ARTICLE

OPEN

Check for updates

An automated system for the assessment and grading of adolescent delinquency using a machine learning-based soft voting framework

Abhinash Jenasamanta [1✉] & Subrajeet Mohapatra[1]

Adolescent (or juvenile) delinquency is defined as the habitual engagement in unlawful behavior of a minor under the age of majority. According to studies, the likelihood of acquiring a deviant personality increases significantly during adolescence. As a result, identifying deviant youth early and providing proper medical counseling makes perfect sense. Due to the scarcity of qualified clinicians, human appraisal of individual adolescent behavior is subjective and time-consuming. As a result, a machine learning-based intelligent automated system for assessing and grading delinquency levels in teenagers at an early stage must be devised. To solve this problem, a soft voting-based ensemble classification model has been developed that includes a Decision Tree, Multi-layer Perceptron, and Support Vector Machine as base classifiers to accurately classify teenagers into three groups based on severity levels, viz., low, medium, and high. Over the normalized structured behavioral data, the proposed soft voting-based model outperforms all other individual classifiers with 87.50% accuracy, an AUC of 0.94, 0.81 Kappa value, and an F-score of 0.88.

[1] Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, India. ✉email: samantajena99@outlook.com

Introduction

A person under the age of 18 who violates the law is considered a juvenile delinquent. Conduct disorder is a clinically recurrent pattern of antisocial behavior in which a person frequently violates social rules and participates in deviant behavior that causes others to be upset (Krohn and Lane, 2015). Adolescents during this phase of life develop more advanced patterns of reasoning. Hence, there is a higher risk of the development of such deviant behavior in juveniles. In India, the latest report published by NCRB (National Crime Records Bureau) states that a total of 31,591 crimes were committed by juveniles in the year 2018. The report states that boys make up more than 99 percent of the juveniles arrested for these offenses. This trend is concerning not just because the offense trends represent more serious offenses committed by these children, but also because these young criminals are more likely to continue their criminal activities. As deviant behavior usually develops during adolescence (Howell, 2003), it is very much essential for early identification (screening) of such individuals. Early intervention helps to prevent delinquent activity and promotes the growth of a young person's assets and resilience. Screening such individuals by a psychologist is quite difficult in many cases since multiple behavioral and environmental dimensions need to be correlated before arriving at a final decision (Siegel and Welsh, 2014). Considering this problem, efforts must be made to identify at-risk adolescents with delinquent behavior early. Conventional diagnosis (Feld and Bishop, 2011) of deviant behavior in adolescents is based on psychological questionnaire measurements and the personal clinical experience of the psychologist.

Figure 1 depicts the steps followed for the conventional diagnosis of juvenile delinquency.

The diagnostic evaluation of individuals with suspected delinquency begins with a careful review of the social and behavioral history, through psychiatric assessment or psychological screening. A formal assessment of delinquent behavior (Gearhart and Tucker, 2020) in adolescents is based on semi-structured diagnostic interviews and validated questionnaires (psychometric scales) under the supervision of a psychologist. Moreover, manual analysis becomes questionable not only because of the amount of work but also with regard to precision and the reproducibility of the results. The motivation to automate comes from the fact that besides being time-consuming, the results of manual diagnosis vary with the psychologist's skill, experience, workload, and stress level. Automated systems (Rathinabalan and Naaraayan, 2017) can help overcome the dearth of trained psychologists. Owing to the existence of such ambiguity in the conventional diagnosis of deviant behavior, the traditional evaluation techniques need to be enhanced by using a quantitative approach in psychology. The principle of quantitative psychology (Ibrahim, 2016) focuses on the use of statistical and machine-learning models for the analysis of psychological patterns of human behavior. Additionally, computer-aided diagnostic systems can support the screening of adolescents in remote and rural parts of the country. In the last few years, various researchers have been attracted to digital psychology and have contributed to the area of modern quantitative psychology. Machine learning is one of the most efficient techniques for predictive data analysis, which is being implemented in multiple research dimensions (Alpaydin, 2021; Greener et al., 2022). Although extensive research has been carried out to understand different risk and protective factors of child delinquency, studies on the machine-learning methods on early risk assessment of delinquent behavior in juveniles are limited. Many schemes are reported to use individual risk factors along with statistical models to assess the significance of individual risk factors in detecting delinquency (Zhang et al., 2014; Svensson et al., 2013; Sciandra et al., 2013). However, most of them fail to provide a generic output. Few schemes have considered delinquency risk assessment as a multivariate pattern classification problem (Lansford et al., 2007; Bor et al., 2001; Britt, 1997). Empirical studies on delinquency showed that machine learning-based model outperforms statistical models in terms of accuracy. Some of the reported methods using single self-report scales claim to be reliable but they are not truly reliable since delinquency is multifaceted (Gray, 1987; Vaux and Ruggiero, 1983; Meldrum et al., 2015). These schemes generally use a single category of the risk factor for the prediction of delinquency in children (Castellana et al., 2014; Jacobsen and Zaatut, 2022).

The accuracy of early detection of delinquency mostly depends on these risk factors and the use of a single category of risk factor may not deliver satisfactory performance in all situations. Thus, to improve the delinquency detection capability a multivariate approach involving different risk factors should be used.

The effect of sibling delinquency on adolescent delinquency on a broader scale with the least effects of other social domains on delinquency was analyzed (Huijsmans et al., 2019). They used hierarchical linear modeling (HLM) techniques using six waves of data for analysis. They suggested that the effect of peers, parents, and school are important factors to be analyzed in future research and interventions targeted to adolescent delinquency analysis. A study in the Philippines using a dataset collected from the City Social Welfare Development Office of Butuan City to develop predictive models to analyze the children at risk as well as children in conflict with the law was conducted (Castro and Hernandez, 2019). They applied algorithms, viz., decision tree, Naive Bayes, generalized linear model, and Logistic regression to the dataset. The Naive Bayes algorithm gave the best classification accuracy and the least classification error. The study showed that a large number of children around the age limit of 12−17 are
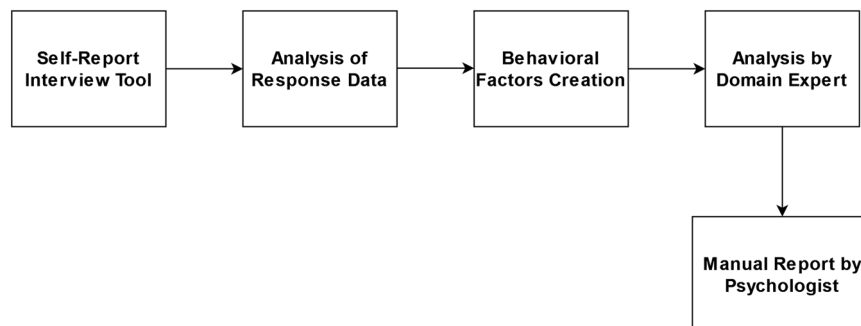
Fig. 1 The procedure followed for the conventional or traditional diagnosis of juvenile delinquency. This type of manual process is dependent upon the experience of the psychologist.
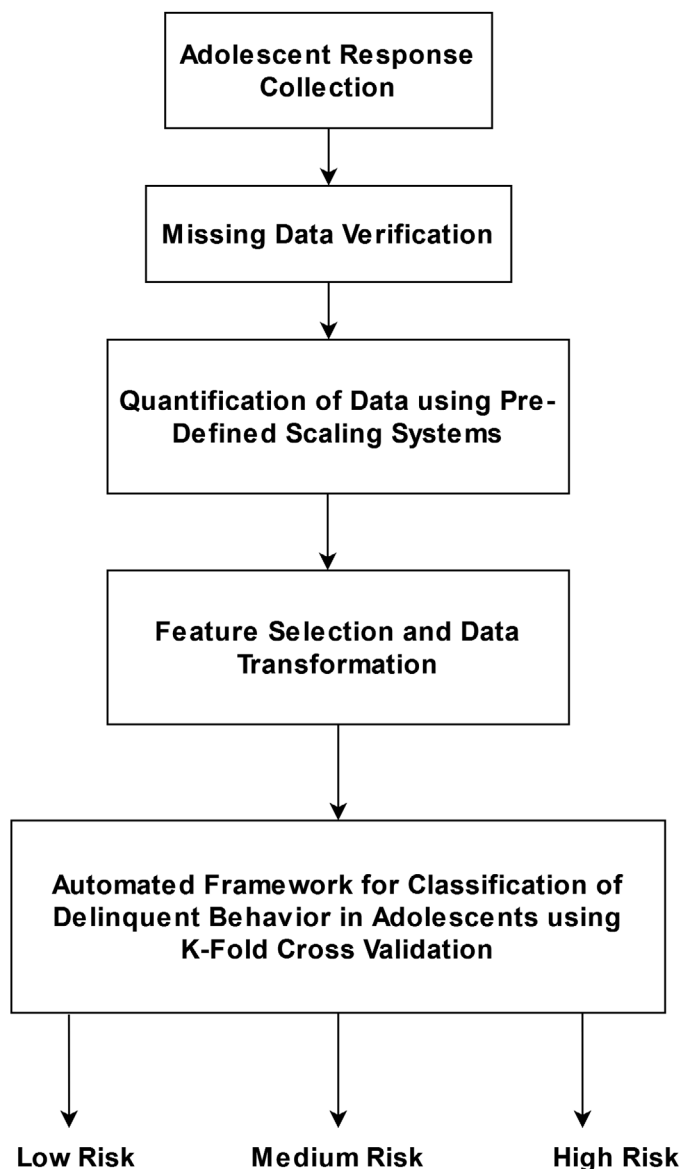
Fig. 2 The overall step-by-step computational approach for the early diagnosis of delinquent behavior in juveniles. The developed model can classify adolescent behavior into three classes, i.e., low, medium, and high risk.

affected by maltreatment and children in the age group 15—17 committed severe crimes.

Hence, exercising multiple risk factors and advanced machine-learning models in the automated detection of delinquency remains an active field of research to attempt this problem. Keeping the research directions in view, it has been realized that there exists enough scope to improve the delinquency detection performance for deviant individuals in different social and family environmental conditions. To improve delinquency detection capability, a multi-dimensional approach involving various independent risk factors has been carried out in this research.

Materials and methodology

In the present work, an effort has been made to perform delinquency behavioral risk assessment for adolescents using soft voting-based machine-learning ensemble modeling. In this section, we describe the details of the layout of the proposed system as well as the step-by-step approach followed. The block diagram of the proposed computational approach for the early diagnosis of delinquent behavior in adolescents is shown in Fig. 2, which is explained in detail in the following sections.

After carrying out the literature review, identifying the research, setting research objectives and research hypothesis following step-by-step methodology has been formulated.

Study area selection. The state of Jharkhand, India was chosen purposively for the study because in the recent past many juveniles have been arrested for heinous activities due to the high poverty rate.

Source of data and survey element. The primary data was collected by interviewing the adolescents studying in Class VII to Class XII of the selected schools. Secondary data of each individual includes his previous social behavior in the school and was obtained by interviewing the associated teachers. A standard delinquency questionnaire (ISRD-3 India Version) (Enzmann et al., 2018; Barranco et al., 2022; Marshall et al., 2019) with required rationalization according to regional conditions was used for the interview.

Participants selection and data collection strategy. A total of 182 students from the Ranchi district of Jharkhand, India with an age variation of 12 to 17 years participated in this study. The required consent to participate in the survey was sought from each school's head (principal) and subsequently from the students in the classroom. The team of three members consisting of both the authors and a behavioral psychologist carried out data collection in 15 schools (both Government and private) of the Ranchi district of Jharkhand. Permission to survey each school was granted by the head of the school. Consent to participate in the survey was sought subsequently from the students in the classroom. Finally face to face interview with the students was carried out by the two members. A total of 182 students (96 boys and 86 girls) participated in the survey and their response was recorded. The first segment of planning involved an overview and discussion of the questionnaire so that students could understand the nature and objectives of the research and answer any questions that were likely to be asked since the interview was being carried out in the classroom. The second segment of planning focused on the data collection plan and discussed coordinating the team of the survey with teachers in each school, along with the procedures to be followed in the classroom. The response options obtained from the adolescents for each factor are converted into a numeric score based on various pre-defined scales as explained in the following sections.

Behavioral attributes creation and quantification. Suitable representation of behavioral attributes of adolescents is essential for computer-aided diagnosis of delinquency. The questionnaire set based on ISRD-3 is quantified into 53 attributes based on specifications provided by a two-member panel of clinical Psychologists. Each data field of a particular factor/feature measurement has been converted to a proportionate numeric value. Standard scaling methods in psychology (modified Likert Scale) have been used to quantify individual, parental, and environmental factors as per the recommendation of the two-member panel. Adding to the above factors, other features, viz., victimization and gang relationships are also analyzed, measured, and represented numerically for computer-based analysis.

Feature selection. To determine meaningful qualities for the current dataset, a standard statistical feature selection technique called analysis of variance (ANOVA) was applied (Saefi et al.,

Table 1 Significant socio-demographic risk factors and predictors for juvenile delinquency detection and classification.

Factors	Significant features ($p < 0.05$)
Individual factors	Age, sibling count
Geographical factors	Regional influence (urban/rural), norm transmission strength
Family and parental factors	Parental supervision, parental attachment level, parental education, parental, and family crime history
Social and religious factors	The language used, belief system
Financial factors	Income source and poverty level
Behavioral and personality factors	Aggression and temperament, animal cruelty, offending
Academic factors	School bonding, school environment, teacher attachment, unexplained absenteeism, academic performance
Peer factors	Peer attachment, effect of delinquent peers, delinquent group or gang involvement, age group of peers
Internal trauma	Victimization, perception of shame, traumatic injury
Environmental factors	Neighbor association, neighboring environmental activities.
Substance abuse	Alcohol abuse, drug abuse, substance abuse integrity
Individual accepting factors	Acceptance of illegal activities, personal arrest record
Public safety	Procedural justice
Self analysis	Subjective well-being, future plans, leisure activities, attitudes toward delinquency
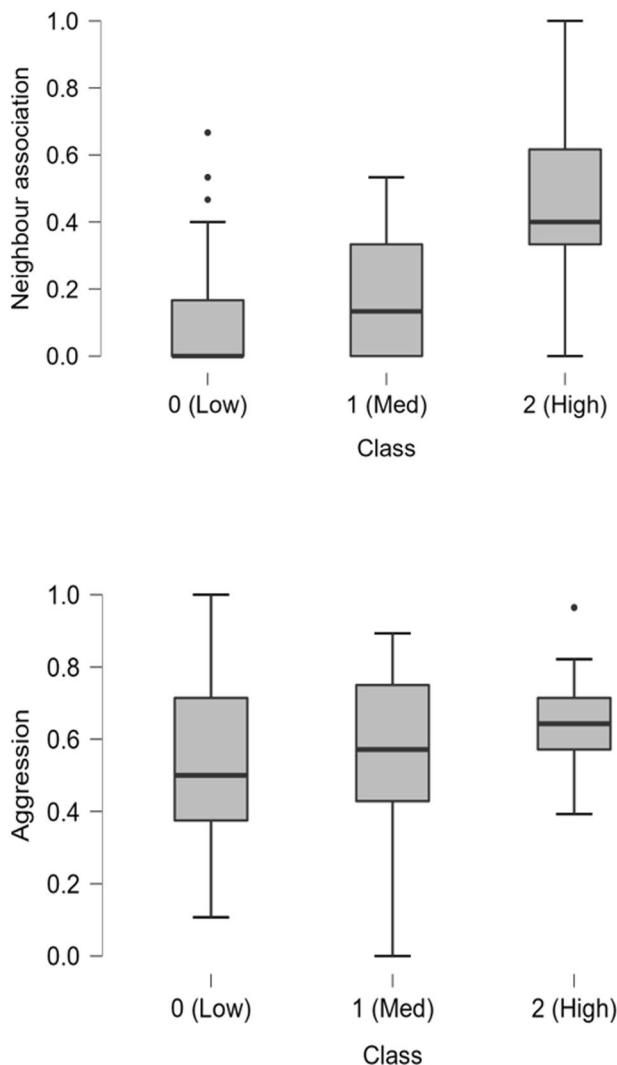
Fig. 3 The box plots of two major attributes viz. neighbourhood association and aggression. X-axis indicates the 3 classes and Y-axis indicates the given feature.

2020). The ANOVA method analyzes group variances and means to see if they are overlapping. The features are considered statistically significant if they have a smaller p-value. The p-value was set at 0.05 in this study to indicate that the features are statistically significant. Out of the 53 features tested, 47 were found to be statistically significant ($p < 0.05$).

One-way ANOVA test. The one-way ANOVA is applied to verify whether any significant differences exist between the means of three or more independent (unrelated) groups. The null hypothesis is rejected if any of the group means deviate considerably from the overall mean (Pauly and Smaga, 2020). Table 1 shows the major statistically significant risk factors and predictors analyzed for juvenile delinquency detection.

Figure 3, Fig. 4, and Fig. 5 illustrate the box plots for some significant features, which were tested and analyzed for juvenile delinquency detection.

Reliability test (using Cronbach's alpha). The degree to which a test or a measuring technique produces consistent results over time is referred to as reliability (Sobri et al., 2019). Reliability is the trend toward consistency exhibited in repeated measurements of the same phenomena. Exploratory factor analysis is a major technique for validating dimensionality. Cronbach's Alpha (Taber, 2018) is commonly used to evaluate internal consistency or reliability. Nunnally and Bernstein, on the other hand, generalize an alpha coefficient of 0.70 or above to be generally acceptable (Nunnally, 1975). Cronbach's alpha is mathematically explained as a function of the test items count and the intercorrelation mean among the features. It is defined as:

$$\alpha = \left[n\,C/(V + (n - 1)\,C \right] \qquad (1)$$

where n is the number of features, C is the average covariance among the features, and V is the average variance. Table 2 shows the result of applying the above reliability test on all features of our dataset. Table 3 shows the values of Cronbach's alpha of individual feature items.

Feature scaling using Min–Max normalization. Implementation of machine-learning algorithms over measured behavioral attributes/features is unfeasible as the raw data vary widely. Hence scaling of features is essential to normalize the range of independent attributes. Different scaling methods were applied and a comparative study was done to identify the outstanding approach. It is important to normalize the dataset with a wide range of values before classification. In this component, normalization (Jain et al., 2005) of each of the 47 attributes present in the dataset has been carried out using Min–Max normalization within a range of 0 to 1.
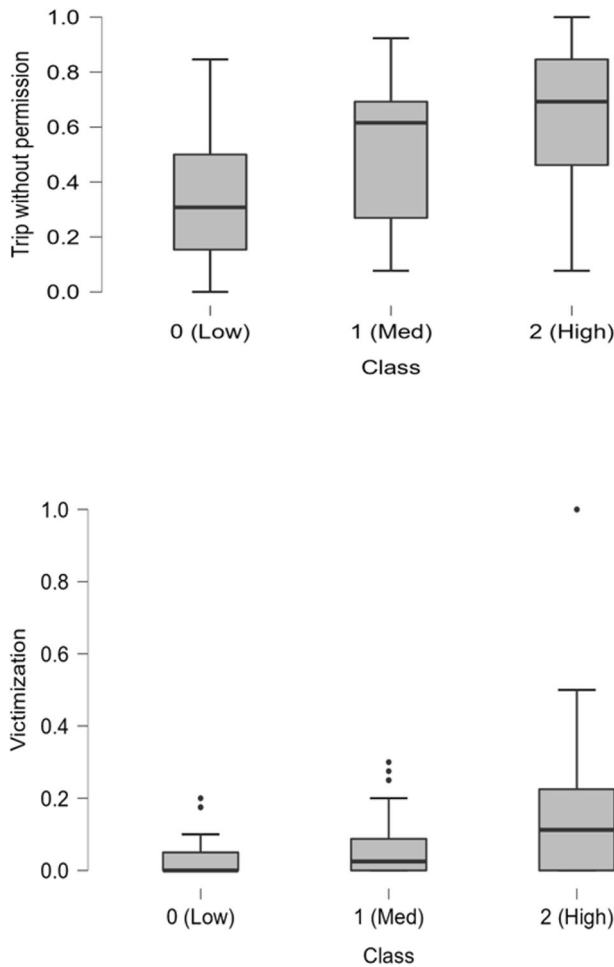
Fig. 4 The box plots of two major attributes viz. trip without permission and victimization. *X*-axis indicates the 3 classes and Y-axis indicates the given feature.
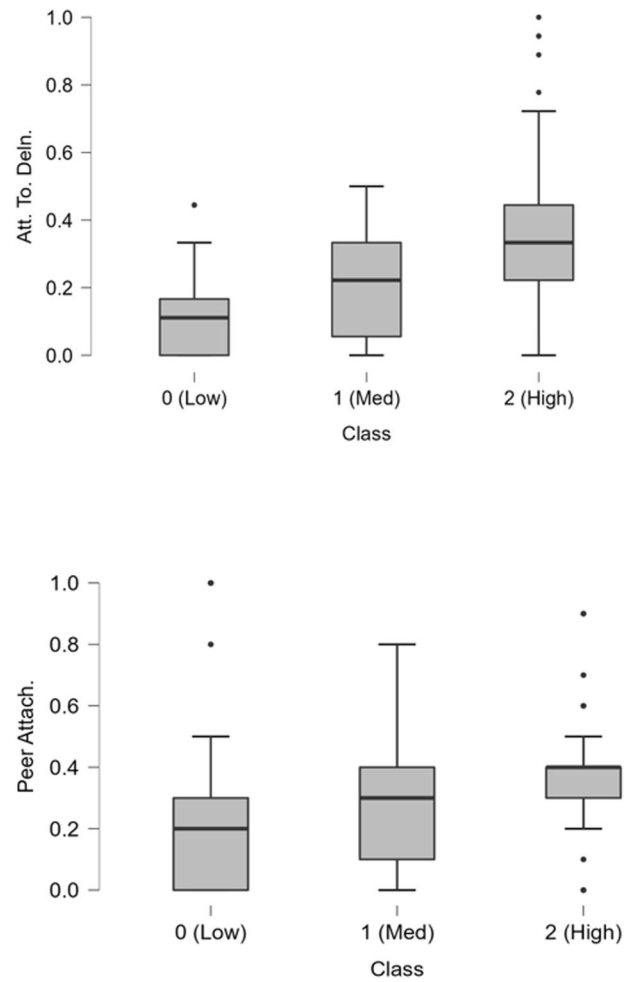
Fig. 5 The box plots of two major attributes viz. attitude towards delinquency and peer attachment. *X*-axis indicates the 3 classes and Y-axis indicates the given feature.

Classification of delinquent behavior of juveniles using machine learning. The problem of child behavior characterization is considered here a supervised machine-learning (or classification) problem. Moreover, adolescents are categorized based on the level of risk of showing delinquent behavior, i.e., low, medium, and high risk. Classification is a category of machine learning, which comes under supervised learning (Aggarwal, 2015). It specifies the data element's class and is appropriate when the output has both finite and discrete values. In multi-class classification, instances are allocated to one among a range of pre-defined classes. This can be achieved by fitting several binary classification models for each class versus all other classes (One-vs.-Rest) or by using a single model for each pair of classes (called One-vs.-One) (Wang et al., 2021). In this work, a multi-class classification model using ensemble learning has been created to characterize adolescent behavior into three classes, i.e., low, medium, and high based on multiple factors. Specific risk and protective indicators of juvenile delinquency, viz., psychological factors and socio-demographic factors have been studied based on the literature review and the ISRD3 questionnaire (India version) has been used along with required modifications to record the student's responses through interviews.

Decision tree (C4.5). In this research, the C4.5 technique has been used which is a widely used algorithm in machine learning. There are several parameters in C4.5, i.e., the parameter to test the

Table 2 Reliability statistics for frequentist scale.

Measure	Cronbach's alpha
Point estimate	0.805
95% CI lower bound	0.745
95% CI upper bound	0.854

The following items are negatively correlated with the scale: atr7, atr8, atr18.

effectiveness of the Decision tree with pruning (Polat and Guneş 2009; Al-qazzaz et al., 2021). The number of folds (*n*) determines the amount of data used for pruning and the minimum count of instances per leaf is denoted by *l*. The confidence Factor is set at 0.25. All the parameters are correctly set and pruning has been applied.

Support vector machines (SVM). SVM does classification by generating n-dimensions and then by maximizing the margin to obtain the best classification results. SVMs are based on the concept of hyperplane classifiers or linear separability. In general, the SVM classifier (Tharwat and Gabel, 2020) is a binary classifier that can be used to classify data instances. It may be readjusted to handle a greater number of classes (>2) by implementing one vs. rest or one vs. one methodology. SVMs may also be readjusted to generate a non-linear decision boundary. This is accomplished by

Table 3 Reliability statistics of frequentist individual item.

Item	If item dropped Cronbach's α
Age	0.807
Single parent	0.802
Personal belief system	0.808
Spiritual opinion	0.827
Groups	0.811
Income source	0.805
Poverty level	0.804
Parental attachment	0.789
Parental supervision	0.794
Family disturbance	0.798
School bonding	0.789
School environment	0.807
Teacher attachment	0.792
Unexplained absenteeism	0.816
Academic performance	0.795
Future planning	0.805
Victimization	0.799
Trip without permission	0.795
Subjective well-being	0.800
Leisure activities	0.804
Peer attachment	0.803
Delinquent peers	0.791
Attitudes towards delinquency	0.791
Perception of shame	0.798
Aggression	0.805
Traumatic injury	0.801
Neighbor association	0.792
Neighborhood environment	0.807
Offending	0.792
Animal cruelty	0.789
Alcohol abuse	0.799
Substance abuse integrity	0.810
Norm transmission strength	0.795
Procedural justice	0.793
Gang	0.802
The age group of peers	0.805
Response integrity	0.806
Behavioral observation	0.773

converting the input from its current dimension to a higher one. The goal is to obtain a non-linear decision boundary since the link between the input space and converted space is non-linear (Xian and Yang, 2021). Kernel functions are functions that are used to transform data into a non-linear format. SVM's main goal is to find the best (or optimal) separating hyperplane. The optimal hyperplane has the highest distance from the margins and they always give a better generalization rate. SVM's objective function (Ricci and Perfetti, 2007) can be expressed as:

$$L_P = (1/2)\|w\|^2 - \sum_{i=1}^{n} \alpha_i \{y_i(w_i^T x_i + w_0) - 1\} \qquad (2)$$

where w indicates the weight vector and x_i refers to the input vector. The transformation function that is used determines the degree of non-linearity of the decision boundary. On data that is non-linearly separable as well as overlapping data, SVMs proves to be extremely robust and efficient. To handle this, a slack variable (Vapnik and Izmailov, 2021), viz., e_i is utilized. The SVM optimization is also termed by a factor $(\|w\|^2/2) + C(\sum_i e_i)^k$.

A kernel $K(x_i, x_j)$ is considered "valid" if there exists feature space "Φ" so

$$K(x_i, x_j) = \Phi(x_i) * \Phi(x_j) \qquad (3)$$

where $\Phi(x_i)$ is the depiction of x_i in higher dimensional space.

The polynomial kernel which is a non-linear kernel (Morales et al., 2021) is used for classification in this work and is given by $(X_i X_j + 1)^h$. The complexity parameter is an indication of the extent of avoidance of misclassification for each training example in SVM optimization. The complexity parameter is set at value one.

Multi-layer perceptron (MLP). It is one of the most efficient systems of neural networks. The efficiency of the multi-layer perceptron arrives (Taud and Mas, 2018) from non-linear activation functions. In maximum cases, sigmoid activation functions are used, which are given by:

$$f(s) = 1/(1 + e^{-s}) \qquad (4)$$

The learning process is based on the minimization of errors between network outputs and desired outputs. Hence a backpropagation of error value through a network similar to that which is learned is implemented. The main task is to find the minimum error function e(w) about the connecting weights. In this research, backpropagation MLP has been used. The learning rate for weight updates and the momentum applied to weight updates are selected properly. There are 2 hidden layers with two nodes each in the network.

Ensemble of classifiers. Ensemble-based methods (Kasim, 2021) are techniques that create multiple models and then combine them to produce better model performance in terms of, viz., accuracy, *F*-score and Kappa value, and other evaluation parameters. Ensemble methods generally produce more accurate solutions than a single model. One of the main issues in combining different classifiers is the framing of appropriate combination rules. Voting is one of the efficient ways of combining the predictions from multiple machine-learning algorithms.

Soft voting. Soft voting (Kumar and Batra, 2021) produces the best outcome by averaging the probabilities calculated by individual techniques. In soft voting (Kieu et al., 2020) the final prediction is the class with the highest class probability averaged over all individual classifiers.

It is mathematically given by:

$$Y' = \text{argmax}_i \sum_j^m w_j, p_{ij} \qquad (5)$$

where w_j is the weight that can be assigned to the jth classifier.

In this research, soft voting technique has been used to classify juvenile delinquents into three classes, viz., low, medium, and high. For base classifiers, Decision Tree (C4.5), Support Vector Machines (Polynomial Kernel of degree 3 using One Versus One technique) and Multi-Layer Perceptron (2,2), i.e., two hidden layers with two nodes each have been utilized.

Proposed soft voting ensemble model for multi-class classification of juvenile delinquency. The proposed approach is to build a soft voting ensemble with 3 base classifiers, viz., C4.5 (Decision Tree), MLP (2,2), and SVM (Polynomial kernel with degree 3 using One vs. One technique). The voting model (soft voting) model has been developed to classify each data instance of the dataset which is shown in Fig. 6.

All three base learners can effectively handle non-linear data as well as do multi-class (>2 class) classifications. After identifying the base learners, the next step was to create a voting ensemble. The voting ensemble that was created combined the base learners by calculating the average predicted probability of each class label for each base classifier's prediction. In the average predicted probability technique (soft voting) the final (target) class label is derived from the class label having the highest average probability
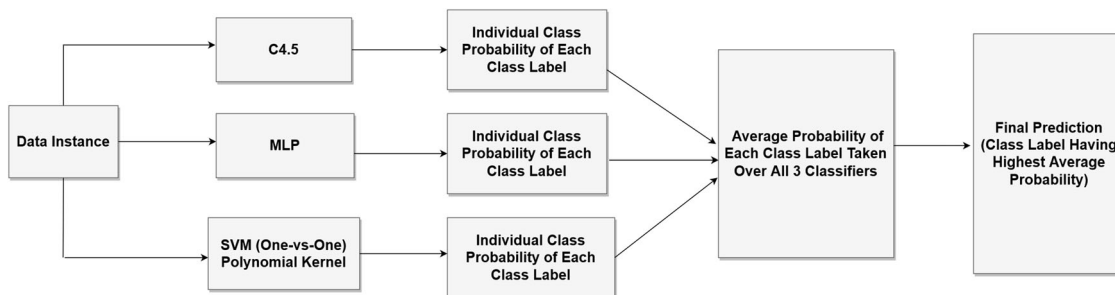
Fig. 6 The proposed soft voting ensemble architecture developed for automated assessment of adolescent delinquency. The base classifiers utilized are, C4.5 (Decision Tree), MLP and SVM.

Table 4 Tuning hyperparameters for machine-learning framework.	
Classifier	**Hyperparameter value**
Decision tree (C4.5)	clf C4.5 = Set Param (Confidence Factor = 0.25, No. of Objects = 3, No. of Folds = 5, Rules = 13)
Multi-layer perceptron (MLP)	clf MLP = set Param (Layers = (2,2), LearningRate = 0.3, momentum = 0.2)
Support vector machine (SVM)	clf SVM = set Param (complexity Parameter C = 1.0, tolerance Parameter = 0.001)
Proposed soft voting ensemble classifier	clf Soft Voting Ensemble = Voting (set estimators = [("C4.5 clf", C4.5(Set Param (Confidence Factor = 0.25, No. of Objects = 3, No. of Folds = 5, Rules = 13))), ("MLP clf", MLPClassifier (set Param (Layers = (2,2), LearningRate = 0.3, momentum = 0.2))), ("SVM clf", Poly SVMClassifier (set Param (complexity Parameter C = 1.0, tolerance Parameter = 0.001))),], voting = "soft")

estimate. This process takes into account the confidence of each voter. Each of the three class probabilities (low, medium, high) is generated by each classifier and the ensemble calculates the average of each class label (overall 3 classifiers) and chooses the class label with maximum average probability as the target class for each instance. The training and testing of the ensemble model will be done using 10-fold cross-validation to find a suitable technique for classifying juvenile delinquency (Marcot and Hanea, 2021).

Hyperparameter tuning. Hyperparameters are learning parameters that are set before the system is trained and have a direct impact on the model's performance and efficiency (Schratz et al., 2019). For our analysis, we have implemented four machine learning-based algorithms, viz., C4.5 (decision tree), MLP (multi-layer perceptron), SVM (support vector machines), and the proposed soft voting ensemble-based classifier are explained in detail in section "Validation". Before utilizing these machine learning-based models, the hyperparameters were effectively tuned so that they could predict and analyze more precisely and efficiently than other models (Duan et al., 2003). Hyperparameters and their tuning values for each model were illustrated in Table 4.

Validation
In this work, 10-fold cross-validation has been utilized for evaluating the machine-learning algorithms. K-fold Cross-validation is a resampling technique used to implement machine-learning models on a small data sample (Grimm et al., 2017). The term K denotes the number of groups into which a given data sample should be divided. Compared to alternative methods such as a straightforward train/test split, cross-validation is simple to grasp and typically yields a less-biased assessment of the model's performance. In K-fold cross-validation, the initial sample is randomly divided into k subsamples of equal size. The remaining k – 1 subsamples are utilized as training data, and one subsample from the total of k subsamples is kept as the validation data for testing the model. Thereafter, the cross-validation procedure is carried out K times and each of the K subsamples is utilized once

as the validation set. To create a single estimation, the K findings are finally averaged (Saud et al., 2020). This technique of validation helps to avoid overfitting and bias.

Evaluation metrics
Proposed ensemble models are evaluated based on different standard performance metrics, i.e., Accuracy (Benussi et al., 2021), MCC (Matthews Correlation Coefficient) (Chicco and Jurman, 2020), and Kappa Value (Alizad et al., 2020). Moreover, ROC-area (Bowers and Zhou, 2019), Precision, Recall, and F1 score (DeVries et al., 2021)-based results have also been provided for comparative performance validation.

$$\text{Classification Accuracy} = (TP + TN)/(TP + TN + FP + FN) \tag{6}$$

$$TPR = TP/(TP + FN) \tag{7}$$

$$FPR = FP/(TN + FP) \tag{8}$$

$$\text{Precision} = TP/(TP + FP) \tag{9}$$

Results
Experimental simulation and analysis were conducted using an Intel Core i7-8700 CPU @ 3.20 GHz PC, along with 16 GB RAM on Windows 10 operating system. Python version 3.9.5 with Pandas and Scikit-learn libraries and Weka have been used for simulation and analysis. For statistical analysis, JASP software version 0.16 has been used. An automated ensemble classification system for the early diagnosis of juvenile delinquency has been developed, experiments were performed using the above configuration, and the corresponding results have been provided in this section. The overview of the performance of the proposed model and its comparative analysis with individual classifiers over 10-fold cross-validation is provided in Table 5.

The results show that the multi-class classification using the proposed soft voting ensemble model outperforms the performance of individual classifiers with higher accuracy, Kappa value, and MCC.

The classification performance of the proposed ensemble model in terms of TPR, FPR, Precision, and F-score and its comparison with different classifiers are shown in Table 6.

The achieved F-score value indicates that the proposed model has successfully been able to classify juvenile delinquency into three classes with substantial accuracy.

Table 5 Comparative classification performance of proposed soft voting ensemble model in terms of accuracy, Kappa, and MCC over 10-fold CV.

Method	Accuracy (%)	Kappa value	MCC
C4.5	76.92	0.65	0.66
SVM (One vs. One)	82.96	0.74	0.75
MLP	85.16	0.78	0.78
Proposed ensemble model	87.50	0.81	0.82

Table 6 Comparative classification performance of proposed soft voting ensemble model in terms of TPR, FPR, Precision, and F-score.

Method	TPR	FPR	Precision	F-score
C4.5	0.77	0.11	0.77	0.77
SVM (One vs. One)	0.83	0.08	0.84	0.83
MLP (2,2)	0.85	0.07	0.85	0.85
Proposed soft voting ensemble model	0.87	0.06	0.88	0.88

Table 7 Class-wise AUC comparison of proposed soft voting ensemble model with standard classifiers for detection of juvenile delinquency.

Method	Low	Med	High	Weighted avg. ROC area (AUC)
C4.5	0.91	0.76	0.84	0.84
SVM (One vs. One)	0.96	0.82	0.94	0.90
MLP	0.96	0.75	0.94	0.88
Proposed ensemble model	0.97	0.89	0.97	0.94

ROC analysis of proposed soft voting ensemble model and comparative performance evaluation with different classifiers. A depiction of a true-positive rate (TPR) versus a false-positive rate (FPR) is known as a ROC curve. ROC is a probability curve and AUC (area under ROC-curve) is used to measure the degree of separability of the classes (Espasandin et al., 2021). In our case of multi-class classification, the ROC curve is plotted and analyzed separately for each class label. AUC is also calculated separately for each class label and then finally averaged as shown in Table 7.

These values of AUC indicate that the proposed soft voting ensemble model is highly effective in differentiating between the class labels and outperforms individual classifiers. Overall multi-class classification using the soft voting ensemble technique proves to be an efficient mode of juvenile delinquency detection at an early stage.
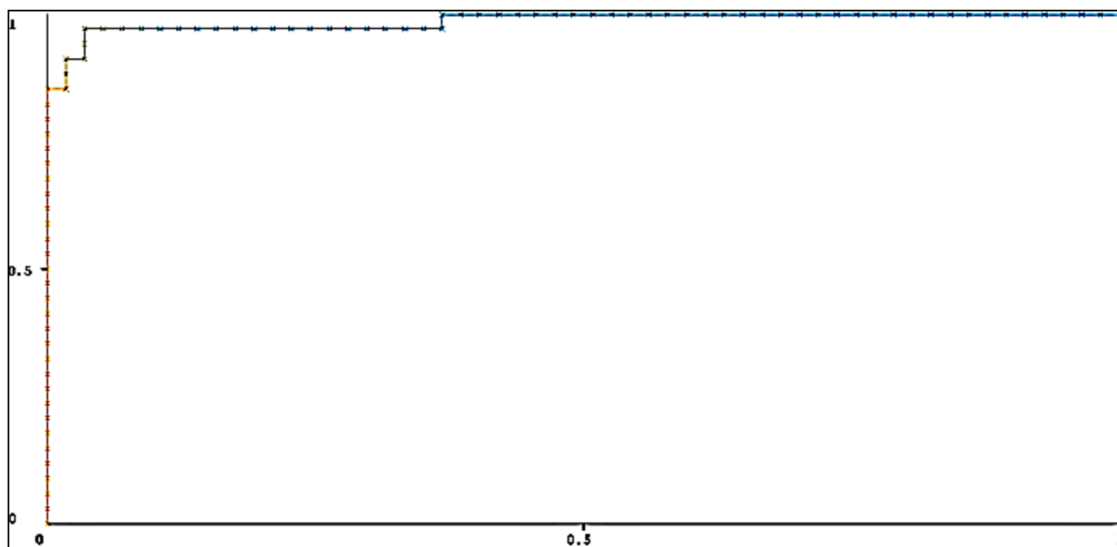
The individual class-wise plots of ROC for the proposed soft voting ensemble model showing the classification performance for classes low, medium, and high are illustrated in Fig. 7, Fig. 8, and Fig. 9, respectively.

These values of AUC and analysis of ROC-plots indicate that the proposed soft voting ensemble model is highly effective in differentiating between the class labels and outperforms individual classifiers. Overall multi-class classification using the soft voting ensemble technique proves to be an efficient mode of juvenile delinquency detection at an early stage.

**Performance validation of proposed architecture using confusion matrix.** A confusion matrix or error matrix is a special tabular structure that permits the visualization of the classifier performance in the case of supervised learning. The instances in the actual class are represented in each row of the matrix, whereas the instances in a predicted class are represented in each column or vice versa. The confusion matrix is an $n*n$ matrix, with n denoting the number of target class labels, which is shown in Fig. 10. For 2 class labels, a $2*2$ matrix is generated. In our case for three class labels $3*3$ confusion matrix is generated for each classifier after 10-fold cross-validation, which is explained in Fig. 10.

All the above results and analysis using confusion matrix, ROC-plots and indicate that the proposed soft voting model using 10-fold CV outperformed all individual machine-learning techniques with 87.50% accuracy, 0.82 MCC, AUC of 0.94, Kappa value of 0.81, and F-score of 0.88. C4.5 (Decision Tree) showed less efficiency when used as an individual classification
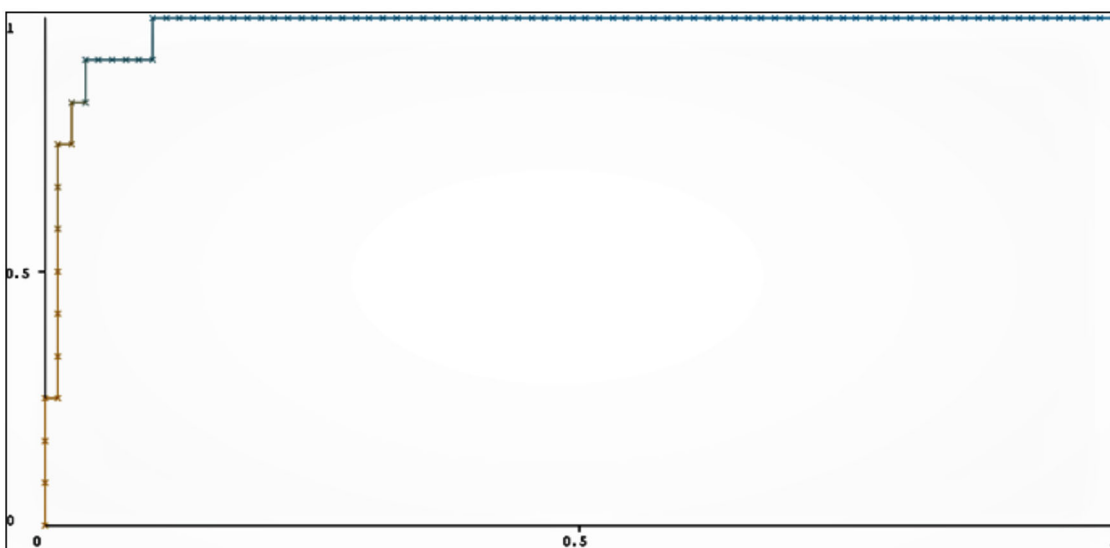

Fig. 7 The individual class-wise ROC plot for the proposed soft voting ensemble model showing the classification performance for class 'Low'.
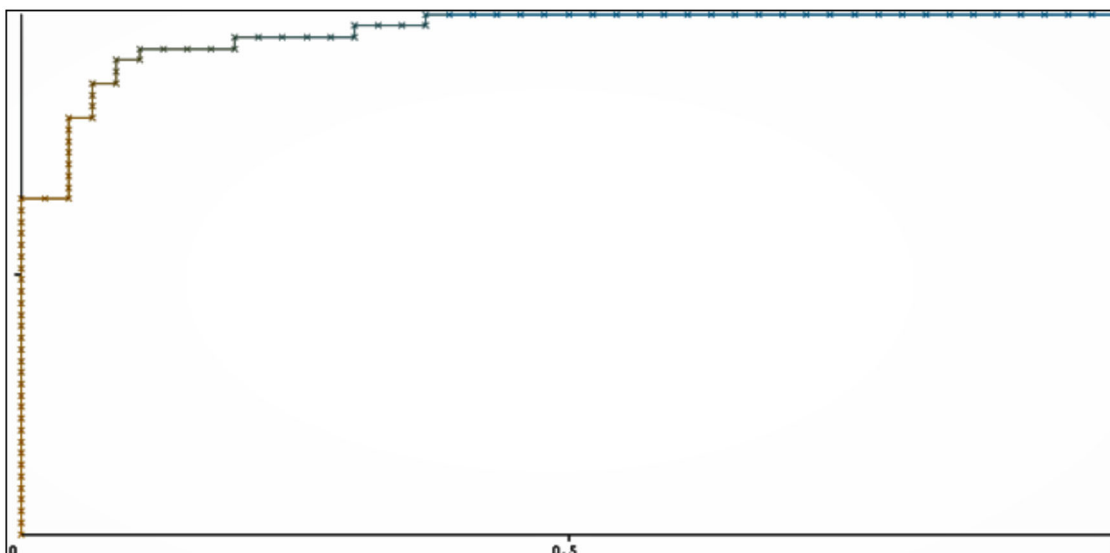
Fig. 8 The individual class-wise ROC plot for the proposed soft voting ensemble model showing the classification performance for class 'Med'.

Fig. 9 The individual class-wise ROC plot for the proposed soft voting ensemble model showing the classification performance for class 'High'.

technique but it was observed that the overall performance of the soft voting ensemble model increased when C4.5 was used as the base learner in the ensemble. MLP emerged as the best individual classifier. Overall multi-class classification using soft voting ensemble modeling emerged as the best technique in terms of model performance.

The comparison of classification performance of proposed soft voting ensemble model with classical ensemble frameworks, viz., AdaBoost (Hastie et al., 2009), Random Forest (Speiser et al., 2019), and CatBoost (Jabeur et al., 2021) in terms of Classification performance accuracy is presented in Table 8.

From the comparison of the proposed model's performance with the above-mentioned classical ensemble techniques, it can be inferred that the proposed soft voting ensemble model outperforms other ensemble techniques in terms of performance accuracy.

Conclusion

Individual factors and family effects are the most important risk factors leading to deviant attributes in adolescents. A single

Table 8 Comparison of classification performance of proposed soft voting ensemble model with classical ensemble frameworks.

Ensemble techniques	Classification performance (accuracy)
CatBoost	80.43%
AdaBoost	80.76%
Random Forest	82.54%
Proposed ensemble model	87.50%

identified risk variable or risk factor cannot be blamed for adolescent delinquency. Specific derived indicators, such as stress levels and aggressiveness are the product of multiple influences across time, ranging from heredity to the child's surroundings. Automated quantitative assessment of delinquent behavior is essential for the early diagnosis of delinquency among adolescents. Such type of quantitative delinquency risk assessment will

Fig. 10 Generated 3*3 confusion matrix for each classifier and proposed architecture after 10-Fold CV.

facilitate clinicians in the early identification of deviant behavior among adolescents for individual timely therapeutic interventions. In this study, we have proposed a novel soft voting-based ensemble model comprising three base classifiers, viz., Decision Tree (C4.5), Multi-Layer Perceptron, and Support Vector Machine for early detection of delinquency among juveniles using socio-behavioral features of adolescents as input. Encouraging results in terms of multiple evaluation metrics for quantitative grading of delinquent behavior in juveniles were obtained using the proposed ensemble classification model. Furthermore, the established classification model has provision to include additional behavioral attributes to perform advanced identification of complex behavioral disorders among juveniles. In the future, such quantitative behavioral studies will further expedite the process of automated juvenile delinquency detection involving a huge number of clinical parameters using deep-learning-based models.

Data availability
The dataset analyzed during the current study is available from the corresponding author upon reasonable request.

References
Aggarwal CC (2015) Data classification. In: Data mining, Springer, pp. 285–344
Alizad K, Medeiros SC, Foster-Martinez MR et al. (2020) Model sensitivity to topographic uncertainty in meso-and microtidal marshes. IEEE J Sel Top App Ear Obs Rem Sen 13:807–814
Alpaydin E (2021) Machine learning. MIT Press, MA
Al-qazzaz S, Sun X, Yang H (2021) Image classification-based brain tumour tissue segmentation. Multimed Tool Appl 80(1):993–1008
Barranco R, Gatti U, Verde A, Rocca G (2022) Psychosocial factors of risk and protection associated with juvenile cyberbullying victimization: results from an international multi-city study (International Self-Report Delinquency Study 3, ISRD3). Psy Psych Law 1–21
Benussi A, Grassi M, Palluzzi F (2021) Classification accuracy of TMS for the diagnosis of mild cognitive impairment. Brain Stimul 14(2):241–249
Bor W, Najman J et al. (2001) Aggression and the development of delinquent behaviour in children. Australian Institute of Criminology, Canberra
Bowers AJ, Zhou X (2019) Receiver operating characteristic (ROC) area under the curve (AUC): a diagnostic measure for evaluating the accuracy of predictors of education outcomes. J Educ Stud Place Risk 24(1):20–46
Britt CL (1997) Reconsidering the unemployment and crime relationship: Variation by age group and historical period. J Quant Criminol 13(4):405–428
Castellana GB, Barros DMD, Serafim ADP (2014) Psychopathic traits in young offenders vs. non-offenders in similar socioeconomic condition. Braz J Psychiat 36:241–244
Castro ET, Hernandez AA (2019) Developing a predictive model on assessing children in conflict with the law and children at risk: a case in the Philippines. In: IEEE 15th International Colloquium on Signal Processing & Its Applications, Penang, 8–9 Mar 2019
Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom 21(1):1–13
DeVries Z, Locke E, Hoda M (2021) Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and F1-score for the assessment of prognostic capability. Spine J 21(7):1135–1142
Duan K, Keerthi SS, Poo AN (2003) Evaluation of simple performance measures for tuning SVM hyperparameters. Neurocomputing 51:41–59
Enzmann D, Kivivuori J, Haen Marshall I et al. (2018) Self-reported offending in global surveys: a stocktaking. In: A global perspective on young people as offenders and victims, Springer, pp. 19–28
Espasandin AG, Cipolini MF, Forletti A (2021) Comparison of serological techniques for the diagnosis of equine infectious Anemia in an endemic area of Argentina. J Virol Method 291:114101
Feld BC, Bishop DM (2011) The Oxford handbook of juvenile crime and juvenile justice. Oxford University Press
Gearhart MC, Tucker R (2020) Criminogenic risk, criminogenic need, collective efficacy, and juvenile delinquency. Crime Just Behav 47:1116–1135
Gray PA (1987) The relative effects of family factors and opportunity factors on juvenile delinquency. Doctoral dissertation, Iowa State University
Greener JG, Kandathil SM, Moffat L et al. (2022) A guide to machine learning for biologists. Nat Rev Mol Cell Biol 23(1):40–55
Grimm KJ, Mazza GL, Davoudzadeh P (2017) Model selection in finite mixture models: a k-fold cross-validation approach. Struct Equat Model Multidiscip J 24(2):246–256
Hastie T, Rosset S, Zhu J, Zou H (2009) Multi-class adaboost. Stat Interf 2(3):349–360
Howell JC (2003) Preventing and reducing juvenile delinquency: a comprehensive framework. Sage
Huijsmans T, Eichelsheim VI, Weerman F et al. (2019) The role of siblings in adolescent delinquency next to parents, school, and peers: do gender and age matter? J Dev Life-Course Criminol 5(2):220–242
Ibrahim S (2016) Causes of socioeconomic cybercrime in Nigeria. In: IEEE International Conference on Cybercrime and Computer Forensic, Canada, 1–9 Jun 2016
Jabeur SB, Gharib C, Mefteh-Wali S, Arfi WB (2021) CatBoost model and artificial intelligence techniques for corporate failure prediction. Technol Forecast Soc Chang 166:120658
Jacobsen SK, Zaatut A (2022) Quantity or quality?: assessing the role of household structure and parent-child relationship in juvenile delinquency. Deviant Behav 43(1):30–43
Jain A, Nandakumar K, Ross A (2005) Score normalization in multimodal biometric systems. Pattern Recog 38(12):2270–2285
Kasim O (2021) An ensemble classification-based approach to detect attack level of SQL injections. J Inform Secur Appl 59:102852
Kieu LM, Ou Y, Truong LT et al. (2020) A class-specific soft voting framework for customer booking prediction in on-demand transport. Transport Res Emerg Technol 114:377–390
Krohn MD, Lane J (2015) The handbook of juvenile delinquency and juvenile justice. John Wiley & Sons, United Kingdom
Kumar D, Batra U (2021) Breast cancer histopathology image classification using soft voting classifier. In: Proceedings of 3rd International Conference on Computing Informatics and Networks. Springer, Singapore, pp. 619–631
Lansford JE, Miller-Johnson S, Berlin LJ et al. (2007) Early physical abuse and later violent delinquency: a prospective longitudinal study. Child Maltreat 12(3):233–245
Marcot BG, Hanea AM (2021) What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? Comput Stat 36(3):2009–2031
Marshall IH, Neissl K, Markina A (2019) A global view on youth crime and victimization: results from the International Self-Report Delinquency Study (ISRD3). J cont cri jus 35(4):380–385

Meldrum RC, Barnes JC, Hay C (2015) Sleep deprivation, low self-control, and delinquency: a test of the strength model of self-control. J Youth Adolesc 44(2):465–477

Morales J, Bor P, Tes D (2021) Linear and non-linear quantification of the respiratory sinus arrhythmia using support vector machines. Frony Physiol 12:58

Nunnally JC (1975) Psychometric theory—25 years ago and now. Educ Res 4(10):7–21

Pauly M, Smaga L (2020) Asymptotic permutation tests for coefficients of variation and standardised means in general one-way ANOVA models. Stat Met Med Res 29(9):2733–2748

Polat K, Guneş S (2009) A novel hybrid intelligent method based on C4. 5 decision tree classifier and one-against-all approach for multi-class classification problems. Exp Syst Appl 36(2):1587–1592

Rathinabalan I, Naaraayan SA (2017) Effect of family factors on juvenile delinquency. Int J Contemp Pediatr 4(6):2079–2082

Ricci E, Perfetti R (2007) Retinal blood vessel segmentation using line operators and support vector classification. IEEE Transact Med Imag 26(10):1357–1365

Saefi M, Fauzi A, Kristiana E et al. (2020) Survey data of COVID-19-related knowledge, attitude, and practices among indonesian undergraduate students. Date Brief 31:105855

Saud S, Jamil B, Upadhyay Y, Irshad K (2020) Performance improvement of empirical models for estimation of global solar radiation in India: a k-fold cross-validation approach. Sustain Energ Technol Assess 40:100768

Schratz P, Muenchow J, Iturritxa E, Richter J, Brenning A (2019) Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecol Model 406:109–120

Sciandra M, Sanbonmatsu L, Duncan GJ et al. (2013) Long-term effects of the moving to opportunity residential mobility experiment on crime and delinquency. J Exp Criminol 9(4):451–489

Siegel LJ, Welsh BC (2014) Juvenile delinquency: theory, practice, and law. Cengage Learning

Sobri AY, Bafadal I, Nurabadi A et al. (2019) Validity and reliability of questionnaire problematics leadership beginner school principals. In: 4th International Conference on Education and Management, Universitas Negeri Malang, Indonesia

Speiser JL, Miller ME, Tooze J, Ip E (2019) A comparison of random forest variable selection methods for classification prediction modeling. Expert Syst Appl 134:93–101

Svensson R, Weerman FM, Pauwels LJ et al. (2013) Moral emotions and offending: do feelings of anticipated shame and guilt mediate the effect of socialization on offending? Eur J Criminol 10(1):22–39

Taber KS (2018) The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 48(6):1273–1296

Taud H, Mas JF (2018) Multilayer perceptron (MLP). In: Geomatic approaches for modeling land change scenarios, Springer, pp. 451–455

Tharwat A, Gabel T (2020) Parameters optimization of support vector machines for imbalanced data using social ski driver algorithm. Neural Comput Appl 32(11):6925–6938

Vapnik V, Izmailov R (2021) Reinforced SVM method and memorization mechanisms. Pattern Recog 119:108018

Vaux A, Ruggiero M (1983) Stressful life change and delinquent behavior. Am J Community Psychol 11(2):169

Wang P, Fan E, Wang P (2021) Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern Recog Lett 141:61–67

Xian Z, Yang H (2021) An early warning model for the stuck-in medical drilling process based on the artificial fish swarm algorithm and SVM. Distrib Parall Databas 1–18

Zhang W, Chen J, Feng Y et al. (2014) Evaluation of a sexual abuse prevention education for Chinese preschoolers. Res Soc Work Pract 24(4):428–436

Competing interests

The authors declare no competing interests.

Ethical approval

This is to certify that the research work on "An automated system for assessment and grading of adolescent delinquency using machine learning-based soft voting framework" is an original research work carried out by both authors following all necessary guidelines. This research work does not involve any human clinical studies and hence is exempted from the institute ethics approval committee.

Informed consent

In the present study, the authors have taken the required permission from the principal of each school for interviewing each student. Necessary permissions for conducting the survey were granted to the authors with a condition that the interview will be conducted in presence of the class teacher. In addition, informed consent was also taken from each participant after explaining to them about the research itself as well as their rights to participate or not. Moreover, verbal informed consent was also taken from the parents of individual participants through voice calls after explaining to them the research objectives.

Additional information

Correspondence and requests for materials should be addressed to Abhinash Jenasamanta.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.