2022 IEEE 7th International conference for Convergence in Technology (I2CT)
Pune, India. Apr 07-09, 2022

Prediction of Depression, Anxiety and Stress Levels Using Dass-42

Srinath K S[1], Kiran K[2], Pranavi S[3], Amrutha M[4], P Deepa Shenoy[5], and Venugopal K R[6]
[1,2,3,4,5]Department of Computer Science and Engineering.
University Visvesvaraya College of Engineering, Bangalore University, Bengaluru, India.
[6]Bangalore University, Bengaluru, India.
[1]solansrinath@gmail.com

2022 IEEE 7th International conference for Convergence in Technology (I2CT) | 978-1-6654-2168-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/I2CT54291.2022.9824087

***Abstract*—Over the past few decades, many people are suffering from psychological health issues such as anxiety, depression, and stress. It is crucial to detect mental health condition on a timely basis and cure before it turns into a severe problem. In this paper, we analyze the performance of Machine learning (ML) algorithms to predict severity levels of depression, anxiety, and stress (DAS). Severity levels can be measured by the Depression, Anxiety, and Stress Scale (DASS) which consists of a set of questionnaires (DASS42). The intention of this research is to enhance and compare the performance of two different ML algorithms namely Support Vector Machine (SVM) and Logistic Regression (LR) based on classification accuracy with other algorithms. After parameter tuning SVM attains a classification accuracy of 97.35%, 97.49%, and 97.20% for Depression, Anxiety, and Stress. LR attains a classification accuracy of 98.15%, 98.05%, and 98.45% for Depression, Anxiety, and Stress dataset respectively. Parameter tuning is done for SVM and LR to obtain better accuracy with the best suitable parameters, and the results revealed that LR achieved better performance in terms of accuracy compared to SVM.**

Keywords—ML, DAS, DASS, DASS42, SVM, LR

I. Introduction

According to World Health Organization (WHO), health is not only related to the absence of disease but also associated with mental and social well-being. In the action plan report (2013-2020) WHO stated that a good mental health condition is one that allows people to know their potential, resist their daily stresses, and work productively to contribute to themselves, family and society [21]. It also reported that approximately four hundred and fifty million people suffer from a mental disorder. Globally, 264 million people are affected by depression, 284 million people experienced an anxiety disorder in which more women are affected than men [1].

Depression is a mental disorder that depends on a person's mood, like a depressed state of mind, loss of interest, loss of pleasure, low energy, guilty feeling, difficulty in sleeping, appetite change, less concentration, and hunger. Moreover, it comes with symptoms of anxiety. This problem can be recurrent or chronic that affects the regular activities of the individual's ability to its extreme stage. It can also lead to self-injury or even suicide [3]. Yearly, around one million people died due to suicide with approximately three thousand suicide deaths happening every day (WHO 2012).

Anxiety is one of the mental disorders, which represents how one's body and mind react to stress, fear, uneasiness or unfamiliar situations. It usually causes restlessness, tension, sweat and rapid heartbeat [22] [23]. Stress is a feeling of emotional tension. Nowadays, as the challenges are growing in every field, anxiety and stress are becoming serious issues, that seriously affects people's wellbeing and additionally causes genuine health problems [2]. It impacts severely on an individual's education outcome, productivity at the workplace and personal relationships. It also impacts society by an increase in crime rate and alcohol abuse.

Psychiatrists find the level of DAS of a patient through testing, examining and other factors like health history. Different measuring scales are available to measure the level of DAS such as DASS-(Depression Anxiety Stress Scales) commonly known as DASS-21 or DASS-42, BDI-(Beck Depression Inventory), BAI-(Beck Anxiety Inventory) [24] [25], HDRS-(Hamilton Rating Scale for Depression) [26] HADS- (Hospital Anxiety and Depression Scale) [27], CES-D-(Center for Epidemiological Studies Depression) [28].

DASS42 scale is selected for our study because people suffering from depression, anxiety, and stress are not actually open to express their psychiatric problems with close friends, relatives, or even to doctors. The dataset is taken from the open-source Internet survey data set. It is originally obtained from the URL (//openpsychometrics.org/_rawdata) [4], which contains the data collected from online questionnaires filled by different users between the years 2017 - 2019. A sample of the online questionnaire used for collecting data is shown in fig 1.

Fig. 1. Sample online Questionnaire for Depression, Anxiety and Stress

To the collected data set, ML algorithms are applied to classify data into different severities such as normal, mild, moderate, severe, and extremely severe (Fig 2). In [5], "Assessment of anxiety, depression, and stress" are done using different ML models and found that Radial Basis Function Network (RBFN) algorithm gives the highest accuracy compared to other algorithms: 96.02% for depression, 97.48 % for anxiety, and 96.17% for stress. This work is indented to enhance and compare the performance of SVM and LR algorithms, based on classification accuracy with other algorithms in [5].

Severity	Depression	Anxiety	Stress
Normal	0-9	0-7	0-14
Mild	10-13	8-9	15-18
Moderate	14-20	10-14	19-25
Severe	21-27	15-19	26-33
Extremely severe	28+	20+	34+

Fig. 2. Rating index of severities.

II. LITERATURE REVIEW

Many researches have carried out work on mental illness using ML algorithms, such as Bayes Net, Naive Bayes, Multilayer Perceptron (MP), RBFN, K-Star, K-Nearest Neighbor, J48, Random Forest and hybrid for classification of DASS42 data. They found that among the mentioned ML models, RBFN gives highest accuracy of 97.48% for anxiety, 96.02% for depression and 96.17% for stress [5].

Tyshchenko. Y used classification algorithms such as Naive Bayes, SVM, Random Forest Tree (RFT) and Convolution Neural Network (CNN) for data posted in the blog [6]. Different encoding techniques were applied to the collected data such as Bag of Words encoding, Term Frequency Inverse Document Frequency encoding and CNN classification algorithm. An accuracy of 78% and recall score as 0.72 were obtained.

The automated depression-detection algorithm is used to detect depression, which utilizes 142 samples of individual suffering from depression and the features extracted from combination of audio and text data is stored in "long short term memory" neural network model. Three levels of experiments are done to predict depression with the feature extracted from audio and text data [7].

To predict anxiety and depression in adult patients, in paper [8], authors collected data from hospital attached with medical college in Kolkata, India. 630 adult individual's samples were collected and applied to different classification algorithms such as Naive Baye's, Random tree, Multiple layer perceptron, J48, Sequential random optimization, K-star, Random sub-space and found that Random Forest gives more accuracy of 91%, 89% precision.

To measure the deference of stress, depression and anxiety in school going adolescents (12-19 years age) and dropouts, kamlesh et al. [9] collected two sets of samples. In the first set a total of 1812 adolescent samples were collected which contained 1054 males and 752 females and 674 were from rural, 1088 from urban with some missing values. The second set contained a total of 120 school dropouts with 65 males and 55 females. To find the accurate difference 498 samples from first set that belonged to same locality were compared with dropouts and found varying result among them.

A computational model is developed to predict Post-Traumatic Stress Disorder and the emergence of depression among Twitter users by Andrew G. Reece et al. collecting data containing depression history and microblog activities of approximately two hundred individuals samples (105 depressed, 99 healthy i.e. 204 samples). A ML model with hidden markov model (HMM) is designed to obtain 31.4% and 24 % high in probability of PTSD [10].

The suicide related stream data is collected from twitter platform by Jingcheng Du et al. [11], and CNN based algorithm was used to build a binary classifier to this dataset. Compared to traditional ML algorithms CNN produced precision of 78% and F-1 83%.

To know the anxiety and depression condition among the seafarers, Sau et al. Screened 470 seafarers through interview and collected socio-demographics and health data. The collected data was applied to ML algorithms such as Naive Baye's, LR, Catboost, Random Forest Tree (RFT), SVM. Among all the algorithms, Catboost gave more accuracy 82.6%, and precision 84.1% [12].

Faneva et al. [13] carried out two tasks for early detection of depression and anorexia by representing text in linguistic feature and text vectorization. These text data were trained using ML model and optimum result were obtained. Yujiao et al. [14] used big data approach to find the connectivity between reading habits and depressive tendency among university students. University library data and mental health questionnaires were collected and compared with ML algorithms and found logistic regression gives higher accuracy.

Sheldon et al. [15] have done a pilot study to analyze DAS level of 90 employees in an industry located at Bangalore to inspect its effects on productivity. Cross-sectional design was used and found that 36% of the workers suffered from anxiety and 18% of them suffered from stress and none of them suffered from depression.

Health related data are not only collected from individuals or from the hospitals but nowadays everyone is connected to social media and share their emotions in it. Saha et al. collected data from online social media posts related to psycho linguistic attributes from 620,000 posts made by 80000 users in 247 online communities. These data are fed into modeling framework to classify the mental problems, specially depression arising in the online communities [16].

To find the presence of DAS in medical students, Shawaz et al. [17] used DASS 42 questionnaire and found 53% suffering from stress, 66.9% were anxious and 51.3% in depressed state. They also identified among all the undergraduates, the 5th semester students were more affected than 2nd semester students and that female students were affected more than the male students. To curb this morbidity, they requested for immediate medical counseling.

III. DATASET COLLECTION AND PREPROCESSING

Data collection plays a vital role in building ML models. There are few possible ways to collect data which purely depends on the type of research carried out. The data collection methods are: observation, interview, document scanning, measurement, questionnaire or combination of these different methods. In our work, data is collected from online questionnaire prepared by author (Lovibond & Lovibond, 1995) [19]. A total of 39,776(thirty nine thousand seven hundred and seventy six) samples are collected from online questionnaires between the year 2017 and 2019 [20].

The dataset contains answers for 42 questions with demographic data. The answers are recorded in the range of 0 to 3. The meaning of the scale values is shown in Fig 3. Final score is obtained after adding all the filled values and it is valid for one week only; users have to assess themselves every week.

scale	Meaning
0	Did not apply to me at all
1	Applied to me to some degree, or some of the time
2	Applied to me to a considerable degree, or a good part of the time
3	Applied to me very much, or most of the time

Fig 3. Rating scale.

DASS42 consists of totally 42 questions. It does not contain any right or wrong answers. Among the 42 questions, 14 questions are categorized for anxiety which evaluate symptoms like the persistent worrying, difficulty in concentrating, situations and events such as threatening and Fatigue. 14 questions are categorized for depression which evaluate symptoms like underestimation of life, criticizing themselves, lost interest, sad, grief, inactivity, and tearfulness. Remaining 14 questions are categorized for stress which evaluate symptoms like difficulty in calmness, Trouble sleeping, sweating, over-reacting and tolerate interrupts [18]. Fig 4 shows all the 42 simple and straightforward questions.

In fig 4, the depression related questions are "q3, q5, q10, q13, q16, q17, q21, q24, q26, q31, q34, q37, q38, q42" (marked as `d`). The anxiety related questions are "q2, q4, q7, q9, q15,q19, q20, q23,q25,q28, q30, q36, q40, q41" (marked as `a`) and the stress related questions are "q1, q6, q8, q11, q12, q14, q18, q22, q27, q29, q32, q33, q35, q39" (marked as `s`).

Scores of DAS are calculated by adding the scores of the relevant questions. Finally, the obtained score is labeled according to the rating index i.e., Extremely Severe, Severe, Moderate, Mild and Normal as in fig 2.

The DASS calculation is shown in figure 5, 6, 7. For depression score calculation, questions 3,5,10,13,16,17,21,24,26,31,34,37,38,42 are considered and the summation is applied horizontally for the values entered (between 0-3) to each question. The total summation value is mapped with severity-rating index that is [0-9] is normal, [10 – 13] is mild, [14 – 20] is moderate, [21 – 27] is severe and 28 and above is extremely severe. Similarly For anxiety and stress score calculation, questions 2,4,7,9,15,19,20,23,25,28,30,36,40,41 (anxiety) and 1,6,8,11,12,14,18,22,27,29,32,33,35,39 (stress) are considered and the summation is applied horizontally for the values entered (between 0-3) to each question. The total summation value is mapped with severity-rating index that is [0 - 7] is normal, [8 - 9] is mild, [10 - 14] is moderate, [15 – 19] is severe and 20 and above is extremely severe for anxiety and [0 - 14] is normal, [15 - 18] is mild, [19 – 25] is moderate, [26 – 33] is severe and 34 and above is extremely severe for stress.

Question	Score	Q.no	Question	S
found myself getting upset by quite ivial things (s)	0 1 2 3	22	I found it hard to wind down (s)	0 1
was aware of dryness of my mouth (a)	0 1 2 3	23	I had difficulty in swallowing (a)	0 1
couldn't seem to experience any ositive feeling at all (d)	0 1 2 3	24	I couldn't seem to get any enjoyment out of the things I did (d)	0 1
experienced breathing difficulty (a)	0 1 2 3	25	I was aware of the action of my heart in the absence of physical exertion (a)	0 1
just couldn't seem to get going (d)	0 1 2 3	26	I felt down-hearted and blue (d)	0 1
tended to over-react to situations (s)	0 1 2 3	27	I found that I was very irritable (s)	0 1
had a feeling of shakiness (a)	0 1 2 3	28	I felt I was close to panic (a)	0 1
found it difficult to relax (s)	0 1 2 3	29	I found it hard to calm down after something upset me (s)	0 1
found myself in situations that made e so anxious I was most relieved when ey ended (a)	0 1 2 3	30	I feared that I would be "thrown" by some trivial but unfamiliar task (a)	0 1
felt that I had nothing to look forward (d)	0 1 2 3	31	I was unable to become enthusiastic about anything (d)	0 1
found myself getting upset rather isily (s)	0 1 2 3	32	I found it difficult to tolerate interruptions to what I was doing (s)	0 1
felt that I was using a lot of nervous iergy (s)	0 1 2 3	33	I was in a state of nervous tension (s)	0 1
felt sad and depressed (d)	0 1 2 3	34	I felt I was pretty worthless (d)	0 1
found myself getting impatient when I as delayed in any way (s)	0 1 2 3	35	I was intolerant of anything that kept me from getting on with what I was doing(s)	0 1
had a feeling of faintness (a)	0 1 2 3	36	I felt terrified (a)	0 1
felt that I had lost interest in just about verything (d)	0 1 2 3	37	I could see nothing in the future to be hopeful about (d)	0 1
felt I wasn't worth much as a person l)	0 1 2 3	38	I felt that life was meaningless (d)	0 1
felt that I was rather touchy (s)	0 1 2 3	39	I found myself getting agitated (s)	0 1
perspired noticeably in the absence of gh temperatures or physical exertion)	0 1 2 3	40	I was worried about situations in which I might panic and make a fool of myself (a)	0 1
felt scared without any good reason (a)	0 1 2 3	41	I experienced trembling (a)	0 1
felt that life wasn't worthwhile (d)	0 1 2 3	42	I found it difficult to work up the initiative to do things (d)	0 1

Fig 4. DASS42 Questionnaires

Q3	Q5	Q10	Q13	Q16	Q17	Q21	Q24	Q26	Q31	Q34	Q37	Q38	Q42	Total score	Severity
3	3	3	3	1	2	1	3	0	3	3	3	2	3	33	Extremely severe
1	3	3	0	0	0	1	3	1	3	0	1	3	3	22	Severe
2	1	2	0	0	2	0	0	2	2	2	1	1	2	17	Moderate
0	1	2	1	0	2	1	1	0	0	0	3	1	0	12	Mild
0	0	2	0	1	0	0	3	0	1	0	3	0	0	9	Normal

Fig 5. Depression- Score Calculation

Q2	Q4	Q7	Q9	Q15	Q19	Q20	Q23	Q25	Q28	Q30	Q36	Q40	Q41	Total score	Severity
3	3	0	0	2	1	0	1	0	3	2	0	1	1	17	Severe
3	3	3	1	2	2	3	3	3	3	3	1	1	2	33	Extremely severe
1	0	2	2	0	1	1	0	1	0	0	1	2	1	12	Moderate
0	2	0	0	0	0	3	0	0	0	1	0	0	1	7	Normal
3	0	0	0	0	0	1	0	2	0	0	2	1	0	9	Mild

Fig 6. Anxiety- Score Calculation

Q1	Q6	Q8	Q11	Q12	Q14	Q18	Q22	Q27	Q29	Q32	Q33	Q35	Q39	Total score	Severity
2	3	3	0	3	0	3	2	3	0	2	3	3	3	30	Severe
2	3	3	1	3	1	2	3	0	1	2	0	2	1	24	Moderate
2	2	0	2	0	0	0	0	0	1	2	3	1	3	16	Mild
3	0	0	1	2	0	0	0	0	1	0	0	0	2	9	Normal
1	2	3	3	2	2	3	3	3	2	3	3	3	3	36	Extremely severe

Fig 7. Stress- Score Calculation

IV. MACHINE LEARNING MODELS

Levels of DAS are predicted using ML model for the five severities. The ML model were build in python using the Scikit-Learn library. Two ML algorithms selected here are SVM and LR. The following subsections describe the working principles of the LR and the SVM algorithms.

A. Support Vector Machine (SVM)

SVM is a supervised ML algorithm used for classification and also for regression. It is mainly used for classification, but sometimes it can be used for regression as well. Basically, it finds optimal hyperplane that creates a boundary between the distinguished data (Represented by blue dots and green squares). In a two-dimensional space, this optimal hyperplane is nothing but a line that separates data points as shown in fig 8. Maximum margin is defined as the maximum distance between the data points [29].

Basically, SVM does not support classifying multiclass data points. It is meant for supporting only binary classification. For classifying multiclass data points, it splits/separates the data point into two classes, hence it breaks down the multiclass classification problem into multiple binary classification problems. To support classification of multiclass data points, One-to-One and One-to-Rest approaches are used.

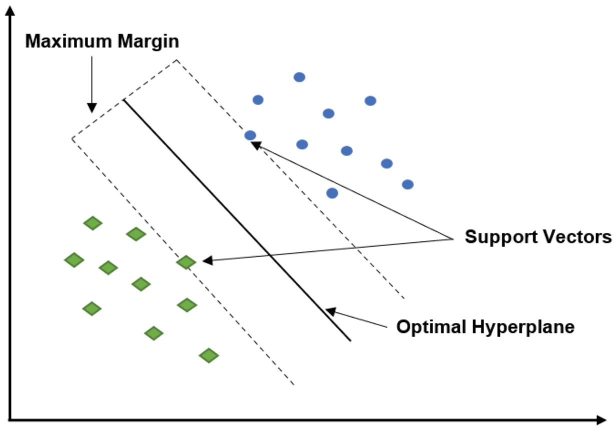

Fig. 8. Hyperplane and Support Vectors of SVM

SVM plots each data item to our dataset in an N-dimensional space, where N is the number of features/attributes in the data. N is 14, i.e. each dataset (for depression, anxiety and stress) consists of 14 features/attributes. Next, it finds the optimal hyperplane to separate the data. Kernelized SVM is commonly used for non-linearly separable data. The kernel function in a kernelized SVM indicates the similarity between points in the newly transformed feature space and for the two given data points in the original feature space. There are various kernel functions available, but Radial Basis Function Kernel (RBF) and Polynomial Kernel are popular [32].

a) RBF Kernel: It is one of the popular kernel functions. It expands and presses the dataset, so that the data are linearly separated and if required it projects the dataset to higher dimensions. It is the default kernel commonly used in SVM and it takes C and gamma as parameters. C is a hyperparameter that is tuned to avoid misclassification while training the model and Gamma is tuned before training. Gamma decides how much curvature is needed in a decision boundary. If Gamma is set to high, more curvature is formed else less curvature is formed. The formula of the RBF kernel is shown in equation 1.

$$\mathbf{K\ (x, y) = exp\ (-\ gamma * sum\ (x - y\ \hat{}\ 2))} \quad (1)$$

Where, x and y are two vectors. The value of gamma ranges from 0 to 1.

b) Polynomial Kernel function: Its output depends on the direction of two vectors. It is commonly used along with SVM. It defines the similarity of vectors (to training data) in a low dimensional space over polynomials of the original variables and allows learning of non-linear models. The formula of the Polynomial kernel is shown in equation 2

$$\mathbf{K\ (x, xi) = 1 + sum\ (x * xi)\ \hat{}\ d} \quad (2)$$

Where, d is the degree of polynomial.

RBF kernel function with SVM shows better performance when compared to Polynomial kernel function with SVM, hence RBF with SVM is used in our work.

B. Logistic Regression (LR)

Logistic Regression (LR) is a supervised ML technique used for predicting the dependent data variable by analyzing an existing set of independent variables. Therefore, the outcome must be either categorical or discrete. It can be either yes or no, 0 or 1, true or False. It is much similar to the Linear Regression [33].

The logistic function is shown in equation 3

$$P(Y = 1/X) = \frac{e^{(b_0 + b_1 x)}}{e^{(b_0 + b_1 x)} + 1} \quad (3)$$

where:-
e, b_0, b_1 are constant value 2.72, constant from data sample, beta-coefficient of data.
On the basis of categories, LR is classified into Binomial and Multinomial LR:

a) Binomial LR: **:** In binomial LR, there are only two possible types of response variables available, such as false or true, 0 or 1. By default, LR is limited to classify only two classes either zero or one type of problem and modeling targets is done using "binomial probability distribution function".

b) Multinomial: In multinomial LR, there are more than two possible types of response variables available, such as 0, 1 or more. To use this LR for multiclass classification problem, the multiclass classification problem should be split into multiple binary classification problems and fit the standard LR model to each sub problem. As our dataset has multiple classes, Multinomial Logistic Regression is performed on it.

V. Results and Discussion

Overview of result obtained are discussed in this section. The implementation of SVM and LR models on a DASS42 dataset are described. It is important to analyze and compare the performance with other algorithms. Every ML algorithm will not work best for all the use cases, as it depends on many factors such as the structure, size and usage type of the dataset. Also, every model gives different performance characteristic.

SVM and LR are applied to mental health states such as DAS and resulting confusion matrices are obtained as shown in fig 12. To visualize the performance of ML models, we make use of confusion matrix. Confusion matrix is taken into consideration because it is simple to predict and display the results of the selected algorithm [30]. In the obtained confusion matrix, rows and columns used to indicate actual and predicted classes. The numbers from 1 to 5 in the matrix represent severities that are from normal to extreme severe as shown in fig 2.

Prince Kumara et al. in [5] worked on nine ML algorithms. They used open source "WEKA" tool to classify depression, anxiety and stress for different severity levels for the dataset [4] taken in a proportion of 75% training and 25% test data. Among all algorithms, they concluded that RBFN algorithm gave highest accuracy: 96.02%, 97.48 % and 96.17% for DAS.

We implemented SVM and LR in Python using the Scikit-Learn library on the same dataset [4] with the same proportion for training and testing. SVM library contains built-in classes for various kernel functions [31].

ML Models	Anxiety	Depression	Stress
Support Vector Machine (SVM)	1 2 3 4 5 1 2419 16 0 0 0 2 9 674 6 0 0 3 0 8 1662 82 0 4 0 0 11 1451 76 5 0 0 0 42 3488	1 2 3 4 5 1 2905 7 0 0 0 2 5 1226 57 0 0 3 0 49 2072 42 0 4 0 0 45 2054 29 5 0 0 0 20 1433	1 2 3 4 5 1 2897 15 0 0 0 2 23 1225 40 0 0 3 0 25 2036 102 0 4 0 0 34 2064 30 5 0 0 0 9 1444
Logistic Regression (LR)	1 2 3 4 5 1 2435 0 0 0 0 2 16 523 150 0 0 3 0 6 1735 11 0 4 0 0 11 1527 0 5 0 0 0 0 3530	1 2 3 4 5 1 2181 39 0 0 0 2 15 867 56 0 0 3 0 16 1686 30 0 4 0 0 26 1601 0 5 0 0 0 2 3425	1 2 3 4 5 1 2905 7 0 0 0 2 5 1226 57 0 0 3 0 49 2072 42 0 4 0 0 45 2054 29 5 0 0 0 20 1433

Fig 12. The confusion matrix for SVM and LR

In order to enhance the accuracy, parameter tuning was done using GridSearchCV and cross-validation. In LR, the best parameters obtained were maximum for iterations of 70, Limited memory Broyden Fletcher Goldfarb Shanno (LBFGS) solver, and multinomial multiclass. In SVM, the best parameters obtained were for maximum iterations of 80, Radial Basis Function (RBF) kernel and gamma value of 1. Obtained best parameters were applied to the models and results were generated. The obtained results are in tabulated form shown in fig 12, it show that LR obtained good

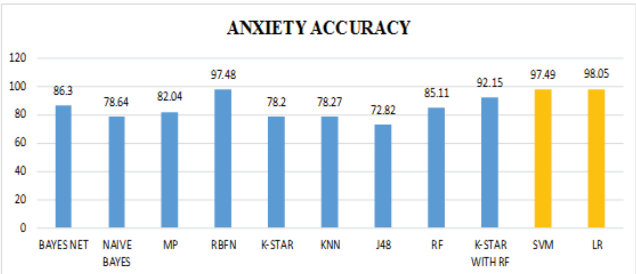

Fig. 9. Accuracy Comparison of Anxiety

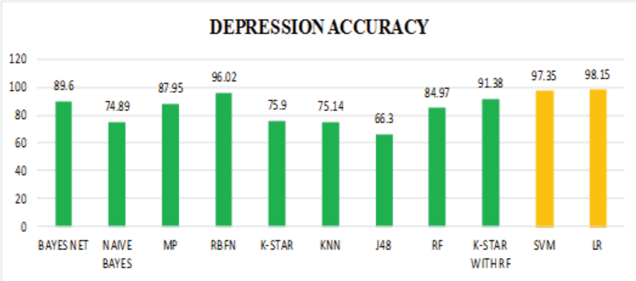

Fig. 10. Accuracy Comparison of Depression

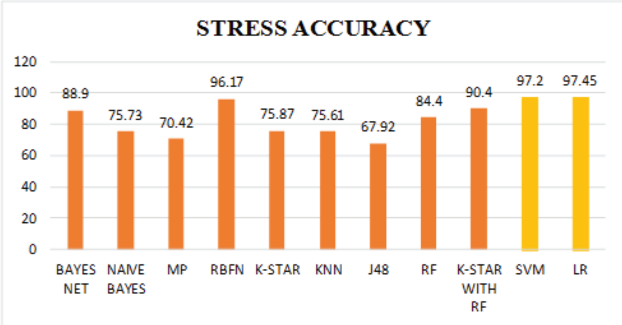

Fig. 11. Accuracy Comparison of Stress

performance for predicting DAS for the DASS42 dataset. Therefore, LR outperforms SVM and other ML models in [5]. Fig 9, fig 10 and fig 11 show the comparison with respect to accuracy in anxiety, depression, and stress of all models respectively.

VI. Conclusions

The main goal of this work was to predict the level of DAS using DASS42. Two classifiers, namely Support Vector Machine (SVM) and Logistic Regression (LR) were selected. These classifiers were implemented using python. We chose the correct configuration of the algorithm for a dataset by tuning the hyper-parameters and obtained the best performance compared to other algorithms. To tune the parameters, in SVM, RBF kernel was used, gamma was set to 1, and 80 maximum iterations were applied whereas to tune in LR, LBFGS solver, multinomial multiclass was used, and 70 maximum iterations were applied. After tuning parameters in these models, LR gave better accuracy than other models. The obtained accuracy rate (98.15% for depression, 98.05% for anxiety, and 97.45% for stress) shows that the LR achieves better performance compared to SVM and other ML models. Mental illness depends not only on psychological facts but also depends on the quality of life, the locality where people live, and other societal factors. The future work is to focus on working with a dataset collected from diverse localities.

References

[1] R.S. Murthy, A. Haden, B.Campanini, editors. "Mental Health: New Understanding, New Hope", World Health Report, Geneva. pp.9, 2001.

[2] G. Mikelsons, M. Smith, A. Mehrotra, M. Musolesi, "Towords deep learning models for psychological state prediction using smartphone data", Challenges and opportunities, 2007.

[3] World Federation for Mental Health (2012), "Depression: a global crisis", World Health Organization. Accessed 4 Dec 2017, Available at:

https://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf

[4] https://openpsychometrics.org/_rawdata/

[5] Prince Kumara, Shruti Garg, Ashwani Garg, “Assessment of Anxiety, Depression and Stress using ML Models”, Procedia Computer Science, 171:1989–1998, 2020.

[6] Tyshchenko, Y., “Depression and anxiety detection from blog posts data.” Nature Precis. Sci., Inst. Comput. Sci., Univ. Tartu, Tartu, Estonia. 2018.

[7] Al Hanai, T., Ghassemi, M. M., Glass, J.R., "Detecting Depression with Audio/Text Sequence Modeling of Interviews." InInterspeech, 1716-1720, 2018.

[8] Sau, A, Bhakta I., "Predicting anxiety and depression in elderly patients using ML technology." Healthcare Technology Letters 4 (6): 238-43, 2017.

[9] K. Singh, M. Junnarkar, and S. Sharma, “Anxiety, stress, depression, and psychosocial functioning of Indian adolescents,” vol. 57, no. 4, 2017.

[10] Reece A. G., Reagan A. J., Lix K. L. M., Dodds P. S., Danforth C. M., Langer E. J., "Forecasting the Onset and Course of Mental Illness with Twitter Data." Scientific reports 7 (1): 13006, 2016.

[11] Du J., Zhang Y., Luo J., Jia, Y., Wei Q., Tao C., Xu, H, “Extracting psychiatric stressors for suicide from social media using deep learning”, BMC medical informatics and decision making, 18 (2): 43, 2018.

[12] Sau, A., Bhakta, I.,"Screening of anxiety and depression among the seafarers using ML technology" Informatics in Medicine Unlocked: 100149, 2018.

[13] Ramiandrisoa F, Mothe J, Benamara F, Moriceau V, "IRIT at e-Risk 2018." E-Risk workshop: 367-377, 2018.

[14] Hou Y, Xu J, Huang Y, Ma X, "A big data application to predict depression in the university based on the reading habits.”3rd IEEE International Conference on Systems and Informatics (ICSAI): 1085-1089, 2016.

[15] S. Rao and N. Ramesh, “Depression, anxiety and stress levels in industrial workers: A pilot study in Bangalore, India,” Ind. Psychiatry J, vol. 24, no. 1, pp. 23–28, 2015.

[16] Saha B, Nguyen T, Phung D, Venkatesh S, "A framework for classifying online mental health- Related communities with an interest in depression", IEEE journal of biomedical and health informatics 20 (4): 1008-1015, 2016.

[17] S. Iqbal, S. Gupta, and E. Venkatarao, “stress, anxiety & depression among medical undergraduate students & their socio-demographic correlates”, pp. 354–357, March, 2015.

[18] Hatton CM, Paton LW, McMillan D, Cussens J, Gilbody S and Tiffin PA, “Predicting persistent depressive symptoms in older adults: A ML approach to personalised mental healthcare”, Journal of affective disorders, 246: 857-60, 2019.

[19] Lovibond, S.H.; Lovibond, P.F., “Manual for the Depression Anxiety Stress Scales”, Psychology Foundation: Sydney, Australia, 1995.

[20] Lovibond, P.F.; Lovibond, S.H., “The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories”, Behav. Res. Ther. 33, 335–343, 1995.

[21] https://www.who.int/publications/i/item/9789241506021.

[22] Chen X, Li H, Zheng X, Huang J. “Effects of music therapy on COVID-19 patients’ anxiety, depression, and life quality: a protocol for systematic review and meta-analysis”, Medicine, Volume 100, Issue 26, pp. 26419, 2021.

[23] https://medlineplus.gov/anxiety.html

[24] Beck, A. T., Ward, C. H., Medelson, M., Mock, J., and Erbaugh, J., “An inventory for measuring depression”, Arch. Gen. Psychiatry, 4: 561–571, 1961.

[25] Beck, A. T., Steer, R. A., and Garbin, M. G., “Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation”, Clin. Psychol. Rev., 8: 77–100, 1988.

[26] Hamilton, M., “A rating scale for depression”, J. Neurol. Neurosurg. Psychiatry, 23: 56–62, 1960.

[27] Kaur, G. H. Tee, S. Ariaratnam, and A. S. Krishnapillai, “Depression, anxiety and stress symptoms among diabetics in Malaysia: a cross sectional study in an urban primary care setting,” 2013.

[28] Radloff, L.S., “The CED-D scale: A self-report depression scale for research in the general population”, Applied Psychological Measurement, 1, 385-401, 1977.

[29] Cortes, Corinna Vapnik, Vladimir N., “Support-vector networks (PDF)”, ML, 20 (3): 273–297, 1995.

[30] R. Andonie, “Hyperparameter optimization in learning systems", J. Mem-brane Comput, vol. 1, no. 4, pp. 279-291, Dec. 2019.

[31] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

[32] Patle, A., & Chouhan, D. S., “SVM kernel functions for classification”, 2013 International Conference on Advances in Technology and Engineering (ICATE), pp 102, 2013.

[33] Tolles, Juliana, Meurer William J, “Logistic Regression Relating Patient Characteristics to Outcomes”, JAMA. 316 (5): 533–4, 2016.