Article   Open Access

Proposing Two Hybrid Data Mining Models for Discovering Students' Mental Health Problems

Shabnam Shadroo [1] (ID)**, Mohsen Yoosefi Nejad** [2] (ID)**, Samira Tavanaiee Yosefian** [3] (ID)**, Morteza Naserbakht** [4] (ID)**, Mehdi Hosseinzadeh** [5] (ID)

[1] Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran
[2] Department of Computer Engineering and Information Technology, Payame Noor University, Tehran, Iran
[3] Department of Counseling, School of Psychology & Training Sciences, Allameh Tabatabai University, Tehran, Iran
[4] Mental Health Research Center, Psychosocial Health Research Institute, Iran University of Medical Sciences, Tehran, Iran
[5] Pattern Recognition and Machine Learning Lab, Gachon University, 1342 Seongnamdaero, Sujeonggu, Seongnam 13120, Republic of Korea

Corresponding author: Mehdi Hosseinzadeh (mehdi@gachon.ac.kr)

Abstract

Mental health is an important issue for university students. The objective of this article was to apply and compare the different classification methods for students' mental health problems. Furthermore, it presents an ensemble classification method to improve the accuracy of classifiers and assist psychologists in the decision making process. For this, 10 different classifiers were applied for classifying students into two groups. In addition, two methods of combining the classifiers are presented. In the first proposed method, the classifiers were selected based on their accuracy, and then voting was carried out based on maximum probability. In the second proposed method, the methods were combined based on the fields of the confusion table, and the voting was carried out based on majority voting scheme. These two methods were evaluated in two ways. Focusing on the accuracy and the maximum probability voting, the accuracy of the first method was 92.24%, whereas in the second method, it was 95.97%. Further, using confusion table and majority voting applied to the entire dataset, the accuracy reached 96.66%. The results are promising to assist the process of mental health assessment of students.

Keywords

Classification, NBTree, ADTree, Random Forest, SVM, Mental health, Majority voting, Data mining.

1   Introduction

Mental hygiene is part of the health evaluation of various societies. Mental health is of great significance as it affects all aspects of daily life. In particular, the mental well-being of undergraduate students is imperative, as it greatly affects their academic success (Deziel et al., 2013). The enrollment to a university is often accompanied by many changes in social and human relations. In such a situation that often abounds with stress and worry, the operation and efficiency of people will be affected. Among the causes of the existing concerns, mental discomforts, and efficiency decline, are the unfamiliarity of so many students with university campus, being away from their families, not being enthusiastic about their major, incompatibility with other people in their dormitories, and the inadequacy of welfare and economic facilities. Therefore, it is impossible to avoid stressful factors and increasing changes. In such a situation, a person who has prepared their physical and mental health is ready to encounter these stressful situations in life. Thus, recognizing the factors conducive to mental health (particularly in students) is of special importance. For this reason, the program of mental health evaluation of new students is carried out nationally by all university counseling centers in Iran every year. This article looks into the mental health of students from two perspectives, health and disease. This research also tried to obtain more information about the students' mental well-being in terms of family atmosphere, efficacy, and social support.

This research investigates the mental health of students for three academic years (2013 to 2015) in one Iranian university. The questionnaire was published by the Ministry of Science, Research, and Technology of the Islamic Republic of Iran. The questionnaire assesses positive emotions, depression, anxiety, obsession, social anxiety, sleep disorders, academic depression, educational anxiety, family relations, perfectionism, and suicidal tendency, using 102 questions. The responses to the questions along with the mentioned aspects are studied in this research. First, the data was manually classified into two categories by a group of psychologists according to the following criteria: students who require a consultation (CR) and students who do not require a consultation (NCR).

After extracting features related to health and disease, the data mining methods were compared. The aim of our study is to compare the data mining algorithms for analyzing mental health. In addition, the article shows that it is possible to achieve good results by combining classifiers that do not have very good results in different aspects such as accuracy, false positive rate (FP rate) and true negative rate (TN rate). Another purpose of this article is to encourage psychologists to use the data mining algorithms and combine them to improve the outcome.

The proposed method in this paper combines classifiers using the voting method. Two methods of combining are used. The first method combines classifiers with the least accuracy criteria, and the second combines classifiers based on the least value of specific fields in the confusion table. Majority and maximum probability voting are used in this proposed method.

Using data mining techniques and learning through classification models, intelligent decision support systems can assist and facilitate the process of decision making in complex, human-centric environment. This scope suffers from incomplete and imprecise information provided by humans (students in this study) and the uncertain knowledge of experts (in the science of psychology) in the process of the assessment of students' mental health. On the other hand, variety in the types of symptoms and disorders and a huge number of samples make it difficult to extract the hidden knowledge and general patterns of mental health, which can be assisted using data mining models. The goal of this paper is to propose hybrid data mining models to discover students at risk of mental health problems.

The remainder of this paper is organized as follows: section 2 presents related works in the area of mental health data mining algorithms. In section 3, the preprocessing and the proposed method are presented. The experimental evaluation and results are shown in section 4, section 5 presents our discussion, and conclusion is presented in section 6.

1.1   Literature review

There is a considerable amount of literature on the mental health of students in different countries. Some studies have evaluated the learning method and learning effective factors for students. Deziel et al. (2013) studied the mental health of undergraduate engineering students in one of Canadian universities. Their research investigates effective factors on the mental health of students by using an online survey questionnaire. This questionnaire was online for 7 days. Five aspects of students' mental health were considered in their research, which are the most effective factors in the academic development process. The sums of the five factors in the questionnaire were modeled by Regression method as general student features. The prism algorithm predicts the rules in various positions. The results of this study suggested a number of recommendations to help improve the mental health of undergraduate engineering students. In this paper, a wider variety of mental-health-related features are considered, including positive emotions, depression, anxiety, obsession, social anxiety, sleep disorders, academic depression, educational anxiety, family relations, perfectionism, and suicidal tendency (Deziel et al., 2013).

Diederich et al. (2007) used machine learning techniques to classify texts related to mental health problems on the basis of speech. In their research, inputs were the used words and the number of repetitions in the transcribed texts of the speech samples of different people, and the output was psychiatric categories. In order to classify the data binaurally, Support vector machine and decision tree classifiers were applied. They concluded that the results of the categories depend on data sizes and that when there is a considerable amount of data, the SVM operates well (Diederich et al., 2007).

The differences in healthcare coverage are an important issue in the United States. Few studies have been published on the factors of healthcare coverage. Unfortunately, in the United States, many people are not covered by healthcare and treatment policy, and several researches have been carried out in order to identify the factors. Delen et al. (2009) studied healthcare coverage using machine learning techniques in a wide range of predicting factors. The data for this investigation was obtained from the survey system Data Set of behavioral risk factors in 2004, whose principal section contains 84 questions associated with individual health. These include: environmental factors, smoking or not, quality of life in terms of health, etc. In their study, the MLP artificial neural networks and CART decision tree model were used for predicting coverage or non-coverage. The neural network was more accurate in general classification, whereas the decision tree worked well in classifying people who were not covered.

Superby et al. (2006) carried out a study to investigate the causes of dropout in first-year students in three universities in Belgium, and divided the students into three groups during the educational year. The "low risk" students who are likely to be successful, the students with "medium risk" that may be successful with the facilities a university provides, the third category named "at risk" who are likely to fail or drop out. This study applied different methods including the decision tree, neural network, Random Forest, and linear discriminant. The final comparison of the results indicates that these methods can be successful for predicting the status of the student.

Soet and Sevig (2006) investigated the effect of ethnicity and sexual orientation on various mental health problems such as depression, eating disorders, substance use, etc. One conclusion from their study was that African American students were found to be less distressed than others. Tomar and Agarwal (2013) compared different classifiers related to health, mentioning their advantages and disadvantages. Another research studied students' efficiency using a new feature called behavioral features. The method uses some data analysis methods for evaluating the students' efficiency by applying behavioral characteristics. The collected features are in three categories (demographic features, educational background, behavioral features). The article applied current data analysis methods such as neural network, decision tree, and Bayesian network. Finally, their results were combined in terms of boosting and bagging (Amrieh et al., 2016).

Ahmadi et al. (2018) demonstrated the contribution of fuzzy logic methods in the diagnosis of diseases. For this purpose, eight databases were selected, limited to the works found from January 2005 to June 2017, identifying 46 articles that met the inclusion criteria (Ahmadi et al., 2018).

Burke et al. (2019) conducted a systematic literature review on the application of machine learning techniques to predict thoughts and behaviors regarding self-inflicted wounds (suicidal and non-suicidal) from five databases up to February 2018 (Burke et al., 2019). Dwivedi et al. (2018) critically analyzed all existing approaches to automatic identification and classification of heart sounds based on 117 peer-reviewed articles found for the period ranging from 1963 to 2018 (Dwivedi et al., 2018). Laijawala et al. (2020) predicted mental health problems using several classification algorithms such as Decision Tree, Random Forest and Naïve Bayes. The target population was in the working class , i.e. people above the age of 18. After building the model, they integrated it in a website to predict the outcome as per the details provided by the user.

Vanlalawmpuia and Lalhmingliana (2020) highlighted and revealed users' mental health status and condition by analyzing social network sites. Analyzing data involves training each user's data to get an output and also have a test set to get efficiency and accuracy. Their study identified many depressive-indicative words, which play an important role in bringing this study success. They also proposed several methods that were applied in their study. Efficiency analysis was performed using a proposed method which contains clusters of emotional words and can also increase the accuracy, efficiency and reduce analysis time (Vanlalawmpuia & Lalhmingliana, 2020). Table 1 describes the summary of data mining approaches that are used in the mental health domain.

Table 1. *Summary of data mining approaches used in the mental health domain.*

Author	Approaches	Evaluation Criteria	Accuracy
(Deziel et al., 2013)	Least-Squares Linear Regression	10-fold cross-validation Accuracy	75 – 85 %
	PRISM Algorithm Prediction Rules		
(Delen et al., 2009)	MLP	Accuracy	78.45%
	Decision Tree		74.11%
(Superby et al., 2006)	Random Forest	Accuracy	51.78%
	Linear Discriminant		57.35%
	Decision Tree		40.63%
	Neural Network		51.88%
(Diederich et al., 2007)	SVM	Accuracy	77%
	Decision Tree	ROC	78%
(Amrieh et al., 2016)	Neural Network	Accuracy	79.1%
	Decision Tree	Recall	75.8%
	Naïve Bayesian	Precision	67.7%
	Random Forest	F-measure	75.6%

There are several methods for data mining, some of which are described in this article. Some researchers, e.g. (Hosseini et al., 2011; Wei et al., 2016; Wei et al., 2017), apply soft computing models such as fuzzy and genetic algorithm for classification. These models have the capability to classify patterns even into

classes with un-sharp boundaries. In summary, various researches have tried to classify mental health data using data mining techniques. However, few researches shed light on the topic of combining classifiers to improve accuracy. This research focuses on comparing classifiers to improve the accuracy using the area under the ROC curve, and combining the classifiers based on voting approaches.

2   Research methods

This section describes the proposed method for predicting first-year students' mental health at university. The data for this research was collected from the engineering college in Mashhad, Iran. The first-year students filled the questionnaire. The questionnaire had four sections. The first section contained personal data such as gender, marital status, and field of study. Other sections came from three different questionnaires, which were labeled "A", "B", and "C". Questionnaire "A" estimated positive emotions and health; questionnaire "B" considered prospective disease including depression, anxiety, obsession, social anxiety, and sleep disturbances; and questionnaire "C" assessed health and mental structures including educational depression, educational anxiety, family environment, perfectionism, and suicidal tendencies. The number of students who filled this questionnaire was 3679 during the years 2013 to 2015.

After collecting data, some preprocessing mechanisms were used to improve the quality of the dataset. Data preprocessing is considered an important step in the data mining process and includes data cleaning, feature selection, data reduction, and data transformation.

Data Cleaning

The first step was data cleaning. K means clustering method was used for predicting clusters and outliers were removed from the dataset. After removing outliers, the miss value was investigated. In the data set used in this paper, about 55 people had a field of no value in their response. Since this value was about 1.5% of the data, we removed the answers of those who had empty fields from the dataset. The actual data number was 3619 after the preprocessing step.

Feature Selection

Feature extraction is a fundamental step in the data preprocessing phase. The objective of the feature selection process is to select an appropriate subset of features which can efficiently describe the input data, reduce the dimensionality of feature space, and remove redundant and irrelevant data. Features used in this article are divided into two parts. Table 2 summarizes the two parts of features: part one includes personal features, and part two includes features extracted from the data. These features contain positive emotions, depression, anxiety, obsession, social anxiety, sleep disorders, academic depression, educational anxiety, family relations, perfectionism, and suicidal tendencies. The criterion for computing extracted features (values in the range 6-30) is the sum of two kinds of response; the first one is the sum of questions that indicated positive mental health such as positive emotions, family atmosphere, perfectionism, religious orientation, social support and efficacy. The second is the sum of negative mental health indicators. Each positive/negative feature value is computed as follows. Six questions were asked in the questionnaire, the answer was a number ranging from 1 to 5, (1 for totally disagree and 5 for totally agree), the value of extracted feature is the sum of numeric answers to the six questions. For instance, questions 1, 3, 5, 7, 9, 11 from questionnaire A are related to positive emotions, and the answers are summed up to a number regarded as a feature.

The questionnaire was completed by students in the first semester of university. These students were usually entered at the college by exam at the age of or above 18. Some of these students had taken the entrance examination many times to be accepted at the university. Therefore, they may not have been in a good mental condition or had not been accepted in the preferred field of study. Thus, the best time to investigate the mental conditions of students is the first semester, so that they can improve their mental and psychological conditions by participating in the necessary consulting.

Data Reduction and Transformation

The questionnaire was filled in the first term of university in Iran from 2013 to 2015, and the data were collected. Then the data were divided into two categories of CR and NCR by a group of active psychologists at the university, according to the criteria set out in the previous paragraph. Out of 3,619 students, 2,040 students who needed consulting were placed in the CR category, and 1,579 students who did not need consulting were placed in the NCR category.

***Table 2.** Summary of features extracted from input data.*

Attribute		Passible Value
Personal Features	Gender	Male, Female
	Marital Status	Single, Married
	Field Of Study	Electrical Engineering, Computer Engineering, Civil Engineering, Industrial Engineering, Materials Engineering, Medical Engineering, Robotics Engineering
Negative Features	Suicidal tendencies	6-30
	Depression	6-30
	Anxiety	6-30
	Obsession	6-30
	Social anxiety	6-30
	Sleep disorder	6-30
	Academic depression	6-30
	Educational anxiety	6-30
Positive Features	Family atmosphere	6-30
	Perfectionism	6-30
	Positive emotions	6-30

In order to detect the attributes that have the greatest influence on the accuracy of prediction, we applied the classification algorithms on three kinds of attributes.

These three kinds of attributes are as follows:

1. Responses to the questionnaire and personal features used in classification.
2. Features of mental health which were derived from the questionnaires and used in classification.
3. All of the features mentioned in 1 and 2 used in classification.

The following classifiers were evaluated in this article. These classifiers include:

1. **Bayes Net:** Bayes network was learned by the use of different search algorithms and quality actions. The main class of a Bayes Net classifier provides the data structure (network structure, conditional probability distributions, etc.) and the common capabilities for learning algorithms of Bayes Network (George-Nektarios, 2013).

2. **Logistic:** A class to build and use a multinomial logistic regression model with a ridge estimator (George-Nektarios, 2013).

3. **RBF Network:** A class which implements a normalized Gaussian radial basis function network. It applies the algorithm of k-means clustering to prepare the basic functions and learns either logistic regression (discrete class issues) or linear regression (numeric class problems) on top of that. Symmetric multivariate Gaussians are proper to data from each cluster. The class uses the number that is given of clusters per class if it is nominal. This class standardizes all numerical features to the mean zero and unit variance (George-Nektarios, 2013).

4. **NBTree:** A class to produce the decision tree with classifiers of naive Bayes in the leaf (George-Nektarios, 2013).

5. **RandomForest:** A class to make a random tree forest (George-Nektarios, 2013). In this article, we used Random Forest in two statuses: 500 and 800 trees.

6. **ClassificationViaRegression:** Class for classifying by use of regression methods. The class is binary and a regression model is constructed for each class value.

7. **ADTree:** A class to produce alternating decision trees (George-Nektarios, 2013)

8. **Support vector machines (SVM):** SVM is a classifier that originated from the theory of statistical learning first introduced by Boser. The main benefits of SVM include: (1) the ability to work with high-dimensional data, and (2) high-performance of popularization without the need to add background knowledge, even when the input space dimensions are very high (Fakhlai et al., 2011). In this article, SVM is used with two kernels – polynominal and RBF.

Considering the three kinds of attributes, we could conclude that using the extracted features alone could reduce the accuracy of the system, but if it was used along with the responses to the questionnaires, it would result in an increase in the accuracy of the classifiers.
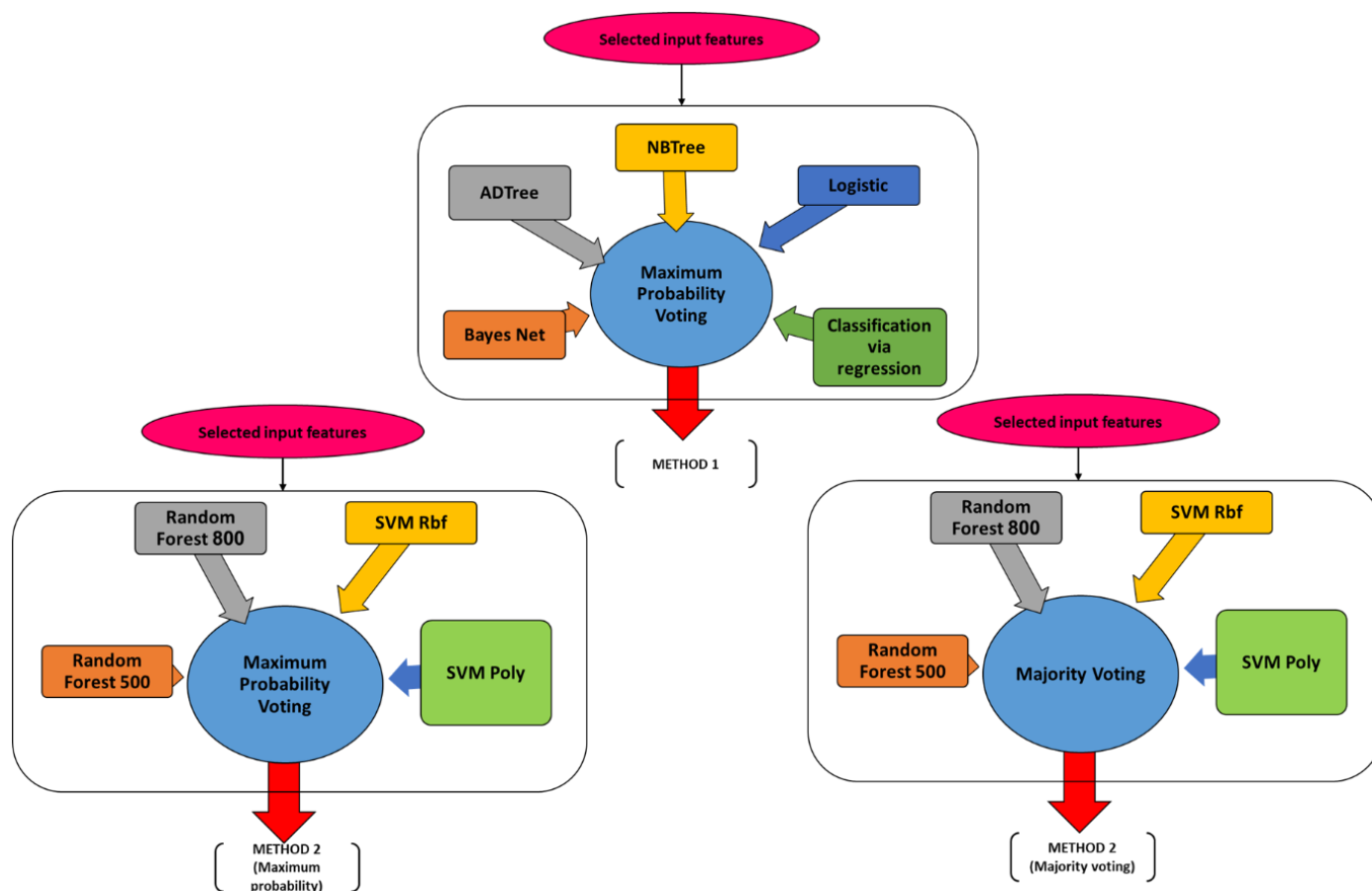

Figure 1. *The outline of proposed methods.*

The proposed method in this paper is applicable when the answers to the questionnaire and the extracted features are used together with personal features, and its goal is to decrease wrong recognition. This paper tries to combine classifiers using the voting method. The outline of the proposed scheme is shown in Figure 1.

As can be seen in Figure 1, two combining methods were used. The first one combined the classifiers with the criteria of the least accuracy (method 1), and the second combined classifiers based on the lowest number of FN and FP fields in the confusion table. The majority and maximum probability voting were used in this proposed method (method 2).

3   Solutions and Results

This section describes the results of the three stages mentioned in previous sections. The WEKA data mining toolkit (George-Nektarios, 2013) was used to analyze the study.

3.1   Results of classifiers for responses to the questionnaire and personal features

The results of different classifiers when the inputs are the responses to the questionnaire and the personal features are shown in Table 3. We used Bayes Net, Logistic, RBF network, SVM with RBF kernel, SVM with polynomial kernel, Random Forest with 500 trees, Random Forest with 800 trees, classification via regression, decision tree, and Naïve Bayes tree for classifications.

K-fold cross validation method was used to have a steady and impartial view of classifier performance. This method divides the data set randomly into groups of k samples. In each of the run algorithms, K-1 folds are used for training and one-fold for testing. Hence, all divisions would have an opportunity for the training and testing process. The process of cross validation is repeated k times, and each of the divisions of k is used exactly once for validity. Then, the k results obtained from the divisions can be averaged to generate the final estimate. 10-fold cross validation is used in each of the classifiers.

Table 3. Results of ten classifiers for responses to the questionnaire and personal features.

	NBTree	ADTree	Classification Via Regression	Forest 800	Forest 500	SVM (RBF)	SVM (poly)	RBF	Logistic	Bayes Net
Correctly Classified Instances	90.91	87.98	88.39	93.17	93.17	**95.27**	95.10	89.86	90.90	88.75
Incorrectly Classified Instances	9.09	12.02	11.61	6.83	6.83	**4.72**	4.90	10.14	9.1	11.26
Specificity	0.909	0.88	0.884	0.933	0.933	**0.953**	0.951	0.899	0.909	0.887
Sensitivity	0.909	0.88	0.884	0.932	0.932	**0.953**	0.95	0.899	0.909	0.888
Kappa statistic	0.8146	0.7551	0.7637	0.8601	0.86	**0.9038**	0.9005	0.7926	0.8153	0.7712

As can be seen in Table 3, SVM showed the acceptable performance compared to other classifiers, and that AD Tree had a lower performance. In order to have a further comparison of the results, we compared

the area under the ROC curve. Figure 2 shows significant differences in the area under the curve (AUC). The Random Forest was the highest value among the classifiers.
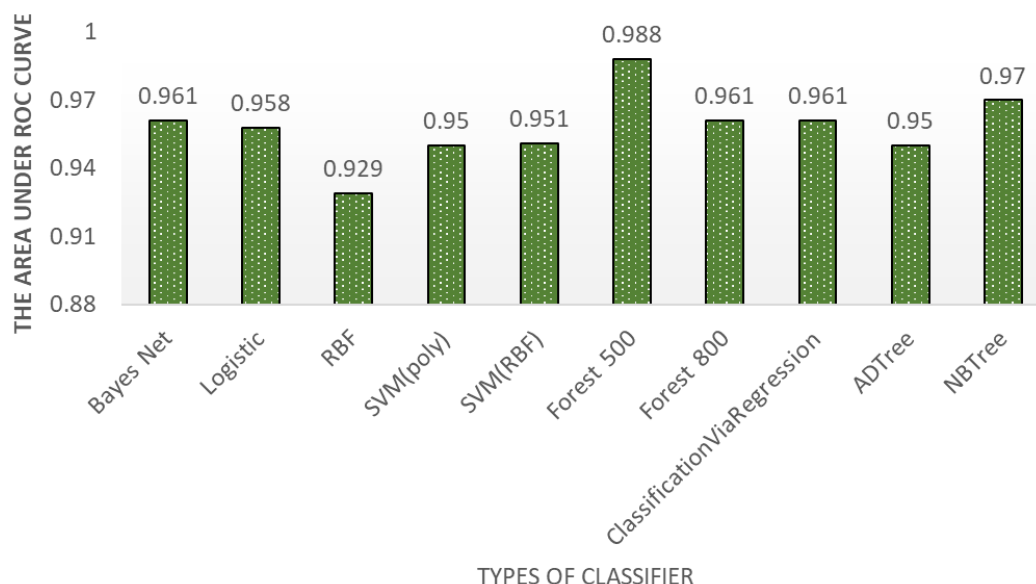

Figure 2. Comparing the area under the ROC curve of classification models.

Table 4 shows the confusion matrix of classifiers, where "P" is the first class and "N" is the second one. The confusion matrix is a summary of the predicted outcomes in a classification issue. The number of predictions is determined by counting the correct and the incorrect predictions, and is summarized in a table. We use this matrix for a further analysis of classifiers. A comparison of the confusion matrix of each classifier indicates that with 500 and 800 trees, Random Forest had the minimum amount of error, and AD Tree had the maximum error in identifying class "P" as class "N" (FN criteria). In identifying class "N" as class "P" (FP criteria), SVM classification with Poly kernel had the least error, and the maximum error was for AD Tree classifier.

Table 4. Comparing confusion tables for responses to the questionnaire and personal features.

	ADTree	NBTree	Classification Via Regression	Forest 800	Forest 500	SVM (RBF)	SVM (poly)	RBF	Logistic	Bayes Net
TP	1839	1896	1840	1976	1979	1964	1955	1897	1871	1843
FP	235	185	220	183	186	95	92	224	160	210
FN	201	144	200	64	61	76	85	143	169	197
TN	1345	1393	1359	1396	1393	1484	1487	1355	1419	1369

3.2 Results of classifiers for features of mental health

This section considers the features extracted by psychologists and personal features. The accuracy percentage and error of applied systems are given in Table 5. In this section, after extracting features, we normalized them with t-score normalization. Z-score normalization is needed for the t-score normalizing. Equation 1 shows the t-score normalization.

$$t - score = z - score * 10 + 50 \qquad (1)$$

The features of age, marital status, field of education and gender were not normalized. Table 5 shows the results of different classifiers.

From Table 5, it can be concluded that different classifiers did not have more than 90% accuracy, which demonstrates that the accuracy decreases when employ only these features (i.e. features extracted by psychologists and personal features).

Table 5. Results of ten classifiers for features of mental health.

	NBTree	ADTree	Classification Via Regression	Forest 800	Forest 500	SVM (RBF)	SVM (poly)	RBF	Logistic	Bayes Net
Correctly Classified Instances	84.75	85	83.78	83.14	83.33	84.80	85.10	84.19	85.22	82.86
Incorrectly Classified Instances	15.25	15	16.22	16.86	16.67	15.20	14.90	15.81	14.78	17.14
Specificity	0.852	0.854	0.846	0.833	0.835	0.856	0.859	0.842	0.856	0.834
Sensitivity	0.844	0.85	0.838	0.831	0.833	0.845	0.851	0.842	0.852	0.829
Kappa statistic	0.6841	0.6896	0.6622	0.6525	0.6565	0.6838	0.6904	0.6755	0.6943	0.6443

Table 6 shows the confusion matrix of classifiers. A comparison of the confusion matrix for each classifier shows that SVM with polynomial kernel had the least error and RBF had the maximum error in FN criteria. FP criterion is approximately above 350 samples in the table.

Table 6. Comparing confusion tables for features of mental health.

	ADTree	NBTree	Classification Via Regression	Forest 800	Forest 500	SVM (RBF)	SVM (poly)	RBF	Logistic	Bayes Net
TP	1890	1893	1914	1833	1836	1924	1925	1822	1891	1875
FP	393	405	461	403	399	434	424	354	386	455
FN	150	147	126	207	204	116	115	218	149	165
TN	1186	1174	1118	1176	1180	1145	1155	1225	1193	1124

3.3 Results of classifiers for responses to the questionnaire, personal features, and features of mental health

In this part, all of the features described in section A and B were used for classification. Table 7 shows the results of different classifiers. SVM with the RBF kernel has the highest accuracy, and ADTree has the lowest accuracy.

Table 7. Results of ten classifiers for responses to the questionnaire, personal features, and features of mental health.

	ADTree	NBTree	Classification Via Regression	Forest 800	Forest 500	SVM (RBF)	SVM (poly)	RBF	Logistic	Bayes Net
Correctly Classified Instances	**88.50**	89.96	88.73	93.46	93.40	**95.44**	95.20	90.20	89.58	88.86
Incorrectly Classified Instances	**11.50**	10.04	11.27	6.54	6.60	**4.56**	4.80	9.80	10.42	11.14
Specificity	**0.885**	0.9	0.887	0.936	0.936	**0.954**	0.952	0.902	0.896	0.889
Sensitivity	**0.885**	0.9	0.887	0.935	0.934	**0.954**	0.952	0.902	0.896	0.889
Kappa statistic	**0.7655**	0.7951	0.7705	0.8656	0.8645	**0.9071**	0.9022	0.7994	0.788	0.7728

Table 8 shows the confusion matrix of classifiers. A comparison of the confusion matrix of every classifier shows that, in FN criteria, Random Forest tree with 500 and 800 trees had the smallest amount of error and classification via regression had the largest amount of error. The classifier SVM (RBF) and SVM (poly) had the least amount of error in FP criteria, where ADTree had the highest. Considering the sum of FP and TN, SVM method with RBF kernel had the least error.

Table 8. Comparing confusion tables for responses to the questionnaire, personal features and features of mental health.

	ADTree	NBTree	Classification Via Regression	Forest 800	Forest 500	SVM (RBF)	SVM (poly)	RBF	Logistic	Bayes Net
TP	1857	1891	1845	1986	1984	1971	1955	1902	1859	1863
FP	**233**	214	213	183	183	**96**	**89**	217	196	226
FN	183	149	**195**	**54**	-	69	85	138	181	177
TN	1346	1365	1366	1396	1396	1483	1490	1362	1383	1353

Figure 3 shows the ROC curve for classifiers. This shows all the agreement between the true (TP) and the false positives (FP) values of the classifier that is usually used to estimate the classification cost in an analysis of the ROC curve, the boundary of classifiers is specified by the threshold value. Each threshold generates a pair (FP rate, TP rate) that demonstrates a point in the ROC curve. The values of TP and FP are described as FP rate = FP/ N, and TP rate = TP/ P, where FP is the number of negative instances that are classified positively by mistake, TP is the number of positive instances that classified correctly, and N and P are the numbers of positive and negative instances in the test dataset, respectively (Latifi et al.,

2015). Random forest tree algorithms have a bigger area under the curve and Logistic classifier has a smaller area.
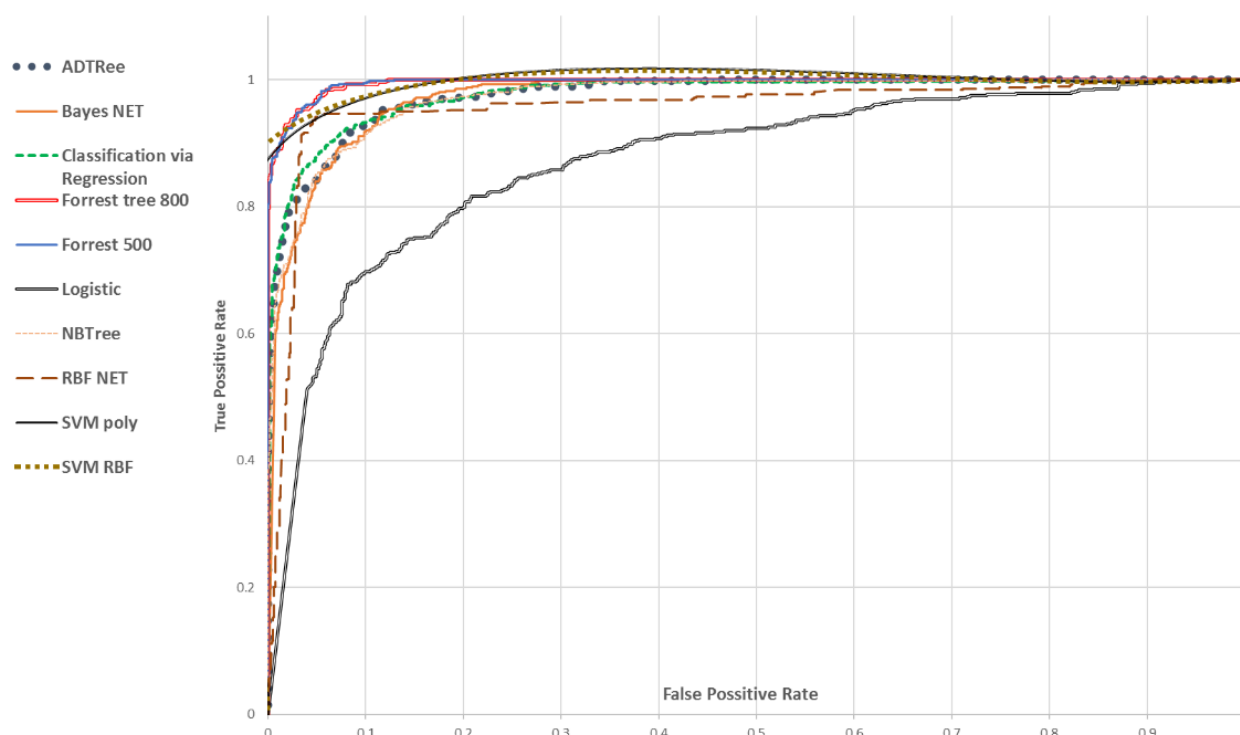

Figure 3. *Comparing the ROC curve of different classifiers.*

3.4   Results of the proposed method

The proposed method was implemented using two types of data sets. First, the construction and evaluation of the model were implemented using the proposed method and 10-fold cross validation was applied to the whole dataset from 2013 to 2015.

In the first proposed method, using voting, the classifiers with lower accuracy were combined together. The output of BayesNet, Logistic, ADTree, NBTree, which according to Table 7 had the lowest accuracy, was combined using voting with the maximum probability. Using this method, the sum of FP and FN could be lowered for each of the classifiers. Accuracy, confusion matrix, and the area under the ROC curve for this combining method are shown in Table 9.

Table 9. *Results of the first proposed method for 2013 to 2015 data.*

combine_BayesNet_ Logistic_Classification via Regression_ADTree_NBTree		Confusion table parameter				ROC	Sensitivity	Specificity	Kappa statistic
		TP	FP	FN	TN				
Correctly Classified Instances	92.24	1903	170	137	1409	0.981	0.922	0.922	0.8416
Incorrectly Classified Instances	7.76								

As table suggests, the accuracy of this combining method was 3.32 percent higher than the average accuracy of all classifiers employed in the method. Likewise, the area under the ROC curve showed a

0.022 increase compared to the average area under the ROC curve for classifiers. False recognition decreased in the confusion matrix for each of the classifiers.

In the second proposed method, the classifiers were combined with lower FN and FP using voting. Considering the confusion matrix in Table 8, Random forest, and SVM with polynomial kernel had the least amount of FN, whereas RBF and SVM with RBF kernel had the lowest FP. Therefore, the voting method of maximum probability and majority voting were used to combine these classifiers. The accuracy and confusion matrix for this combining method are shown in Table 10.

Table 10. Results of the second proposed method for 2013 to 2015 data.

Type of Voting	combine_Random Forest800,500_ SVMpoly_SVMrbf		Confusion				ROC	Sensitivity	Specificity	Kappa statistic
			TP	FP	FN	TN				
Maximum Probability	Correctly Classified Instances	95.16	2004	139	36	1440	0.993	0.952	0.953	0.901
	Incorrectly Classified Instances	4.84								
Majority Voting	Correctly Classified Instances	94.64	1988	142	52	1437	0.942	0.946	0.947	0.8903
	Incorrectly Classified Instances	5.36								

The accuracy of the voting method with maximum probability showed a 0.785 percent increase compared to the average accuracy of the classifiers used in it. Also, in the voting method with majority voting the accuracy showed a 0.265 increase compared to the average accuracy of the classifiers used in it.

In the second type, the proposed model was constructed using a dataset from 2013 and 2014, and evaluated on the 2015 dataset. Out of 2842 students in 2013 and 2014, 1794 students were NCR (63%) and 1046 were CR (37%) The number of 2015 students was 777, out of which 533 students were NCR (68.5%) and 244 students were CR (31.5%). In the first proposed method, the basis for selecting the classifiers was the criterion of accuracy, and those with low accuracy were used for the combination. In this type, in order to identify low accuracy classifiers, all classifiers were implemented on 2013 and 2014 data, and the model was also evaluated on the same data using a 10-fold cross validation.

Table 11 shows the value of NBTree, ADTree, Classification via Regression, Forest 800, Forest 500, SVM (RBF), SVM (poly), RBF, Logistic, and Bayes Net.

Table 11. *The result of ten classifiers for 2013 and 2014 data (training data) and 2015 data (test data).*

A. Results of ten classifiers for 2013 and 2014 data.

	NBTree	ADTree	Classification Via Regression	Forest 800	Forest 500	SVM (RBF)	SVM (poly)	RBF	Logistic	Bayes Net
Correctly Classified Instances	**96.30**	**88.17**	97.04	99.33	99.33	99.22	99.23	**93.30**	98.27	**89.09**
Incorrectly Classified Instances	**3.70**	**11.83**	2.96	0.67	0.67	0.77	0.77	**6.69**	1.73	**10.91**
Specificity	**0.963**	**0.881**	0.975	0.993	0.939	0.992	0.922	**0.933**	0.983	**0.896**
Sensitivity	**0.963**	**0.882**	0.978	0.993	0.947	0.992	0.922	**0.933**	0.983	**0.891**
Kappa statistic	**0.9207**	**0.7432**	0.9364	0.9856	0.9856	0.9834	0.9834	**0.843**	0.963	**0.7706**

B. The results of ten classifiers for 2015 data.

	NBTree	ADTree	Classification Via Regression	Forest 800	Forest 500	SVM (RBF)	SVM (poly)	RBF	Logistic	Bayes Net
Correctly Classified Instances	81.85	90.60	91.37	94.73	94.852	96.39	94.98	90.74	90.60	92.79
Incorrectly Classified Instances	18.15	9.40	8.63	5.27	5.148	3.61	5.02	9.26	9.40	7.21
Specificity	0.854	0.849	0.916	9.951	0.952	0.939	0.95	0.908	0.907	0.928
Sensitivity	0.819	0.852	0.914	0.948	0.949	0.947	0.95	0.907	0.906	0.928
Kappa statistic	0.6182	0.7822	0.8031	0.808	0.8836	0.9165	0.8834	0.8024	0.7836	0.8281

As shown in Table 11, NBTRee, ADTRee, RBF, and Bayes net classifiers have a low accuracy. Therefore, we selected these classifiers to combine with each other. First, we constructed a model using data from 2013 to 2014, and then evaluated the constructed model on 2015 data. Accuracy, confusion matrix, and the area under the ROC curve for this combining method are shown in Table 12.

Table 12. *Results of the first proposed method on 2015 data.*

combine_BayesNet_RBF_ADTree_NBTree		Confusion table parameter				ROC	Sensitivity	Specificity	Kappa statistic
		TP	FP	FN	TN				
Correctly Classified Instances	94.07	215	17	29	516	0.968	0.941	0.94	0.8607
Incorrectly Classified Instances	5.93								

The output of this method confirmed that there was a 5.08 percent rise in its accuracy compared to the average accuracy (88.99) of standalone classifiers used in the method. Likewise, the area under ROC curve showed a 0.009 increase compared to the average area under ROC curve (0.959) for classifiers. The wrong recognition in confusion matrix for each of the classifiers decreased.

In order to investigate the second proposed method in this type, after implementing the classifiers on 2013 and 2014 data, we compared the values of the confusion matrix to each other, so that we could identify the classifiers that had lower values of FN and FP. The values of confusion matrix are shown in Table 13.

Table 13. *Comparison of confusion matrix after the implementation of classifiers on 2013 and 2014 data.*

	ADTree	NBTree	Classification Via Regression	Forest 800	Forest 500	SVM (RBF)	SVM (poly)	RBF	Logistic	Bayes Net
TP	1654	1740	1757	1792	1791	1784	1784	1647	1764	1586
FP	194	49	45	**15**	**14**	**10**	**10**	114	17	100
FN	142	56	39	**4**	**5**	**12**	**12**	149	32	210
TN	852	997	1001	1031	1033	1036	1036	932	1029	946

As shown in Table 13, Forest classifiers with 500 and 800 trees as well as SVM with the RBF and polynomial kernels have the lowest FN and FP values. Therefore, to reduce these values and improve the output of the classifiers, we combined them together. According to the second proposed method in this article, we used the voting method with maximum probability and majority voting. In this step, after constructing the model using the voting method, we evaluated it on 2015 data. The results of this step are shown in Table 14.

Table 14. Results of the second proposed method on 2015 data.

Type of Voting	combine_Random Forest800,500_ SVMpoly_SVMrbf		Confusion				ROC	Sensitivity	Specificity	Kappa statistic
			TP	FP	FN	TN				
Maximum Probability	Correctly Classified Instances	95.63	233	23	11	510	0.993	0.956	0.957	0.8998
	Incorrectly Classified Instances	4.37								
Majority Voting	Correctly Classified Instances	96.53	232	15	12	518	0.961	0.965	0.965	0.9196
	Incorrectly Classified Instances	3.47								

The accuracy of the voting method with maximum probability showed a 0.392 percent increase compared to the average accuracy (95.238) of the classifiers used in it. Also, in the voting method with majority voting, the accuracy showed a 1.292 increase compared to the average accuracy (95.238) of standalone classifiers employed in the method. The false recognition decreased in the confusion matrix for each of the classifiers.

4 Discussion

The analysis described in this section was carried out in two phases. The first phase was when the model was constructed and evaluated on data from 2013 to 2015 using the 10-fold cross validation method. The second phase was when the model is constructed using the 10-fold cross validation method on 2013 and 2014 data, and the model was evaluated with 2015 data.

4.1 The first phase of analysis

For a further examination of the accuracy of the classifiers used in the article and an investigation of the proposed method, we compared the region under the ROC curve and the F-Measure criterion.

The ROC curve is a two-dimensional image of classifier efficiency. To compare classifiers, we may want to reduce the ROC's efficiency to a singular scalar value that indicates the expected efficiency. A current method for calculating the area under the ROC curve is AUC. Because the AUC is part of the unit square area, its value is always between 0 and 1. None of the realistic classifiers should have an AUC less than 0.5. The AUC has a significant statistical characteristic: the AUC of a classifier is the possibility that the classifier will grade a randomly selected positive sample higher than a randomly selected negative sample.

Figure 4 shows the area under the curve for all classifiers used in the three sections (Section 1: Dataset containing personal features and answers to the questionnaire; Section 2: Dataset, including personal and extracted features; Section 3: Includes personal feature, answers to the questionnaire, and the extracted features).

This comparison was done on the data set of the problem, including the period from 2013 to 2015, and the classifier was evaluated on the entire dataset using the 10-fold cross validation method.
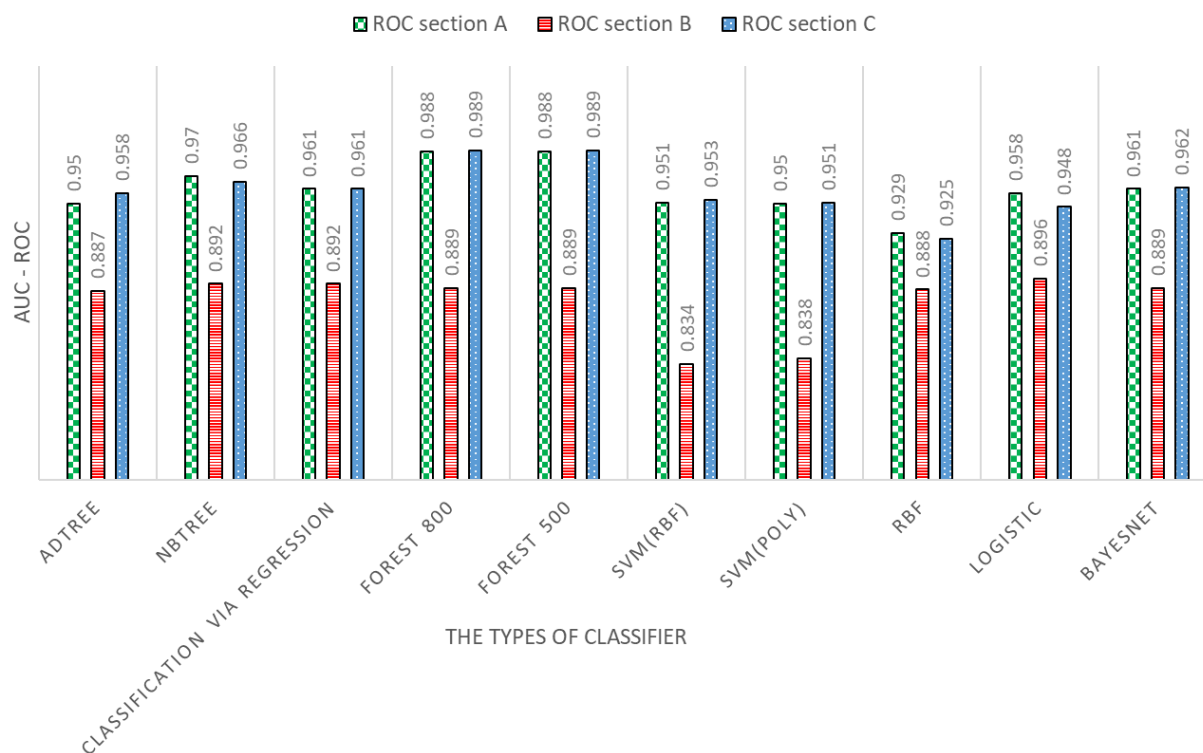
Figure 4. *Comparing the area under the ROC curve of all classifiers used in the three sections.*

Another criterion that can be analyzed is the F-Measure. The F-measure is the harmonic mean of precision and sensitivity. F-measure has an intuitive meaning. It indicates how precise a classifier is (how many instances it classifies correctly), as well as how robust it is (that it does not miss a significant number of instances). Figure 5 shows the comparison of the f-measure criterion.
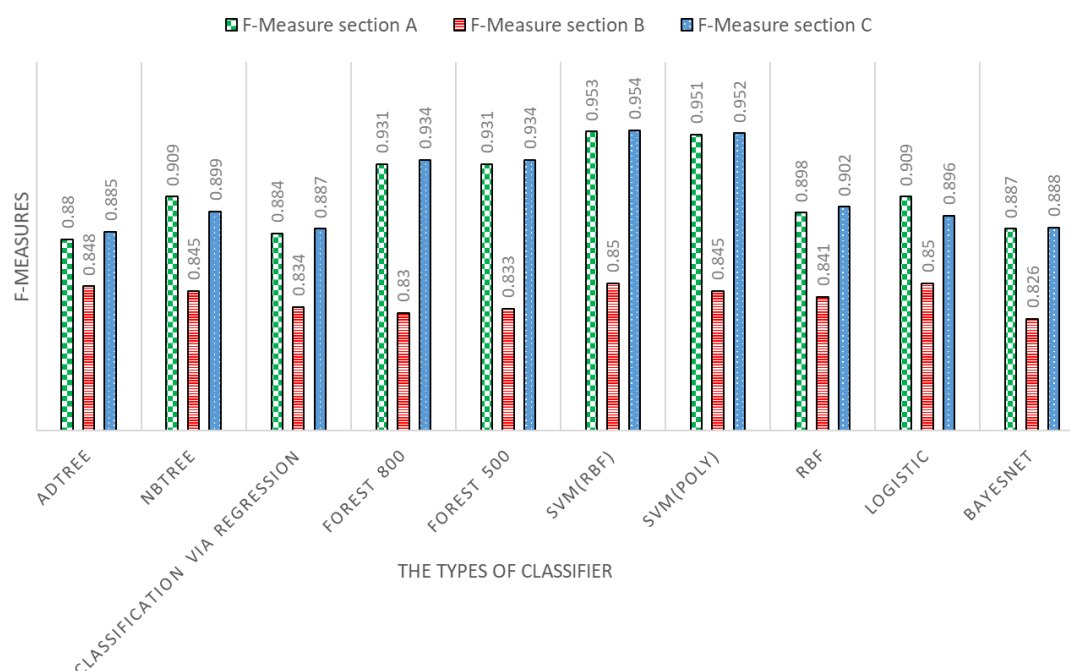

Figure 5. *Comparison of F-Measure.*

As shown in Figure 5, when classifiers perform classification for the data of section two, due to the fact that the data only include extracted features and personal information, they do not work well, the area under their curve is small and their F-measure value is also low. On the other hand, the area under the

curve of Random Forest classifiers has the highest value. Moreover, SVM classifiers have higher values than other classifiers in Figure 5. By observing Figures 4 and 5, and Tables 3, 5 and 7, we would conclude that the SVM classifier can classify the data with greater accuracy.

Figure 6 shows the ROC graph of the classifier used in the first proposed method, along with the proposed method for the data of section 3. The first proposed classifier has the highest point in the graph, and larger area under the curve than its own classifier, so the combination of these classifiers has improved the result.
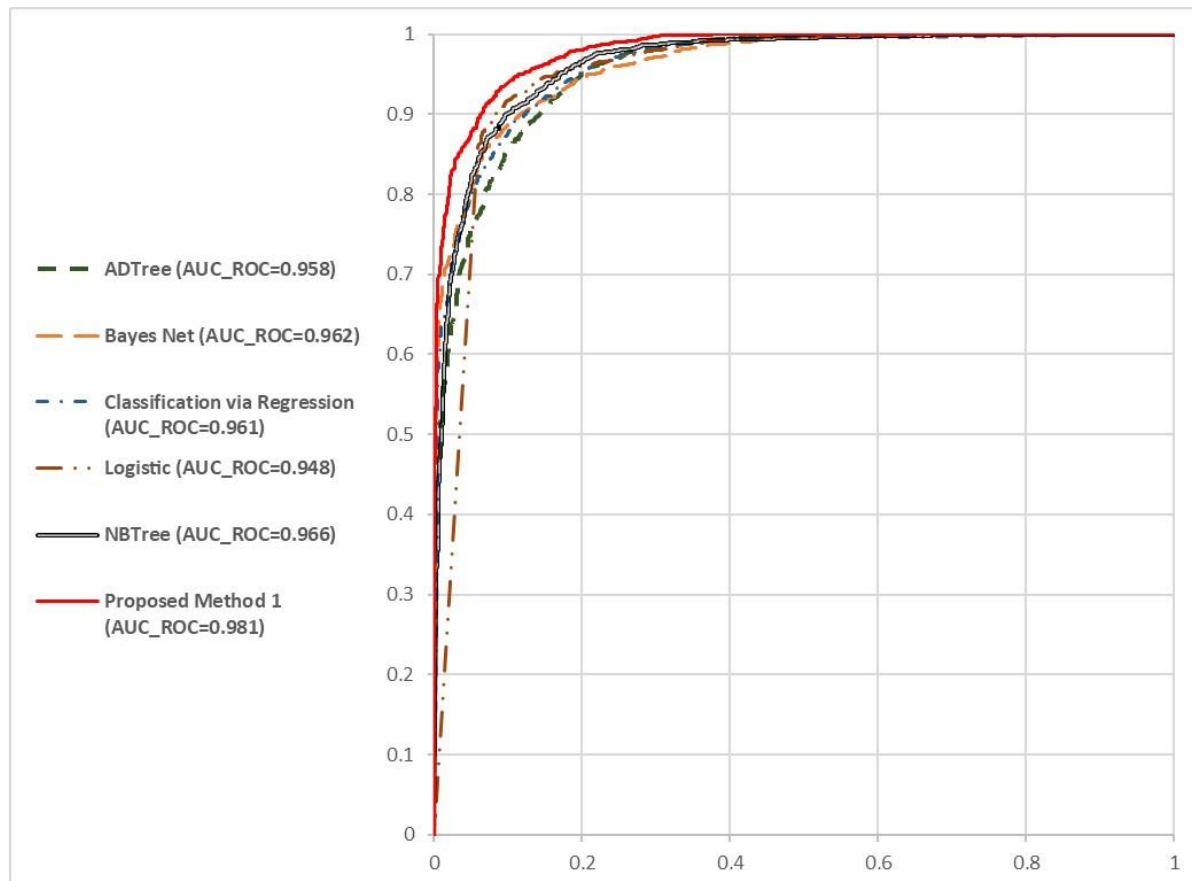
Figure 6. ROC curves for the first proposed method and standalone classifiers employed in the method.

As can be seen from the figure, the first proposed method had a bigger area under the curve. It could also decrease FN error as much as 21 units compared to the average error for the four classifiers. In addition, FP error can be decreased as much as 20 units compared to the average of FP error of the four classifiers.
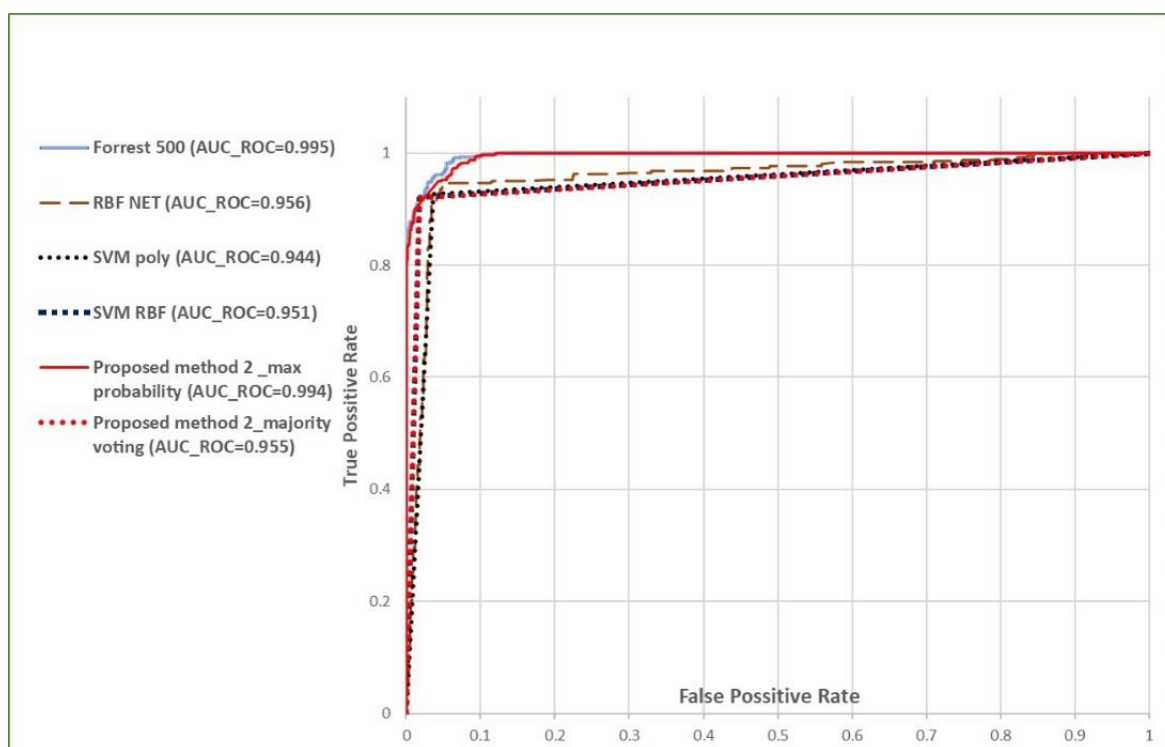
Figure 7. ROC curves for the second method and classifiers used in it.

Figure 7 shows the ROC curve in the second proposed method and standalone classifiers used in the method. As the figure shows, in this method there was an increase in ROC and the voting method with maximum probability had a bigger area under the curve. It was also able to decrease FN error as much as 7 units compared to the average error of the four classifiers. However, the average error of FP was unchanged compared to the average FP error of the four classifiers.

The voting method with majority voting had a larger area under the ROC curve. It was also able to reduce FN error compared to the average error of the four classifiers for 3 units. Furthermore, it reduced the FP error compared to the average FP error of the four classifiers for 16 units.

By comparing the ROC and AUC graph, we conclude that the methods proposed in this issue had greater accuracy, and the combination of classifiers based on the low accuracy criterion or the low FP and FN criteria was able to increase the accuracy and reduce FN and FP values. As a result, the system performance has increased.

4.2 The second phase of analysis

In this phase, we analyzed the ROC graph and the time complexity of the first and second proposed methods. In addition, the classifier model was constructed based on data from 2013 and 2014 and evaluated based on data from 2015.

Figure 8 shows the ROC curve used in the first proposed method, along with the proposed method. The first proposed classifier has the highest graph and a larger area under the curve than its own classifier used, thus the combination of these classifiers has improved the result.
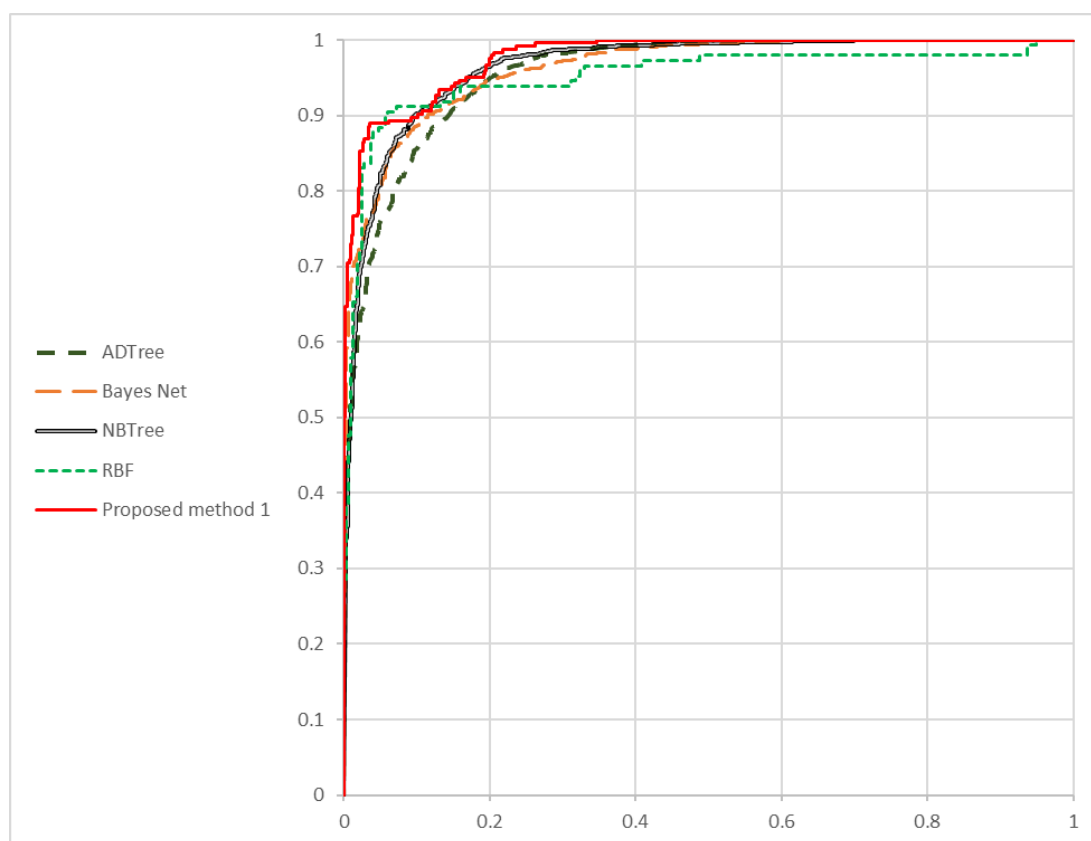
Figure 8. *The ROC curves for the first method and standalone classifiers employed in the method.*

The first proposed method had a larger area under the curve. It can also decrease FP error as much as 13 units compared to the average error (40) for the four classifiers. Furthermore, it can decrease FN error as much as 2 units, compared to the average FP error (28) of the four classifiers.
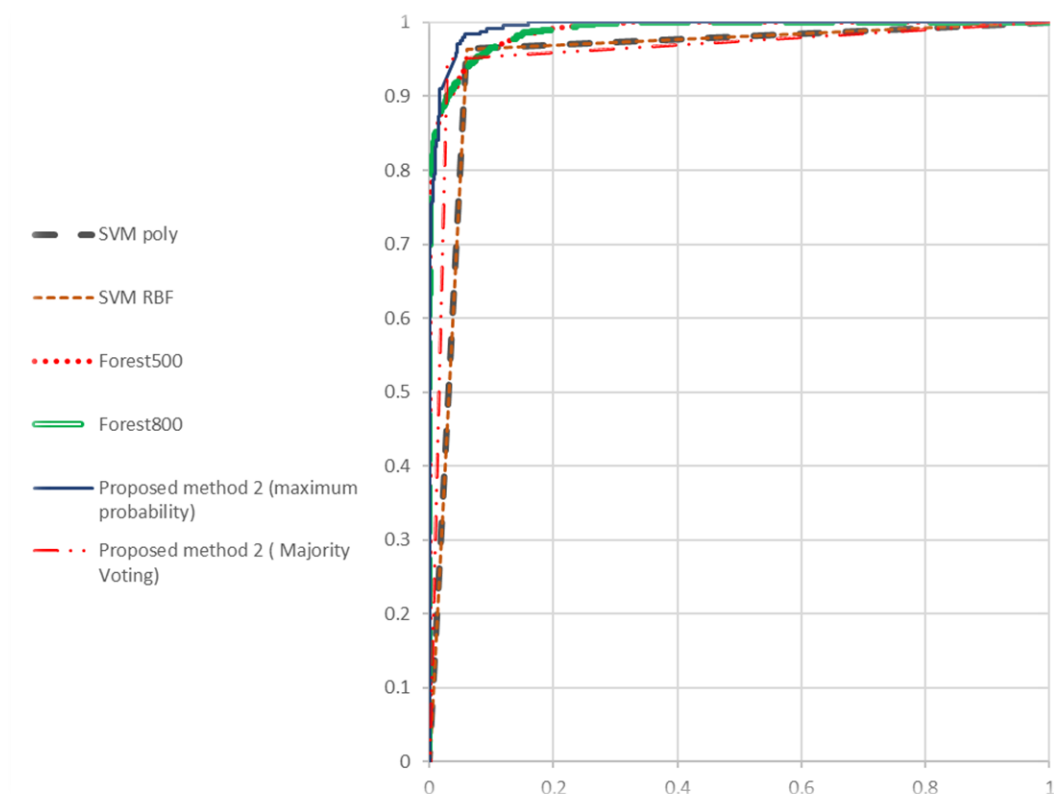

Figure 9. *The ROC curves for the second method and standalone classifiers employed in the method.*

Figure 9 shows the ROC curve in the second proposed method and standalone classifiers participating in the method. In this method there was an increase in ROC and the voting method with the maximum probability had a bigger area under curve. It was also able to decrease FN error as much as 2 units compared to the average error of the four classifiers. Furthermore, it can decrease FP error as much as 2 units compared to the average FP error (25) of the four classifiers.

In order to investigate the applicability of the method, the time complexity is compared for model construction and evaluation in the second type (when the construction data of the model is for 2013 and 2014, and evaluation data is for 2015). These analyses were carried out on a computer with Intel® Core ™ i7-2630QM processor with 6 GB of RAM on Windows 7, 64-bit.

Table 15. The time complexity of the second proposed method.

Classifiers	Model construction Time	Model evaluation Time	Accuracy
ADTree	10 s	Less than 1 s	90%
NBTree	475 s	Less than 1 s	91%
BayesNet	1 s	Less than 1 s	92%
RBF	Less than 1 s	Less than 1 s	93.30%
Proposed method 1	475 s	Less than 1 s	94.07%

In order to construct the model faster, it is possible to skip the NBTree classifier and create the model again with ADTree, NBTree and BayesNet, using the voting method with maximum probability. In this condition, the model was constructed in 3 seconds, but the accuracy of the model was 93.43 on the test data, which makes its difference from the average of the three classifiers (92.23) equal to 1.2%.

Table 16 shows the time complexity of the second proposed method. Although the model was constructed in about 105 seconds, this method could increase the percentage of accuracy compared to the average of the classifiers by about 3.13%.

Table 16. The time complexity of the second proposed method.

Classifiers	Model construction Time	Model evaluation Time	Accuracy
Random Forest 500	10 s	1s	84.852%
Random Forest 800	19 s	1 s	94.73%
SVM (Poly)	14 s	Less than 1 s	94.98%
SVM (RBF)	59 s	2 s	95.39%
Proposed method 2	105 s	2 s	95.62%

Although the proposed method increases accuracy, it has disadvantages. In this method, it is possible to consider other criteria such as time complexity of the classifier, or combine the classifiers based on fuzzy methods. These changes can be applied in later works, and investigate the accuracy of this type of classifier.

5   Conclusion

The article tried to study and compare different data mining methods in mental health assessment of Iranian students. The results indicate that the features which the psychologists used for classifying the students in addition to the answers of students to questionnaires influence the accuracy level of the classification model. The output of classifiers was combined based on different criteria. If the accuracy is considered, the second proposed method that uses voting method with majority voting is appropriate. By examining the area under the ROC curve, it was determined that the voting method with maximum probability has appropriate performance. One of the challenges in this research was the disproportionality of the amounts of data in the two groups of students. It can be suggested for future studies to apply fuzzy methods with this data to improve accuracy in the process of students' mental health assessment and to estimate the empty fields in questionnaires, the probability of the right answer and their side effects in the classification results.

Additional Information and Declarations

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: S.S.: Data curation, Formal analysis, Writing – original draft. M.J.N.: Formal analysis, Writing – review & editing. S.T.Y.: Data curation, Formal analysis, Visualization. M.N.: Validation. M.H.: Supervision, Project Administration.

Data Availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ahmadi, H., Gholamzadeh, M., Shahmoradi, L., Nilashi, M., & Rashvand, P. (2018). Diseases diagnosis using fuzzy logic methods: A systematic and meta-analysis review. *Computer Methods and Programs in Biomedicine*, 161, 145–172. https://doi.org/10.1016/j.cmpb.2018.04.013

Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining educational data to predict student's academic performance using ensemble methods. *International Journal of Database Theory and Application*, 9(8), 119–136. https://doi.org/10.14257/ijdta.2016.9.8.13

Burke, T. A., Ammerman, B. A., & Jacobucci, R. (2019). The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. *Journal of affective disorders*, 245, 869–884. https://doi.org/10.1016/j.jad.2018.11.073

Delen, D., Fuller, C., McCann, C., & Ray, D. (2009). Analysis of healthcare coverage: A data mining approach. *Expert systems with applications*, 36(2), 995–1003. https://doi.org/10.1016/j.eswa.2007.10.041

Deziel, M., Olawo, D., Truchon, L., & Golab, L. (2013). Analyzing the mental health of engineering students using classification and regression. In *Proceedings of the Educational Data Mining 2013* (paper no. 34). International Educational Data Mining Society.

Diederich, J., Al-Ajmi, A., & Yellowlees, P. (2007). Ex-ray: Data mining and mental health. *Applied Soft Computing*, 7(3), 923–928. https://doi.org/10.1016/j.asoc.2006.04.007

Dwivedi, A. K., Imtiaz, S. A., & Rodriguez-Villegas, E. (2018). Algorithms for automatic analysis and classification of heart sounds–a systematic review. *IEEE Access*, 7, 8316-8345. https://doi.org/10.1109/ACCESS.2018.2889437

Fakhlai, M., Pourreza, H. R., Moarefdost, R., & Shadroo, S. (2011). Offline signature recognition based on contourlet transform. In *Proceedings of the International Conference on Machine Learning and Computing IPCSIT*. IACSIT Press. http://www.ipcsit.com/vol3/036X241.pdf

George-Nektarios, T. (2013). *Weka classifiers summary*. Athens University of Economics and Bussiness Intracom-Telecom.

Hosseini, R., Qanadli, S. D., Barman, S., Mazinani, M., Ellis, T., & Dehmeshki, J. (2011). An automatic approach for learning and tuning Gaussian interval type-2 fuzzy membership functions applied to lung CAD classification system. *IEEE Transactions on Fuzzy Systems*, 20(2), 224–234. https://doi.org/10.1109/TFUZZ.2011.2172616

Laijawala, V., Aachaliya, A., Jatta, H., & Pinjarkar, V. (2020). Mental Health Prediction using Data Mining: A Systematic Review. In *Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST) 2020*. KJ Somaiya Institute of Engineering and Information Technology.

Latifi, F., Hosseini, R., & Mazinai, M. (2015). A fuzzy expert system for diagnosis of acute lymphocytic leukemia in children. *International Journal of Information, Security and Systems Management*, 4(2), 424–429.

Soet, J., & Sevig, T. (2006). Mental health issues facing a diverse sample of college students: Results from the College Student Mental Health Survey. *NASPA Journal*, 43(3), 410–431.

Superby, J.-F., Vandamme, J., & Meskens, N. (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods. In *Presented on the 8th international conference on intelligent tutoring systems, Educational Data Mining Workshop*. https://www.educationaldatamining.org/ITS2006EDM/superby.pdf

Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. *International Journal of Bio-Science and Bio-Technology*, 5(5), 241–266.

Vanlalawmpuia, R., & Lalhmingliana, M. (2020). Prediction of Depression in Social Network Sites Using Data Mining. In *Proceedings of the 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE. https://doi.org/10.1109/ICICCS48265.2020.9120899

Wei, Y., Qiu, J., & Karimi, H. R. (2016). Reliable output feedback control of discrete-time fuzzy affine systems with actuator faults. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 64(1), 170–181. https://doi.org/10.1109/TCSI.2016.2605685

Wei, Y., Qiu, J., & Karimi, H. R. (2017). Notice of Violation of IEEE Publication Principles: Fuzzy-affine-model-based memory filter design of nonlinear systems with time-varying delay. *IEEE Transactions on Fuzzy Systems*, 26(2), 504-517. https://doi.org/10.1109/TFUZZ.2017.2686352