group is analyzed using Artificial Flora to pick features. An SVNN is employed to predict the efficacy of error rate detection, sensitivities, selectivity, accuracy, and timeliness. The effectiveness of the novel approach surpasses that of the existing one. The Artificial Flora employs established techniques such as the Particle Swarm Optimization (PSO), Firefly Algorithm (FA), and Genetic Algorithm (GA) to identify the minimum attribute for selection, which are shown in Table 1.

Table 1. Selected features from datasets

S. No.	Technique	CDP	CI	CD	ACD
1	GA	490	485	476	465
2	PSO	373	377	419	402
3	FA	314	379	310	291
4	AF	147	115	117	103

The results of the feature selection process for the gathered crime dataset are presented in Table 1. The obtained result is depicted in the graph shown in Figure 6. Based on the AF method effectively selects the minimum number of crime features from multiple datasets, as demonstrated in Figure 6. The AF method determines 147 features from the CDP dataset, 115 features from the CI Select dataset, 117 features from the CD dataset, and 103 features from the ACD dataset.

Traditional methods, including GA, PSO, and FA, exhibit greater criminal characteristics. As stated in the preceding sections, the AF algorithm effectively determines the optimal crime feature from the diverse datasets. The AF method strives to minimize the error rate in the selected features. The error mentioned above rate will enhance the efficacy of the crime rate forecast.

For the minimum number of features, the AF method determines the optimal number and minimal error rate in the various crime datasets, as shown in Figure 6. The CDP has a minimum error rate of 0.175, the CI guarantees an error rate of 0.168, the ACD has 0.132, and the CD has attained an error rate 0.153. From Figure 7, the root square error rate is minimized for the HCP dataset at 0.163, the CDP at 0.163%, the CI at 0.154, and the ACD at 0.122. The proposed technique exhibits a significantly lower error rate than other methods, including GA, PSO, and FA.

It is evident from the outcome that the AF method produces a reduced error rate when analyzing features associated with criminal activities. Criminal activity predictions are more precise due to the reduced error rate. As a result, the sensitivity and specificity of the SVNN crime prediction method are more accurate. It is evident from Figure 8 that the sensitivity rate of the SVNN method is greater across the various crime datasets. The SVNN method predicts characteristics from the CDP with a sensitivity rate of 96.31%, while the CI database achieves 97.42%, the CD 96.31%, and the ACD 97.41%.

It is evident from Figure 8 that the proposed method outperforms the ANN, SVM, and Whale Optimisation Algorithm (WOA) in terms of prediction accuracy. In contrast to alternative methods currently in use, SVNN exhibits a superior sensitivity rate when predicting across diverse crime datasets. Although the SVNN strategy demonstrates effective prediction of wrongdoing patterns, it is required to have a higher specificity rate than other expectation techniques across different datasets.

This is because the forecast technique must analyze wrongdoing information and select unnecessary negative highlights at a high rate, ultimately improving the recognition rate accuracy [17]. Consequently, the precision value obtained is depicted in Figure 9.

Based on the preceding discourse and the data presented in Figure 9, it is evident that the SVNN technique exhibits a greater degree of specificity when examining patterns associated with criminal activities across various crime datasets.

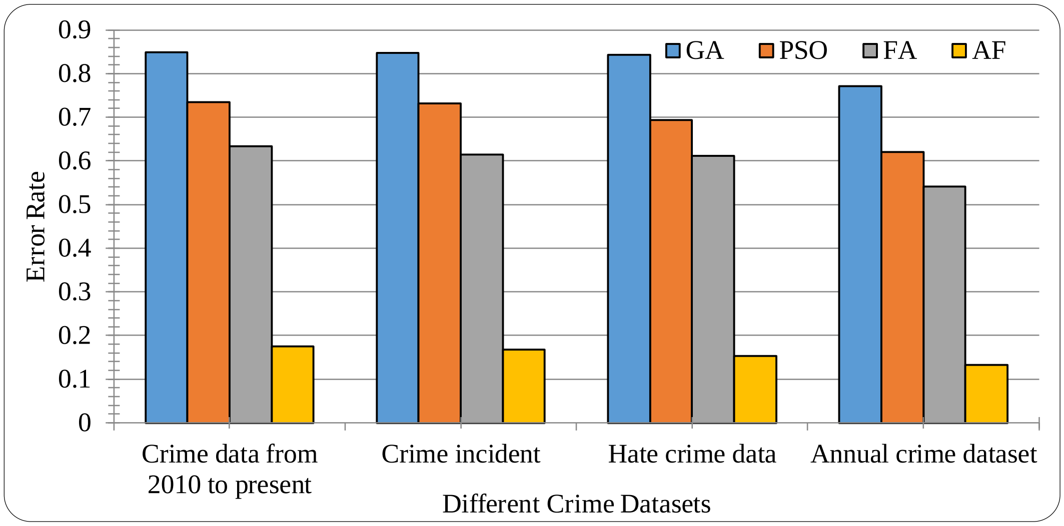

Fig. 6 Mean Square Error comparison

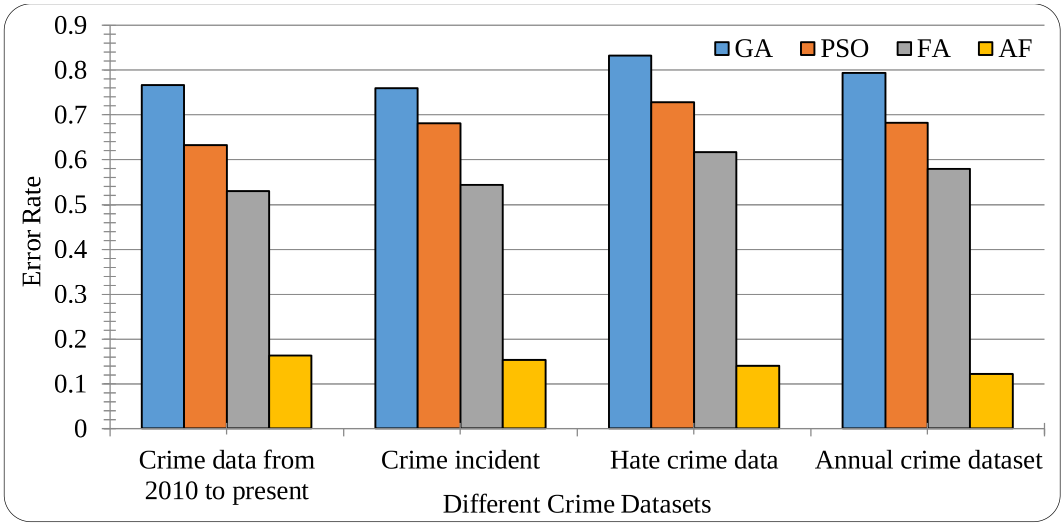

Fig. 7 Graphical representation of Root Mean Square Error rate

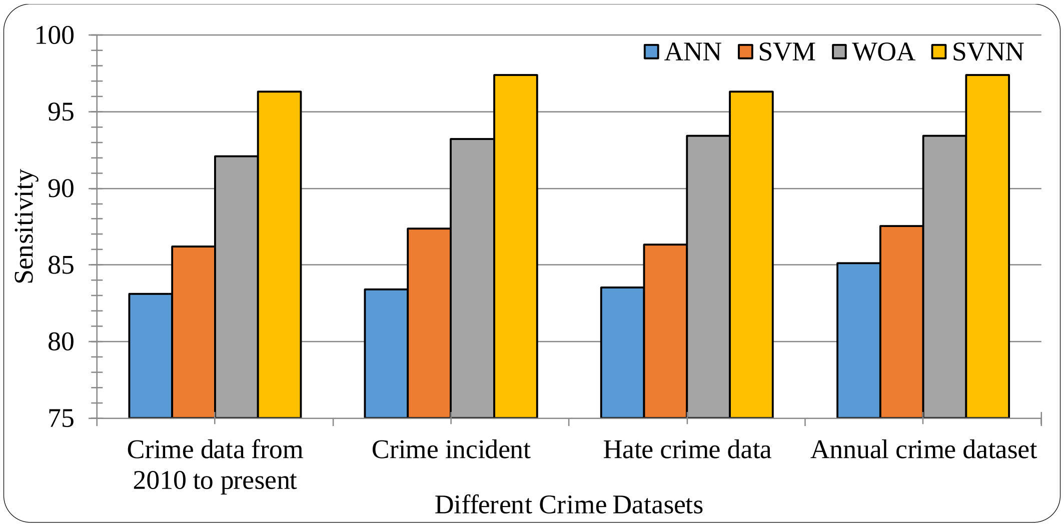

Fig. 8 Sensitivity plot

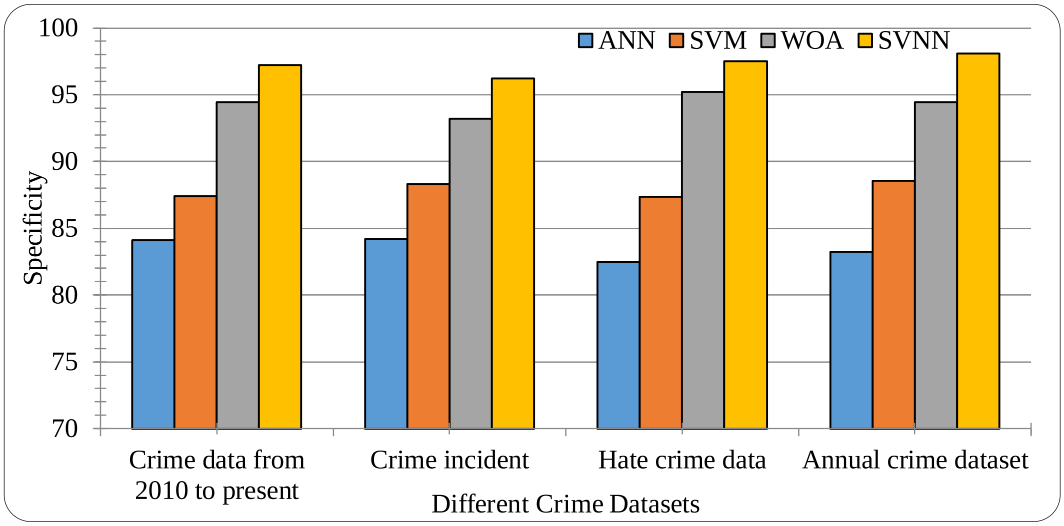

Fig. 9 Specificity plot

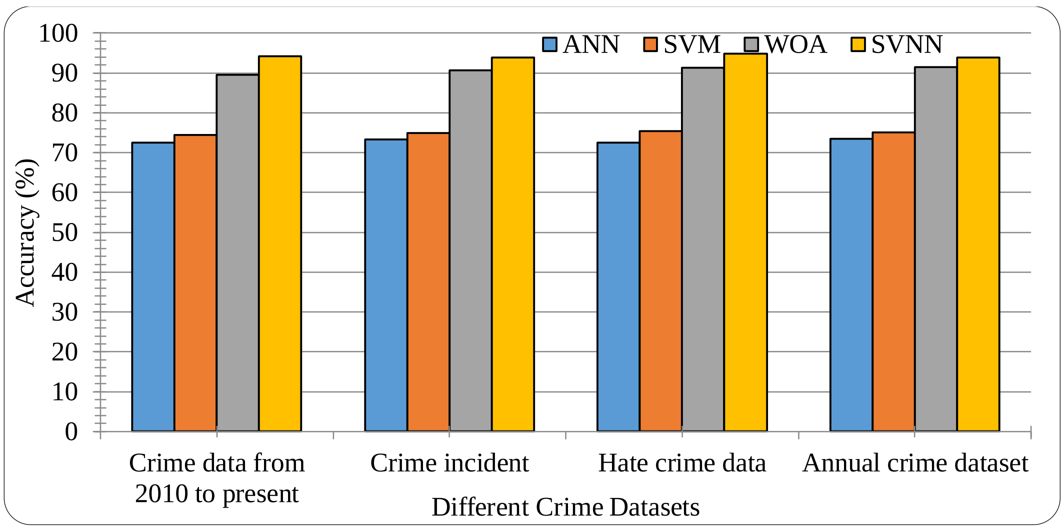

Fig. 10 Accuracy plot

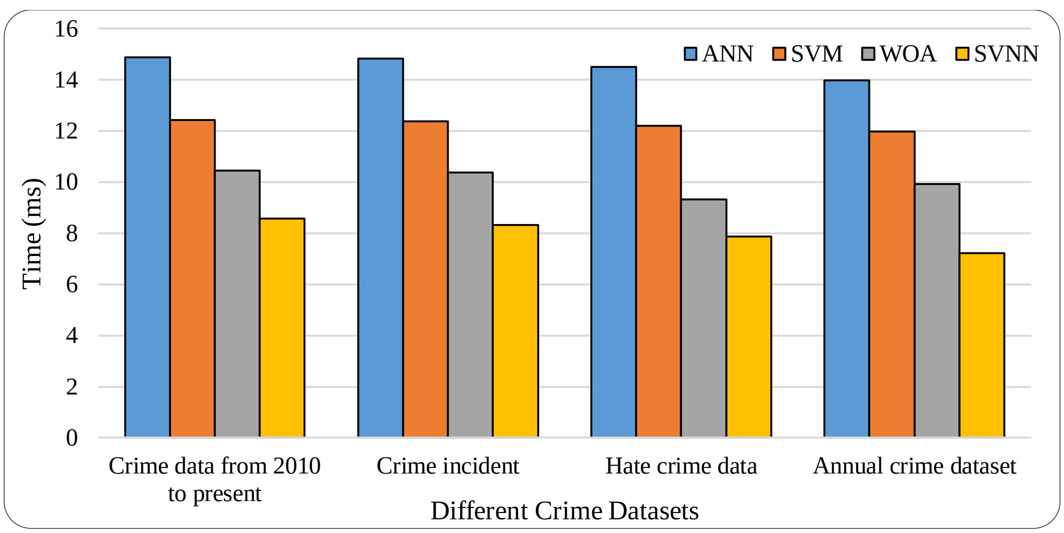

Fig. 11 Plot of time (ms)

With a sensitivity rate of 97.21%, the SVNN approach identifies the characteristics of the CDP, while the CI database achieves 96.21%, the CD 97.52%, and the ACD 98.1%. It is evident from the diagram that the expected precision is significantly higher in comparison to alternative forecasting methods such as ANN, SVM, and WOA. Based on the preceding discussions, the SVNN method employs the selected features and predicts wrongdoing-related activities with a high rate of dataset specificity compared to other methods. The enhanced level of specificity and sensitivity results in a corresponding increase in the overall precision of the crime detection metric. As the SVNN method retrieves crime-related patterns with a significantly higher degree of accuracy (94.028% on average) than existing methods, including ANN (72.9925%), SVM (74.9875%), and WOA (90.797%). The accuracy obtained is subsequently illustrated in Figure 10.

The SVNN technique, as illustrated in Figure 10, employs the features extracted from the various informational indexes with a high degree of accuracy. This means it effectively selects the feature corresponding to the client's interest. The SVNN predicts the characteristics of criminal activity based on the CDP with the following rates: 94.245% for the CI information, 94.93% for the CD, and 93.97% for the ACD.

The predicted criminal model's performance is enhanced by implementing more effective features compared to existing methods, such as ANN, SVM, and WOA. Based on the investigations above, it can be concluded that the SVNN strategy predicts crime-related information from diverse crime informational collections with greater accuracy and in less time than other methods. The time required to obtain an understanding of the characteristics that the SVNN strategy differentiates in this manner.

The SVNN exhibits a significantly shorter time needed to predict incorrect information in the future compared to other established techniques such as ANN, SVM, and WOA [18]. Subsequently, the time effectiveness is delineated in Figure 11. As depicted in Figure 11, the SVNN method differentiates criminal designs from crime information collection in the shortest time possible. The SVNN technique predicts qualities from the CDP in 8.58 milliseconds, while the CI predicts attributes in 8.32 milliseconds. Anticipating the model from the CD requires 7.36 milliseconds, and expecting the model from the ACD requires 7.21 milliseconds. The efficacy of the recovered crime models in generating future decisions is more significant when compared to alternative techniques such as ANN, SVM, and WOA.

Thus, it is evident from the preceding research that the analysis of AF and SVNN predict significant patterns (crime) with minimal time and error from a vast quantity of crime data. Due to this error rate reduction, crime forecasting is now more precise compared to other established conventional methods [7, 8, 9]. The ultimate outcomes of the research endeavor will enhance the law enforcement system's decision-making process.

4. Conclusion

This research paper presents an efficient and effective hybrid classifier for crime rate prediction. In this contribution, the AF optimization algorithm has been developed for feature selection, and a hybrid classifier has been developed for prediction. SVM and ANN classifiers have been combined to form the hybrid classifier. Four sets of datasets were developed for experimental analysis: crime data spanning the years 2010 to the present, hate crime data, annual crime data, and crime incident data. The efficacy of the proposed method in terms of precision, sensitivity, and specificity has been evaluated. For the annual crime dataset, the method achieved a maximal accuracy of 93.97%, a sensitivity of 97.41%, and a specificity of 98.1%. Implementation of the proposed method was performed using the JAVA program. Based on the findings, this hybrid approach is suggested to be the most effective crime prediction model compared to the current alternatives.

Acknowledgments

The authors acknowledge Noorul Islam Centre for Higher Education for supporting the research.

References

[1] Anupam Mukherjee, and Anupam Ghosh, "Predictive Framework for Crime Data Analysis Using A Hybrid Logistic Regression-Support Vector Machine Based Ensemble Classifier Powered by CART (LR-SVM CART)," *Multimedia Tools and Applications*, vol. 82, pp. 35357-35377, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Ruaa Mohammed Saeed, and Husam Ali Abdulmohsin, "A Study on Predicting Crime Rates through Machine Learning and Data Mining Using Text," *Journal of Intelligent Systems*, vol. 32, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Angel Gonzalez-Prieto et al., "Hybrid Machine Learning Methods for Risk Assessment in Gender-Based Crime," *Knowledge-Based Systems*, vol. 260, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Vinothkumar K. et al., "Predicting High-Risk Areas for Crime Hotspot Using Hybrid KNN Machine Learning Framework," *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, pp. 848-852, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Ning Wang et al., "A Hybrid Ensemble Learning Method for the Identification of Gang-Related Arson Cases," *Knowledge-Based Systems*, vol. 218, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Bhavna Saini et al., "A Study of Lightweight Approaches to Analyze Crime Conditions in India," *Journal of Applied Security Research*, vol. 18, no. 2, pp. 221-245, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Wim Hardyns, and Robin Khalfa, "Predicting Crime across Cities and Regions: A Comparative Analysis of Predictive Modelling in Three Belgian Settings," *Applied Spatial Analysis and Policy*, vol. 16, pp. 485-508, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Rabia Musheer Aziz, Prajwal Sharma, and Aftab Hussain, "Machine Learning Algorithms for Crime Prediction under Indian Penal Code," *Annals of Data Science*, pp. 1-32, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Rabia Musheer Aziz, Aftab Hussain, and Prajwal Sharma, "Cognizable Crime Rate Prediction and Analysis under Indian Penal Code Using Deep Learning with Novel Optimization Approach," *Multimedia Tools and Applications*, pp. 1-38, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Pooja Mithoo, and Manoj Kumar, "Social Network Analysis for Crime Rate Detection Using Spizella Swarm Optimization Based BiLSTM Classifier," *Knowledge-Based Systems*, vol. 269, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Ali Raza et al., "Preventing Crimes through Gunshots Recognition Using Novel Feature Engineering and Meta-Learning Approach," *IEEE Access*, vol. 11, pp. 103115-103131, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Urvashi Gupta, and Rohit Sharma, "Analysis of Criminal Spatial Events in India Using Exploratory Data Analysis and Regression," *Computers and Electrical Engineering*, vol. 109, Part A, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] S. Geetha, and Malathi K., "SVM vs Decision Tree Algorithm Cost Effective Comparison to Enhance Crime Detection and Prevention," *Journal of Survey in Fisheries Sciences*, vol. 10, no. 1S, pp. 2814-2822, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Monika, and Aruna Bhat, "DAC-BiNet: Twitter Crime Detection Using Deep Attention Convolutional Bi-Directional Aquila Optimal Network," *Multimedia Tools and Applications*, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Moruf Akin Adebowale, Khin T. Lwin, and M.A. Hossain, "Intelligent Phishing Detection Scheme Using Deep Learning Algorithms," *Journal of Enterprise Information Management*, vol. 36, no. 3, pp. 747-766, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Yifei Gong, Mengyan Dai, and Feng Gu, "CARESim: An Integrated Agent-Based Simulation Environment for Crime Analysis and Risk Evaluation (CARE)," *Expert Systems with Applications*, vol. 214, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Anitha Thangasamy, Bose Sundan, and Logeswari Govindaraj, "A Novel Framework for DDoS Attacks Detection Using Hybrid LSTM Techniques," *Computer Systems Science & Engineering*, vol. 45, no. 3, pp. 2553-2567, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Balajee R.M., and Jayanthi Kannan M.K., "Intrusion Detection on AWS Cloud through Hybrid Deep Learning Algorithm," *Electronics*, vol. 12, no. 6, pp. 1-21, 2023. [CrossRef] [Google Scholar] [Publisher Link]