Semiconductor Optoelectronics
ISSN: 1001-5868

HYBRID CRIME PREDICTION MODEL USING SUPPORT VECTOR NEURAL NETWORKS

Santhosh S, Dr. N.Sugitha
Research Scholar, Department of Information Technology, Noorul Islam Centre for Higher Education, Kumaracoil, Kanyakumari, Tamilnadu, India
Scientist-E/Joint Director, Communications and Software Technology Group, Centre for Development of Advanced Computing (C-DAC), Vellayambalam, Thiruvananthapuram, Kerala, India
Professor, Department of Electronics and Communication Engineering, Saveetha College of Engineering , Chennai, Tamil Nadu, India
Email: santhosh.vattappara@gmail.com, sugithavinukumar@gmail.com
Corresponding: sugithavinukumar@gmail.com

Abstract - *Criminal activity is an inherent component of the everyday dangers we face. The causes, dangers, and preventative measures of an isolated criminal act were the subject of numerous publications. A novel hybrid classifier-based approach to predicting criminal behavior is offered. Many different criteria were used to evaluate the effectiveness of the suggested strategy. This newly developed hybrid prediction model is compared with existing models like Particle Swarm Optimization (PSO), Genetic Algorithm (GA), as well as Firefly Algorithm (FA) in terms of diverse performance metrics like error rate, sensitivity, specificity, precision, and execution time*
Keywords- *Crime; hybrid classifier; Feature selection; pre-processing; prediction; Artificial neural network (ANN) and support vector machine (SVM).*

I. INTRODUCTION

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 they are 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 given as follows:

- A novel Support Vector Neural Network (SVNN) is created to decrease classification complications while maximizing accuracy. This combines SVM as well as ANN techniques. Since this is done, things will get simpler.

- 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.
- The accuracy, sensitivities, and precision of the suggested method are evaluated, among other metrics.

II. Proposed Methodology

There are four steps in the suggested methodology: information collection, pre-processing, feature extraction, and predictions. Information is initially gathered through the internet. Finally, in the pre-processing phase, we get rid of the 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 characteristics that were previously chosen. The 'Support Vector Machine (SVM) using Artificial Neural Network (ANN) technique is integrated into a hybrid classifier to create a new approach.

This suggested methodology includes four stages

- Data Assortment
- Data Preprocessing
- Feature collection using the Artificial Flora algorithm
- Hybrid SVM-ANN categorization

In what follows, a detail about the steps that make up the suggested approach.
(1) Data collection: The Internet is mined for crime statistics during processing, yielding a trove of knowledge useful for putting into practice an efficient crime-prediction system. The gathered information is provided as a parameter.
(2) Pre-processing: 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 pre-processing phase was used to get rid of any extraneous information. The information is cleaned up by removing all of the noise. That's why the forecast is more precise now.
(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 algorithm simulates the process of flora migration and reproduction in nature. Although plants cannot move, they can scatter seeds inside a specific range to allow progeny to select the best habitat for them. Originally, plants dispersed their seeds for a specific purpose, and the initial plants to successfully separate and modify 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 at all adapt to its environment, it will die, just like its relatives did [3, 4, 5]. If a seed is successful, it will grow into a new plant,

which in turn will produce 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, as well as 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 at which a seedling can germinate and propagate [6, 7]. Transformative behavior, contagious behavior, 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

(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. For this purpose, we employ the Supporting Vector Neural Network (SVN). The goal of this technique is 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 as well as SVM, the hybrid method that emerged was SVNN. The rate of crime is accurately predicted using SNVV in the suggested technique.

(5) Support vector Neural Network (SVNN)

There isn't any 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 SVNN classifiers, ANN, as well as SVM classifiers, are combined.

Choose the feature just after the process has been completed. choose the features that provide the best SVNN record. Combining SVM and ANN produces a classification selection. In the hybrid method, ANN & SVM classifiers are combined to form SVM-ANN classifiers. In this case, the artificial neural network (ANN) classifier is initially shown the chosen features. Hidden layer ANN output is fed into SVM to drive the same goal. Figure 4 depicts the SVM-formed hybrids ANN classifier once training is complete.

Step 1: ANN has 3 layers called the input, hidden, and output layers. 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 among input layer Xn and hidden layer Yn is described as W_{ij}.
Step 2: Every node I within the query layer was first increased by such a weight value shared either by the hidden layer as well as the query layer. The hidden layer output was shown in the equation

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

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

Step 3: Further, the hidden layer output Yj was passed via the activated function. The activation function was given in equation (13).

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

Step 4: We queried the SVM having output data that was 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.

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

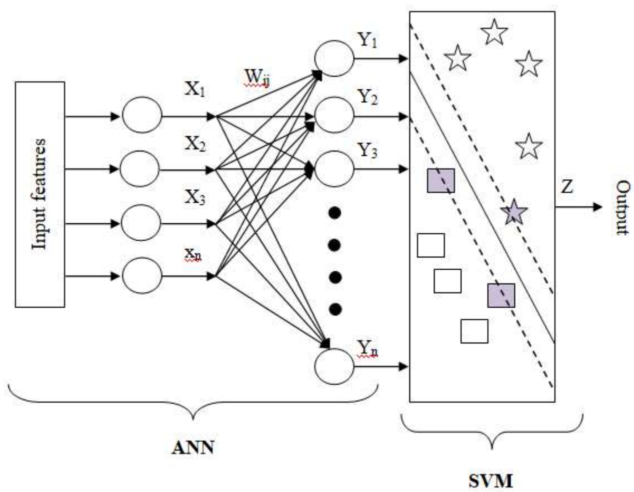

Figure 1: Overall structure of optimal SVMANN

III. RESULT AND DISCUSSIONS

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

a) DataSet Explanation:

Every piece of crime statistics is a compilation of statistics and other details about criminal activity in Los Angeles between 2010 to the current moment. This data obtained is more reliable because it comes from actual crime study following, and it can be used to predict criminal behavior. 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 gathered mostly by Bloomington Police Department, abbreviated "CD." Countries belonging to the African, Indian, Caucasian, as well as Pacific Islander groups are just as different as the behaviors of their inhabitants. The Annual Crime Dataset (ACD) is a database that is regularly updated with information about crimes committed in the Austin, Texas area and the surrounding state of Texas. This information is then used to examine whether or not particular people are involved in those crimes.

b) Result Explanation:

Artificial flora is used to choose features for the acquired sample group, and a Support Vector Neural Network is used to forecast the effectiveness in terms of detecting error rate, sensitivities, selectivity, accuracy, and timeliness. The effectiveness of the new method is improved over that of the current one. The artificial flora uses known methods like Firefly Algorithm (FA), Particle Swarm Optimization (PSO), and Genetic Algorithm (GA), and to determine the minimum characteristic to select. To achieve the lowest possible error rate with the fewest possible features, the AF technique chooses the ideal number. Error rates for CDP are at least 0.175, for CI they are 0.168, for the CD they are 0.153, and even for ACD, they are 0.132. The Sensitivity Rates for predicting characteristics from the CDP using the SVNN approach is 96.31%, 97.42%, 96.31%, and 97.42%, respectively, for the CI database, the CD, and the ACD, respectively. When compared to other available methods, SVNN can more accurately predict a greater sensitivity rate across a wider range of crime datasets. When comparing the SVNN technique to other methods for assessing crime trends, the testing results found that it had a higher Selectivity Rate. Selectivity rates for predicting characteristics using CDP range from 97.21% for the SVNN technique to 96.21% for the CI database, 97.52% for the CD, as well as 98.1% for the ACD. The outcomes are depicted in figures 2 and 3 below.

Relative to other current approaches like WOA (90.797%), SVM (74.9875%), and ANN (72.925%) experimental findings reveal that the SVNN technique has a good Precision (94.028% on average) when retrieving patterns connected to the crime. In light of this, it is evident that the suggested method outperforms the SVM, ANN, and WOA in terms of prediction accuracy. The outcomes are shown in Figure 4 beneath.

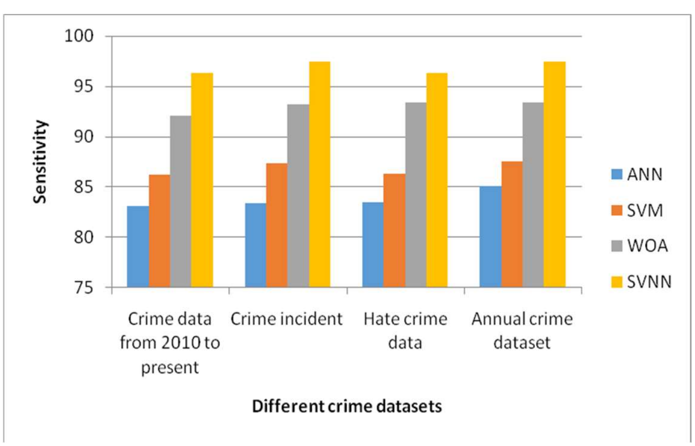

Figure2: Sensitivity

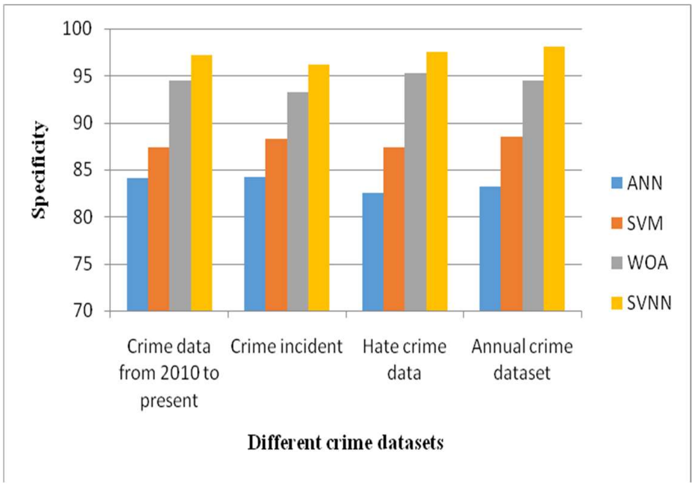

Figure 3: Specificity

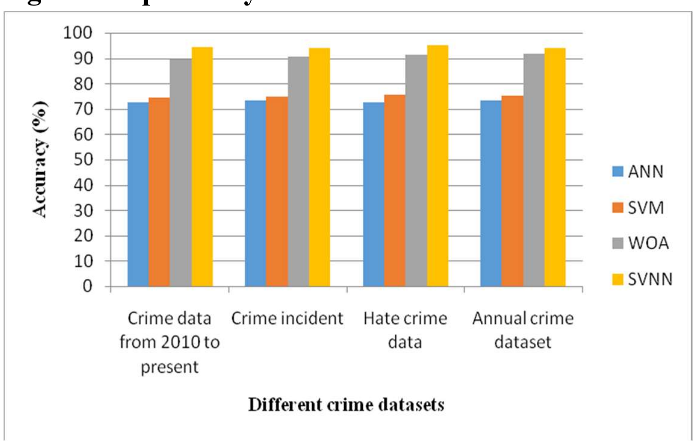

Figure 4: Accuracy

Predicting significant patterns (crime) from large quantities of crime data using AF and SVNN requires little time and error compared to traditional methods. The precision of crime prediction over more traditional approaches is a direct outcome of the decreased error rate. The research methodology, when completed, will yield results that enhance the justice system's judgment.

IV. CONCLUSION

To predict future crime rates, the authors of this study introduce a hybrid classifier that is both accurate and effective. This work contributes an Artificial Flora (AF) optimization approach for feature extraction and a hybrid classifier for prediction. This new classifier, called a hybrid, combines SVM and ANN techniques. Crime statistics from 2010–the present, incident statistics, statistics on incidents of hate, as well as an annual crime dataset were used in the study. The accuracy, sensitivities, and selectivity of the suggested approach have all been evaluated. For such an Annual crime database, the technique obtained a maximum efficiency of 93.97%, a sensibility of 97.41%, and selectivity of 98.1%. The JAVA programming language has been used to actualize the proposed solution. Based on the findings, this combination model is recommended over the other current crime forecasts.

REFERENCES

[1] Mathew, Jincy C., B. Nithya, C. R. Vishwanatha, Prathiksha Shetty, H. Priya, and G. Kavya. "An Analysis on Fraud Detection in Credit Card Transactions using Machine Learning Techniques." In *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, pp. 265-272. IEEE, 2022.

[2] Kshatri, S.S., Singh, D., Narain, B., Bhatia, S., Quasim, M.T. and Sinha, G.R., 2021. An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach. *IEEE Access*, *9*, pp.67488-67500.

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

[4] Das, P., Das, A.K., Nayak, J., Pelusi, D. and Ding, W., 2021. Incremental classifier in crime prediction using bi-objective particle swarm optimization. *Information Sciences*, *562*, pp.279-303.

[5] Nitta, G.R., Rao, B.Y., Sravani, T., Ramakrishiah, N. and Balaanand, M., 2019. LASSO-based feature selection and naïve Bayes classifier for crime prediction and its type. *Service Oriented Computing and Applications*, *13*(3), pp.187-197.

[6] Malik, A., Maciejewski, R., Towers, S., McCullough, S. and Ebert, D.S., 2014. Proactive spatiotemporal resource allocation and predictive visual analytics for community policing and law enforcement. *IEEE transactions on visualization and computer graphics*, *20*(12), pp.1863-1872.

[7] Lin, Y.L., Yen, M.F. and Yu, L.C., 2018. Grid-based crime prediction using geographical features. *ISPRS International Journal of Geo-Information*, *7*(8), p.298.

[8] Nguyen, T.T., Hatua, A. and Sung, A.H., 2017. Building a learning machine classifier with inadequate data for crime prediction. *Journal of Advances in Information Technology Vol*, *8*(2).

[9] Yerpude, P., 2020. Predictive modelling of crime data set using data mining. *International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol*, *7*.

[10] Rumi, S.K., Deng, K. and Salim, F.D., 2018. Crime event prediction with dynamic features. *EPJ Data Science*, *7*(1), p.43.

[11] Abdulrahman, N. and Abedalkhader, W., 2017. KNN classifier and naïve bayse Classifier for crime prediction in san francisco context. *Int. J. Database Manag. Syst*, *9*(4), pp.1-9.

[12] Gupta, V.K., Shukla, S.K. and Rawat, R.S., 2022. Crime tracking system and people's safety in India using machine learning approaches. *International Journal of Modern Research*, *2*(1), pp.1-7.

[13] Kumar, R. and Nagpal, B., 2019. Analysis and prediction of crime patterns using big data. *International Journal of Information Technology*, *11*(4), pp.799-805.