Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 171 (2020) 1989–1998

Procedia Computer Science

www.elsevier.com/locate/procedia

Third International Conference on Computing and Network Communications (CoCoNet'19)

Assessment of Anxiety, Depression and Stress using Machine Learning Models

Prince Kumar[a], Shruti Garg[a]*, Ashwani Garg[b]

[a]BIT,Mesra,Ranchi,835215, INDIA

[b]Griffith University, Queensland, Australlia

Abstract

Over the last few decades, psychological health issues have become very common in people worldwide. In this paper, prediction of the occurrence of psychological problems such as anxiety, depression and stress has been made by applying eight machine learning algorithms to data taken from the online DASS42 tool. Five different severity levels of anxiety, depression and stress have been predicted using eight algorithms. The algorithms are grouped into four categories: probabilistic, nearest neighbor, neural network and tree based. A hybrid classification algorithm was also applied for prediction of different severity level anxiety, depression and stress. The same methods were also applied to another dataset, DASS21 collected by authors. The prediction accuracy found by using the hybrid algorithm was greater than by using single algorithms, but the highest accuracy was found by use of the radial basis function network, which comes under the category of neural network.

Keywords: DASS42, DASS21, J48, MLP, RBFN

1. Introduction

Modern lifestyles are causing different types of psychological health problems in many people. Psychological problems like anxiety, depression and stress have some overlapping features, for example, a person feels low and

* Corresponding author. Tel.: +919430730791
E-mail address: gshruti@bitmesra.ac.in

1877-0509
10.1016/j.procs.2020.04.213

lonely in all three. Generally, psychiatrists assess anxiety, depression and stress through questionnaires such as DASS42 and DASS21[1] because people suffering with anxiety, depression and stress are often not open to sharing their feelings with doctors, relatives or friends.

Therefore, in current work, the author has attempted to assess levels of anxiety, depression and stress by using computers without the help of any medical experts or face to face interaction. The data has been taken from https://openpsychometrics.org/_rawdata/, collected by online questionnaires filled in by different users between 2017 and 2019. A sample snapshot of the online questionnaire [2] is shown in figure 1.

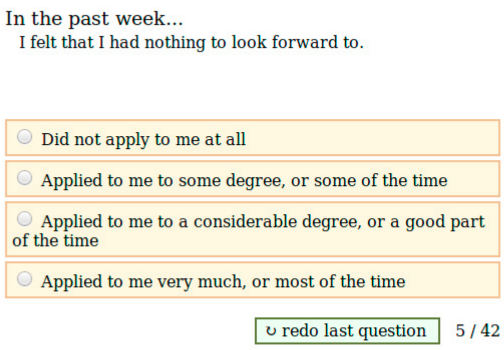

Fig.1. Sample snapshot of online questionnaire

Then, eight machine learning algorithms along with hybrid techniques were applied to classify data into five different Likert scales. Supervised machine learning algorithms were applied here as our data can be labelled by calculating scores of DASS42. Many researchers [3,4,5,6,7] applied machine learning (ML) for diabetes prediction. But the classification of diabetes is easier than that of psychological health because it has only two classes of outcome, whereas the prediction of anxiety, depression and stress has different severity levels for each psychological condition. Thus, this study comes under multiclass classification; five classes have been found in this work corresponding to five severity levels. A meta-analysis and review of machine learning techniques applied to depression is demonstrated in [8].

Mary, in [9] undertook the assessment of depression, stress and anxiety using different ML methods and logistic regression was found to be the most efficient with the highest accuracy: 90.33 percent for depression, 92 percent for anxiety and 90.33 percent for stress. The DASS21 questionnaire was used for data collection. Kessler et. al., in [10] surveyed 5,877 English speaking residents of the USA and assessed the severity of depression using ML and conventional methods. They found that ML methods performed better than conventional techniques.

Anxiety and depression were also screened among seafarers in [11]. Boosting, along with traditional ML methods, were applied to interviewed seafarer data and boosting was found to be better than other approaches. Multiple kernels of support vector machine (SVM) along with other ML methods were applied to classify depression on twitter data are shown in [12]. Multi kernel SVM was found to be the best method, with accuracy of 83.46 percent. Different levels of anxiety, depression and stress were further assessed through ML in [13] from the DASS21 questionnaire. The shortcoming of this work arises from the dataset having a small number of data items. The actual performance of ML methods can be more successfully assessed on large data. Persistent depression among elderly adults was assessed by extreme gradient boosting in [14]. Data was collected through the PHQ-9 questionnaire.

All of the work mentioned above suffers from the problems associated with a small number of data items. Another problem with interviews or questionnaire-based studies is that participants are not willing to respond to many of the questions. As previously noted, people suffering with anxiety, depression and stress are often not open to close relatives, friends or medical experts and they generally share their feelings by anonymous means. So, the internet is the best platform to collect their responses. That is why the data set taken here is from online questionnaires.

The organization of the remaining sections of this paper is as follows: Section 2 consists of a brief description of the dataset and methods used. The results and discussions are shown in section 3. Section 4 describes the conclusions of this work.

2. Methodology

2.1 Data Collection

In total, total 39,776 instances were collected through online questionnaires between 2017 and 2019 by different methods. The dataset consists of 42 questions taken from the standard form of DASS42. The responses are scaled between 1 and 4 and are coded as:
1 = Did not apply to me at all
2 = Applied to me to some degree, or some of the time
3 = Applied to me to a considerable degree, or a good part of the time
4 = Applied to me very much, or most of the time
DASS42 consist of 42 questions of which 14 questions relate to each of anxiety, depression and stress [15]. Although the responses have been collected from 1 to 4, the calculation has been on 0 to 3 by subtracting 1 from each.

A quality comparison of data collected has been performed. This dataset is found better than or equivalent to the data available on Amazon Mechanical Tuck. The comparison of quality of data is given in [21].
The scores for anxiety, depression and stress were calculated by adding the values associated with the answers to each question of the particular class. Once the final scores had been calculated, they were labelled according to severity, i.e. Extremely Severe, Severe, Moderate, Mild and Normal.

The extremely severe class for anxiety, depression and stress was given a score of 20+, 28+ and 33+ respectively. The severe class for the three was scored at 15 to 19, 21 to 27 and 26 to 33 respectively. The moderate class was assigned scores of 10 to14 for anxiety, 14 to 20 for depression and 19 to 25 for stress. The score range for mild is 8 to 9 for anxiety, 10 to 13 for depression and 15 to 18 for stress. Data falling below these scores comes under the normal category for all three conditions.

A snapshot of the score calculation sheet is shown in figure 2.

Q2	Q4	Q7	Q9	Q15	Q19	Q20	Q23	Q25	Q28	Q30	Q36	Q40	Q41	Score	Anxiety
3	3	0	0	1	1	0	1	0	2	3	1	0	0	15	SEVERE
3	3	3	1	3	2	3	3	3	3	3	1	1	2	34	EXTREMELY SEVERE
1	0	2	1	0	1	1	0	1	0	0	1	2	0	10	MODERATE
0	2	0	0	0	0	1	0	0	0	1	0	0	1	5	NORMAL
3	0	0	0	0	0	1	0	2	0	0	2	0	0	8	MILD

(a)Score Calculation of Anxiety

Q3	Q5	Q10	Q13	Q16	Q17	Q21	Q24	Q26	Q31	Q34	Q37	Q38	Q42	Score	Depression
3	3	3	3	1	2	1	3	1	3	3	3	0	3	32	EXTREMELY SEVERE
1	2	1	0	0	0	3	3	1	3	0	1	3	3	21	SEVERE
3	1	2	0	0	2	0	1	2	2	2	1	1	2	19	MODERATE
0	1	2	1	0	2	1	1	0	0	0	3	1	0	12	MILD
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	NORMAL

(b) Score Calculation of Depression

Q1	Q6	Q8	Q11	Q12	Q14	Q18	Q22	Q27	Q29	Q32	Q33	Q35	Q39	Score	Stress
3	2	3	0	3	3	0	2	3	1	3	3	3	3	32	SEVERE
3	2	1	3	3	1	1	3	0	1	1	0	1	2	22	MODERATE
2	0	0	2	0	0	0	1	0	3	2	3	1	1	15	MILD
2	3	0	0	1	2	0	0	0	0	1	0	0	0	9	NORMAL
3	1	3	3	3	3	3	3	3	2	3	3	3	3	39	EXTREMELY SEVERE

(c) Score Calculation of Stress

Fig.2. Score calculation sheet for Anxiety, Depression and Stress

2.2 Classification

Five different severity level have been predicted using eight ML algorithms. The algorithms applied here are naïve Bayes (NB), Bayes network (BN), k-star, local nearest neighbor (LNN), multilayer perceptron (MLP), radial basis function network (RBFN), random forest (RF) and J48. All eight algorithms belong to four broad categories. The description of each category is explained in the following subsections:

2.2.1 Bayes classification

The naïve Bayes classifier calculates conditional probability using Bayes theorem to divide into different classes. This theorem depends on the naïve assumption, in which input factors are independent of each other. A detail description of this method is available in [16]. Another method applied to this category was Bayesian network, a neural network which updates weights based on conditional probability. The method description is available in [17].

2.2.3 K-Nearest Neighbor

This algorithm finds similarity between predefined classes and the classes to be classified using Euclidean distance. Another algorithm used in this category is K-star, uses similarity measure as entropy distance. The detailed description of this method is available in [18].

2.2.4 Neural network

The neural network classifier works on the principle of error correction learning. The networks learn from the training dataset and the networks evolve until an acceptable error is not found. The neural networks used in this work are multilayer perceptron (MLP) and radial basis function network (RBFN). The RBFN is more efficient because it used gaussian kernel for the separation of patterns. The description of these methods is found in [19].

2.2.5 Tree-based classification

Two tree-based classifiers are used here: J48 and random forest. J48 constructs a decision tree using information gain. Random forest creates a forest of multiple decision trees. J48 and random forest are described in [20].

3. Results and Discussions

The eight method of ML were applied using the WEKA data mining tool to classify three psychological disorders of five different severity levels. The dataset is divided by a ratio of 75:25 to create, train and test cases. A five-fold cross validation has been applied after training to improve accuracy.

Table 1 Confusion Matrix obtained by different ML methods on Anxiety, Depression and Stress

Method Name	Anxiety	Depression	Stress
BayesNet	a b c d e	a b c d e	a b c d e
	4249 612 0 0 0	2044 331 0 0 0 0	1754 150 0 0 37
	26 1535 56 0 0	34 2003 1340 0 0	72 2193 44 0 0
	0 128 1425 0 0	0 103 2074 26 0	0 203 1067 9 0
	0 0 1 1245 122	0 0 178 1005 15	0 0 429 2986 0
	0 0 388 29 128	0 0 0 208 1789	151 0 0 0 849
NaiveBayes	a b c d e	a b c d e	a b c d e
	3381 497 0 0 0	1630 271 0 0 0	1404 128 0 0 24
	19 1236 52 0 0	28 1604 93 0 0	62 1762 41 0 0
	0 93 1129 0 0	0 84 1682 26 0	0 167 848 5 0
	0 0 1 984 102	0 0 143 793 11	0 0 348 2373 0
	0 0 332 27 102	0 0 0 167 1423	115 0 0 0 678
Multilayer Perceptron	a b c d e	a b c d e	a b c d e
	4861 0 0 0 0	2375 0 0 0 0	0 1941 0 0 0
	0 1617 0 0 0	0 2171 0 0 0	0 2309 0 0 0
	0 0 313 1240 0	0 0 2203 0 0	0 0 1279 0 0
	0 0 0 1368 0	0 0 622 0 576	0 0 0 3415 0
	0 0 0 545 0	0 0 0 0 1997	0 1000 0 0 0
RBFN	a b c d e	a b c d e	a b c d e
	4810 51 0 0 0	2330 45 0 0 0	1853 63 0 0 25
	83 1512 22 0 0	87 2018 66 0 0	88 2183 38 0 0
	0 20 1507 0 26	0 73 2092 38 0	0 36 1174 69 0
	0 0 0 1355 13	0 0 54 1132 12	0 0 30 3385 0
	0 0 19 16 510	0 0 0 20 1977	31 0 0 0 969
K-Star	a b c d e	a b c d e	a b c d e
	4610 237 14 0 0	2302 73 0 0 0	1286 132 0 0 523
	251 806 494 46 20	550 1280 340 1 0	491 1475 263 75 5
	2 91 904 382 174	3 203 1512 313 172	0 154 433 692 0
	0 0 0 1342 26	0 0 71 489 638	0 1 46 3368 0
	0 0 66 355 124	0 0 0 31 1966	17 0 0 0 983
K-nearest neighbour	a b c d e	a b c d e	a b c d e
	4587 257 17 0 0	2282 93 0 0 0	1404 138 0 0 399
	339 831 408 13 26	585 1272 313 1 0	598 1390 277 44 0
	4 169 894 277 209	9 319 1418 359 98	3 259 516 501 0
	0 0 3 1326 39	0 0 128 567 503	0 6 143 3266 0
	0 0 92 308 145	0 0 0 64 1933	57 0 0 0 943
J48	a b c d e	a b c d e	a b c d e
	4459 333 64 3 2	1981 360 34 0 0	1251 484 26 6 174
	603 693 286 11 24	441 1160 549 19 2	485 1337 325 157 5
	111 362 816 139 125	45 530 1312 245 71	44 431 381 423 0
	3 7 117 1137 104	2 38 414 445 299	3 121 261 3030 0
	4 30 210 164 137	0 0 73 229 1695	240 5 0 0 755
RandomForest	a b c d e	a b c d e	a b c d e
	4806 55 0 0 0	2254 121 0 0 0	1705 169 0 0 67
	497 955 165 0 0	203 1713 255 0 0	230 2010 63 6 0
	4 224 1235 48 42	0 216 1919 66 2	0 406 429 444 0
	0 0 5 1334 29	0 0 339 618 241	0 7 38 3370 0
	0 1 196 214 134	0 0 0 51 1946	121 0 0 0 879
K-Star with Random Forest	a b c d e	a b c d e	a b c d e
	4753 108 0 0 0	2280 95 0 0 0	1728 134 0 0 79
	200 1340 77 0 0	132 1909 130 0 0	144 2020 145 0 0
	0 127 1361 1 64	0 152 1983 68 0	0 125 994 160 0
	0 0 0 1309 59	0 0 102 999 97	0 0 111 3304 0
	0 0 81 63 401	0 0 0 81 1916	56 0 0 0 944

a, b, c, d and e in the confusion matrix represent the normal, mild, moderate, severe and extremely severe classes respectively. After deriving the confusion matrix for anxiety, depression and stress shown in table 1, equations 1 to 6 were used to calculate the efficiency of different algorithms.

$$AccuracyRate = \frac{Sum of diagonals(TP)}{Total number of instances\ to\ classify} \tag{1}$$

$$ErrorRate = 1 - AccuracyRate \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$Kappa = \frac{Total\ Accuracy - randomAccuracy}{1 - randomAccuracy} \tag{5}$$

$$F1\ Measure = 2 * Precision * Recall/(Precision + Recall) \tag{6}$$

Where,
TP (True positive) = Diagonals of matrix
FN (False Negative) = Sum of the corresponding row for class (excluding TP of that class)
FP (False Positive) = Sum of the corresponding column for class (excluding TP of that class)
TN (True Negative) = Sum of the all row and column (excluding row and column of that class

Table 2. Statistical measures of different classification methods for DASS42

Classifier	Mental illness	Accuracy	Error Rate	Precision	Recall	Kappa	F-Measure	ROC Area
Bayes Net	Anxiety	86.3	13.7	0.877	0.863	0.8055	0.860	0.990
	Depression	89.6	10.4	0.904	0.897	0.8693	0.898	0.992
	Stress	88.9	11.1	0.901	0.890	0.8569	0.892	0.992
Naïve Bayes	Anxiety	78.64	21.36	0.779	0.786	0.6877	0.781	0.973
	Depression	74.89	25.11	0.747	0.749	0.6818	0.747	0.966
	Stress	75.73	24.27	0.752	0.757	0.6799	0.753	0.967
Multilayer Perceptron	Anxiety	82.04	17.96	0.858	0.820	0.7384	0.735	0.993
	Depression	87.95	12.05	0.889	0.880	0.8456	0.937	0.955
	Stress	70.42	29.58	0.813	0.704	0.6019	0.870	0.964
RBFN	Anxiety	97.48	2.52	0.975	0.975	0.9634	0.975	0.999
	Depression	96.02	3.98	0.960	0.960	0.9498	0.960	0.998
	Stress	96.17	3.83	0.962	0.962	0.9499	0.962	0.999
K-Star	Anxiety	78.2	21.8	0.781	0.783	0.6845	0.773	0.966
	Depression	75.9	24.1	0.759	0.759	0.6945	0.746	0.963
	Stress	75.87	24.13	0.755	0.759	0.6809	0.743	0.964
K-nearest neighbour	Anxiety	78.26	21.74	0.775	0.783	0.6832	0.773	0.930
	Depression	75.14	24.86	0.746	0.751	0.6851	0.740	0.926
	Stress	75.61	24.39	0.749	0.756	0.6793	0.746	0.926
J48	Anxiety	72.82	27.18	0.712	0.728	0.5982	0.719	0.861
	Depression	66.30	33.7	0.657	0.663	0.5727	0.659	0.847
	Stress	67.92	32.08	0.670	0.679	0.5766	0.673	0.854
Random Forest	Anxiety	85.11	14.89	0.840	0.851	0.778	0.837	0.982
	Depression	84.97	15.03	0.851	0.850	0.809	0.844	0.979
	Stress	84.40	15.6	0.842	0.844	0.792	0.829	0.978
K-star with Random Forest	Anxiety	92.15	7.85	0.920	0.922	0.885	0.921	0.994
	Depression	91.38	8.62	0.913	0.914	0.891	0.914	0.993
	Stress	90.40	9.6	0.90.	0.904	0.874	0.904	0.992

Table 2 shows the different statistical measures calculated from the confusion matrix obtained after applying different ML classification techniques. It is very much evident from table 1 that Bayes net performed better than naïve Bayes in the Bayes group. As Bayes net is a neural network, the network evolved over different iterations, so its performance was better than normal classifiers. RBFN performed best in the neural network category, because RBFN uses kernels which separate classes in higher dimensions. The performance of K-Star and KNN was almost equal. Both work on the same principle with different distance metrics. Among tree classifiers, the performance of random forest was better than J48 because it searched deeper than the decision tree. And last, the hybrid method, that is, a combination of K-star and random forest, increased the accuracy of random forest, which is greater among random forest and k-star. But the hybrid method is very much time consuming. Although the hybrid method increased accuracy, it was not greater than RBFN. The capacity for learning makes neural networks better than other classification methods. Among all classification methods, the performance of RBFN was best and the worst performance was given by J48, as shown in figure 3.

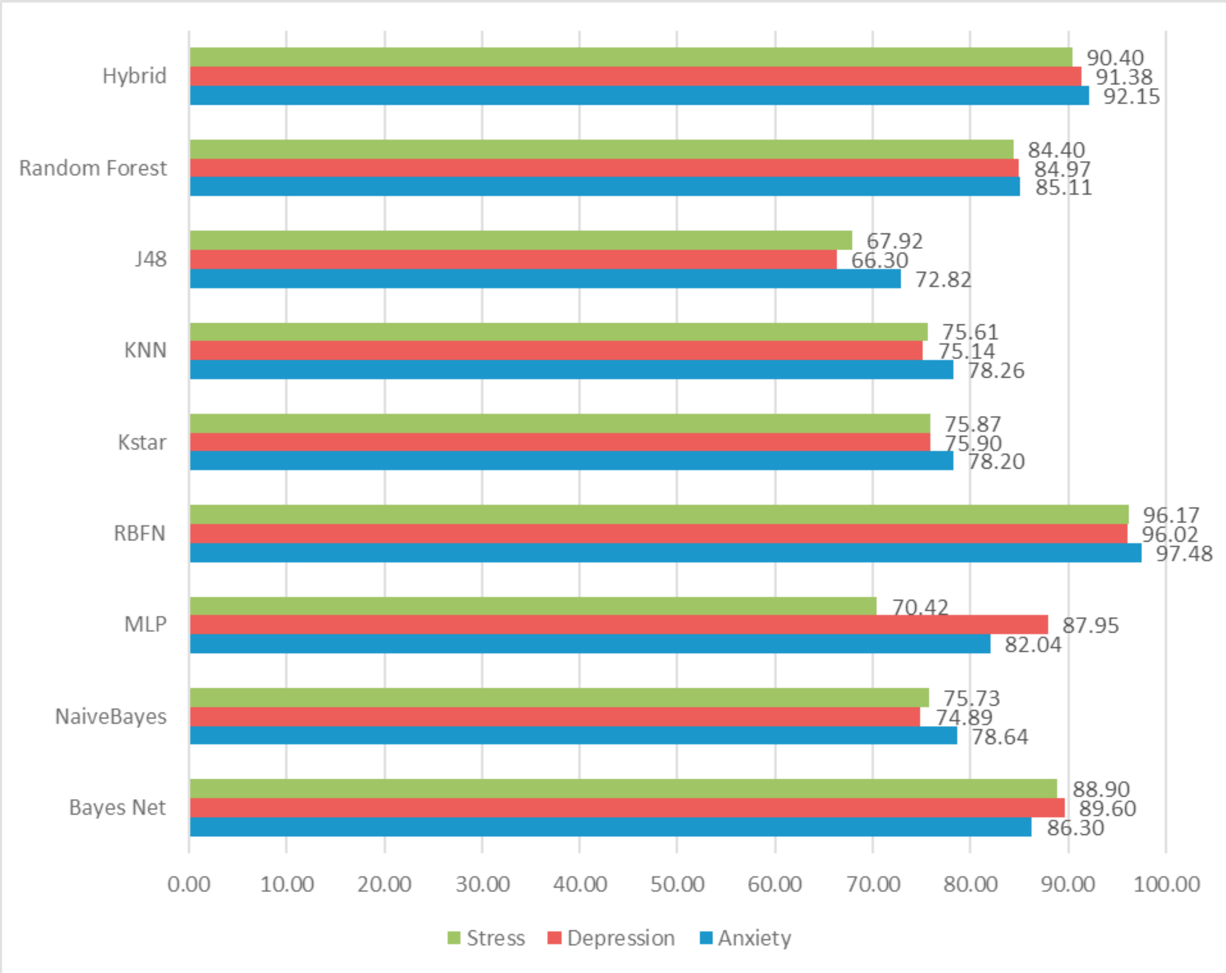

Fig.3. Accuracy of classification for different ML algorithm for DASS42

The same classification techniques were also applied on a different dataset from DASS21. This data was collected through Google forms completed by 349 participants from various parts of north India. The dataset consisted of 349 adults aged between 18 and 60 years with five severity levels of anxiety, depression and stress. The DASS21 score was calculated in same manner as DASS42. This score needed to be multiplied by two for each condition: anxiety, depression and stress because DASS21 consisted of only 21 items, with seven questions in each category i.e. anxiety, depression and stress. The results of the application of eight different ML methods on DASS21 is shown in table 3.

The results in table 3 show that MLP gives best performance for anxiety, depression and stress and RBFN gives the best performance for depression in DASS21. The accuracy is 100 percent and the precision is one for anxiety in random forest. These fluctuations may have arisen because there was only a small number of instances in dataset. Otherwise RBFN outperformed for this dataset also. Figure 4 shows the accuracy of classification of different ML methods for DASS21.

Table 3. Statistical measures of different classification methods for DASS21

Classifier	Mental illness	Accuracy	Error Rate	Precision	Recall	Kappa	F-Measure	ROC Area
Bayes Net	Anxiety	80.45	19.55	0.856	0.805	0.725	0.782	0.955
	Depression	83.90	16.1	0.834	0.839	0.770	0.831	0.975
	Stress	79.31	20.69	0.818	0.793	0.695	0.794	0.970
Naïve Bayes	Anxiety	83.90	16.1	0.835	0.839	0.781	0.826	0.975
	Depression	85.05	14.95	0.783	0.863	0.851	0.835	0.974
	Stress	79.31	20.69	0.818	0.793	0.695	0.794	0.970
Multilayer Perceptron	Anxiety	98.85	1.15	0.989	0.989	0.984	0.988	0.992
	Depression	93.10	6.9	0.942	0.931	0.921	0.927	0.999
	Stress	96.55	3.45	0.972	0.966	0.948	0.966	0.999
RBFN	Anxiety	82.75	17.25	0.847	0.828	0.771	0.834	0.954
	Depression	96.55	3.45	0.974	0.966	0.951	0.964	0.995
	Stress	89.65	10.35	0.916	0.897	0.846	0.896	0.976
K-Star	Anxiety	89.65	10.35	0.904	0.897	0.860	0.898	0.984
	Depression	93.10	6.9	0.942	0.931	0.901	0.927	0.995
	Stress	93.10	6.9	0.956	0.931	0.897	0.934	0.992
K-nearest neighbour	Anxiety	89.65	10.35	0.912	0.897	0.861	0.899	0.951
	Depression	95.40	4.6	0.952	0.954	0.934	0.952	0.973
	Stress	93.10	6.9	0.956	0.931	0.897	0.934	0.974
J48	Anxiety	91.66	8.34	0.916	0.917	0.889	0.915	0.989
	Depression	87.35	12.65	0.879	0.874	0.816	0.864	0.960
	Stress	87.35	12.65	0.892	0.874	0.813	0.877	0.962
Random Forest	Anxiety	100	0	1	1	1	1	1
	Depression	93.10	6.9	0.942	0.931	0.901	0.927	0.995
	Stress	91.95	8.05	0.935	0.920	0.88	0.918	0.994
K-Star with Random Forest	Anxiety	90.80	9.2	0.908	0.908	0.875	0.906	0.989
	Depression	91.95	8.05	0.933	0.920	0.884	0.913	0.998
	Stress	94.25	5.75	0.955	0.943	0.913	0.942	0.994

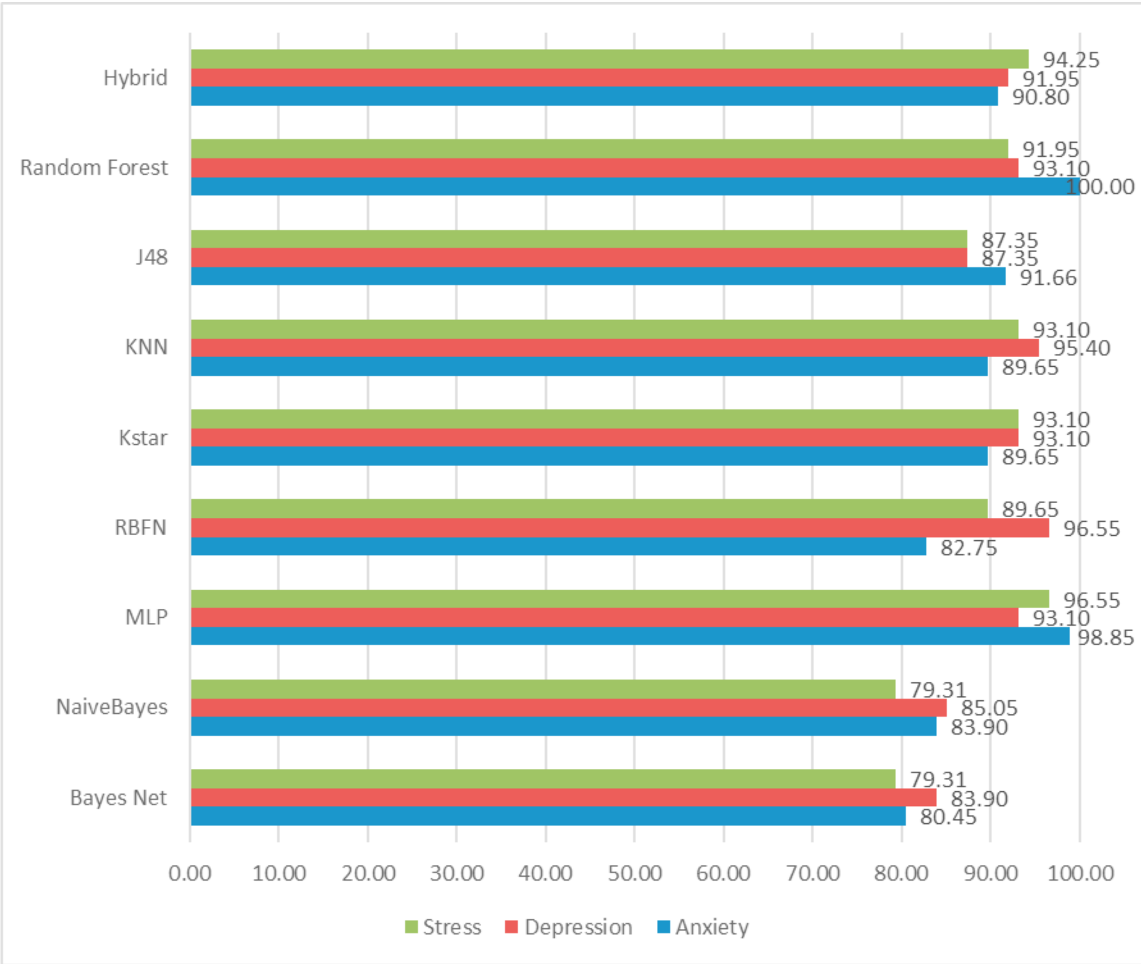

Fig.4. Accuracy of classification for different ML algorithm for DASS21

4. Conclusions

This research focused on the prediction of five severity levels of anxiety, depression and stress using eight different machine learning models. These methods fall in to four different categories: Bayes, neural network, lazy and tree. Last is a hybrid technique of K-star and random forest method. The hybrid approach improved the accuracy of single algorithm, but it took 30 to 45 minutes to execute, whereas single algorithms were executed in a maximum of five minutes. All the methods were applied to two different databases, DASS42 and DASS21, collected from different sources. After application of all the techniques, the results showed that neural networks performed better than all the others. Among the category of neural networks, RBFN performed the best for depression in both the datasets. However, the result of random forest is 100 percent for anxiety in DASS21. This occurred because of using a small number of instances in the dataset, also the dataset was imbalanced.

References

[1] Parkitny L and McAuley J.(2010)"The depression anxiety stress scale (DASS*)". Journal of physiotherapy* **56(3):**204.
[2] https://openpsychometrics.org/tests/DASS/
[3] Choi BG, Rha SW, Kim SW, Kang JH, Park JY and Noh YK. (2019) "Machine learning for the prediction of new-onset diabetes mellitus during 5-year follow-up in non-diabetic patients with cardiovascular risks."*Yonsei medical journal*;**60(2)**:191-9.
[4] Swapna G, Vinayakumar R andSoman KP. (2018) "Diabetes detection using deep learning algorithms."*ICT Express*.**4(4)**:243-6.

[5] Wu H, Yang S, Huang Z, He J and Wang X. "Type 2 diabetes mellitus prediction model based on data mining."*Informatics in Medicine Unlocked*.10:100-7.
[6] Tigga N. P. and Garg S. "Prediction of Type 2 Diabetes using Machine Learning Classification Methods."Accepted to publish in *Procedia Computer Science Journal*.
[7] Perveen S, Shahbaz M, Keshavjee K andGuergachi A. (2018) "Metabolic Syndrome and Development of Diabetes Mellitus: Predictive Modeling Based on Machine Learning Techniques."*IEEE Access*.7:1365-75.
[8] Lee Y, Ragguett RM, Mansur RB, Boutilier JJ, Rosenblat JD, Trevizol A, Brietzke E, Lin K, Pan Z, Subramaniapillai M and Chan TC.(2018)"Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review." *Journal of affective disorders*.241:519-32.
[9] Mary SA andJabasheela L. (2018) "An Evaluation of Classification Techniques for Depression, Anxiety and Stress Assessment." *In International Conference for Phoenixes on Emerging Current Trends in Engineering and Management (PECTEAM 2018)*. Atlantis Press.
[10] Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Cai T, Ebert DD, Hwang I, Li J, de Jonge P and Nierenberg AA. (2016)"Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports."*Molecular psychiatry.***21(10):**1366.
[11] Sau A and Bhakta I. (2019)"Screening of anxiety and depression among seafarers using machine learning technology."*Informatics in Medicine Unlocked*. 100228.
[12] Peng Z, Hu Q and Dang J. (2019)"Multi-kernel SVM based depression recognition using social media data."*International Journal of Machine Learning and Cybernetics*.**10(1)**:43-57.
[13] Anupriya, Garg S. andTigga N. P. "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms." Accepted to publish in *Procedia Computer Science Journal*.
[14] Hatton CM, Paton LW, McMillan D, Cussens J, Gilbody S and Tiffin PA. (2019) "Predicting persistent depressive symptoms in older adults: A machine learning approach to personalised mental healthcare."*Journal of affective disorders*.;246:857-60.
[15] Lovibond SH and Lovibond PH. (2006)"Manuals for the Depression Anxiety Stress Scales." 2nd ed. *Psychology Foundation*.
[16] Martinez-Arroyo, M., Sucar and L. E. (2006) "Learning an optimal naive bayes classifier."*In18th International Conference onPattern Recognition (ICPR'06) IEEE***3**:1236-1239.
[17] Friedman N, Geiger D and Goldszmidt M. (1997)"Bayesian network classifiers."*Machine learning*. **29(2-3)**:131-63.
[18] Cleary JG and Trigg LE. (1995)"K*: An instance-based learner using an entropic distance measure."*In Machine Learning Proceedings 1995* (pp. 108-114). Morgan Kaufmann.
[19] Khalili M and Ghasemi MM. (2014) "The effect of neural network models on diagnosis of bipolar disorder." *International Journal of Engineering Innovation & Research*.**3(9)**:700-04.
[20] Ali J, Khan R, Ahmad N and Maqsood I. (2012) "Random forests and decision trees." *International Journal of Computer Science Issues (IJCSI)*.**9(5)**:272.
[21] Open-Source Psychometrics Project. (2019) "A quality comparison of data collected on this website to data collected on Amazon Mechanical Turk." *Open-Source Psychometrics Project* viewed on 02 Decmber 2019
< https://openpsychometrics.org/_rawdata/validity/>