IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 5, 2021, accepted April 16, 2021, date of publication April 22, 2021, date of current version May 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3075140

An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach

SAPNA SINGH KSHATRI[1], DEEPAK SINGH[2], BHAVANA NARAIN[1], SURBHI BHATIA[3], MOHAMMAD TABREZ QUASIM[4], (Senior Member, IEEE), AND G. R. SINHA[5], (Senior Member, IEEE)

[1]Department of Computer Science and Application, MATS University, Raipur 492001, India
[2]Department of Computer Science and Engineering, Bennett University, Noida 201310, India
[3]Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, Al-Hasa 36362, Saudi Arabia
[4]College of Computing and Information Technology, University of Bisha, Bisha 67714, Saudi Arabia
[5]Department of Electronics & Communication Engineering , Myanmar Institute of Information Technology (MIIT), Mandalay 05053, Myanmar

Corresponding author: Sapna Singh Kshatri (sapna.chandel.2002@gmail.com)

ABSTRACT Ensemble learning method is a collaborative decision-making mechanism that implements to aggregate the predictions of learned classifiers in order to produce new instances. Early analysis has shown that the ensemble classifiers are more reliable than any single part classifier, both empirically and logically. While several ensemble methods are presented, it is still not an easy task to find an appropriate configuration for a particular dataset. Several prediction-based theories have been proposed to handle machine learning crime prediction problem in India. It becomes a challenging problem to identify the dynamic nature of crimes. Crime prediction is an attempt to reduce crime rate and deter criminal activities. This work proposes an efficient authentic method called assemble-stacking based crime prediction method (SBCPM) based on SVM algorithms for identifying the appropriate predictions of crime by implementing learning-based methods, using MATLAB. The SVM algorithm is applied to achieve domain-specific configurations compared with another machine learning model J48, SMO Naïve byes bagging and, the Random Forest. The result implies that a model of a performer does not generally work well. In certain cases, the ensemble model outperforms the others with the highest coefficient of correlation, which has the lowest average and absolute errors. The proposed method achieved 99.5% classification accuracy on the testing data. The model is found to produce more predictive effect than the previous researches taken as baselines, focusing solely on crime dataset based on violence. The results also proved that any empirical data on crime, is compatible with criminological theories. The proposed approach also found to be useful for predicting possible crime predictions. And suggest that the prediction accuracy of the stacking ensemble model is higher than that of the individual classifier.

INDEX TERMS Boosting classifier, ensemble classifiers, Indian crime prediction, machine learning, statistical classifiers, stacking.

I. INTRODUCTION

Recently, a lot of research and predictions have been attempted on how to curb crimes by various criminologists and researchers using different modeling and statistical tools. As the rate of crime is still on the hike, therefore, there is a potential need of some important research that can help the policy makers and the concerned department about challenges and issues in the area of crime prediction and control mechanisms. Skillset of human fails to keep track of criminal records, if handled manually. So, there is need for identifying in a novel way, which will help in analyzing crime related information. Analysis on crime prediction is currently based on two significant aspects, prediction of crime risk field [1], [2] and crime hotspot forecast [3]. Data processing techniques are applied to facilitate this task. The expanded accessibility of computers and data innovations have empowered law authorization offices to incorporate broad databases

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Moinul Hossain.

with detailed information about major felonies, such as murder, rape, arson etc. In recent years, huge number of crimes is being reported in the world. Violence is a major crime in which a criminal threatens to use force on the victim. It refers to both crimes in which a violent act is the motive, such as murder or rape, robbery, as well as crimes in which violence is used as coercion. Crime may not necessarily, be initiated using weapons, depending on the jurisdiction, and violence crimes can range from murder to harassment. Typically, violence crimes include murders, robberies, rapes, attempt to murder, kidnapping, thefts, riots, dowry death, dowry atrocity etc. The rate of violent crimes is very high in 22 districts of Madhya Pradesh state and 23 districts of Rajasthan state in India; whereas the number of states is 34 and 25 respectively for the above-mentioned states concerning the crimes against women. Commenting on the factual data in terms of the number of murders, it is reported that 17 districts listed among the poorest counties were in Maharashtra, with Andhra Pradesh and Rajasthan (having the next largest concentrations of violence crime districts). The crimes generally fall in the areas of violence criminal activities mostly against women. Similarly, defying public order as a crime was reported in 13, 12 and 11 districts of Rajasthan, Bihar and Tamil Nadu, respectively, out of total 100 districts [4].

The crime predictions are generally suggested by using machine learning techniques with respect to what percentage of future violence is possible in crimes. This research has been done for many years, but with some limited algorithms and small dataset. This research claims its novelty with the help of empirical analysis of machine learning and other contributions listed in this section. Though, machine learning models are widely used in crime prediction, but still despite of its expanding application and its gigantic potential, there are numerous regions, where the new procedures created in the zone of artificial intelligence have not been completely explored and has major drawbacks. The most common approaches which have reported achievable accuracy in machine learning classifiers are Random Tree Algorithm, K-Nearest Neighbor(KNN), Bayesian model, Support Vector Machine (SVM), Neural Network [5].

Among these algorithms, crime prediction technique is proposed by integrating a number of algorithms named as a crime prediction ensemble model using bagging and stacked ensemble techniques, reflecting the beauty of this research work. Ensemble model is a method for constructing a predictive model by combining multiple models to solve a single problem in order to improve predictive efficiency. Ensemble learning techniques have been shown to be important in many applications [6]. Introduction of prediction ensembles in ensemble learning based models is used to combine the output capability of more than one classifier to generate the final result. Ensembles can be compared to the efficiency that are generated with the previous single learning classifiers. This work proposes an approach for constructing an ensemble based classification model called ensemble-approach crime prediction method (SBCPM) for predicting crime, representing a supervised based technique, also known as tree-based or classification approach. The contributions of the research work are explained below:

- Multi-level method as crime stack is presented and tested for crime prediction tested on real-time database.
- Crime stack is developed as the combination of machine learning, and ensemble learning -based techniques.
- Stack generalization is used to minimize error rate.
- The accuracy is improved by applying the principle of piling in order to estimate violent crime.

The organization of the paper is as follows: Literature Review is presented and discussed in section 2. Section 3 describes the Crime Stack methodology and the data set information that is used for experimentation and evaluation. In section 4, descriptions of the evaluation criterion are detailed with the results obtained. Finally, the paper concludes with the future scope in Section 5.

II. LITERATURE SUMMERY

Some methods are reported in current literature for predicting crime data. Lawrence McClendon and Natarajan Meghan than presented a comparative study on linear regression, addictive regression and decision tree and found linear regression algorithm to be very effective and accurate in predicting the crime data based on the training set input for the three algorithms [7].

A survey [8] found that using efficient data collection and data mining techniques to create a better crime prediction using knowledgeable learning to develop multiple models for single problem solving will improve crime forecasting. Prediction output of a single classification that aids in predicting what the next crime may be in a specific district within a given time period and identifies the season and Crime has a time dimension in which it occurs more often.

Babakura *et al.* presented a comparison between Naïve Bayesian and Back Propagation to predict the crime data and classified on various levels of crime rate such as low, medium and high. The accuracy, recall and precision were also calculated. The Naïve Bayesian was found performing better for data classification tested over crime dataset using WEKA [9].

Yadav *et al.* employed different types of machine learning methods supervised as well as unsupervised. Clustering, k-means clustering, Naïve Bayes, Regression methods are studied for analysis of crime rates and their impact based on criminal data [10]. Zhe Li *et al.* presented prediction of crimes of China tested over different season's data [7].

Sivaranjani *et al.* used K-means clustering, hierarchical clustering and DBSCAN clustering for analysis of crime data of cities in Tamilnadu state of India. Liao *et al.* suggested a novel method of prediction of crimes using Bayesian learning based on different geographic data [7].

Hazwani *et al.* presented a comparative study between different machine learning methods such as SVM, fuzzy theory, artificial neural network. The multivariate time series report was presented as a result of an extensive comparison of crime prediction methods. The futuure scope still explained the limitation of current methods for obtaining better accurate results and good performance by optimizing and tuning the parameters [11].

LuizG. *et al.* suggested random forest method for prediction and analysis of crimes using some urban indicators and homicides [12]. This study is based on forecasting crime rate, and number of criminal activities in different indian states A number of regression techniques were used but the prediction results claimed on limited dataset were not upto mark. The future scope opened direction on applying several data engineering, Creating an Intrusion Detector Network, for Creating a Prediction Model [12], Super learning approaches are used in a machine learning environment to predict drug use disorder care [13] and filtration tasks for appropriate and more accurate analysis.

Xiangyu *et al.* utilized spatiotemporal patterns in urban data in one borough in a city, and then leveraged the transfer learning techniques to reinforce the crime prediction of other boroughs. A novel transfer learning framework is used to integrate these features and model patio-temporal patterns for crime prediction [14]. The Indian numerical plates [13] were introduced using fuzzy based approach and 95% accuracy was achieved. Tilted and noisy images have not been checked, but blurred images were used. The authors also proposed second model for Myanmar numbers and characters as well [15], further elaborating the concept for identifying vehicles used in crime prediction.

Ensemble learning offers robust methodologies to handle the uncertainties in most complex industrial problems. [16]. In another side Anifowose, F *et al.* present ANN introduces an ensemble model in a different direction that combines various outputs of seven 'weak' learning algorithms and compared to the individual ANN, ANN-bagging and Random Forest to create an ensemble solution for the prediction of petroleum reservoir porosity and permeability [17]. In another research They suggest the Artificial Neural Networks ensemble model (ANN). Using standard decision rules, the performance of the ensemble model was evaluated and compared to those of ANN-Ensemble with the traditional Bootstrap Aggregation method and Random Forest. The results showed that the suggested method outperformed the others with the highest coefficient of correlation and the least errors [18]. The ANN ensemble model was developed with ten simple ANN model learners, each using a randomly generated number of hidden learners. For neurons, Each student contributed to improved prediction accuracies of reservoir properties for improved hydrocarbon exploration, development, and management activities, and demonstrated the principle of randomization of the number of hidden neurons, demonstrating the great potential for applying this learning paradigm to the characterization of petroleum reservoirs [19].

The number of research work presented above focuses mainly on the analysis of crime data, and is actually very restricted in the predictive modeling of crime and its rate. Also, the studies which presented researches in terms of prediction and classification using machine learning techniques do not discuss decent accurate results and this gives a major research gap and opens space of new work in crime prediction and its analysis.

III. EXPERIMENT

The results of the data analysis are reported in this section. The dataset is integrated by collecting data from various domains like murder, robbery, kidnapping, and criminal data. The prepared dataset is preprocessed and 5-fold cross validation test is applied for evaluating using the proposed model. The technique for testing predictive models is obtained by dividing the original data set into a model training sample and a test set for assessment.

A. DATA SOURCE AND SELECTION

The dataset is prepared collecting crime data from all states in India. Collected data contains factual reports on murder, rape and theft, which are considered as main elements. Basically, crime records from NCRB (National crime record bureau) are taken which are considered as easily accessible All violent crimes in crime data have been recorded in India over the last 15 years from 2001 to 2015, for a simple yet effective demonstration of our method, without distinction between our forms.

A total of around 60,000 cases of crimes have occurred which are divided into different states (7 Association domains and 28 states with the corresponding regions) of India. Information on the nature of crime is used in the proposed prediction model. Every offence has two connections: the beginning and the end times. In order to minimize, operation confusion, the starting time of each event are considered, which the time slot of corresponding year. In the dataset, various types of offences (crimes) is seen in the data set such as murder, assault, stealing, dacoits, theft, robbery, cheating, related crime, and so on.

- To validate our approach, we have used a dataset with 180 instances.
- The figure consists of 36 attribute fields, which are characterized in the next section on the left by the Attributes and Information sub window that shows the various fields in the database. The crime database includes 36 fields including several states on India such as Andhra Pradesh, Assam, Bihar Chhattisgarh, Goa etc. as shown in Figure 1.

The datasets from different stages and union territories are used for analysis and visualization as shown with different attributes in different regions. The visual data representation is shown in Figure 2. The objective of this paper is to examine the evaluation performance on the basis of advanced data mining techniques with 5-fold cross validation technique, with the proposed assemble-stacking based crime prediction method (SBCPM).

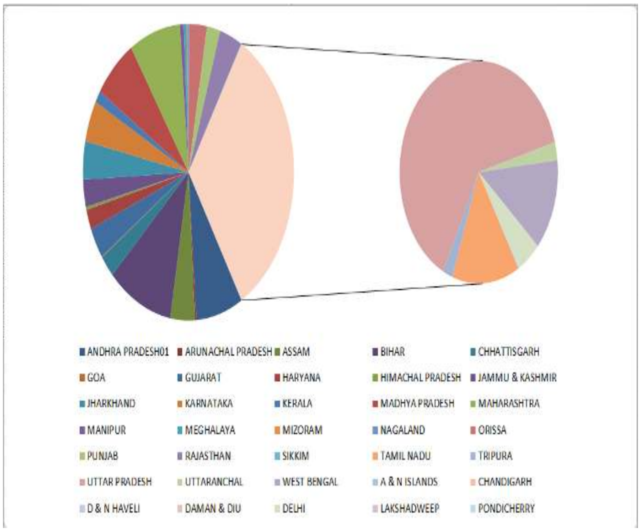

FIGURE 1. Crime distribution over the selected regions.

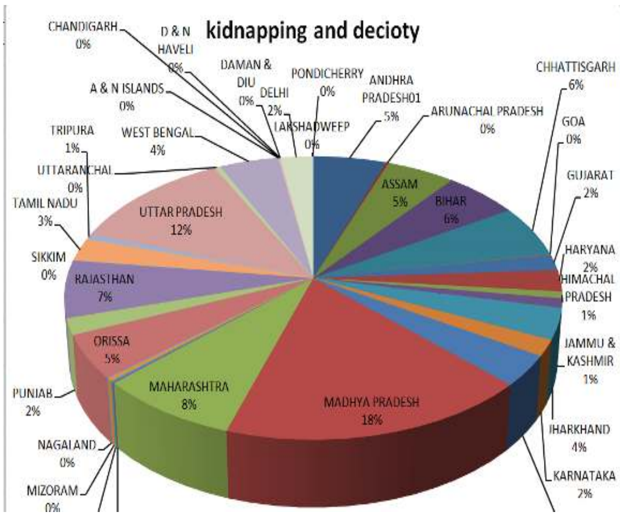

FIGURE 2. Visual representation of data of kidnapping and dacoits.

The crime historical data has been provided by National crime record bureau (NCRB). Quite comprehensive, it was chosen to challenge crimes, primarily to predict crime. It is well visible that it is divided into many outlets, but it is difficult to see and explore all of these crimes. Therefore, the dataset has been mixed with coherence by incorporating some data related to crimes. Violence crimes and FIRs have been combined with similar characteristics of violent crime outcomes so that the percentage of overall crime of violence is known.

B. EXTRACTION, TRANSFORMATION AND REDUCTION OF DATA

The primary aspect is the description of the class of each record and the class characteristics of the individuals arrested during the year, as well as the distribution of crimes as a characteristic from levels 1–5. When data set contains non-existent data that is incomplete (missing), noisy (external), and inconsistent, pre-processing of the data set is necessary. This was referred to as ‘‘murder, robbery, kidnapping, theft and FIR’’ as elements. Listing year is added as a new feature to the year of listing and to reduce the size of the dataset, and all records with missing values in the dataset have been closed. All records with missing values were eliminated in the dataset to reduce the size of the dataset. The final dataset contains 1,860 elements in datasets.

C. ANALYSIS PROCEDURE

A major problem in this particular study is to understand the classification of crime rates as low and high. There are machine learning classifiers available to answer this problem, which would identify the ‘‘crime rate’’ based on the characteristics described by the algorithm. The six algorithms are used in implementation as classifiers such as single classifier J48, SMO, Naive Bayes and random forest, bagging classifier in ensemble, and the proposed algorithm stacking crime prediction model with MATLAB.

D. CLASSIFICATION

Classification is a technique in which one can categorize data into given number of classes. The fundamental objective of a grouping problem is to identify the class. It is important to categories different forms of intrusions. Stacking is an effective pattern recognition and ensemble-based machine learning algorithm [20]. There are several crimes, and each one has been labeled differently based on its type, such as murder, dowry death, and attempted murder, which are all considered to be of the same nature, so labels 1 through 5 were assigned. Similarly, in Table 1, the collection is organized by the type of the offence.

In this classification method, data mining is performed to predict the class name of a hierarchical event based on a

TABLE 1. Types of violence crime.

Number	crimes
1	C.H. not amounting to murder
2	Attempt to murder
3	Murder
4	Dowry death
5	Rape
6	Kidnapping & Abduction
7	Dacoity
8	Preparation and assembly for dacoity
9	Assembly for robbery
10	Robbery
11	Riots
12	Arson
13	Total no of FIR

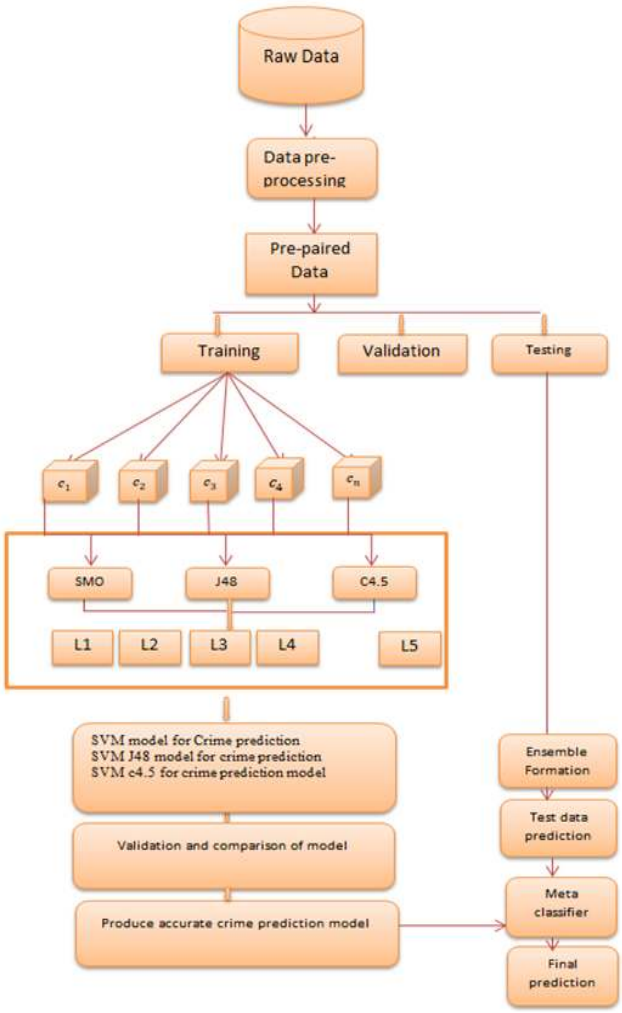

FIGURE 3. Proposed classifier for ensemble stack based crime prediction model.

classification model using attribute estimation. classification is divided into two stages 1) development of model and 2) implementation of model. The test model is established in the initial process. In the second stage, the crime prediction process is completed by selecting the best model from the model created. These models will be chosen based on the accuracy and timing of the model.

IV. PROPOSED APPROACH

The workflow of the studies, the computational measures, and the methods involved in the analysis are schematically established and are defined in Figure 3, consisting of three steps:

(1) The preparation of a combined machine learning stack-based crime prediction model.
(2) To identify and examine the nature of crime Investigation of computer empiric research and building crime prediction models by applying multiple basic simple and aggregate machine learning estimate the crime rate, pre-determined factors of crime, implement multiple ensemble machine learning (ML) models (bagging, boosting and stacking);
(3) Comparing and validating the output by checking the performance of the model.

The following flowchart describes the complete method, and the workflow of the proposed model is described as shown in Figure 3. The initial step of the model is data pre-preparing (DP) which involves filling the missing qualities, information cleaning, and change of information.

Five algorithms used in the data set are selected and checked the accuracy of the explorers: "classifier"

- Classifiers-are a template for predicting numeric quantities.
- Decision trees and lists-e.g., classifiers, vector machines, multi-layer assumptions, logistic regression, Bayes' support.
- "Meta"-Classifiers include: Bagging, Boosting, Error Code Error Correction, and Private Weighted Learning.

A. MODIFIED ENSEMBLE STACKING CLASSIFIER

Support vector machines (SVM) are one of the promising machine learning tools that have performed very well in most prediction problems [21]. The most commonly used mixing rule is the dominant part of casting a ballot which has two variants, i.e., straightforward and weighted lion's share voting procedure. For ensemble development, the proposed NSGA-II-Stacking approach uses the stacked speculation approach with multi-target streamlining calculation. For joining the meta-data, three kinds of meta-learners are used such as specific direct SMO (L-SVM), decision tree (DT), and j48. We investigated various meta-learners where a mixture of the chosen base-learners and each meta-learner were chosen and tested.

Stack normalization is a common approach among machine learning algorithms. several of experiments were performed to align the suggested method with some well-known data mining dataset ensemble techniques. This technique is often applied to learning sets for a cost-sensitive data mining problem in the real world. The results of the experiment show that better stacking assemblies can be created by the new approach.

The proposed model is based on J48 classifiers and C4.5 classifier, which are consolidated to form a crossover model. Defects of individual classifiers are addressed by the assembled classifier used for prediction purposes.

V. COMPARISON OF DIFFERENT CRIME PREDICTION MODELS

In the proposed protocol, we consider the problem of predicting the type of crime in the grid cell as the next step.

The benefit of an approach to data mining is that it can predict crime, such as crime data being trained to model from the machine learning process. Various data mining algorithms that can help create a classifier are capable of detecting the crime rate of different crimes [5]. The research goal focuses on the following objectives:

- An empirical analysis of machine learning to obtain accurate predictions by correcting errors in previous years
- The main objective of the research is to correctly classify the existence of the fault and prediction
- Enhance the efficiency of prediction of crime.
- Identify the impact on incarceration due to an increase or decrease in crime. Predictive accuracy will increase.
- Choosing the best classifier from multiple classifiers for accurate prediction, which requires precise timing and is an effective objective.
- Generating data with an effective research approach to resolving gaps in the analysis so that better outcomes can be achieved.
- Develop a model that gives a new direction to the prediction of crime.

Generalization or Stacking suggested by Wolpert is an ensemble model [22]. Ensemble methods are used to improve statistical machine learning. This also deals with estimating the low-performance raw classifiers. Independent Ensemble stacking is referred to as blending because of mixed sub-model predictions. It can be blended to produce an approximation or classification. As in fig. 4, the stacking theory gives a clear sketch. In the SVM and J48 and C4.5 contract, the stacking method increases Predictive power for the classifier

Ensemble learning methods intend to make a meta-classifier by consolidating a few classifiers, usually by casting a vote, using similar data, and enhancing their presentation [23], the data structure and the criminal estimation process can be summarized in the following steps. The whole dataset was split into, 18 years (2001-2015) and 3 years (2016-2018) training and testing, Also, we call the model used in the second layer of the meta-model as shown in fig. 4 [24]. This model is divided into three steps:

- Aggregator and sub-models.
- Predictions combining.
- Data set on organized crime.

The two models are used as sub-models for stacking and an SVM model as an aggregator model. These models provide the basis for understanding and implementing the stacking of our own predictive modeling Problems. This section is divided into 3 parts:

- Sub model # 1: J48.
- Sub model # 2: C4.5.
- Model aggregator: support vector machine.

Each model is described in terms of the functions used to train a model as well as the functions that help make accurate predictions. Vector Machine Support (SVM) is a supervised machine learning Algorithm. The goal of SVM is to learn optimally. Hyperplane that separates samples, like crimes according to class. It is built to find the widest possible margin.

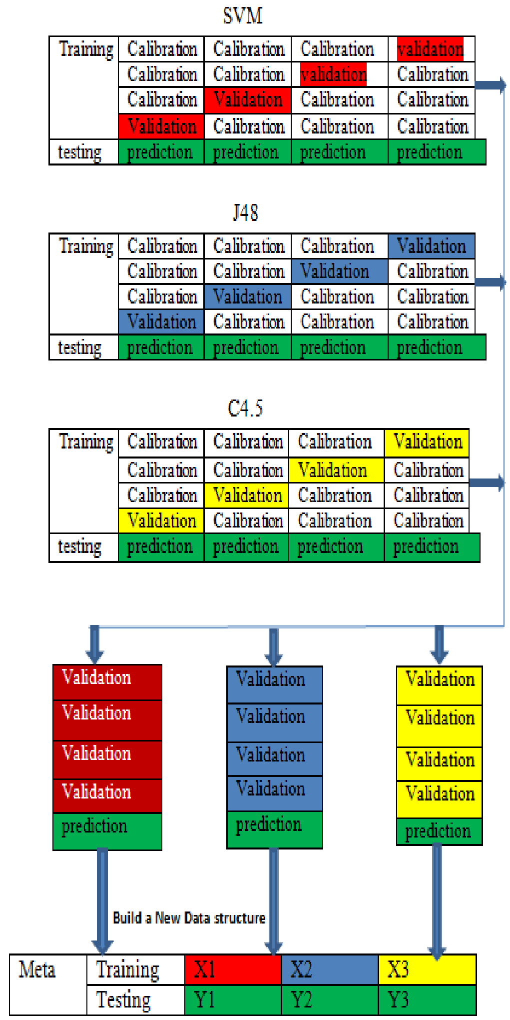

FIGURE 4. Structure of the proposed stack based crime prediction model.

Move 1: The three models mentioned above (SVM, j48, C4.5) are independently configured with the help of training the dataset during the cycle followed by the cross validation (loocv) technique that generates the validation values. Specifically, a duration of 15 years is used to calibrate the models and the remaining 1 year is used for validation. We can therefore obtain long validation results of 18 years (blue green, yellow) for a certain model.

Move 2: Use the entire training dataset (15 years) in the testing cycle to calibrate the model and then produce the forecasts (15 years, red components).

Move 3: For the second layer, all validations are composed sequentially into a new training package. The result of the 18 predictions is significantly different; as there is only one test prediction (RE is less than 1 percent, not shown). Consequently, by means of a simple average process, all the forecasts are integrated into a new testing package for the second layer.

Move 4: The dataset consists of observations and four validated values of the model (validation in the first layer) that are 18 years in the second training period.

Move 5: Similarly, in the second research cycle of study, the dataset consists of observations and three models' expected values that are 15 years long (prediction in the first layer). After that, the Meta model is used to recalibrate and predict the values of the final crime.

Move 6: The meta-model is then used to calibrate the aggregate multimodal simulation (using Step 4's second training period dataset) and to apply the final aggregate multimodal forecast to the second test dataset (Step 5).

In this section, the framework of model selecting using stack SVM algorithm are introduced. We employ the SVM (SVM-J48, SVM and SVM-C4.5) to select the most suitable meta-learners by maximizing accuracy and diversity as two conflicting objectives. SVM is regarded as one of the most excellent multi objective evolutionary algorithms and has been widely used in many fields so far. It features for its fast and accurate search performance. The schematic diagram of Stack based crime prediction for model selecting is illustrated in Fig. 4. The fivefold cross-validation is to get the predicted class of the training set. The evaluate objective functions are accuracy and diversity based on the predicted class and the actual class.

As mentioned earlier, accuracy and diversity are two crucial factors that decide the success of the stacking. Accuracy measures the difference between the predicted class and actual class, while diversity measures the differences between meta-learners.

VI. MACHINE LEARNING MODELS

A. J48 CLASSIFIER

The j48 algorithm was built in [25], an improved algorithm for IDE3 which is based on Hunt's algorithm and is implemented sequentially like IDE3, with the internal node turning into a leaf node and pruning, which makes the rate a lot of work [26]. Like IDE3, C4.5 acknowledges both unceasing and the different level aspects of tree development. Have an advanced tree pruning technique that eliminates premature distribution errors due to the amount of detail or duplication in the planning of the information package. Like IDE3, knowledge is organized on each tree hub such that the best distribution characteristics can be solved. It utilizes the addition proportion polluting influence technique to assess the parcel trademark.

To explain the working of the J48 algorithm, take an example of crime prediction in order to produce a decision tree that can help to determine the mark considered as a specific crime. Both previously available crimes and state data have been reviewed and a training dataset has been developed. Table 3 displays the training dataset, which is used to construct a classification model. The attributes of this training dataset include murder, attempt to murder, rape, kidnapping, riots arson, etc. The primary objective of data visualization is to convey knowledge clearly and efficiently. Effective visualization helps users to analyze and justify the data and evidence. It makes complex data more accessible, understandable and, usable.

First of all, the knowledge gained from all the attributes are determined. The attribute with the highest knowledge gained is selected as a root node [27].

Internal node (himachal Pradesh) denotes a test on the following:

- Branches (damn&diu and raunchily Pradesh) represent an outcome of the test.
- All records in a branch have the same value for the tested attributes.
- Leaf node represents (class 1, class 2, class 3, etc.) class label or class label distribution.

To produce the model, training data was used, and a data set with known output values used this data set to build this model. However, this type of model takes an entire training set and splits it into two parts, i.e., about 80 percent of the data is taken and placed into a training set, used to build the model, and then the remaining 20 percent data set is put into a test data set.

The test data was created to monitor the fitting after the model was created; it was tested to ensure that the accuracy of the model manufactured does not decrease with the test set. In order guarantee the above-listed feature, the proposed model will reliably predict potential mysterious qualities.

B. SMO CLASSFIER

SMO is an algorithm that is widely used for the training of Support vector machine (SVM) [25]. Sequential minimal optimization (SMO) is a (QP) quadratic programming problem algorithm that arises during the preparation of support vector machines (SVM). It was invented by John Platt in 1998 at Microsoft Research. SMO is widely used for training support vector machines and is implemented by the popular LIBSVM tool; and is an iterative calculation, which refreshes just two Lagrange multipliers in each progression in a way that ensures the intermingling to the ideal arrangement. For the induction of SMO calculation, the corresponding structure type of the arched non-smooth improvement is used. [25], [23].

SMO is an iterative algorithm, which updates at each point. Here, two Lagrange multipliers are used so that the optimal solution will converge. For the derivation of SMO algorithm, the following matrix form of the convex non-smooth optimization problem is considered.

$$\min_{\alpha \in R} \mathrm{J}(\alpha) = \frac{1}{2}\alpha^T Qa + p^T \alpha + \varepsilon||\alpha||_1$$
$$\text{Subject to} \quad u^T a = 0$$

The basic algorithm for finding solutions to help vector machines is called the SML algorithm, and a very stripped-down version of this is performed in order to find the weight vector. The SMO algorithm consists of:

- One heuristic, to a select Lagrange multiplier and then second multiplier of Lagrange

- The code to solve the optimization problem analytically for the two chosen multipliers

A vector is taken at the beginning of the algorithm, with a vector: a($a_1, a_2, a_3 \ldots\ldots .a_m$).

The idea is to pick two elements of this vector, which are named as name α_1 and α_2, and to change their values so that the constraints are still respected. The first constraint $0 \leq a_1 \leq C$ *for any* $i = 1 \ldots .m$ means that $0 \leq a_1 \leq C$ and $0 \leq a_2 \leq C$.

The second constraint is a linear constraint

$$\sum\nolimits_{i=1}^{m} a_i y_i = 0$$

This makes the values to lie on the red diagonal, and the first few selected a_1 and a_2 values should have different labels ($y_1 \neq y_2$)

- The SMO algorithm selects two, α parameters, a_i and a_j and advances the target esteem mutually for both these α's. At long last it alters the b parameter dependent on the new α's. This procedure is rehashed until α's merge.
- On the off chance that αi doesn't satisfy the KKT conditions to inside some numerical resistance, we select αj at arbitrary from the rest of the m − 1 α's and endeavor to mutually streamline a_i and a_j.

SMO classification is strong algorithm. However, it can be seen that when training on data is done in the classifier, then the accuracy rate is only 41.667. This is very low for any crime prediction.

C. NAÏVE BAYES CLASSIFIER

The Naive Bayes algorithm is a machine learning algorithm that uses a training dataset for classification problems that is mainly used for text classification which further involves high dimensional training data sets. An example of this algorithm is spam filtration sentimental analysis which was first introduced in 1995 by "John and Langely [25], [28]. This is popular due to its simplicity and usefulness with the Naive Bayes algorithm since the models are constructed easily and fast predictions are obtained. The algorithm learns the likelihood of an object with some features belonging to a specific category of class, and a probabilistic classifier assumes the independence of the class attributes [29]. The so-called naïve Bayes algorithm is called naïve because it assumes that the occurrence of a certain function is independent of the occurrence of other characteristics, for example [30].

1) BAYESIAN PROBABILITY THEORY

- Proposition is called A and evidence is called the B
- $P(A)$ is called the prior probability of proposition and P (B) is called the prior probability of evidence
- P (A | B) is called the posterior
- P (B | A) Is the probability?

$$Posterior = \frac{(likelihood)\,.(Proposition\ prior\ probability)}{Evidence\ Prior\ Probability}$$

There are several features and classes in the problem of machine learning classification, such as C_1, $C_{1,\ldots\ldots\ldots}$ $C_k \ldots\ldots$. It's C¬k. The main objective of the Naive Bayes algorithm is to measure the conditional probability of a feature vector object $X_{1,}$ $X_{2,} \ldots\ldots X_n$ which belongs to a particular class C.

$$\mathrm{P}\left(\mathrm{C_i} \mid \mathrm{r_1,r_2}\ldots\ldots,\mathrm{x_n}\right) = \frac{\mathrm{P}\left(\mathrm{x_1,x_{2,}}\ldots.\mathrm{x_n} \mid \mathrm{C_i}\right)\mathrm{P(C_1)}}{\mathrm{P}\left(\mathrm{r_1,r_1}\ldots\ldots,\mathrm{r_n}\right)}\mathrm{For1} \leq \mathrm{i} \leq \mathrm{k}$$

Now the numerator of the fraction on the right-hand side of the equation above is

$$\begin{aligned}
&\mathrm{P}\left(\mathrm{r_{1,}r_{2\ldots\ldots,}}\ \mathrm{r_n}|_{\mathrm{C_1}}\right),\\
&= \mathrm{P}\left(\mathrm{r_{1,}r_{2\ldots\ldots,}r_{n,}C_i}\right)\\
&\mathrm{P}\left(\mathrm{r_{1,}r_{2\ldots\ldots,}r_{n,}C_i}\right)\\
&= \mathrm{P}\left(\mathrm{r_1} \mid \mathrm{r_{2,\ldots\ldots,}r_{n,}C_i}\right).\mathrm{P(r_{2,\ldots\ldots,r_{n,}}C_1)}\\
&= \mathrm{P}\left(\mathrm{r}| \mathrm{r_{2,\ldots\ldots,}r_{n,}C_i}\right).\mathrm{P(\ r_2| r_{3,\ldots\ldots,}r_nC_1)}P\left(\mathrm{r_{3,\ldots\ldots,}r_{n,}C_i}\right)\\
&= \ldots\ldots\ldots.\\
&= \mathrm{P}\left(\mathrm{r_1} \mid \mathrm{r_{2,\ldots\ldots,}r_nC_i}\right).\mathrm{P}\left(\mathrm{r_2}| \mathrm{r_{3,\ldots\ldots,}r_nC_1}\right)\ldots\ldots P\left(\mathrm{r_{n-1}}\|\ \mathrm{r_{n,}C_i}\right)\\
&\quad .\mathrm{P}\left(\mathrm{r_n}| \mathrm{C_i}\right).\mathrm{P}\left(\mathrm{C_i}\right)
\end{aligned}$$

The conditional probability term P of R J given R of J plus 1 R of J plus 2 R n CI becomes P of RJ given CI because of the assumption that features are independent.

The Conditional probability term P of X J given X of J plus 1 R of J plus 2R n CI becomes P of RJ given CI because of the assumption that features are independent. Formulate the above calculation and the independent calculation.

The assumption base theorem comes down to a very simple expression and makes the assumption that the documents' terms are independent from each other.

$$\begin{aligned}
&\mathrm{P}\left(C_i \mid r_{1,}r_2\ldots\ldots,x_n\right)\\
&\quad = \prod_{j=1}^{j=n} P\left(x_j| \ C_i\right)\frac{\mathrm{P}\left(C_i\right)}{\mathrm{P}\left(x_1, x_2\ldots\ldots.,x_n\right)}\ For\ 1\\
&\quad \leq i \leq k
\end{aligned}$$

The articulation $\mathrm{P}\left(x_1, x_2\ldots\ldots.,x_n\right)$ is steady for all the classes, we can essentially say that

$$\mathrm{P}\left(C_i \mid r_{1,}r_2\ldots\ldots,x_n\right) \infty \prod_{j=1}^{j=n} P\left(x_j| \ C_i\right).P\left(C_i\right) For1 \leq i \leq k$$

These are the basics and mathematical concepts behind the Naive Bayes algorithm. Thus, the following calculation has been rendered based on the above algorithm in MATLAB, summarizing the effects of the use of the Naïve Bayes classifier to identify the documents. It is surprising to find the results of the pre-processed dataset (67.222 percent)

D. BAGGING CLASSIFIER

The Bagging classifier is the ensemble learning method proposed by Leo Breiman in 1994 and used to construct a community of learners [31]. A similar learning algorithm can train each student on an alternative arrangement of the knowlwdge. This is called bootstrap collection or bagging [25], [23]. It was invented by Bremen in the late '1980s, the early 1990s. The working of how the bagging works is explained here, which is done by having a variety of subsets of the results.

Multiple models are constructed by making and using bootstrap replicas for the training package. The training subsets are drawn at random with substitution to avoid trainers from being trained in the same subset [32]. The basis is chosen for standard classification rankings and bagging aims to improving the accuracy of the base by classifying the predictions of the learned classifier into a single prediction by constructing a composite classifier. The voting system is the method used by bagging [33].

The Bagging Algorithm
Input: Training set S
Base Learning Algorithm B
Number of bootstrap samples T
Procedure: For i = 1 to T
{
S' = S' = bootstrap test from (S' is an example With substitution from S)
Ci = B (S') (create a new classifier from S')
}
C*(x) = (the most often predicted label y)
Output
Classifier C*

The adjectives are more predictive of sentiment classification and predicted the sentiment of adjectives by inspecting them in conjunction with "and," "or," "but," "either/or," and "neither/nor." However, this approach may overestimate the importance of adjectives and underestimate some. Bagging classifier is quite good classifier as compared to SMO and Naïve byes. However, this cannot be considered as the best classifier for violent crime prediction.

E. RANDOM FOREST CLASSIFIER

A supervised learning technique that creates a forest with a number of trees is the Random Forest algorithm [34]. It can be used by constructing multiple decision trees during training for regression, classification, and other assignments. The variables can be ranked on the base of their priority, using Random Forest.

Brief description Random forest algorithm fills in like an enormous assortment of connected decision trees. We use lot of selected trees, but this irregular random forest algorithm makes a bunch of selected trees and uses them to classify [35]. This is why the random technique-dependent procedure is implemented.

$$M = \begin{matrix} f_{A1} & f_{B1} & f_{C1} \\ f_{A2} & f_{B2} & f_{C3} \\ f_{An} & f_{Bn} & f_{n} \end{matrix}$$

shows the matrix M in this example assume matrix S is a matrix of training samples that we can give to the algorithm to construct a classification model in this case f A1, f B1, f C1, there are a set of characteristics. If A1 is the feature A of the first sample and we proceed in all the samples up to N So, the f B N is the feature B of the Nth sample and also, we have got in the last columns here the C1 and CN, which means we have got a lot of features and we have got a training class So the objective is to create an arbitrary random forest to set this example, how does the algorithm work, in detail?

Build a random subset, as:

- The creator right off the bat shows the Random Forest creation pseudo-code:
- Haphazardly select "M" highlights from all-out "N" highlights where M $\ll$ k
- Among the "M" highlights, figure the hub "b" utilizing the best part point
- Split the hub into little girl hubs utilizing the best split
- Rehash the a to c ventures until "l" number of hubs has been reached
- Construct woods by rehashing stages a to do for "n" number occasions to make "n" number of trees

The model (random forest) is trained from year to year as the accuracy of the model is increasing as the accuracy of 2001 was 0 percent, it increased to 79.37 percent in 2015. Similarly, the rate of true positive value is also increased and the rate of FP declined.

VII. MODEL PERFORMANCE AND DISCUSSION

A. PROPOSED CRIME PREDICTION MODEL

Once a model is learned, it can be used to classify new unseen data. These notes describe the process of doing the same thing both graphically and from the command line. The proposed model consists of violence crime data from 2001 to 2015.

It is only possible to use a qualified classifier model for forecasts if it provides good precision in the testing process. By using the training dataset, the precision of the classifier model is obtained. The missing values are included in the research dataset. The classifier estimates these missing values and thus produces a matrix of uncertainty. The classifier model that previously saved, loaded and re-evaluated on the test dataset. 180 instances are in the research dataset used here. More than 60000 of them are data.

B. CRIME RESULT DISCUSSION

In the first layer, the best-output machine learning processes replace the machine learning algorithms in the second layer as the meta-model. A total of three forms of prediction findings are equated to make the conclusions more convincing. Discussion of the result of predictor dataset is based on twelve violence crime. It contains five categories: Since the publication time of the Abuse Crime Index is the first of each year, it also predicts stack-based crime for the next three years.

The structure of the prediction can be defined as follows:

$$P(c) = f(P(c-1)M\mathrm{I}_i, (c-1)M\mathrm{I}_i, (t-2) \\ \ldots\ldots\ldots\ldots\ldots M\mathrm{I}_i(t-3))$$

Y where the crime expected for the year y is P(c).

The predictors are composed of the previous yearly observations P(c-1) and the preceding 3 years' climate indices.

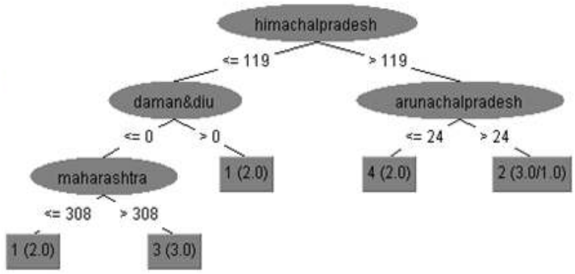

FIGURE 5. Tree representation of j48.

P(c)=f(P(c-1) MI_i,(c-1) MI_i,(t-2)............ MI_i(t-3))
MI_i is a vector of 5 crime indices
$MI_i = [MI_1 MI_2 \ldots\ldots\ldots\ldots MI_{180}]$

So, there are 192 key predictors. Out of these Out of these 180 predictors, a subset of the most important ones for effective modeling results must be chosen. As a model construction, predictor selection is an equally essential step. In this study, the classification inherent in the three predictor selection models is used since for the method of selection and selection of the number of predictors. All the schemes and the outcomes are estimated with the number of predictors selected ranging from 1 to 5. The models have the best simulation in the Loocv period where the number of predictors is between 1 and 5.

If the all model is ready, the best model is used to produce the desired results in any form we choose. In this case, a graph shown in Fig. 8 will be plotted according to the criteria, as discussed earlier in this paper. This information was collected from 2016 to 2018, in which the class is not defined.

The performance of several models (J48, Naive bayes, SMO, Bagging and random forest) is shown in the Table 3 as given below.

Figure 8 shows that the proposed model uses an assembly-approach crime prediction method (SBCPM) and produces good prediction results with small mean absolute error and root mean squared error compared to other machine learning algorithms. These algorithms were tested with a full training dataset with 80 percent training and 20 percent cross-validation review testing. The correctly categorized situation, the correct intense value is 179 out of 180 actual values and the predicted value, 99.5 percent correct in the feigning of vice versa was the model developed in this paper and it was the assembled model. Mean absolute error (MAE) is known as the sum used to calculate, how close the inevitable outcomes are to expectations or conjectures. The root mean square error (RMSE) is defined as a proportion of the contrasts between the values by the model or an estimator and the true accuracy observed for most of the time used.

The accuracy, root mean square error (RMSE) and absolute total error (MAE) are used for the forecast error estimation. The study suggested that in this case the root means square error (RMSE) is anything but a good marker of normal error, so MAE would be a higher measure for this function. Fundamentally, although that they have rather special figuring

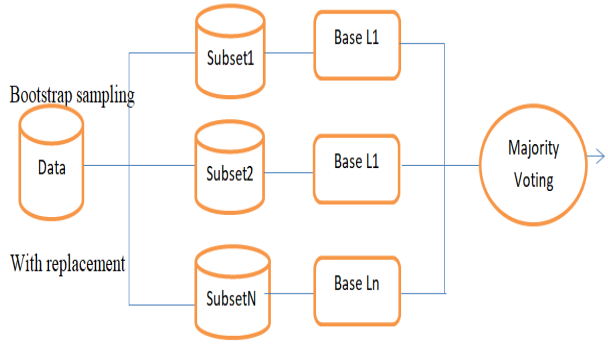

FIGURE 6. Bagging classifier model.

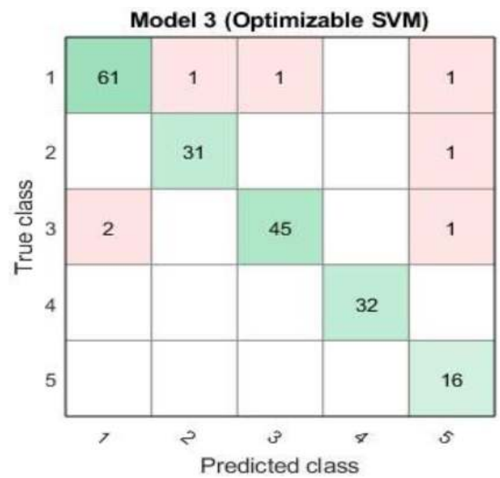

FIGURE 7. Correlation matrix between true class and predicted class.

TABLE 2. Experiment results of proposed model.

Performance measures	Description
Accuracy	99.5
Total misclassification cost	1
Prediction speed	~9000 obs/sec
Training time	139.95 sec
Standard data	true

equations, root mean squared error has a similar function to mean total error. The standard distinction between the actual outcome and the prescient outcome of the entire dataset as a whole is the mean absolute error (MAE) itself.

C. PERFORMANCE OF MEASURE METRICS

The efficiency is calculated in order to measure the performance of the model Metrics of measurement is used. Any model with evaluation the output of the regarding these metrics the pattern. The metrics for the following are the evaluation of results.

1. TP rate $= \frac{currect\ class}{actual\ class}$
2. FP $= \frac{TN}{N}$
3. Precision rate $= \frac{PT}{(TP+FP)}$

TABLE 3. Experiment performance of machine learning models.

Performance measureS	J48	Random Forest	SMO	Begging	Naive Bayes	stacking
time	0.01 Sec	0.03 sec	0.09 sec	0.08 sec	0.08 sec	135.95sec
Accuracy	94.44 %	97.21%	41.667 %	95.55%	67.22%	99.5
TP rate	0.944	0.978	0.417	0.956	0.672	0.99
FP rate	0.022	0.009	0.292	0.014	0.125	0.005
Kappa	0.921	0.970	0.151	0.942	0.570	0.998
Precision	0.945	0.982	0.322	0.952	0.701	0.99
recall	0.944	0.982	0.417	0.956	0.672	.998

4. Recall $= \frac{PT}{(TP+FN)}$
5. F measure $= \frac{2*recall*precision}{(recall*precision)}$

Here TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative. By using this formula, we calculate Specificity, Sensitivity for all algorithms and then compared to check which one gives better result.

D. MODEL PERFOMANCE AND COMPARISON

In this research, six classifier models are put to find out the best model for crime prediction. It is implemented using MATLAB, Current ML algorithms results are compared with the proposed model ensemble stack-based crime prediction model.

Results analysis was presented in Table 3 and Fig. 9 to compare the efficiency and accuracy of the forecast algorithms. During the training process, the overall performance is in favor of SVM-stacking (especially TPR, FPR, precision, sensitivity, RMSE, MAE, MCC, and Kappa). The Random forest model comes in second, followed by the Bagging model.

The model J48, it is visually evident that the evaluation stage displays a similar ranking pattern for the success of the model to the training stage.

1) RESEARCH OUTCOME

Many machine learning algorithms are available for research. As studied in paper shown, six algorithms have been chosen for the analysis named as J48, Naïve Bayes, SMO, and together stacking learning Bagging classifier and Random Forest. Based On efficiency, the Classify-based Stacking Ensemble is the best giving 99.5 % accuracy while with other classifiers only 95.55% and 97.21 % accuracy is reported. While calculating the time for all three classifiers, J48 giving 0.01 sec, random forest 0.03 sec and bagging as 0.08 sec. Begging and random forest takes longer for a large number of samples, while J48 requires less time to estimate.

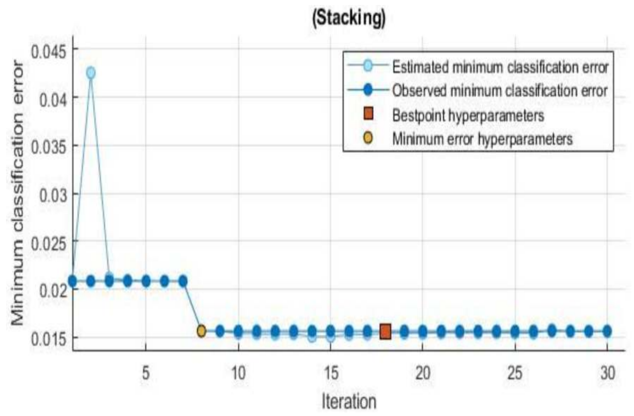

FIGURE 8. Optical parameter tuning of the conventional learning on feature extraction.

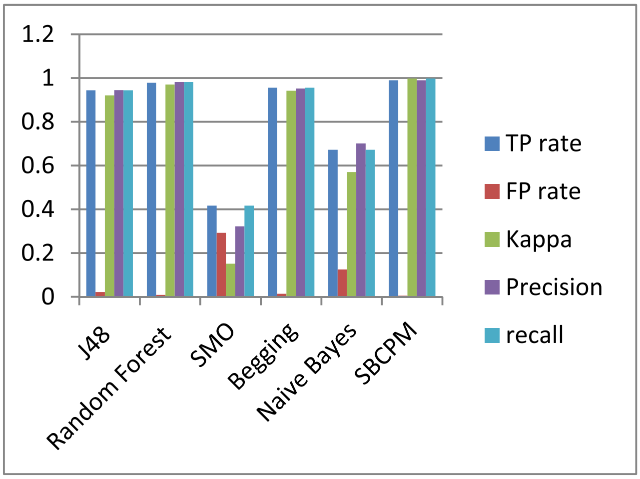

FIGURE 9. Graph representations with different outcomes.

E. RESULT ANALYSIS OF CRIME PREDICTION

In the first phase, five classifier models are applied to find out the best model for crime prediction. It is implemented in MATLAB, and the results of all ML algorithm models are compared. Based on the findings of the above experiments, it was determined that the random forest algorithm produces robust prediction results on crime data in the ensemble approach. However, in the second stage, we suggest a stack-based crime prediction method (SBCPM) and we found that, the stacking method is more reliable than the Ensemble method.

VIII. CONCLUSION AND FUTURE WORK

In the present study, ML models with machine learning algorithms (ensemble and simile), i.e., SMO, SVM-bagging, SVM-Random forest, SVM-stacking J48, and Naive Bias, were designed and were implemented. Each predetermined factor was feed into a violent crime training dataset (murder, rape, robbery, etc.). We discovered a major conclusion after successfully training and validating six models.

We found major conclusions:

1. The prediction accuracy of the SBCPM model is better than other models. This can be better extracted from the violence crime data pattern and regularity.
2. The SBCPM employs one-against-five classifiers with the transmission of predictive labels. It provides a new labelled classification method where SBCPM machine learning classifiers at different tree levels can effectively work together to define the association between the labels and to predict labels and predict instance labels more exactly. The 15-year criminal ML-Tree approach is taken and compared with the other classifiers as J48, Naïve Bayes, SMO, and ensemble random forest bagging classifiers.
3. The experimental results demonstrate the efficacy of the proposed SBCPM model. The novel SBCPM has a core working period in the training phase, with a high growth rate of 99.5 percent. The efficient method of finding an appropriate method in crimes can be predicted in such a way that the measurement is unpredictable. Also, planning to investigate additional violations and exceptions by incorporating the SBCPM method into the selection process is explained. During the analysis work, when formulating the theory, veracity of the current research problems and hypotheses was formulated.

In the future, the whole model will be converted into an opensource library and connected to the crime site, allowing it to function at the highest level of expertise Function on a framework where the threshold for class crime rates can be set. Instead of a limited crime, the highest crime rate can be measured. A small grouping is used to assess the success of all of the criminal figures examined in the proposal. A Calculation can be measured on a sufficiently broad scale of local or cloud-based crime with heavy datasets will pay forecast of multi-label charging, expand more possibilities, and realistically increase our research

REFERENCES

[1] M. Cahill and G. Mulligan, "Using geographically weighted regression to explore local crime patterns," *Social Sci. Comput. Rev.*, vol. 25, no. 2, pp. 174–193, May 2007, doi: 10.1177/0894439307298925.

[2] J. M. Caplan, L. W. Kennedy, and J. Miller, "Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting," *Justice Quart.*, vol. 28, no. 2, pp. 360–381, Apr. 2011, doi: 10.1080/07418825.2010.486037.

[3] A. Almehmadi, Z. Joudaki, and R. Jalali, "Language usage on Twitter predicts crime rates," in *Proc. 10th Int. Conf. Secur. Inf. Netw.*, Oct. 2017, pp. 307–310, doi: 10.1145/3136825.3136854.

[4] V. K. Borooah and N. Ireland, "Deprivation, violence, and conflict: An analysis of Naxalite activity in the districts of India," *Int. J. Conf. Violence*, vol. 2, no. 2, pp. 317–333, 2008, doi: 10.4119/UNIBI/ijcv.42.

[5] A. Babakura, M. N. Sulaiman, and M. A. Yusuf, "Improved method of classification algorithms for crime prediction," in *Proc. Int. Symp. Biometrics Secur. Technol. (ISBAST)*, Aug. 2014, pp. 250–255, doi: 10.1109/ISBAST.2014.7013130.

[6] P. Pławiak, M. Abdar, and U. R. Acharya, "Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring," *Appl. Soft Comput.*, vol. 84, Nov. 2019, Art. no. 105740, doi: 10.1016/j.asoc.2019.105740.

[7] Z. Li, T. Zhang, Z. Yuan, Z. Wu, and Z. Du, "Spatio-temporal pattern analysis and prediction for urban crime," in *Proc. 6th Int. Conf. Adv. Cloud Big Data (CBD)*, Aug. 2018, pp. 177–182, doi: 10.1109/CBD.2018.00040.

[8] A. Almaw and K. Kadam, "Survey paper on crime prediction using ensemble approach," *Int. J. Pure Appl. Math.*, vol. 118, no. 8, pp. 133–139, 2018. [Online]. Available: https://internal-pdf://107.93.182.66/18.pdf%0Ahttp://www.ijpam.eu

[9] T. B. Hyde, H. Dentz, S. A. Wang, H. E. Burchett, S. Mounier-Jack, and C. F. Mantel, "The impact of new vaccine introduction on immunization and health systems: A review of the published literature," *Vaccine*, vol. 30, no. 45, pp. 6347–6358, 2015, doi: 10.1016/j.vaccine.2012.08.029

[10] S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma, and N. Yadav, "Crime pattern detection, analysis & prediction," in *Proc. Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Apr. 2017, pp. 225–230, doi: 10.1109/ICECA.2017.8203676.

[11] N. H. M. Shamsuddin, N. A. Ali, and R. Alwee, "An overview on crime prediction methods," in *Proc. 6th ICT Int. Student Project Conf. (ICT-ISPC)*, May 2017, pp. 1–5, doi: 10.1109/ICT-ISPC.2017.8075335.

[12] S. S. Roy, V. M. Viswanatham, P. V. Krishna, N. Saraf, A. Gupta, and R. Mishra, "Applicability of rough set technique for data investigation and optimization of intrusion detection system," in *Proc. Int. Conf. Heterogeneous Netw. Quality, Rel., Secur. Robustness*, in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 115, 2013, pp. 479–484, doi: 10.1007/978-3-642-37949-9_42.

[13] L. Acion, D. Kelmansky, M. D. Van Laan, E. Sahker, D. S. Jones, and S. Arndt, "Use of a machine learning framework to predict substance use disorder treatment success," *PLoS ONE*, vol. 12, no. 4, pp. 1–14, 2017, doi: 10.1371/journal.pone.0175383.

[14] V. Ingilevich and S. Ivanov, "Crime rate prediction in the urban environment using social factors," *Procedia Comput. Sci.*, vol. 136, pp. 472–478, Jan. 2018, doi: 10.1016/j.procs.2018.08.261.

[15] G. R. Sinha, K. S. Raju, R. K. Patra, D. W. Aye, and D. T. Khin, "Research studies on human cognitive ability," *Int. J. Intell. Defence Support Syst.*, vol. 5, no. 4, pp. 298–304, 2018.

[16] F. A. Anifowose, J. Labadin, and A. Abdulraheem, "Ensemble model of non-linear feature selection-based extreme learning machine for improved natural gas reservoir characterization," *J. Natural Gas Sci. Eng.*, vol. 26, pp. 1561–1572, Sep. 2015, doi: 10.1016/j.jngse.2015.02.012.

[17] F. Anifowose, J. Labadin, and A. Abdulraheem, "Predicting petroleum reservoir properties from downhole sensor data using an ensemble model of neural networks," in *Proc. Workshop Mach. Learn. Sensory Data Anal.*, 2013, pp. 27–34, doi: 10.1145/2542652.2542654.

[18] F. Anifowose, J. Labadin, and A. Abdulraheem, "Ensemble learning model for petroleum reservoir characterization: A case of feed-forward back-propagation neural networks," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 7867, 2013, pp. 71–82, doi: 10.1007/978-3-642-40319-4_7.

[19] F. Anifowose, J. Labadin, and A. Abdulraheem, "Ensemble model of artificial neural networks with randomized number of hidden neurons," in *Proc. 8th Int. Conf. Inf. Technol. Asia (CITA)*, Jul. 2013, pp. 1–5, doi: 10.1109/CITA.2013.6637562.

[20] S. S. Roy, P. V. Krishna, and S. Yenduri, "Analyzing intrusion detection system: An ensemble based stacking approach," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2014, pp. 307–309, doi: 10.1109/ISSPIT.2014.7300605.

[21] F. Anifowose, J. Labadin, and A. Abdulraheem, "Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines," *Appl. Soft Comput.*, vol. 26, pp. 483–496, Jan. 2015, doi: 10.1016/j.asoc.2014.10.017.

[22] D. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.

[23] X. Zhou, Y. Peng, and B. Liu, "Text mining for traditional Chinese medical knowledge discovery: A survey," *J. Biomed. Inform.*, vol. 43, no. 4, pp. 650–660, 2010, doi: 10.1016/j.jbi.2010.01.002.

[24] B. Zhai and J. Chen, "Development of a stacked ensemble model for forecasting and analyzing daily average PM2.5 concentrations in Beijing, China," *Sci. Total Environ.*, vol. 635, pp. 644–658, Sep. 2018, doi: 10.1016/j.scitotenv.2018.04.040.

[25] Y. Hu, J. Li, and L. E. Holloway, "Towards modeling of resilience dynamics in manufacturing enterprises: Literature review and problem formulation," in *Proc. IEEE Int. Conf. Automat. Sci. Eng.*, 2008, pp. 279–284, doi: 10.1109/COASE.2008.4626539.

[26] H. Mouratidis, M. Ouedraogo, E. Dubois, and D. Khadraoui, "Information systems security criticality and assurance evaluation," in *Advances in Computer Science and Information Technology* (Lecture Notes in Computer Science), vol. 6059, T. Kim and H. Adeli, Eds. Berlin, Germany: Springer, 2010, doi: 10.1007/978-3-642-13577.

[27] C. Walgampaya and M. Kantardzic, "Cracking the smart ClickBot," in *Proc. 13th IEEE Int. Symp. Web Syst. Evol. (WSE)*, Sep. 2011, pp. 125–134, doi: 10.1109/WSE.2011.6081830.

[28] A. Misal and G. R. Sinha, "Denoising of PCG signal by using wavelet transforms," *Adv. Comput. Res.*, vol. 4, no. 1, pp. 46–49, 2012.

[29] M. Zareapoor, K. R. Seeja, and M. A. Alam, "Analysis on credit card fraud detection techniques: Based on certain design criteria," *Int. J. Comput. Appl.*, vol. 52, no. 3, pp. 35–42, Aug. 2012.

[30] F. Harahap, A. Y. N. Harahap, E. Ekadiansyah, R. N. Sari, R. Adawiyah, and C. B. Harahap, "Implementation of Naïve Bayes classification method for predicting purchase," in *Proc. 6th Int. Conf. Cyber IT Service Manage. (CITSM)*, Aug. 2018, pp. 1–5, doi: 10.1109/CITSM.2018.8674324.

[31] L. G. A. Alves, H. V. Ribeiro, and F. A. Rodrigues, "Crime prediction through urban metrics and statistical learning," *Phys. A, Stat. Mech. Appl.*, vol. 505, pp. 435–443, Sep. 2018, doi: 10.1016/j.physa.2018.03.084.

[32] M. Reza, S. Miri, and R. Javidan, "A hybrid data mining approach for intrusion detection on imbalanced NSL-KDD dataset," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 6, pp. 1–33, 2016, doi: 10.14569/ijacsa.2016.070603.

[33] D. Stalling, M. Westerhoff, and H.-C. Hege, "Amira: A highly interactive system for visual data analysis," in *The Visualization Handbook*, vol. 38. 2005, pp. 749–767.

[34] L. J. Muhammad, E. A. Algehyne, and S. S. Usman, "Predictive supervised machine learning models for diabetes mellitus," *Social Netw. Comput. Sci.*, vol. 1, no. 5, pp. 1–10, Sep. 2020, doi: 10.1007/s42979-020-00250-8.

[35] M. Amrehn, F. Mualla, E. Angelopoulou, S. Steidl, and A. Maier, "The random forest classifier in WEKA: Discussion and new developments for imbalanced data," 2018, *arXiv:1812.08102*. [Online]. Available: http://arxiv.org/abs/1812.08102

SAPNA SINGH KSHATRI received the M.C.A. degree from the Rungta College of Engineering and Technology. She is currently a Research Scholar of Ph.D. (computer science and application) with the MSIT Department, MATS University, Raipur. She has two years of teaching and academic experience. Her Ph.D. dissertation was on crime prediction. She served as a reviewer of several journals.

DEEPAK SINGH received the Ph.D. degree from the Department of Computer Science and Engineering, National Institute of Technology Raipur. He is currently an Assistant Professor with the School of Engineering and Applied Science, Bennett University, Greater Noida. He has over six years' experience in various academic institutes. He has published over ten refereed articles. His research interests include evolutionary computation, machine learning, and data mining. He served as a reviewer of several journals.

BHAVANA NARAIN received the M.Phil. degree in computer science, the M.C.A. degree, the L.L.B. degree, two P.G. Diploma courses, and the Ph.D. degree in computer application and IT.

She is currently an Associate Professor with the MSIT Department, MATS University, Raipur. She has 23 years' experience of work in academic area and five years of industrial experience. She has undergone through two minor projects as CO-PI. Since 2014, she has been working as a State Student Coordinator of Region 4 Chhattisgarh state with the Computer Society of India. She has published two books in Lambert publication and Amazon. She has published more than 56 articles in national and international journals. She has presented papers in national and international conferences, two patents, and life membership of four societies. She is guiding eight Ph.D. research scholars and two Ph.D. scholars are awarded. Under her supervision, four M.Phil. (CS) has been awarded. Her research interests include wireless networks, machine learning and digital image processing, and machine learning and cyber security.

SURBHI BHATIA is currently pursuing the Ph.D. degree in computer science and engineering from Banasthali Vidypaith, India. She is also an Assistant Professor with the Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Saudi Arabia. She has rich eight years of teaching and academic experience. She has been granted and published seven national and international patents. She has published more than 30 articles in reputed journals and conferences in high indexing databases. She has successfully authored two books from Springer and Wiley. She is editing three books from CRC Press, Elsevier, and Springer. She has been an active researcher in the field of data mining, machine learning, and information retrieval. She is in the Editorial Board Member of the *International Journal of Hybrid Intelligence* (Inderscience Publishers), *SN Applied Sciences* (Springer), and also in several IEEE conferences. She is also serving as a guest editor for special issues in reputed journals. She has delivered talks as keynote speaker in IEEE conferences and in faculty development programs. She is a PMP®.

MOHAMMAD TABREZ QUASIM (Senior Member, IEEE) received the Master of Computer Application degree from Punjab Technical University, Kapurthala, India, and the Ph.D. degree in computer science from Tilkamanjhi Bhagalpur University, India. He is currently working as an Assistant Professor with the University of Bisha, Saudi Arabia. He has more than ten years of experience in his research area. He has published many journal articles, edited book, book chapters, and conference papers in various internationally recognized academic databases. He is contributing to the research community by various volunteer activities. He has served as conference chair in various reputed IEEE/Springer conferences.

G. R. SINHA (Senior Member, IEEE) received the Ph.D. degree. He is currently an Adjunct Professor with the International Institute of Information Technology Bangalore (IIITB). He is also deputed as a Professor with the Myanmar Institute of Information Technology (MIIT), Mandalay, Myanmar. He has been the Dean of the Faculty and Executive Council Member of CSVTU. He has delivered more than 50 keynote/invited talks and chaired many technical sessions in international conferences across the world. He has supervised eight Ph.D. Scholars, 15 M.Tech. Scholars, and has been supervising one more Ph.D. Scholar. He has more than 200 research articles, edited books, and books into his credit. He has edited books for reputed International publishers. He has teaching and research experience of 21 years. His research interests include biometrics, cognitive science, medical image processing, computer vision, outcome-based education (OBE), and ICT tools for developing employability skills. He is a member of Senate of MIIT. He is a Fellow of the Institute of Engineers India and the IETE India. He was a recipient of many awards and recognitions at national and international level. He is an Associate Editor of three important SCI journals, including IEEE ACCESS. He is an active reviewer and an editorial member of more than 12 reputed International journals in his research areas, such as IEEE TRANSACTIONS, Elsevier journals, and Springer journals. Since 2017, he has been delivering ACM lectures as an ACM Distinguished Speaker in the field of DSP across the world. He served as a Distinguished IEEE Lecturer in IEEE India Council for Bombay Section.