Synergy of Artificial Intelligence and Big Data in Criminal Investigations

Sulabh Mahajan[1], Sudhakar Reddy[2], Dr. Satish Upadhyay[3], Dr. Arvind Kumar Pandey[4], Dr. Trapty Agarwal[5], Yuvraj Parmar[6]

[1]Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India sulabh.mahajan.orp@chitkara.edu.in https://orcid.org/0009-0004-5048-7352
[2]Professor, Department of Physics, School of Sciences, JAIN (Deemed-to-be University), Karnataka, India, Email Id- r.sudhakar@jainuniversity.ac.in, Orcid Id- 0000-0001-8207-3526
[3]Assistant Professor, Department of uGDX, ATLAS SkillTech University, Mumbai, Maharashta, India, Email Id- satish.upadhyay@atlasuniversity.edu.in, Orcid Id- 0000 0002 2865 014X
[4]Associate Professor, Department of Computer Science & IT, ARKA JAIN University, Jamshedpur, Jharkhand, India, Email Id- dr.arvind@arkajainuniversity.ac.in, Orcid Id-0000-0001-5294-0190
[5]Associate Professor, Maharishi School of Engineering & Technology, Maharishi University of Information Technology, Uttar Pradesh, India, Email Id- trapty@muit.in, Orcid Id- 0009-0007-4081-4999
[6]Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh-174103 India yuvraj.parmar.orp@chitkara.edu.in https://orcid.org/0009-0007-1619-9885

The development and maturation of criminal investigation have been highlighted by their use in security and criminal investigation throughout the previous three decades. Examine their advanced software for use in criminological analytics and the attendant difficulties that have arisen there. It expands our focus to include Artificial intelligence (AI) and big data, offering cutting-edge tools that can revolutionize police work. Recent developments in the field examine how big data can be used to collect security intelligence in the framework of criminal analytics. In this study, we suggested the Optimized Honeybee Boosted Support Vector Machine (OH-BSVM) to use big data in criminal investigations. First, we collect the data on big data from criminal investigations. Then, the Z-score normalization was applied during the data preprocessing step. Different feature extraction parameters were examined using Independent Component Analysis (ICA). In contrast, the OH-BSVM method outperformed the comparison in several performance requirements, including F1-score (92%), accuracy (97%), precision (95%), and recall (93%). Our result shows the contribution by promoting a revolutionary big data analytics-based criminal analytics technique and discusses the benefits and difficulties of using big data analytics for investigation reasons.

Keywords: Big Data, Criminal Investigation, Optimized Honeybee Boosted Support Vector Machine (OH-BSVM); Artificial intelligence (AI).

1. Introduction

There is a significant rise in crime investigation, which presents a challenge to a city's police force. Various law enforcement agencies have collected enormous amounts of data over the past few decades. Data encompasses information on multiple crimes investigated in different places throughout a given nation rather than limited to a single type of crime [1]. The proper method of bringing the perpetrator to justice has been solved the offenses. Furthermore, data presents more opportunities and problems for analysts and researchers alike. Researchers can establish connections between the characteristics of the data to assist law enforcement in apprehending the offender [2]. A distinct method must be used when analyzing crime data to identify trends and distinctive patterns in the crime reports. When additional information about the event, including the criminal's background and social network and geographic information like the date and location of the incident, is factored into the analysis, crime prediction becomes difficult and complex task. Finding the perpetrators of a criminal act is the first and most critical stage in any investigation into a criminal incident [3]. Researchers and analysts use big data analytics to perform the analysis, enabling the identification process. The study is done with the data gathered from the crime scene. To make the most efficient use of human and technical resources, it is necessary to locate the individuals believed to be responsible for the accidents [4]. Crime-analysis technologies have been developed, and the security forces are using them to solve occurrences like these. However, with the assistance of the tool, the most crucial criterion that is needed for the software to function is the date and the geographical location of the incident [5]. Utilizing the instrument and the available human resources, it is not possible to review such a massive amount of data. Consequently, where big data comes into play, which can be utilized in crime investigation to discover trends and forecast future crimes. The degree of lives kept in data generation is at a point where the complexity of the data is increasing. This is true whether the data is in social posts or a straightforward multimedia message [6]. The data were organized or unstructured in the past, but they are in various forms, all mixed together, making understanding them significantly more difficult. The term big data refers to the process of compiling information on an extremely broad scale, whether for conducting business, providing web services, or dealing with retail concerns. Databases can store massive amounts of data using the benefits vendors' offer [7].

The rest of the paper is divided into Section 2, which provides relevant studies; Section 3, which describes the methodology; Results and discussion are covered in Sections 4 and 5; and the end of the paper concluded in Section 6.

1.1 Research Gap

AI-Big Data integration in criminal investigations has examined AI and Big Data's separate contributions, but they did not read their seamless integration and cooperation in criminal investigations. AI algorithms for large, heterogeneous datasets, notably criminal data, have been optimized in research. Ethical, privacy and standardized frameworks for law enforcement AI and Big Data adoption are also needed. This study gap underlines the need for AI and Big Data studies to improve criminal investigations. The novel and effective Optimized Honeybee Boosted Support Vector Machine (OH-BSVM) smoothly integrates AI and Big Data in criminal investigations to fill these research gaps. OH-BSVM optimizes AI algorithms on large, heterogeneous crime datasets. Our strategy increases criminal data analysis accuracy and efficiency, utilizing boosted SVMs and honeybee-inspired optimization. Our research tackles ethical, privacy, and standards challenges to guarantee law enforcement legitimately uses AI and Big Data. By closing these gaps, our technique enables AI-Big Data synergy for ethical and successful criminal investigations.

1.2 Significance of the study

Integrating AI and big data in criminology is paramount in police forces. This multidisciplinary method uses the analytical capacity of AI algorithms to examine huge amounts of data, ranging from surveillance video to social media activity. Investigators can save time by skipping the trial-and-error phase and going straight to the decision-making phase of machine learning. This cooperation helps authorities solve cases faster and prevent crimes before they happen. Law enforcement organizations can better anticipate and counteract criminals' changing strategies via real-time data analysis and predictive modeling.

1.3 Objective of the Study

This study examines how AI and Big Data might improve criminal investigations. The study examines how AI algorithms can investigate various data sets to provide law enforcement with relevant insights. The paper discusses the limits and ethical concerns of employing AI and Big Data in investigations. The study seeks to guide the creation of more effective, honest and privacy-aware law enforcement technologies by evaluating synergies and possible obstacles.

2. Related works

According to the author of, [8] examined how mathematical concepts are opening up a new vocabulary for criminal activity inside criminal justice organizations using machine learning, algorithmic analytics, and big data. It demonstrates how these cutting-edge instruments are eroding the protections built into existing regulatory systems, eliminating subjectivity, and eradicating case-by-case narratives. The study [9] hybridization of cloud computing, data mining, and large amounts of available internet data is a topic of discussion in the study. Data mining visualization refers to analyzing and presenting massive data sets. Conventions and methods in computing have been shown to impact how much space and bandwidth is needed to store and transfer data. Research [10] offered that the most probable perpetrator of a particular crime occurrence can be determined using historical data on comparable cases and incident-level crime data using a Naive Bayes classifier, which is a solution to the issue of crime prediction. A crime dataset detailing incidents is supplied, including information such as the kind of crime, the involved parties, the date and place of the occurrence, and the offender's unique identifier. The author of, [11] examined various real-world examples of clustering and big data analytics might be set to use. The number of people using this technology in e-learning, healthcare, and the Internet of Things (IoT) is fastly growing. After reviewing other studies, this investigation considers the problems associated with clustering and large datasets. For the sake of doing analytics on large datasets, the current study has used a state-of-the-art mechanism, the K-mean mechanism, to create dynamic clusters. Research [12] suggested the latest fingerprint classification algorithm and forensics research. Law enforcement, customs, forensics, and other public security agencies use fingerprints to keep order and investigate crimes. Techniques for capturing, identifying, classifying, and analyzing fingerprints have been reported using machine learning and neural networks. To investigated the impact of organizational justice and emotional intelligence on the climate of care and job happiness and the implications such factors have the performance of Criminal Investigation officers [13]. The suggested framework can extract the social network from chat logs and combine them into topics of conversation [14]. It's possible that crime scene investigators can use data visualization software to analyze the results of their cases. The tests' results on real-world data and the comments from law enforcement officials show that the recommended chat log mining framework is valuable for police agencies and can assist them in solving crimes. The suggested the pressing need in forensic genetics to investigate the DNA methods and data utilized and the underlying models to infer relatedness. The study compiles a wide range of relevant subjects into one place, where their efficacy and operational restrictions are analyzed, and potential next steps in their forensic validation are proposed [15]. Author [16] employed empirical legal research to see how the law operates in the community. This empirical legal research investigates community members in live interactions, making it sociological. The study is based on books, legislation, journals, legal experts' and academics' scientific viewpoints. Study [17] assessed the influence of these investigative variables on murder clearances after controlling for other covariates. Mediation analysis was used to dissect investigative resources' overall, direct, and indirect effects on murder clearances. To identify any differences between killings associated with gangs and drug use and those not associated with either, exploratory group comparisons. To analyze network traffic, a key part of the digital forensics methodology for finding and researching security breaches [18]. Many different models, each with advantages and disadvantages, have been proposed to analyze cybercrime. The research investigates some of the difficulties encountered in digital forensics and compares them to the features of popular digital forensic investigation models.

3. Methods

This paper provides a method for managing large volumes of data during a criminal investigation. Some modules in big data analytics are universal across all tasks in a given issue area. After gathering information from several sources, the next stage is to prepare it for analysis using various big data tools and methodologies. Analysts use different methods for sorting through the cleaned data in search of evidence of wrong. Finally, the criminal information (such as a criminal network) obtained automatically is used to aid in criminal investigations. The methodology for using big data in criminal investigation is shown in Figure 1.

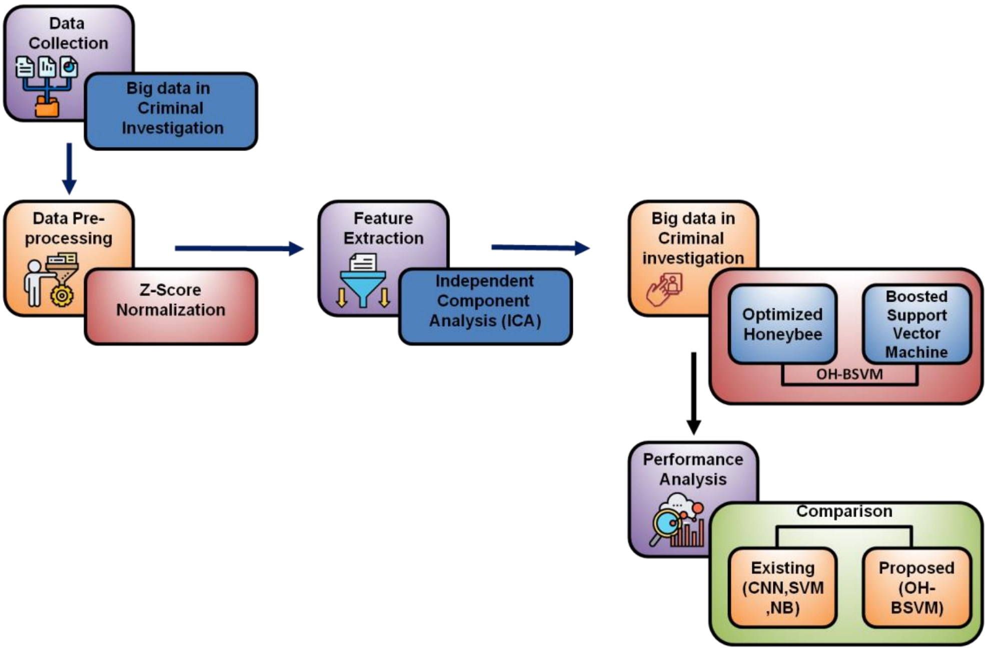

Figure 1: Methodology of Criminal Investigation [Source: Author]

3.1. Data collection

Security intelligence agencies gather information from various data sources to research criminal investigation networks. Primary data sources among the types of data that can be used for the relational analysis are call logs, surveillance footage, social network location data, financial data, and reports of criminal activity [19]. Bank accounts, transaction data, and phone call logs are the main tools used by police and intelligence to take down criminal networks. Textual data (such as emails and SMS texts) has been used in various security intelligence research initiatives to identify correlations and relationships between suspicious entities. Several social media platforms, including Facebook, Twitter, LinkedIn, and blogs, have been analyzed for the positioning analysis. Positional analysis does not focus on location-specific patterns; instead, it looks at the connectivity patterns across network members.

3.2. Data preprocessing using Z-score Normalization

There might be duplicate packets or even blank data with the acquired information. During this phase, any information found to be redundant, unnecessary, or duplicated is removed. Academic systems use sample methods due to the massive amounts of data contained inside them. Due to the vast number of variables in this dataset, researchers performing the study must adopt a feature extraction approach to eliminate the extraneous information. The normalization method is used to maintain consistency across all data. As a result, normalizing the data set is a crucial procedure. The z-score is determined as the initial step in the normalization process. The z-score can be calculated using Equations (1-2), as seen below. In this equation, the mean value is shown by γ, while the symbol for the standard deviations.

$$ZS = \left[\frac{as}{(SM-\gamma)}\right] \qquad (1)$$

$$ZS = \frac{N-\overline{SM}}{as} \qquad (2)$$

Equation (2) gives the Z-scores calculation method, where $\overline{SM}$ represents the sample mean. The collection that was chosen at random adheres to the steps in Equation (3), which $\epsilon_l$stands for the inaccuracies.

$$M_z = \delta_0 + \delta_1 SM_l + \epsilon_l \qquad (3)$$

Therefore, the errors must be unrelated, as Equation (4) shows, where y represents any variable.

$$SM_l \sim \sqrt{D}\frac{y}{\sqrt{y^2+S-1}} \qquad (4)$$

The momentarily scaled variation can be computed using Equation (5, 6, 7, and 8), where ms stands for momentary scaling.

$$ms = \frac{\lambda^{ms}}{\gamma^{ms}} \qquad (5)$$

$$\lambda^{ms} = A(sm-\sigma)\widehat{N}, \qquad (6)$$

$$X^{ms} = \left(\sqrt{A(sm-\sigma)\widehat{N}}\right)\hat{2} \qquad (7)$$

$$C_R = \frac{n}{sm} \qquad (8)$$

E is the anticipated value, and Z is a random variable if C_R is the variability factor. After changing all variables to 0 or 1, characteristic scaling is complete.

$$SM' = \frac{(SM-SM_{min})}{(SM_{max}-SM_{min})} \qquad (9)$$

The unison-based normalizing approach is used to describe this method to show the normalized Equation (9). Controlling the magnitude and the amount of fluctuations is possible with this data management method. Once the data has been normalized, it can be utilized by further steps in the process.

3.3. Feature extraction using Independent Component Analysis (ICA)

ICA is used to separate segment-specific feature vectors for further analysis. After training an ICA network to extract independent components Y, the learned weight matrix y removes the basis function coefficients U from u. The B ICA method presupposes that U is linear in its constituent parts. If E is the transpose of B, then its columns stand in the observation y's basis feature vectors.

$$\mathrm{q} = \mathrm{Y.t}, \mathrm{t} = \mathrm{W.q} \qquad (10)$$

The unfixing matrix w or the mixing matrix w are required to train the mixing matrix, from the basic functions are extracted. The maximization of joint entropy is the basis for the learning rule given by Equation (13) $M(z)$.

$$\Delta \mathrm{Y} \propto \frac{\partial \mathrm{H}(z,\mathrm{t})}{\partial \mathrm{Y}} = \frac{\partial \mathrm{M}(z)}{\partial \mathrm{Y}} \tag{11}$$

$$\Delta \mathrm{Y} \propto [\mathrm{Y}^{S}]^{-1} + \left(\frac{\frac{\partial \mathrm{v}(q)}{\partial \mathrm{q}}}{\mathrm{b(q)}}\right)\mathrm{t}^{S} \tag{12}$$

where $i(c)$is the approximate probability density function,$i(c_h) = \frac{\partial z_h}{\partial c_h} = \partial s(c_h)/\partial c_h$for a part of the voice signal. In this situation, the accumulative distribution function of the source signal C is similar to the nonlinearity function $g(c)$, as shown in Equation (11). To accelerate convergence, augment Equation (12) with a natural gradient. In particular, this method results in the following rule, which does not include the inverse of matrix Z.

$$\Delta \mathrm{Y} \propto \frac{\partial \mathrm{M}(z)}{\partial \mathrm{Y}} \mathrm{Y}^{S}\mathrm{Y} = [1 - \varphi(\mathrm{Q})Q^{S}]Y \tag{13}$$

3.4. Criminal Investigations using optimized Honeybee Boosted Support Vector Machine (OH-BSVM)

As it applies to criminal justice, big data describes the massive volume, variety, and velocity of digital evidence generated by criminal activity. Data from several sources, including social media, security cameras, financial transactions, and conversation logs, are included in this. Advanced machine learning methods, including the optimized Honeybee Boosted Support Vector (OH-BSVM) Machine, are used to analyze this enormous amount of data. A complex process called OH-BSVM combine’s boosted algorithms' effectiveness with support vector machines' strength.

3.4.1. Boosted Support Vector Machine (BSVM)

The architecture of the recommended BSVM is presented in Figure 2. The example shows Various system components: the Eigenvector xi, the kernel function K, the output vector W, and the bias term b. The ultimate decision function of the model is described as follows in Equation (14).

$$\mathrm{w(q)} = \mathrm{sgn}(\textstyle\sum_{\mathrm{h}=1}^{\mathrm{n}} \mathrm{Z_h K(q, Q_h)} + \mathrm{v}) \tag{14}$$

Linear, polynomial, and Multilayer Perceptron with Radial Basis kernel functions are applied to the BSVM to evaluate its final predictive capacity. The results are analyzed.

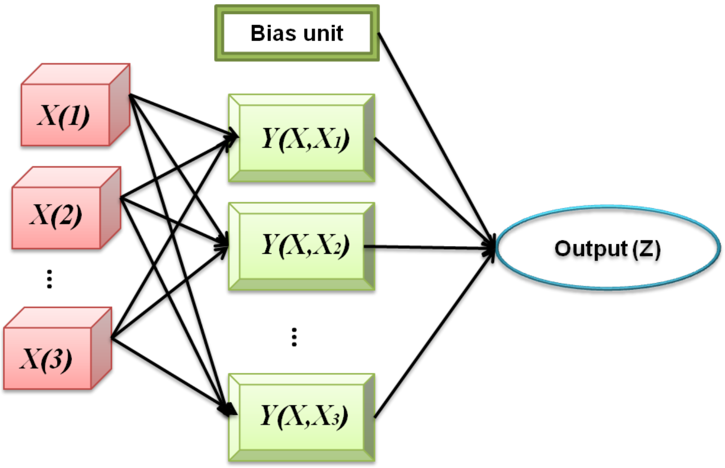

Figure 2: BSVM Structure [Source: Author]

The radial basis function provides the highest predictive power, as shown by the authors. It necessitates absolute improvement in fit quality, mistake rate, or time invested. One of the most prevalent possibilities is the Gaussian function. Therefore, this method uses a radial basis function as the kernel function in Equations (15-18).

$$P(q_i - s_h) = \exp\left(-\frac{1}{2\sigma^2} q_i - s_j^2\right), \quad (15)$$

Where, $q_i - s_h$is centre of the Gaussian function, the variance, and the Euclidean norm. The following is the result of a neural network architecture equipped with radial basis functions:

$$mh = \sum_{h=1}^{f} z_{hu} \exp\left(-\frac{1}{2\sigma^2} q_i - s_h^2\right), u = 1,2, \dots n \quad (16)$$

where$q_i = \left(q_1^i, q_2^i, \dots q_b^i\right)^A$is the oth input sample $(i = 1,2,3, \dots, O)$ is the overall count of pieces, $z_{hu}$is the weight of the connection $(h = 1, 2, 3, \dots, f)$, the total number of nodes is f. Using the least-squares method, can express the variance of the basis function as

$$\sigma = \frac{1}{i} = \sum_h^b x_u - m_h\, s_h^2 \quad (17)$$

The parameter value of the gamma function is set to negative, with the attribute count as a default. With the help of Equation (16 to 18) can be transformed into a novel decision-making function.

$$w(q)\mathrm{sgn}\left(\sum_{h=1}^{n} z_h \exp\left(-\mathrm{gamma} x_h - q^2\right) + l\right), \quad (18)$$

The gamma term (g) and penalty term (c) of the ABVM must be defined for optimal model performance. Interval size and classification accuracy metrics can be modified and enhanced the penalty period to go in the direction of the permissible mistake. When the value of the penalty term is tiny, under fitting is more probable, whereas when the value is large, overfitting is more likely. If the value of the punishment term is too high or too low, the BSVM's ability to generalize will be impaired. In addition, the gamma (g) value of the BSVM is overlooked despite its importance. The method relies on the value of g; many mappings exist between low- and high-dimensional samples. Overfitting is more probable with a higher g value, which reduces the ability to generalize. In Algorithm 1, the BSVM algorithm is shown.

Algorithm 1: Pseudo-code of BSVM algorithm

```
Inputs: Determine the various training and test data.
Outputs: Determine the calculated accuracy.
Select the optimal value of cost and gamma for BSVM
While (the stopping condition is not met) do
Implement the BSVM train step for each data point.
Implement BSVM classification for testing data points.
end while
Return accuracy
```

3.4.2. Optimized Honeybee Algorithm

An example of a metaheuristic that draws inspiration from honey bee foraging strategies is the Optimized Honeybee (OH) algorithm. It is not used for security but rather for optimization problems. However, look at the advantages of OH in the context of confidentiality. Honeybee colonies typically consist of a single queen bee from zero to thousand drones. The main purpose of the queen is to lay eggs. Depending on the conditions, a colony can be monogynous (with one queen) or polygynous (with several queens) throughout its life. Only the queen bee in the colony receives "royal jelly," a milky-white, jellylike substance produced by the "nurse bees" in their glands and fed to her. The queen bee consumes more royal jelly than any worker bee, contributing to her enormous size. There have been reports of queen bees living for up to six years, although worker bees and drones only have a few months and typically survive for around a year. A colony will typically include several hundred drones in addition to the queen and worker bees. Drones have a short lifespan after mating, they perish. Drones first established the settlement. Haploid kids increase the amount of their mother's genome but otherwise do not alter her genetic makeup. Drones are seen as agents because they disperse one of their mother's gametes, allowing females to exhibit male-like genetic characteristics. These employees have a specific skill set in caring for the young, including egg-laying. Eggs do not need to be fertilized to hatch. The former group includes queens and workers, while the latter comprises drones. Smart hunger can be used to create a versatile weight system. The authors optimized the search procedure by factoring in the impacts of desire at each stage. The first phase, as shown in Equation (19), characterizes individuals' interactions and hunting behavior:

$$O(n+1) = \begin{cases} O(n) * (1 + e_s(1)) \\ \omega_r' * O^v(n) + \bar{S}\omega_r' * |O^v(n) - O(n)|^{p1<x} p > x, p_2 < C \\ \omega_r' * O^v(n) + \bar{S} * |O^v(n) - O(n)|^{p1<x} p_1 > x, p_2 < C \end{cases} \tag{19}$$

Where p1 and p2 = Random numbers [0 to 1]

ω_r'andω_r'' = Weight value

x = Control variable

C = Control of variation in all locations

O^p = Position of the best individual

$O(n)$ = Position of all individual

After that, Equation (20) is used to calculate the variation control for each location as follows:

$$C = hyp(D(h) - X^D) \tag{20}$$

Where X (.) represents the cost function of population $h \epsilon 1,2, \ldots ., e,$ and the hyperbolic function can be approximated by applying the Equations (21, 22 and 23) that follows:

$$hyp(z) = \frac{2}{exp^z + exp^{-z}} \tag{21}$$

$$\bar{W} = 2 * v * e_s - v \tag{22}$$

$$v = 2 * \left(1 - \frac{n}{Utr_{max}}\right) \tag{23}$$

4. Results

A Windows 8.1 operating system, a 2.33 GHz CPU, and 4 GB of RAM were employed in the investigation. Throughout the testing procedure, Python was used. In this study, OH-BSVM methods were proposed for analyze the criminal investigations. Accuracy, precision, Recall, and f1-score are measures of a model's capacity to identify the outcomes of a criminal investigation using an AI-based technique. In contrast to the most recent and greatest techniques Convolutional Neural Network (CNN) [20], Support Vector Machine (SVM) [20], and Naive Bayes (NB) [20] and our suggested technique, OH-BSVM, perform much better.

4.1. Accuracy

The accuracy of big data in criminal investigations depends on numerous elements, including the quality of the data, the usefulness of analytical tools and algorithms, and the investigators' experience. Advanced machine learning techniques, such as support vector machines, boosted algorithms, or other optimization methods, can raise accuracy by discovering complicated relationships with the data. A reliable prediction happens when the predicted and actual categorizations align. To quantify accuracy, divide the number of right forecasts by the total number of suggestions. Table 1 and Figure 3 demonstrate the comparison of accuracy.

Table 1: Comparison of Accuracy

Methods	Accuracy (%)
CNN [20]	91
SVM [20]	84
NB [20]	83
OH-BSVM [Proposed]	97

When compared to CNN (91%), SVM (84%), and NB (83%), the OH-BSVM approach suggested here came out on top in terms of accuracy, obtained a phenomenal 97%.

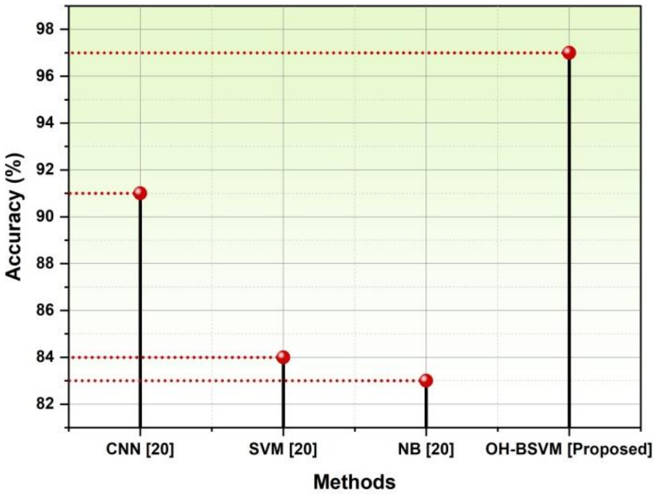

Figure 3: Accuracy

4.2. Precision

Precision means analysis accuracy and the capacity to find important data in a massive data set. With great precision, big data analytics for criminal investigations can distinguish relevant from irrelevant data, giving investigators trustworthy and actionable information. Precision is essential for law enforcement to make educated decisions and take appropriate steps based on big data, improving criminal investigations. A model calculates the proportion of effective positive predictions from the amount of true positive forecasts. Table 2 and Figure 4 compare precision.

Table 2: Comparison of Precision

Methods	Precision (%)
CNN [20]	92
SVM [20]	80
NB [20]	68
OH-BSVM [Proposed]	95

The following is a list of the precision percentages achieved by different methods: According to the research [20], CNN gained 92%, SVM reached 80%, NB obtained 68%, and the suggested OH-BSVM revealed a precision of 95%.

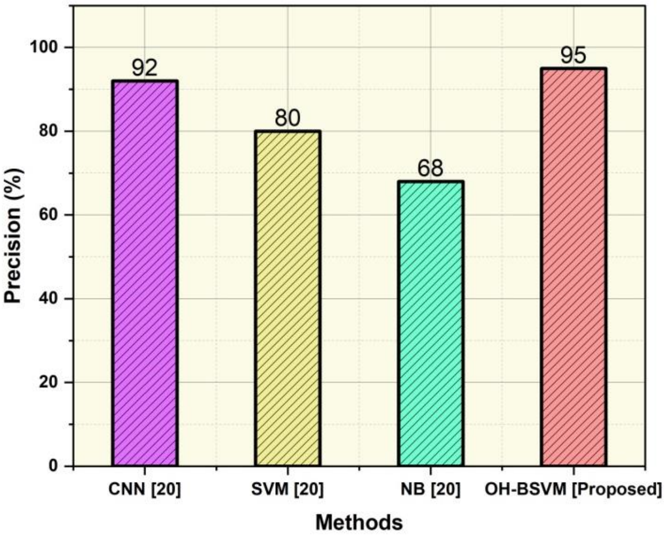

Figure 4: Precision

4.3. Recall

In criminal investigations, recall indicates a big data model identifies and retrieves illegal activity-related data. A high memory suggests that the model captures a considerable part of important data, reducing the risk of missing vital evidence or linkages throughout the investigation. This is significant in law enforcement because it helps investigators to notice key information that might solve a case or prevent crime. Table 3 and Figure 5 show the comparison of Recall.

Table 3: Comparison of Recall

Methods	Recall (%)
CNN [20]	82
SVM [20]	72
NB [20]	80
OH-BSVM [Proposed]	93

According to [20], an evaluation of many different classification strategies revealed that the suggested OH-BSVM beat the alternatives with an exceptional recall rate of 93%, outperforming CNN (82%), NB (80%), and SVM (72%), respectively.

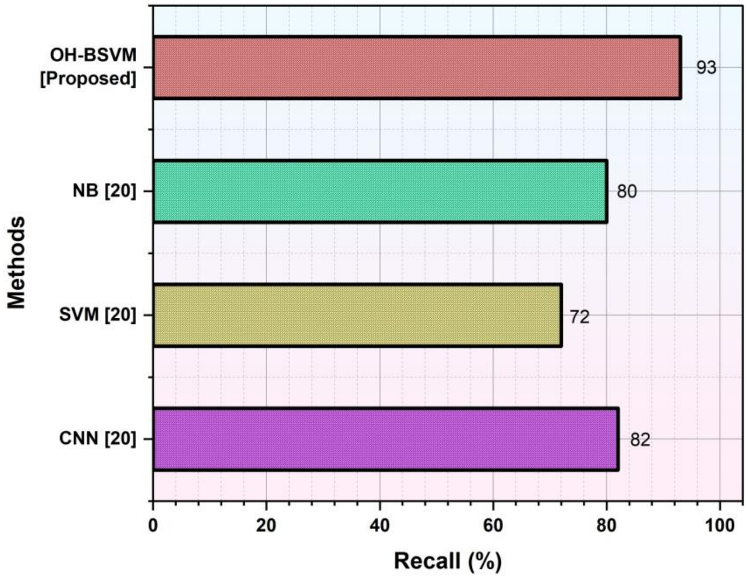

Figure 5: Recall

4.4. F1-Score

The F1-score may evaluate prediction models or algorithms for discovering and categorizing illegal activity in big data criminal investigations. A model's precision (positive prediction accuracy) and recall (capturing all relevant occurrences) are measured by the F1-score. Accuracy and recall are balanced in a model with a high-quality F1 score. Comparisons of the F1-score are shown in Table 4 and Figure 6.

Table 4: Comparison of F1-score

Methods	F1-Score (%)
CNN [20]	84
SVM [20]	74
NB [20]	72
OH-BSVM [Proposed]	92

According to the findings of the research [20], the CNN algorithm (84%), the SVM algorithm (74%), and the NB algorithm (72%) all performed worse than the suggested OH-BSVM approach, which achieved a great F1-Score of 92%.

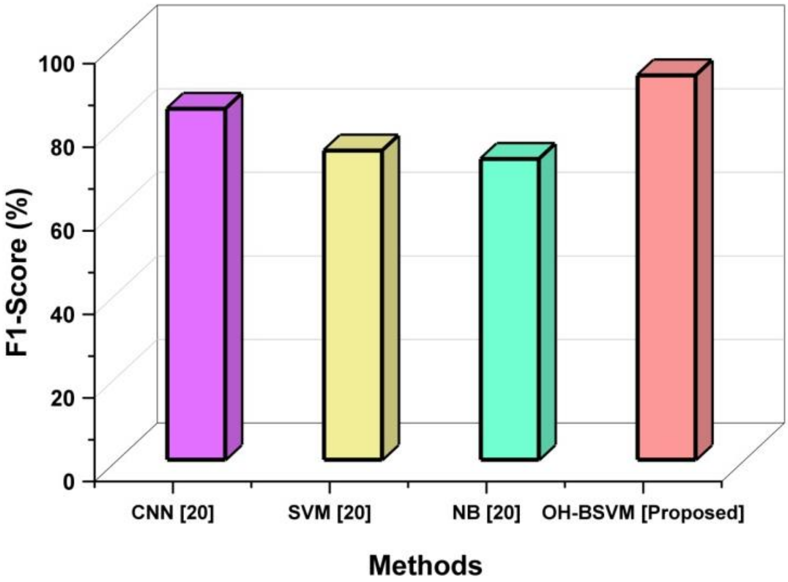

Figure 6: F1-score

The solar cell model parameters are estimated using a HWSSO. Table 1 display the solar cell parameter values WOA, SA, and our proposed HWSSO approach. Figures 2 and 3 show the variation among the I-V characteristics obtained from the WOA method parameter and the actual data. Figures 4 and 5 show the variation between the SA method parameter-derived I-V characteristics and the actual data. The HWSSO system parameter values and the I-V characteristic derived from real data are perfectly matched, as seen in Figure 6.

5. Discussion

The study discussed how machine learning was used to examine intelligent enforcement technologies that can help law enforcement agencies respond to new offenses. The architecture of the prediction model OH-BSVM is developed. Evaluation is done on the prediction models performed. Existing methods such as SVM [20], CNN [20], and Naive Bayes (NB) algorithms [20] are outperformed by criminological investigation prediction models that make use of the recommended designs. Compared to the SVM, a criminal prediction model based on CNN data and naïve Bayes algorithms perform 7% and 8% better, respectively. Our suggested model enhances the accuracy of the already available models. In addition, a functional verification of the GUI-based smart polishing system's functioning in real time is carried out. The OH-BSVM improves classification accuracy, identifying criminal tendencies precisely. This combination enhances decision-making, predictive policing, and crime prevention. As law enforcement uses these new tools, criminal investigations become more efficient and effective.

6. Conclusion

The study provides academics and professionals with a historical perspective on numerous methodologies and technologies, including massive amounts of information, to better police work and national safety. To set up a fully automated system for security and criminal investigations, we find BSVM machine learning approaches that have been widely used. The potential of using big data for policing and security purposes seems promising. However, new techniques and analytic tools must be explored to overcome the fundamental difficulties of large data and enhance criminal investigations. Big data analytics might transform how law enforcement and security intelligence companies gather essential information (such as criminal networks) using several data sources in real-time to speed up and improve investigations. Security intelligence services might use real-time data to prepare for and respond to organized crimes like terrorism and human trafficking. This study presents OH-BSVM methods for recognizing large datasets used in criminal investigations. In this study, the OH-BSVM was generated using our suggested method, and the results showed an F1-score of 92%, 97% accuracy, 95% precision, and 93% recall. In the future, big data analytics to aid security intelligence services seek justice. As analytics platforms, tools, and methodology develop, law enforcement and security intelligence agencies will use big data in the future. This big data-inspired Position says bullets and explosives won't fight crime, extremism, and terrorism.

References

1. Hannah-Moffat, Kelly. "Algorithmic risk governance: Big data analytics, race, and information activism in criminal justice debates." Theoretical Criminology 23, no. 4 (2019): 453-470. https://doi.org/10.1177/1362480618763582
2. Yaksic, Enzo. "Evaluating the use of data-based offender profiling by researchers, practitioners, and investigative journalists to address unresolved serial homicides." Journal of Criminal Psychology 10, no. 2 (2020): 123-144. https://doi.org/10.1108/JCP-09-2019-0032
3. Ariel, Barak, and Matthew Bland. "Is crime rising or falling? A comparison of police-recorded crime and victimization surveys." In Methods of Criminology and criminal justice research, vol. 24, pp. 7-31. Emerald Publishing Limited, 2019. https://doi.org/10.1108/S1521-613620190000024004
4. Hariri, Reihaneh H., Erik M. Fredericks, and Kate M. Bowers. "Uncertainty in big data analytics: survey, opportunities, and challenges." Journal of Big Data 6, no. 1 (2019): 1-16. https://doi.org/10.1186/s40537-019-0206-3
5. Kalynovskyi, Oleksandr, Viktor Shemchuk, Mykhailo Huzela, and Halyna Zharovska. "Fighting crime through crime analysis: The experience of using innovative technologies in European Union countries." Cuestiones Políticas 41, no. 76 (2023). https://doi.org/10.46398/cuestpol.4176.16
6. Moran, Gillian, Laurent Muzellec, and Devon Johnson. "Message content features and social media engagement: evidence from the media industry." Journal of Product & Brand Management 29, no. 5 (2020): 533-545. https://doi.org/10.1108/JPBM-09-2018-2014
7. Horan, Cecelia, and Hossein Saiedian. "Cyber crime investigation: Landscape, challenges, and future research directions." Journal of Cybersecurity and Privacy 1, no. 4 (2021): 580-596. https://doi.org/10.3390/jcp1040029

8. Završnik, Aleš. "Algorithmic justice: Algorithms and big data in criminal justice settings." European Journal of Criminology 18, no. 5 (2021): 623-642. https://doi.org/10.1177/1477370819876762
9. Ageed, Zainab Salih, Subhi RM Zeebaree, Mohammed Mohammed Sadeeq, Shakir Fattah Kak, Hazha Saeed Yahia, Mayyadah R. Mahmood, and Ibrahim Mahmood Ibrahim. "Comprehensive survey of big data mining approaches in cloud systems." Qubahan Academic Journal 1, no. 2 (2021): 29-38. https://doi.org/10.48161/qaj.v1n2a46
10. Kumar, Ravi, and Bharti Nagpal. "Analysis and prediction of crime patterns using big data." International Journal of Information Technology 11 (2019): 799-805. https://doi.org/10.1007/s41870-018-0260-7
11. Gupta, Ankur, Ram Singh, Vinay Kumar Nassa, Rohit Bansal, Priyanka Sharma, and Kartikey Koti. "Investigating application and challenges of big data analytics with clustering." In 2021 international conference on advancements in electrical, electronics, communication, computing and automation (ICAECA), pp. 1-6. IEEE, 2021. https://doi.org/10.1109/ICAECA52838.2021.9675483
12. Win, Khin Nandar, Kenli Li, Jianguo Chen, Philippe Fournier Viger, and Keqin Li. "Fingerprint classification and identification algorithms for criminal investigation: A survey." Future Generation Computer Systems 110 (2020): 758-771. https://doi.org/10.1016/j.future.2019.10.019
13. Sembiring, Nurdin, Umar Nimran, Endang Siti Astuti, and Hamidah Nayati Utami. "The effects of emotional intelligence and organizational justice on job satisfaction, caring climate, and criminal investigation officers' performance." International Journal of Organizational Analysis 28, no. 5 (2020): 1113-1130. https://doi.org/10.1108/IJOA-10-2019-1908
14. Iqbal, Farkhund, Benjamin CM Fung, Mourad Debbabi, Rabia Batool, and Andrew Marrington. "Wordnet-based criminal networks mining for cybercrime investigation." Ieee Access 7 (2019): 22740-22755. https://doi.org/10.1109/ACCESS.2019.2891694
15. Kling, Daniel, Christopher Phillips, Debbie Kennett, and Andreas Tillmar. "Investigative genetic genealogy: Current methods, knowledge and practice." Forensic Science International: Genetics 52 (2021): 102474. https://doi.org/10.1016/j.fsigen.2021.102474
16. Zulyadi, Rizkan. "Police's Role in Investigation Process of Fraud Criminal Act of Civil Servants Candidate (Case Study of Police Station Binjai)." Britain International of Humanities and Social Sciences (BIoHS) Journal 2, no. 2 (2020): 403-411. https://doi.org/10.33258/biohs.v2i2.238
17. Braga, Anthony A., Brandon Turchan, and Lisa Barao. "The influence of investigative resources on homicide clearances." Journal of Quantitative Criminology 35 (2019): 337-364. https://doi.org/10.1007/s10940-018-9386-9
18. Kumar, Saurabh, Suryakant Pathak, and Jagendra Singh. "An enhanced digital forensic investigation framework for XSS attack." Journal of Discrete Mathematical Sciences and Cryptography 25, no. 4 (2022): 1009-1018. https://doi.org/10.1080/09720529.2022.2072424
19. Pramanik, Md Ileas, Raymond YK Lau, Wei T. Yue, Yunming Ye, and Chunping Li. "Big data analytics for security and criminal investigations." Wiley interdisciplinary reviews: data mining and knowledge discovery 7, no. 4 (2017): e1208. https://doi.org/10.1002/widm.1208
20. Baek, Myung-Sun, Wonjoo Park, Jaehong Park, Kwang-Ho Jang, and Yong-Tae Lee. "Smart policing technique with crime type and risk score prediction based on machine learning for early awareness of risk situation." IEEE Access 9 (2021): 131906-131915. 10.1109/ACCESS.2021.3112682