Computers, Materials & Continua
DOI: 10.32604/cmc.2022.026077
Article

Tech Science Press

Decision Level Fusion Using Hybrid Classifier for Mental Disease Classification

Maqsood Ahmad[1,2], Noorhaniza Wahid[1], Rahayu A Hamid[1], Saima Sadiq[2], Arif Mehmood[3] and Gyu Sang Choi[4,*]

[1]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia
[2]Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, 64200, Pakistan
[3]Department of Information Technology, The Islamia University of Bahawalpur, Bahawalpur, 63100, Pakistan
[4]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38542, Korea
*Corresponding Author: Gyu Sang Choi. Email: castchoi@ynu.ac.kr
Received: 15 December 2021; Accepted: 16 February 2022

Abstract: Mental health signifies the emotional, social, and psychological well-being of a person. It also affects the way of thinking, feeling, and situation handling of a person. Stable mental health helps in working with full potential in all stages of life from childhood to adulthood therefore it is of significant importance to find out the onset of the mental disease in order to maintain balance in life. Mental health problems are rising globally and constituting a burden on healthcare systems. Early diagnosis can help the professionals in the treatment that may lead to complications if they remain untreated. The machine learning models are highly prevalent for medical data analysis, disease diagnosis, and psychiatric nosology. This research addresses the challenge of detecting six major psychological disorders, namely, Anxiety, Bipolar Disorder, Conversion Disorder, Depression, Mental Retardation and Schizophrenia. These challenges are mined by applying decision level fusion of supervised machine learning algorithms. A dataset was collected from a clinical psychologist consisting of 1771 observations that we used for training and testing the models. Furthermore, to reduce the impact of a conflicting decision, a voting scheme Shrewd Probing Prediction Model (SPPM) is introduced to get output from ensemble model of Random Forest and Gradient Boosting Machine (RF + GBM). This research provides an intuitive solution for mental disorder analysis among different target class labels or groups. A framework is proposed for determining the mental health problem of patients using observations of medical experts. The framework consists of an ensemble model based on RF and GBM with a novel SPPM technique. This proposed decision level fusion approach by combining RF + GBM with SPPM-MIN significantly improves the performance in terms of Accuracy, Precision, Recall, and F1-score with 71\%, 73\%, 71\% and 71\% respectively. This framework seems suitable in the case of huge and more diverse multi-class datasets. Furthermore, three vector spaces based on TF-IDF (unigram,

bi-gram, and tri-gram) are also tested on the machine learning models and the proposed model.

Keywords: Mental health diagnosis; machine learning; depression; shrewd probing; diagnostic approach

1 Introduction

Mental disorders are highly prevalent in the population and impact on the feelings or mood of affected people. The symptoms range from months to years in terms of duration, and also mild to be cured in terms of their severity. Since the mental illness is increasing at epidemic rates, it affects all aspects of life and people of all ages. It is noteworthy that the risk of mental disorder is increased with the unemployment, poverty, physical illness, and use of alcohol or drugs [1]. According to a World Health Organization (WHO) report, about 450 million people are affected by mental disorders and placing the disease as a leading cause of health illness globally [2]. These mental disorders are diagnosable and health conditions are clearly distinguishable from fear, stress or sadness that one can experience in life.

However, mental disorder diagnosis is not straightforward and it involves various steps. Diagnosis starts with the symptoms, medical history, and specially designed interview questions and, sometimes by physical examination. Different psychological tests are also conducted to ensure mental health reasons behind symptoms. Numerous assessment tools are available for the evaluation of mental disorders. Furthermore, the cooperation of patients is highly required in the diagnosis phase as it is regarded as a complicated task. It is pertinent to mention that the number of practitioners is very less as compared to the number of people suffering from mental illness.

There are a number of different types of the mental disorders that are being discovered namely, Anxiety, Bipolar Disorder, Conversion Disorder, Depression, Mental Retardation, and Schizophrenia. The anxiety problems are mostly chronic and characterized by excessive worries about different work, money, and family-related issues that result in fatigue, restlessness, and lack of concentration [3]. The bipolar disorder is a mental illness that causes frequent mood swings and a shift in energy and concentration levels [4]. On the other hand, the Conversion Disorder commonly affects the nervous system along-with the physical distress or emotional disturbance [5]. The depression is the most common and serious mental health problem that negatively affects the feeling, thinking, and actions that may provoke the desire for suicide [6]. Similarly, the Mental Retardation is associated with the development delay during early childhood and causes impairment in cognitive functions [7]. The last one, the Schizophrenia is one of the most serious mental disorder in which patient abnormally interpret the reality and daily life tasks that is one of the causes for the hallucinations and disordered thinking [8].

In order to gather the mental illness information, electronic health records that are based on disease symptoms, are considered as a valuable source of data that can support an extensive range for secondary informatics usage, such as observational research, decision support and business intelligence. The health records may be able to overcome the cost barriers to address the issues using some appropriate way that would be out of the reach of common patients. This activity helps in using the health related data to reduce the burden of health professionals. Furthermore, it will also provide cost effective solutions to patients without any further examinations or tests. Thus, there is a need to apply non-linear programming (NLP) to develop an appropriate hybrid machine learning-based

approach to capture symptoms from clinical text and to facilitate the use of mental health symptoms data in research.

Nowadays, artificial intelligence (AI) exhibited by machines helps computers to imitate human logic to solve problems. The researchers have developed many machine learning (ML) models to deal with a huge amount of information, and incomplete and, uncertain information. ML models can infer mental states by observing behaviors [9], and can predict depression [10,11], anxiety [12], stress [13], autism [14], as well as the suicide risks [15]. Many ML models such as Naïve Bayes (NB), Decision Tree (DT), Gradient Boosting Machine (GBM), k-Nearest Neighbor (kNN), Random Forest (RF), Support Vector Machine (SVM), and Neural Network (NN) are performing well in mental disease classification using text [16].

The aim of this research is to explore the ML techniques to diagnose mental health issues over a dataset of 1771 observations. These techniques are then compared by utilizing the TF-IDF feature set. The 6 mental health problems such as Anxiety, Bipolar Disorder, Conversion Disorder, Depression, Mental Retarded, and Schizophrenia are considered for our experimental work. The symptoms and the factors observed by medical professionals are then used as input to the models that in turn detects the type of the disease.

Thus, a novel framework of an ensemble model (RF + GBM) using Shrewd Probing Prediction Model (SPPM) is proposed to diagnose mental disease in multi-class recognition problem. The major contributions of this research are two-fold: 1) to propose an ensemble classifier by aggregating the bagging and boosting algorithms namely (RF + GBM) to automatically estimate mental diseases, and 2) to identify a decision level fusion to generate the final prediction of (RF + GBM) using SPPM-MIN. The experimental evaluation shows that the proposed framework, RF + GBM using SPPM-MIN generates comparable results with other ML techniques individually as well as their combinations as Voting Classifiers (VC). ML models used for comparison include RF, GBM, LR, SVM, VC (LR + GBM), VC (LR + SVM), VC (GBM + SVM), VC (RF + LR), VC (RF + SVM), VC (RF + GBM). Following ensemble models: LR + GBM, LR + SVM, GBM + SVM, RF + LR, RF + SVM and RF + GBM are used with SPPM-MIN and SPPM-MAX. The comparative analysis was performed using feature engineering technique TF-IDF involving Unigrams, bi- and tri-grams.

The rest of the paper is organized as follows: Section 2 discusses about related work, data collection is explained in Section 3, Section 4 presents the proposed framework for predicting six mental disorders, Section 5 explains about the proposed framework, Section 6 discusses the experimental results and discussion, and Section 7 finally concludes the work.

2 Related Work

Machine learning techniques have been used since the last decades in mental health diagnosis and also facilitated the clinical psychiatry in prognosis of mental disorders [17] such as depression [18], anxiety [19], and autism [20]. The Diagnostic and Statistical Manual of Mental Disorders (DSM) was published in May 2013, and came up with the standard language by which researchers and public health officials communicate regarding mental disorders [21]. Since that, an expert system was proposed to help psychologists in diagnosis and treatment of mental disorder patients using fuzzy logic, fuzzy genetic algorithms and rule based approach [22]. In addition, current trends and applications in psychiatry by applying artificial intelligence were also been discussed [23].

Rostami et al. [24] applied decision tree as classification algorithm to evaluate neuropsychology in ADHD diagnosis. Chattopadhyay et al. [25] graded adult depression by utilizing a neuro-fuzzy

approach. They differentiated depression grades as mild, severe and moderate using back propagation neural network (BPNN). They generated hierarchical tree from depression parameters and feed to neural network. A hybrid approach was applied to diagnose schizophrenia by integrating structured methodology in the decision support system [26].

Rahman et al. [27] compared various ML techniques such as DT, Bayesian Network, MLP, Rule Based Approach, Fuzzy Inference and Neuro fuzzy Inference with respect to different evaluation measures (accuracy, error rate, Kappa stats). Performance of linear discriminant analysis (LDA), k-NN, NB, regression trees (RT), SVM, radial basis function neural network and mahalanobis distance classifier is compared in Parkinson's disease prediction [28]. SVM outperformed with 92% accuracy than other classifiers on dysphonia symptoms.

Kipli et al. [29] used MRI scans for mental health diagnosis. They deployed four approaches based on feature selection; Information gain (IG), SVM evaluator, ReliefF and OneR. They found SVM evaluator and IG with Random Tree Classifier best in disease diagnosis. In Seixas et al. [30], the authors performed an experiment using Decision based Bayesian Network to diagnose mild cognitive impairment, dementia and Alzheimer's disease. They applied supervised machine learning approach on real clinical dataset and also performed sensitivity analysis using quantitative methods and achieved reasonable results as compared to other well-known classifiers.

Authors compared six models namely Logistic Regression (LR), Radial Basis Function (RBF), Polygenic Scoring, Bayesian Network, SVM, and RF in association of genome to predict mood disorders [31]. Simple polygenic scoring measure performed well with a complete genome set. Yalamanchili et al. [32] diagnosed depression using acoustic features on DIAC-WOZ database. They also deployed future fusion on voice and spectral features. They also used SMOTE technique to deal with class imbalance problem and also tested their model on real time data. Priya et al. [33] predicted anxiety, stress and depression using five machine learning models (DT, NB, RF, SVM & KNN) and also considered five severity levels. In their work RF achieved highest result.

An ensemble learning of ML models is being widely used in many researches for mental disease classification [34]. Idea behind ensemble learning is to combine different ML based models to reduce bias and variance and result in improvement in classification result [35]. Previous studies have shown significant outcome to improve medical decision making process by using ensemble learning. Bashir et al. [36] utilized ensemble framework of multi-layer classifier for disease prediction and showed significant results on disease dataset. Ozcift et al. [37] proposed ensemble of 30 classifiers to predict Parkinson's disease and heart disease and achieved good results. Bagging (RF) and boosting (GBM) models are being extensively used in disease diagnosis [38,39]. Sakr et al. [34] predicted risk of hypertension using RF. Results proved that RF achieved highest AUC score of 0.93 as compared to NB, SVM and neural network. Chekroud et al. [40] predicted clinical remission from depression using GBM and achieved 64% accuracy. Decision level fusion of features and classifiers has been proposed in the literature for disease classification [41–43]. Therefore in this study decision level fusion of RF and GBM is proposed to improve the performance of ML models.

The literature reveals that on one side a number of efforts are going on automation of mental health diagnosis. On another way round, efforts are taken to use ML approaches in an efficient and improved way. Hybrid approach is an important step for text classification tasks in this regard. We analyze ML models in predicting mental health disorders like Anxiety, Bipolar Disorder, Conversion Disorder, Depression, Mental Retarded and Schizophrenia. Early diagnosis of these mental health problems helps professionals to treat in a better way and improve the quality of patient's life.

3 Data Collection

Psychological dataset for predicting mental health problems has been collected from clinical psychologist of Shakoor Mind Care (SMC) Hospital Bahawalpur, Pakistan. The dataset contains 1771 observations/objects observed by the psychological and medical experts of the hospital from 2012 to 2018. The dataset consists of 6 different categories/classes includes anxiety, bipolar disorder, conversion disorder, depression, mental retarded and schizophrenia. Fig. 1 represents the name and count of each disease category in the dataset. There are 467 records in bipolar disorder category which is highest in number and 151 records in mental retardation which is lowest in number. Some contents of the dataset are presented in Tab. 1.

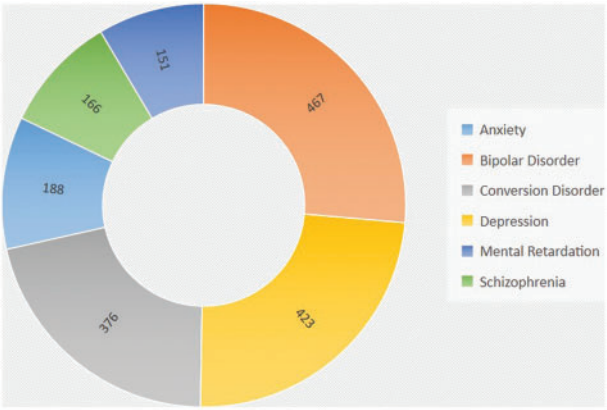

Figure 1: Count of each category of mental disease in the dataset

Table 1: Sample of SMC dataset sample

PID	Presentation	Diagnosis
20120832	Numbness of head, lack of sleep, irritability, headache after awaking, decreased appetite, depressed mood, she feels that something going to happen	ANXIETY
20121088	She was alright before 2 month ago but she having symptoms of back headache, decreased appetite, vomiting, pain tummy, lack of sleep. weeping tendency, irritability	BIPOLAR DISORDER
20120049	Fits, pain in legs and body, pain in chest, problem in breathing	CONVERSION DISORDER
20123827	Headache, dizziness, difficulty in breathing, decrease sleep and appetite, aggression, lack of interest, feeling of weakness, body aches, urine problem during menses	DEPRESSION
20121002	Less attention in studies for 3 year, aggression for 3 year. hyperactivity for 3 year	MENTAL RETARDED

(Continued)

Table 1: Continued

PID	Presentation	Diagnosis
20121019	She was alright before 5 days ago but now she having symptoms of weeping tendency, headache, suspiciousness, fear of death, auditory hallucination	SCHIZOPHRINIA

4 Proposed Framework

This section presents all four modules of proposed framework: (1) pre-processing; (2) feature extraction (3) ensemble classifiers (RF + GBM), and (4) Shrewd Probing Prediction Model (SPPM). Fig. 2 illustrates the data flow model with their components for solving mental disease classification problem. At one end, the text of mental disease symptoms observed by medical professionals is fed as input to the system, and at the other end, predictive results are generated as output. A detail activity of each step in the proposed framework is explained in the following sub-section.

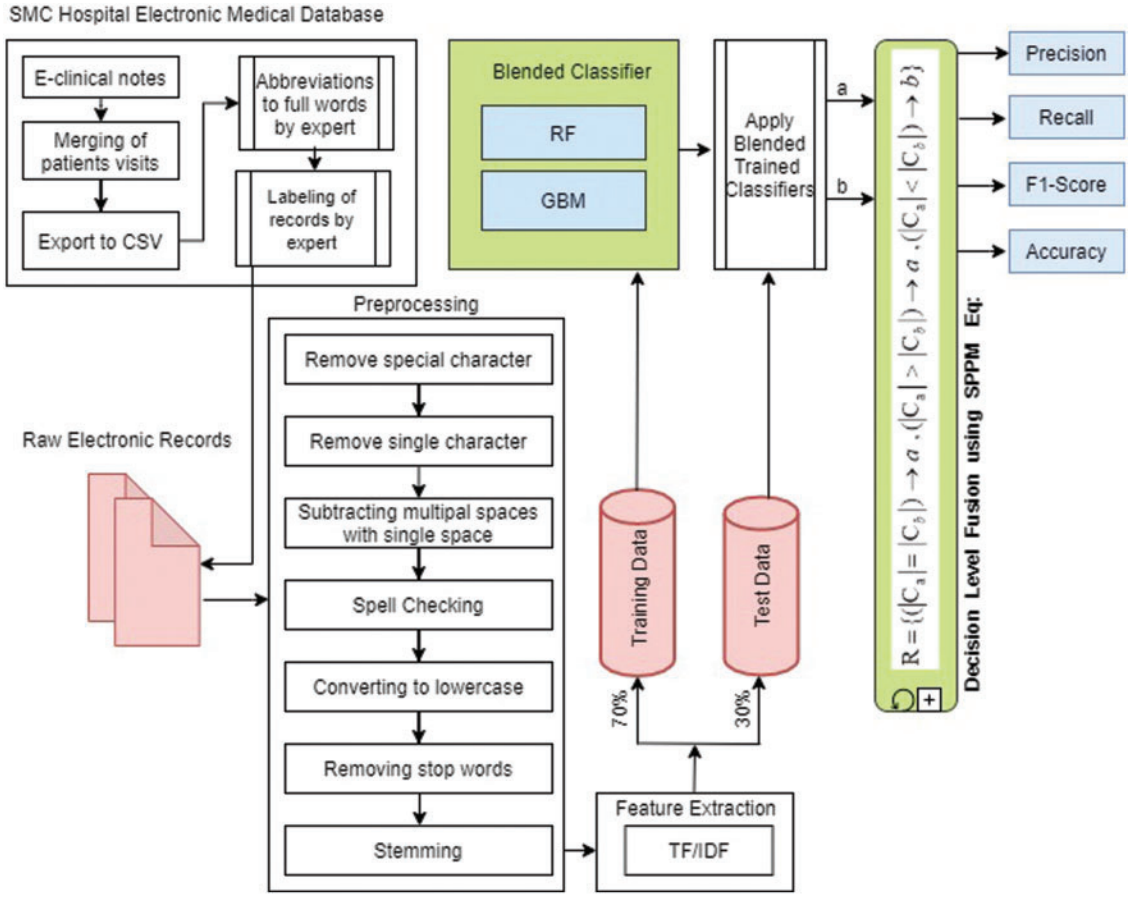

Figure 2: Pipeline of proposed architecture

4.1 Preprocessing

In data preprocessing, useful information from the hospital electronic database were extracted from the text by removing extra and irrelevant text. Dimensions of features were reduced in this way, which also reduced the processing time. Tokenization, data cleaning, spelling correction, conversion to lowercase and stemming were performed. The spelling correction was done by Peter Norvig's spelling corrector [44] who handles the jargon words as well. The spellings were corrected to improve the accuracy of the model. Moreover, stemming was applied using the Porter Stemmer algorithm [45]. The purpose of stemming was to get the root form of any word used in the corpus.

4.2 Feature Extraction

The conversion of raw documents in numerical form is called feature extraction. These features become the input of any classifier for classification and prediction process. Several feature extraction methods are used by researchers to represent text. TF-IDF is being widely used in text classification. TF-IDF assigns weights to the term and most important term is assigned with higher weight [46]. In this experiment, TF-IDF was used in terms of Unigrams, Bi-grams and Tri-grams. The weight of TF-IDF was calculated by using Eq. (1).

$$W_{i,j} = TF_{i,j}\left(\frac{N}{D_{f,t}}\right) \tag{1}$$

where N is the total documents, $D_{f,t}$ is the number of documents D containing term t and $TF_{i,j}$ is the occurrence of term t in document D.

4.3 Machine Learning Classifiers

In this section, the ML models are presented. The task of classification is mostly performed by supervised ML model. Many researchers used ensemble classifiers for medical disease decision support [36]. Most common ensemble techniques are bagging [47] and boosting [48]. RF is a tree-based model and use bagging to make prediction while GBM is a boosting algorithm.

4.3.1 Random Forest (RF)

Random forest (RF) is an ensemble classifier which creates multiple decision trees from boot-strap dataset using bagging technique. Boot-strap sample is obtained by sub-sampling of training data, but the size remains the same as training dataset. Major concern of RF is the selection of root attribute at each level of tree, which is also called as attribute selection. Every time input is passed to each decision tree, that tree predicts output independently and vote for that specific class. Random forest overall prediction is decided on the basis of majority voting by decision trees [49]. This voting technique reduces outliers as compared to the output predicted by single decision tree. Bagging technique improves performance and robustness of model.

RF can deal with various input parameters as it contains built-in feature selection method. Variable importance score can also be determined by RF to handle error due to out of the bag observations. This score is calculated from each tree and averaged across all ensembles by dividing from standard deviation. RF can be explained as in following Eqs. (2) and (3).

$$\mathrm{P} = \mathrm{mode}\left\{T_1(y), \quad T_2(y), \ldots, T_m(y)\right\} \tag{2}$$

$$p = mode\left\{\sum_{m=1}^{m} T_m(y)\right\} \tag{3}$$

where $T_1(y)$, $T_2(y)$, $T_3(y)$ and $T_m(y)$ are the number of decision trees participating in the prediction process and p is the finalized prediction by majority voting of decision trees.

4.3.2 Gradient Boosting Machine (GBM)

The GBM is based on iterative decision tree algorithm and originally designed by Friedman [50]. In ensemble technique, gradient boosting is used for both regression and classification which make prediction on the ensemble of weak models such as decision trees. Boosting refers to conversion of weak-learners to strong-learners. GBM uses gradient in the loss function. The loss function measures the error and loss, which represents the credibility of the classifier. The loss function is used to measure efficiency of model coefficients. Boosting model has several advantages such as generates a tree model on the basis of a previous model, uses a weighted majority vote, and reduces the variance and bias of the base classifier.

In GBM, a base class of learners is presented as B and the target function class which is the linear combination of such base learners is denoted by lin(B). The prediction corresponding to a feature vector x is given as in Eq. (4).

$$f(x) = \left(\sum_{m=1}^{M} \beta_m b_{T,m}(x) \right) \varepsilon \; lin(B) \tag{4}$$

where bT, m(x) ε B is a weak-learner and βm is its corresponding additive coefficient. Here, bTm and βm are chosen in an adaptive fashion. The goal of GBM is to obtain a good estimate of the function f that approximately minimizes the empirical loss, as given in Eq. (5).

$$L* = \min_{f \in lin(B)} \left\{ L(f) := \sum_{i=1}^{n} l(y_i, f(x_i) \right\} \tag{5}$$

where $(y_i, f(x_i))$ is a data-fidelity measure for the i^{th} sample for the loss function.

4.3.3 Logistic Regression (LR)

Logistic regression is a linear classifier based on probabilities. LR predicts class probabilities using logit transform and then converts probability to class. It describes the relationship between dependent variable and independent variable. It is a statistical model and draws a log line that distinguishes between output variables [51]. It is being used efficiently for classification as well as regression tasks and provides low variance. LR can be updated by stochastic gradient descent. A logistic function that is S-shaped curve, as shown in Eq. (6).

$$f(x) = \frac{L}{1 + e^{-m(v - v_o)}} \tag{6}$$

where e is the natural log, v_o is the x-value of the sigmoid midpoint, L is the maximum value of curve and m is the steepness of the curve.

4.3.4 Support Vector Machine (SVM)

Support Vector Machine separates data into different labels or classes on the basis of hyperplane that act as decision boundary. This model can be utilized for linear or non-linear data. It maps input data to high dimensional space using kernel mapping to make the problem linearly separable. SVM has been extensively used for binary classification as well as multi-class classification problems and text classification [52]. SVM is performing well on text and categorical data.

4.3.5 Ensemble Models

Ensemble model is a blending of models that performs prediction by combining probability score values of different base learning models. Final output is obtained by aggregating the result of the base classifiers. In general, models are blended as a voting classifier and the final result can be obtained on the basis of soft voting or hard voting. In hard voting, class labels are predicted on the basis of majority voting of each classifier. On the other hand in soft voting consider the prediction of class labels from each classifier. It determines the class label with high probability after taking the average of each classifier value. In this study, voting classifiers (VC) was used for comparison which based on soft voting and expressed as Eq. (7).

$$\hat{p} = Max\left(\sum_{i}^{n} C1_i,\ \sum_{i}^{n} C2_i\right) \tag{7}$$

where $\sum_{i}^{n} C1_i$ will give prediction probability for first classifier and $\sum_{i}^{n} C2_i$ will give prediction probability for second classifier. Let C1 classifier score are 0.710554 and 0.988724 for class label positive and negative respectively and C2 classifier probability scores are 0.21110 and 0.682451 for positive and negative class respectively. Average probability will be calculated as:

$$\text{Avg (Positive)} = (0.710554 + 0.21110)\ /2 = 0.460827$$

$$\text{Avg (Negative)} = (0.988724 + 0.682451)\ /2 = 0.8355875$$

According to Eq. (7), final prediction is max (Avg(positive), Avg(Negative)) which is negative in this case.

In order to take the advantages of bagging and boosting techniques, this research proposed to ensemble RF and GBM since these values possess different flavor of computations. Besides, boosting performs better on non-noisy data than that of bagging. Since both are different in operations with different limitations, combining them together helps to tackle the deficiency of each other. Due to variation of text data in length, structure and domain, a single model cannot perform better in the experimental datasets.

4.4 Shrewd Probing Prediction Model (SPPM)

The traditional methods to combine results of classifiers are by taking average or majority voting, but there is still lack of the ability to represent all classes in multi-class classification. Therefore, a novel technique namely Shrewd Probing Prediction Model (SPPM) is developed which combines the RF and GBM. The experimental results showed that SPPM could improve the performance and efficiency of the model for mental disorders classification. This study employs SPPM in two scenarios. SPPM-MAX considers the result of the ensemble model matching with majority occurrence while SPPM-MIN considers the ensemble model matching with the least occurrence class of the dataset.

SPPM-MAX works together with the proposed ensemble classifiers of (RF + GBM) and consider the result having majority occurrence in the dataset as presented by Eq. (8).

$$R = \{(|C_a| = |C_b|) \rightarrow a,\ (|C_a| > |C_b|) \rightarrow a,\ (|C_a| < |C_b|) \rightarrow b\ \} \tag{8}$$

where a is the output predicted by RF and b is the output predicted by GBM. While $|C_a|$ is the occurrence of output a in the dataset, $|C_b|$ is the occurrence of output b in the dataset and R is the final result of the instance. This method used RF and GBM to predict the mental illness then their

results are passed through the equation of SPPM-MAX, and the obtained result, R is considered as final output. As shown in Eq. (8), if occurrence of the output of both models are equal in the dataset then any of the output can be considered as final output. If occurrence of the output a in the dataset is greater than the occurrence of the output b in the dataset then a will be considered as final output. If occurrence of the output a is lower than the occurrence of output b in the dataset then b will be considered as final output.

In SPPM-MAX, the result of the model with majority occurrence class in the dataset is considered as more important as compared to the result of the model with lower occurrence class. In general, the results from the two ensemble models are combined by taking their average.

To further increase the diversity in the multi-class classification, SPPM-MIN is used in finalizing the final output from the two ensemble model's result. SPPM-Min works together with the proposed ensemble classifiers of (RF + GBM) and consider the result having least occurrence in the dataset as presented by Eq. (9).

$$R = \{(|C_a| = |C_b|) \rightarrow a,\ (|C_a| < |C_b|) \rightarrow a,\ (|C_a| > |C_b|) \rightarrow b\ \} \tag{9}$$

This technique uses RF and GBM to predict the mental illness then their results are passed through the Eq. (9) of SPPM-MIN, and the obtained result, R is considered as final output. As shown in Eq. (9), if occurrence of the output of both models are equal in the dataset then any of the output can be considered as final output, in this case a will be final output. If occurrence of the output a in the dataset is lower than the occurrence of the output b in the dataset then a will be considered as final output. If occurrence of the output a is greater than the occurrence of output b in the dataset then b will be considered as final output.

5 Experiment and Results

The proposed approach was tested on psychological dataset predicting mental health problems. Dataset is based on 1771 observations observed by medical expert and consist of six different classes. After completing the feature extracting process using TF-IDF, the dataset was split into train/test with 70:30 ratios. After splitting the data, the ensemble classifiers (RF + GBM) were trained on the train data. In this experiment, (RF + GBM) were implemented using NLTK and Scikit-learn library [53]. (RF + GBM) were tuned with different hyper-parameters as shown in Tab. 2.

Table 2: Machine learning models and detail of hyper parameters

ML Algorithms	Hyper parameters
RF	n_estimator = 500, random_state = 22
GBM	n_estimators = 400, max_features = 18, max_depth = 20, random_state = 1

This study implemented RF + GBM with different values of hyper parameters. In RF, *n*_estimator is defined as number of trees used for prediction. For the experimental purpose, the RF was trained on 500 trees and the final prediction was made on voting of all trees. Random_state variable that was used for randomness of samples is set as 22. In GBM, *n_estimator* is set as 400 that enable the GBM to combine 400 *weak-learners* to make final prediction. The *max_features* was set as 18, *max_depth* as 20

and *random_state* as 1 during training of GBM. Next, the ensemble classifiers of RF + GBM trained on the training dataset and calculate the prediction average. Then, the test data was passed through the trained model obtained from RF + GBM where the SPPM-MAX and SPPM-MIN were used to validate the final output of the both models, In Fig. 3, the class label predicted by RF and GBM is called as *p1* and *p2*, respectively. If the occurrence of *p1* is equal to or greater than *p2* in the dataset, then *p1* is considered as final output or, *p2* otherwise.

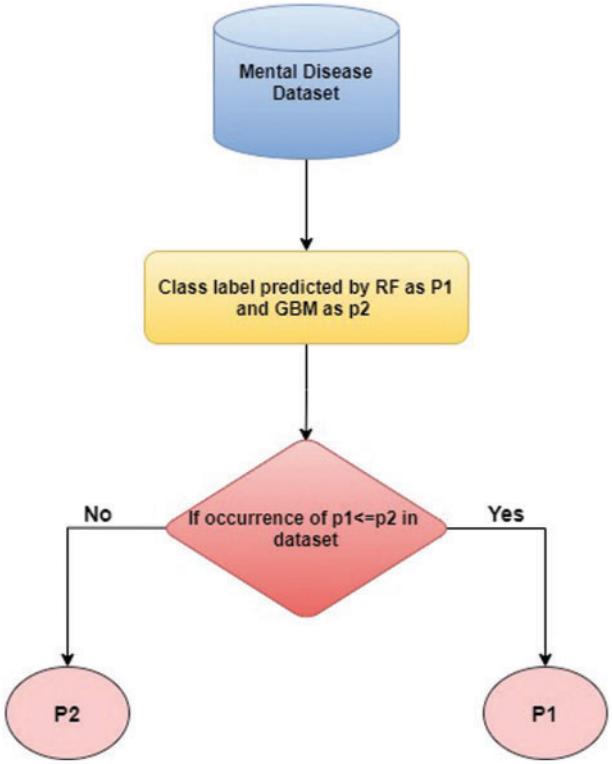

Figure 3: Flowchart of mental disease prediction using SPPM

After completing the task, the models were evaluated on four different evaluation measures namely, Accuracy, Recall, Precision, and F1-score. The mathematical formulation for the evaluation measures are presented in Eqs. (10)–(13), respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$Recall = \frac{TP + TN}{TP + FN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{13}$$

The accuracy scores using different vector space techniques of RF, GBM, LR, SVM and the SPPM-MAX(RF + GBM) model are tabulated in Tab. 3. The highest accuracy for each classifier is shaded in grey color. It is clear that RF with TF-IDF (unigram) is showing highest accuracy with 64%.

The GBM also shows highest accuracy with TF-IDF (unigram) reporting 64%. In addition, the TF-IDF (unigram) shows highest accuracy for LR and SVM with values 62% and 61%, respectively. It is also concluded that, among the classifiers, SPPM-MAX (RF + GBM) shows highest accuracy using TF-IDF with 67%.

Table 3: Accuracy comparison of RF, GBM, LR and SVM with proposed approach using different vector space techniques

Model					
Vector space	RF	GBM	LR	SVM	SPPM-MAX(RF + GBM)
TF-IDF (unigram)	64%	64%	62%	61%	67%
TF/IDF (Bi-gram)	49%	45%	54%	58%	45%
TF/IDF (Tri-gram)	46%	42%	49%	49%	43%

Meanwhile, Tab. 4 presents the overall results of various classifiers using TF-IDF (Unigram) based on various evaluation measures. The highest accuracy for each evaluation measures is shaded in grey color for better visualization.

Table 4: Result of classifiers using TF-IDF (Unigram)

Classifier	Accuracy	Precision	Recall	F1 score	AUC
RF	67%	72%	65%	67%	.88
GBM	67%	69%	65%	66%	.89
LR	62%	65%	61%	62%	.89
SVM	61%	65%	59%	61%	.89
VC (LR + GBM)	65%	67%	63%	64%	.89
VC (LR + SVM)	62%	65%	61%	62%	.89
VC (GBM + SVM)	65%	67%	63%	65%	.90
VC (RF + LR)	63%	65%	61%	62%	.89
VC (RF + SVM)	62%	64%	60%	61%	.90
VC (RF + GBM)	66%	69%	64%	65%	.89

The accuracy is the intuitive evaluation metric, it tells the ratio of correct predictions from total observations. RF and GBM show highest accuracy value of 67% among individual classifiers as shown in Tab. 4. It can be observed that simple voting ensemble of classifiers did not improve the accuracy value for mental disease classification. The proposed SPPM-MIN technique significantly improved the results with 71% accuracy.

The precision is the measure of agreement of class labels among positive labels predicted by classifiers. The values of precision score for each of the classifier used in the experiment are shown in Tabs. 4 and 5. It can be seen in Tab. 5 that the proposed RF + GBM using SPPM-MIN achieve highest precision score of 73%. RF and RF + LR using SPPM-MAX achieve 72% and 70% respectively for

mental disorder classification. GBM, GBM + SVM and RF + SVM using SPPM-MIN achieved 69% precision score.

Table 5: Result of classifiers using TF-IDF (unigram) using SPPM technique

	SPPM-MIN				SPPM-MAX			
Classifier	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
LR + GBM	65%	67%	65%	66%	63%	66%	59%	62%
LR + SVM	61%	63%	60%	61%	62%	67%	59%	63%
GBM + SVM	67%	69%	65%	66%	63%	66%	62%	64%
RF + LR	66%	68%	65%	66%	64%	70%	61%	65%
RF + SVM	66%	69%	66%	67%	62%	68%	58%	63%
RF + GBM	71%	73%	71%	71%	67%	68%	66%	67%

The recall is also called as sensitivity that calculates the effectiveness of model to predict class labels. Sensitivity is of very significant importance that is used in classification purpose. The recall scores of 6 class labels of all models are presented in Tabs. 4 and 5. The proposed approach RF + GBM using SPPM-MIN achieved highest recall value of 71% for mental disorder classification. Similarly, the F1 score is the measure that is calculated by taking harmonic mean of precision and recall. It gives balanced results of classifier and presents their performance. Our proposed model RF + GBM using SPPM-MIN achieved highest F1 score among all other models used in classification and achieved 71% score. RF and RF + GBM using SPPM-MAX achieved 67% F1 score. The F1-Score balances both precision and recall in one value. AUC metric tells how well model can distinguish between classes. AUC values of classifiers are presented in Tab. 4. VC (GBM + SVM) and VC (RF + SVM) achieve highest AUC score with 0.90 value.

Besides, a confusion matrix was used for the classification task. It contains two types of information, predicted class and true/actual class with the following representations: (a) True Positive (TP) are correctly identified class, (b) False Negative (FN) is the incorrect classified class, (c) True Negative (TN) are the correctly classified classes, and (d) False Positive (FP) are the incorrectly classified classes. Tab. 5 illustrates the confusion matrix of proposed approach. Samples classified as true are presented in bold and others as misclassified. The confusion matrix also validates the effectiveness of proposed model, heat map of RF + GBM using SPPM-MIN and SPPM-MAX is presented in Fig. 4.

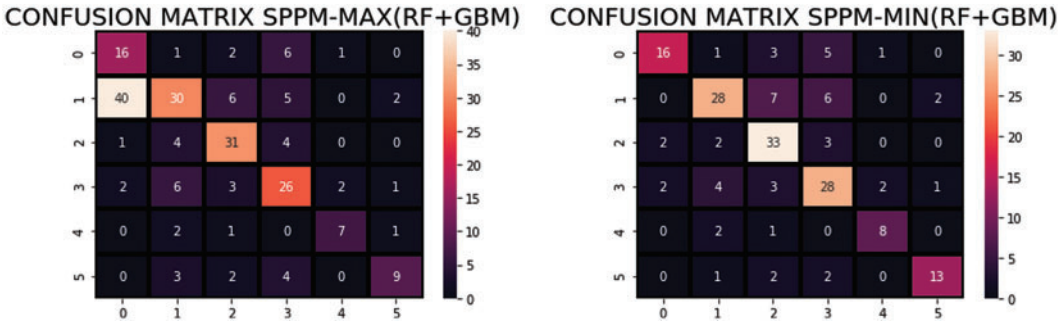

Figure 4: Heat map of the proposed approach using SPPM-MIN and SPPM-MAX

Main objective of the research is to predict the mental disorders in six target groups with high accuracy using RF + GBM supported by SPPM-MIN AND SPPM-MAX. The Performance comparison of all ensemble models using SPPM in terms of Accuracy, Precision, Recall and F1-score is presented in Fig. 5. Among the classifiers RF and GBM as individual classifiers and their combination as voting classifier give reasonable Accuracy, Precision and F1-Score. RF + GBM using SPPM-MIN outperformed in terms of accuracy, recall and F1 score than all other approaches.

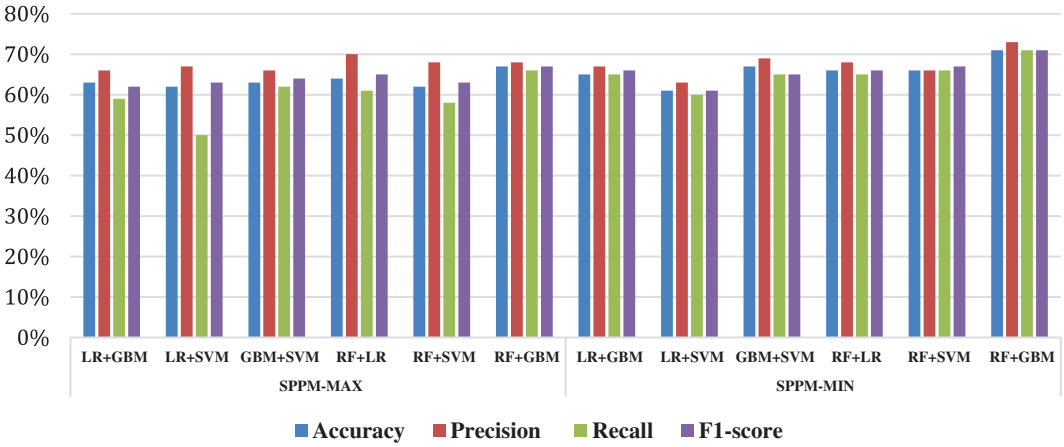

Figure 5: Performance comparison of SPPM-MAX and SPPM-MIN on ensemble models

The SVM does not perform well in multiclass task while the LR shows a fair performance in terms of precision. We also compared the performance of individual base learners and proposed model with difference vector space techniques such as TFIDF (uni-, bi- and tri-gram). As dataset consist of 1771 observations of patient by a medical expert in the form of symptoms, so unigram represent well and show highest result and tri-gram show lowest result.

Regarding the effectiveness of the proposed approach, RF is prominent among the individual classifiers used in the experiment. It surpassed all other methods when applied with TF-IDF (unigram) for mental disease classification. While RF did not show any superiority of performance when applied with TF-IDF (Bi-gram and Tri-gram). Performance comparison using TF-IDF (Unigram, Bi-gram, and Tri-gram) is presented in Tab. 3. This manifests that the appropriate representation of features has significant importance in proving the effectiveness of the classification task.

Finally, results reveal that the tree-based classifiers achieved good classification results using TF-IDF (unigram) in multiclass text classification problems. RF is an ensemble model and works by combining multiple trees and therefore generalize better when applied with appropriate feature extraction techniques. GBM transforms weak learner to strong learner and often perform well in text classification. It can be noticed that when these tree-based models are combined with LR and SVM as voting classifiers, their performance is degraded. Even the ensemble of GBM and RF did not improve the results. To overcome the deficiencies of the classifiers, this study proposes the SPPM technique. However, RF + GBM using SPPM-MAX improved the classification results in terms of Accuracy, Recall, and F1 score. While proposed model RF + GBM using SPPM-MIN outperformed all classifiers and their voting ensembles and improved 4% classification result. Hence, simple ensemble of base learners does not improve performance as there is significant diversity among class labels. Ensemble of bagging (RF) and boosting (GBM) classifiers tends to be more accurate with proposed SPPM-MIN technique as shown in Fig. 5.

6 Conclusion & Future Work

This research provides an intuitive solution for mental disorder analysis among different target class labels or groups. A framework is proposed for determining the mental health problem of patients using observations of medical experts. This framework consists of an ensemble model based on RF and GBM with novel SPPM technique. This proposed decision level fusion approach by combining RF + GBM with SPPM-MIN significantly improves the performance in terms of Accuracy, Precision, Recall and F1-score with 71%, 73%, 71% and 71% respectively and surpassed the other models. This framework seems suitable in case of huge and more diverse multi-class datasets. Furthermore, three vector spaces based on TF-IDF (uni-, bi-gram and tri-gram) are also tested on the machine learning models and proposed model. Experiments revealed that unigram performed better on the experimental dataset. In future, more physiological parameters such as respiratory rate, ECG and EEG signals can be included as features to improve the accuracy. Also, the proposed framework can be tested on a wide range of mental illness categories by adding more mental illness diseases in the dataset which will result in an increase of class labels.

Acknowledgement: This work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2019R1A2C1006159) and (NRF-2021R1A6A1A03039493).

Funding Statement: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2019R1A2C1006159) and MSIT (Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2020-2016-0-00313) supervised by the IITP (Institute for Information & communications Technology Promotion).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] W. H. Organization and others, "Depression and other common mental disorders: Global health estimates," *World Health Organization*, 2017.

[2] W. H. Organization, "The world health report 2001: Mental health: New understanding, new hope," *World Health Organization*, 2001.

[3] C. M. Vicario, M. A. Salehinejad, K. Felmingham, G. Martino and M. A. Nitsche, "A systematic review on the therapeutic effectiveness of non-invasive brain stimulation for the treatment of anxiety disorders," *Neuroscience & Biobehavioral Reviews*, vol. 96, pp. 219–231, 2019.

[4] K. MacDonald, A. Krishnan, E. Cervenka, G. Hu, E. Guadagno *et al.,* "Biomarkers for major depressive and bipolar disorders using metabolomics: A systematic review," *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, vol. 180, no. 2, pp. 122–137, 2019.

[5] J. Ratcliff and C. V. D. Feltz-Cornelis, "Conversion disorder/functional neurological disorder–a narrative review on current research into its pathological mechanism," *The European Journal of Psychiatry*, vol. 34, no. 3, pp. 143–152, 2020.

[6] A. M. Shaw, K. A. A. Hall, E. Rosenfield and K. R. Timpano, "Body dysmorphic disorder symptoms and risk for suicide: The role of depression," *Body Image*, vol. 19, pp. 169–174, 2016.

[7] A. J. Al-Mosawi, "The etiology of mental retardation in Iraqi children," *Autism*, vol. 1, pp. 4–7, 2019.

[8] R. A. McCutcheon, T. R. Marques and O. D. Howes, "Schizophreniaan overview," *JAMA Psychiatry*, vol. 77, no. 2, pp. 201–210, 2020.

[9] C. Burr and N. Cristianini, "Can machines read our minds?" *Minds and Machines*, vol. 29, no. 3, pp. 461–494, 2019.
[10] J. C. Eichstaedt, R. J. Smith, R. M. Merchant, L. H. Ungar, P. Crutchley *et al.,* "Facebook language predicts depression in medical records," *Proceedings of the National Academy of Sciences*, vol. 115, no. 44, pp. 11203–11208, 2018.
[11] A. G. Reece and C. M. Danforth, "Instagram photos reveal predictive markers of depression," *EPJ Data Science*, vol. 6, no. 1, pp. 1–12, 2017.
[12] Y. Fukazawa, T. Ito, T. Okimura, Y. Yamashita, T. Maeda *et al.,* "Predicting anxiety state using smartphone-based passive sensing," *Journal of Biomedical Informatics*, vol. 93, no. 103151, 2019.
[13] D. Leightley, V. Williamson, J. Darby and N. T. Fear, "Identifying probable post-traumatic stress disorder: Applying supervised machine learning to data from a UK military cohort," *Journal of Mental Health*, vol. 28, no. 1, pp. 34–41, 2019.
[14] Y. Nakai, T. Takiguchi, G. Matsui, N. Yamaoka and S. Takada, "Detecting abnormal word utterances in children with autism spectrum disorders: Machine-learning-based voice analysis versus speech therapists," *Perceptual and Motor Skills*, vol. 124, no. 5, pp. 961–973, 2017.
[15] G. Coppersmith, R. Leary, P. Crutchley and A. Fine, "Natural language processing of social media as screening for suicide risk," *Biomedical Informatics Insights*, vol. 10, no. 1178222618792860, pp. 1–11, 2018.
[16] O. Biran and C. Cotton, "Explanation and justification in machine learning: A survey," in *IJCAI-17 Workshop on Explainable AI (XAI)*, Melbourne, Australia, pp. 8–13, 2017.
[17] P. Fusar-Poli, Z. Hijazi, D. Stahl and E. W. Steyerberg, "The science of prognosis in psychiatry: A review," *JAMA Psychiatry*, vol. 75, no. 12, pp. 1289–1297, 2018.
[18] D. A. Beck, H. G. Koenig and J. S. Beck, "Depression. *Clinics in Geriatric Medicine*," vol. 14, no. 4, pp. 765–786, 1998.
[19] C. D. Spielberger, Anxiety and Behavior, Cambridge, Massachusetts, United States, Academic Press, 2013.
[20] O. Bagasra and C. Heggen, in *Autism and Environmental Factors*, Hoboken, New Jersey, U.S., Wiley Online Library, 2018.
[21] D. A. Regier, E. A. Kuhl and D. J. Kupfer, "The DSM-5: Classification and criteria changes," *World Psychiatry*, vol. 12, no. 2, pp. 92–98, 2013.
[22] R. Y. Masri and H. M. Jani, "Employing artificial intelligence techniques in mental health diagnostic expert system," in *2012 Int. Conf. on Computer & Information Science (ICCIS) IEEE*, Kuala Lumpur, Malaysia, pp. 495–499, 2012.
[23] D. D. Luxton, "Artificial intelligence in psychological practice: Current and future applications and implications," *Professional Psychology: Research and Practice*, vol. 45, no. 5, pp. 332, 2014.
[24] M. Rostami, S. Farashi, R. Khosrowabadi and H. Pouretemad, "Discrimination of ADHD subtypes using decision tree on behavioral, neuropsychological, and neural markers," *Basic and Clinical Neuroscience*, vol. 11, no. 3, pp. 359, 2020.
[25] S. Chattopadhyay, P. Kaur, F. abhi and R. Acharya, "An automated system to diagnose the severity of adult depression," in *2011 Second Int. Conf. on Emerging Applications of Information Technology IEEE*, India, pp. 121–124, 2011.
[26] L. N. Comin, P. R. Pinheiro, T. P. Cavalcante and M. C. D. Pinheiro, "Handling diagnosis of schizophrenia by a hybrid method," in *Computational and Mathematical Methods in Medicine 2015*, vol. 2015, Article ID 987298, pp. 13, 2015.
[27] R. M. Rahman and F. Afroz, "Comparison of various classification techniques using different data mining tools for diabetes diagnosis," *Journal of Software Engineering and Applications*, vol. 6, no. 3, pp. 85–97, 2013.
[28] S. Lahmiri, D. A. Dawson and A. Shmuel, "Performance of machine learning methods in diagnosing parkinsons disease based on dysphonia measures," *Biomedical Engineering Letters*, vol. 8, no. 1, pp. 29–39, 2018.

[29] K. Kipli, A. Z. Kouzani and I. R. A. Hamid, "Investigating machine learning techniques for detection of depression using structural MRI volumetric features," *International Journal of Bioscience, Biochemistry and Bioinformatics*, vol. 3, no. 5, pp. 444–448, 2013.
[30] F. L. Seixas, B. Zadrozny, J. Laks, A. Conci and D. C. M. Saade, "A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment," *Computers in Biology and Medicine*, vol. 51, pp. 140–158, 2014.
[31] M. Pirooznia, J. J. F. Seifuddin, P. B. Mahon, J. B. Potash and P. P. Zandi *et al.,* "Data mining approaches for genome-wide association of mood disorders," *Psychiatric Genetics*, vol. 22, no. 2, pp. 55, 2012.
[32] B. Yalamanchili, N. S. Kota, M. S. Abbaraju, V. S. S. Nadella and S. V. Alluri, "Real-time acoustic based depression detection using machine learning techniques," in *2020 Int. Conf. on Emerging Trends in Information Technology and Engineering (ic-ETITE) IEEE*, Vellore, India, pp. 1–6, 2020.
[33] A. Priya, S. Garg and N. P. Tigga, "Predicting anxiety, depression and stress in modern life using machine learning algorithms," *Procedia Computer Science*, vol. 167, pp. 1258–1267, 2020.
[34] S. Sakr, R. Elshawi, A. Ahmed, W. T. Qureshi, C. Brawner *et al.,* "Using machine learning on cardiorespiratory fitness data for predicting hypertension the henry ford ExercIse testing (FIT) project," *PLoS One*, vol. 13, no. 4, pp. e0195344, 2018.
[35] S. Rayana, W. Zhong and L. Akoglu, "Sequential ensemble learning for outlier detection: A bias-variance perspective," in *2016 IEEE 16th Int. Conf. on Data Mining (ICDM) IEEE*, Barcelona, Spain, pp. 1167–1172, 2016.
[36] S. Bashir, U. Qamar and F. H. Khan. "IntelliHealth: A medical decision support application using a novel weighted multi-layer classifier ensemble framework," *Journal of Biomedical Informatics*, vol. 59, pp. 185–200, 2016.
[37] A. Ozcift and A. Gulten, "Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms," *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3, pp. 443–451, 2011.
[38] J. Sun, C. D. McNaughton, P. Zhang, A. Perer A. Gkoulalas-Divanis *et al.,* "Predicting changes in hypertension control using electronic health records from a chronic disease management program," *Journal of the American Medical Informatics Association*, vol. 21, no. 2, pp. 337–344, 2014.
[39] M. Augsburger and T. Elbert, "When do traumatic experiences alter risk-taking behavior? A machine learning analysis of reports from refugees," *PLoS One*, vol. 12, no. 5, pp. e0177617, 2017.
[40] A. M. Chekroud, R. J. Zotti, Z. Shehzad, R. Gueorguieva and M. K. Johnson, "Cross-trial prediction of treatment outcome in depression: A machine learning approach," *The Lancet Psychiatry*, vol. 3, no. 3, pp. 243–250, 2016.
[41] D. Gökçay, A. Eken and S. Baltacı, "Binary classification using neural and clinical features: An application in fibromyalgia with likelihood-based decision level fusion," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1490–1498, 2018.
[42] M. K. Abd Ghani, M. A. Mohammed, N. Arunkumar, S. A. Mostafa, D. A. Ibrahim *et al.,* "Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques," *Neural Computing and Applications*, vol. 32, no. 3, pp. 625–638, 2020.
[43] A. De and A. S. Chowdhury, "DTI based Alzheimer's disease classification with rank modulated fusion of CNNs and random forest," *Expert Systems with Applications*, vol. 169, pp. 114338, 2021.
[44] P. Norvig, *How to Write a Spelling Corrector*, 2007. [Online]. Available: http://norvigcom/spell-correcthtml.
[45] M. F. Porter, *Snowball: A Language for Stemming Algorithms*, 2001.
[46] W. Zhang, T. Yoshida and X. Tang, "A comparative study of TF∗ IDF, LSI and multi-words for text classification," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2758–2765, 2011.
[47] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
[48] R. E. Schapire, "A brief introduction to boosting," in *Proc. Ijcai*, USA, pp. 1401–1406, 1999.
[49] V. Y. Kulkarni and P. K. Sinha, "Pruning of random forest classifiers: A survey and future directions," in *Proc. ICDSE*, Cochin, India, IEEE, pp. 64–68, 2012.

[50] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[51] S. B. Ariffin, H. Midi, J. Arasan and M. S. Rana, "The effect of high leverage points on the maximum estimated likelihood for separation in logistic regression," in *AIP Conference Proceedings American Institute of Physics*, vol. 1643, pp. 402–408, 2015.

[52] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[53] H. V. Halteren, "Teaching NLP/CL through games: The case of parsing," in *Proc. of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics [Internet] Philadelphia*, Pennsylvania, USA, Association for Computational Linguistics, pp. 1–9, 2002.