Predicting Mental Illness using Social Media Posts and Comments

Mohsin kamal[1], Saif Ur Rehman khan[2], Shahid Hussain[3]
Anam Nasir[4], Khurram Aslam[5], Subhan Tariq[6], Mian Farhan Ullah[7]*
Department of Computer Science, COMSATS University, Islamabad, Pakistan[1, 2, 3, 4, 6]
Department of Computer Science, University of Oregon, USA[5]
Electrical Engineering Department, Wah Engineering College, University of Wah, Wah Cantt, Pakistan[7]

Abstract—From the last decade, a significant increase of social media implications could be observed in the context of e-health. The medical experts are using the patient's post and their feedbacks on social media platforms to diagnose their infectious diseases. However, there are only few studies who have leveraged the capabilities of machine learning (ML) algorithms to classify the patient's mental disorders such as Schizophrenia, Autism, and Obsessive-compulsive disorder (OCD) and Post-traumatic stress disorder (PTSD). Moreover, these studies are limited to large number of posts and relevant comments which could be considered as a threat for their effectiveness of their proposed methods. In contrast, this issue is addressed by proposing a novel ML methodology to classify the patient's mental illness on the basis of their posts (along with their relevant comments) shared on the well-known social media platform "Reddit". The proposed methodology is exploit by leveraging the capabilities of widely-used classifier namely "XGBoost" for accurate classification of data into four mental disorder classes (Schizophrenia, Autism, OCD and PTSD). Subsequently, the performance of the proposed methodology is compared with the existing state of the art classifiers such as Naïve Bayes and Support vector machine whose performance have been reported by the research community in the target domain. The experimental result indicates the effectiveness of the proposed methodology to classify the patient data more effectively as compared to the state of the art classifiers. 68% accuracy was achieved, indicating the efficacy of the proposed model.

Keywords—*Machine learning; mental disorders; Reddit; Schizophrenia; Autism; OCD; PTSD*

I. INTRODUCTION

Mental disorders are growing sprightly all over the world and World Health Organization (WHO) predicted that one of four people around the world, at some point in their lives will be afflicted with mental disorders. Furthermore, according to Üstün [1], Depressive disorders will become second largest cause of the global disease burden. Around the world, thousands of Men, women, adults and even children are suffering from mental disorders. Mental illness is a disease that negatively impacts the human behavior and thoughts. Subsequently, disturbing an individual's social and domestic life. A Mental disorder has more than two hundred classified forms, in which some of the more common types are bipolar disorder, schizophrenia, depression, stress, anxiety and dementia. There are symptoms that determine either an individual is suffering from mental illness or not. These symptoms include: strong feelings of anger, dramatic changes in eating and sleeping, excessive fears, feelings of extreme high and low, worries, stress, anxieties and suicidal thoughts [2]. People with serious mental illness have a high mortality rate. The authors identified an excessive mortality rate among an individual of serious mental illness, they listed episodic depression and recurrent depressive disorder. The results showed the mortality rate varies with gender, age, diagnosis and ethnicity [3].

The risk of mental disorder misdiagnosing is decreasing and individuals are being diagnosed on time due to the awareness of symptoms and predominantly on introducing Machine Learning (ML) techniques for the diagnosis of mental disorder. Machine Learning techniques are widely used in the medical field, as they provide high accuracy and effective results. Furthermore, with the use of machine learning techniques, the symptoms of mental disorder are predicted and based on the predictions the individual are informed about the condition of their mental health. Pattern recognition algorithm of ML is well-recognized for the treatment, diagnosis and prediction of complications in the treatment of mental disorder [4]. Machine learning acquired deep data or big data for the best results. ML algorithm such as Decision Tree, Support Vector Machine (SVM), Naïve Bayes (NB), Logistic Regression (LG) and k-Nearest Neighbor (KNN) classifiers are used to analyze the state of mental health of a particular group of individuals. All of the ML techniques performed well by achieving an accuracy up to 85% [5] subject to dataset characteristics, which threat their employment in real world scenarios.

The prediction of mental health from individual's social media posts (e.g. Facebook, Twitter, Instagram and Reddit) is one of the encroachments. Social media is a great source of communication and interaction among people. Where they share their opinions and thoughts with each other, such as (posting their photos, videos and comments) which reflect their feelings, mood and sentiments. Hence, their mood and emotions from their posts and comments can be predicted using machine learning algorithms [6]. In [7] the authors analyze the depression on the data gathered from Facebook posts. According to [8] Instagram is another social network, which is widely used all over the world and daily millions of people share their feelings and opinions on it. The authors predicted the markers of depression from Instagram data. They extracted the statistical features from 43,950 individual Instagram photos, such as, metadata component, color analysis

*Corresponding Author

and algorithmic face detection. The machine learning models well performed and predicted early mental disorder and screening.

In this research work, a methodology is proposed for the identification of the mental illness of a patient via their communication on the social media networks. The data for this study were drawn from a well-known social media platform "Reddit". Though, there are several mental illnesses whose data can be collected from social media network. However, data is mine (i.e. post and comments) of certain widely-known mental illness classes such as Schizophrenia, Autism, OCD and PTSD. Furthermore, the proposed model evaluated in terms of accuracy, precision, recall and f-measure. Two research questions are formulated to conduct this study:

Research Question 1: Is the proposed methodology (utilizing XGBoost) in terms of identifying the metal disorders?

Research Question 2: Which is more accurate ML algorithm for the prediction of mental disorder?

The rest of the paper is organized as follows: Section 2 presents the related work, Section 3 discusses about the proposed methodology. In Section 4, presents results and discussion. Finally, Section 5 presents the conclusion.

II. RELATED WORK

Many techniques have been developed and adopted by different researchers for automated prediction of mental disorders such as, schizophrenia, PTSD, OCD and ASD. Schizophrenia is indicated by thought disorder, hallucinations, cognitive deficits and delusions. In the early stage of schizophrenia authentic diagnosis is very important but remains challenging. In study [9] the classification between the control groups and schizophrenic patients was done on basis of magnetic resonance imaging (MRI) using ML. The authors implemented a multimodal classification technique (*structural MRI, Diffusion Tensor Imaging (DTI) and resting state-functional MRI data*) to differentiate drug-naïve first-episode schizophrenic patients from control groups. The authors used feature selection sparse coding (SC) and multi-kernel SVM for the features combination and classification. Similarly, in study [10] analysis of DNA methylation was implemented by the modern machine learning and Bio-conductor Minfi package. Case-controls studies in methylated patterns were successfully detected and highlighted. The classification of schizophrenia patients and healthy control groups was carried out through extracted features of both source-level and sensor-level from EEG signals taped during an auditory oddball task. The results of the study indicated a higher classification accuracy, as the features of both source-level and sensor-level are used together[11]. Some psychiatric diseases such as bipolar and schizophrenia disorder are very contrasted both in etiology and clinical manifestation [12].

Similarly, for the prediction of post-traumatic stress disorder (PTSD) ML also plays a vital role. As, in study [13] machine learning approach was used to forecast the long-term PTSD and risk indicators in soldiers from Afghanistan. The ML approach performed very well and resulted an outstanding performance to analyze the PTSD in soldiers. In [14], PTSD was predicted using supervised ML in order to crack the interchangeable and increasing risk indicator combinations. SVM and 10-fold cross validation were used for PTSD. Furthermore, [15] presents a ML technique to classify between the child PTSD and healthy group. Results were accurate, showing that ML is much reliable and helpful in predicting mental disorders.

Compulsivity disorders such as obsessive-compulsive disorder (OCD) and stimulant use disorder (SUD) are indicated by a loss in behavioral flexibility. As, in [16] OCD and SUD are identified through probabilistic reversal learning (PRL) paradigms. In this work the author applied hierarchical Bayesian model to PRL data of an individual with an OCD, SUD and the healthy controls. According to [17] the effective treatment for OCD is cognitive behavioral therapy (CBT). In this study, the authors gathered resting state functional MRI of patients with OCD before and after 4 weeks of daily intensive CBT. Cross validation was applied to observe the Functional Connectivity (FC) patterns to predict OCD symptoms of an individual. Similarly, in [18] the patients with pediatric OCD ,who acquired to Internet-delivered cognitive behavioral therapy (ICBT) were tested by four machine learning techniques(Random Forest, SVM, linear model and the L1 elastic Net (Lasso)). In study [19], Deep brain stimulation (DBS) treatment was applied on two patients with tractable OCD using automated face analysis. The activation of DBS device and intraoperative implantation was applied on one patient, while the other one was assessed with 3 month post-implantation. Convolutional Neural Network (CNN) was used to quantify positive and negative valence. This technique yield better results in both situations.

Likewise, Autism spectrum disorder (ASD) is a developmental neuropsychiatric disorder characterized by impairments in communication (both verbal and non-verbal), social interaction and restricted, repetitive behavior. According to [20] the diagnosis of ASD is slowed because it requires administration for standardized examination, such as the Autism Diagnostic Observation Schedule (ADOS). Furthermore, it took several hours to analyze 20 to 100 behaviors of an ASD patient through standard approaches of diagnosis. So, to overcome this, the authors applied 17 unique supervised learning methods to classify dataset in five classes. In [21] the authors evaluate video recording from two standard diagnostic instruments in order to design ML classifiers to optimize interpretability, sparsity and accuracy. In [22], the classification of the low functioning of children aged between 2-4 years was possible. The dataset was classified in two classes ASD vs TD by using ML pattern classification technique. This research shows the maximum classification accuracy with 7 features that belong to the goal-oriented of movement of the body parts.

From the above related work it can be concluded that Machine Learning (ML) plays a significant role in automated prediction of mental disorders. Furthermore, the above studies present SVM and NB as the two most accurate algorithms for the classification of mental illness.

III. METHODOLOGY

The proposed model comprises of three phases (Fig. 1). In the first phase, the data were collected from a social network "Reddit". In the second phase, preprocessing was done on the dataset and then feature selection techniques were applied. Finally, in the third phase the model was trained, and was tested.

A. Data Acquisition

Data collection is the first phase in the proposed model. For the data collection purpose social media platform "Reddit" is chosen, and only clinical posts were selected from "subreddits". Subreddits are the pages that are made by the users of "Reddit" for specified topics. Posts from four clinical subreddits are collected: r/Schizophrenia, r/Autism, r/OCD and r/PTSD. Furthermore, 1000 posts from each subreddit is collected using the Reddit Application Programming Interface (API: reddit.com/dev/api). After collection of posts, the comments are gathered and store in another csv file.

B. Data Preprocessing

Data pre-processing is the second phase of the proposed model. Pre-processing activities such as tokenization, removal of stop words, punctuation removal, word stemming and vectorization are applied to the collected data to make the data useful for test model. The chain of preprocessing activities is presented in Fig. 2.

1) Tokenization: Tokenization is the first activity. It is a process to break down the text data into smaller units called tokens. The token may be the symbols, words, numbers, phrases or other meaningful elements. The list of these tokens are input to other preprocessing activities.

2) Stop word: The second activity is the removal of stop words. Stop words are the words which are commonly used in sentences and phrases and have no significant information, such as "the, is, am, are, a, an".

3) Remove punctuations: The third activity is the removal of punctuation. It is the process to remove all punctuations from the data such as:

"!"$#%&'()*+,-./:;<=>?@[\]^_`{|}~", ☺, ☻, ☹ etc.

4) Word stemming: The fourth activity is word stemming. It is the process to compress each word into a common root or base. It involves cutting off the prefixes or suffixes of the word and change it into an inflected word.

In preprocessing, non-alpha characters and spaces are removed. After completing the preprocessing steps, the data convert into the form of vectors.

5) Encoding data: In this activity the target data are encoded in numeric form. As, here a multi-class problem, the available data are categorical data which are in form of string. So, the data are labelled in the form of numeric values, which is understandable for test model.

6) Word vectorization: Word vectorization is another activity in pre-processing phase. It is a process to convert all the stream of text data into numerical feature vectors. As the proposed model only understand the numeric data so, word vectorization is a necessary task in this phase. To convert the text stream of data into numeric data, a famous technique Term Frequency- Inverse Document Frequency (TF-IDF) is being used. TF-IDF is widely used for weighting the text stream to retrieve the information [23]. It implicates, adjusting the frequencies of the data. Frequency depends upon the number of occurrences of a word. If a word occurs several times, then it is assigned a high frequency numerical value. Basically, TF-IDF offers weight for each word. A TF-IDF score is given by the formula:

$$tfidf(t,d,D) = \frac{tf(t,d)}{df(t,d)} \tag{1}$$

A Python library "scikit-learn" is used to apply this scaling and extract features from the data. A maximum of 5000 features is extracted from the data set.

C. Feature Selection

Feature selection is the last activity of the data pre-processing phase. It is an important step to select the significant features from the data set.
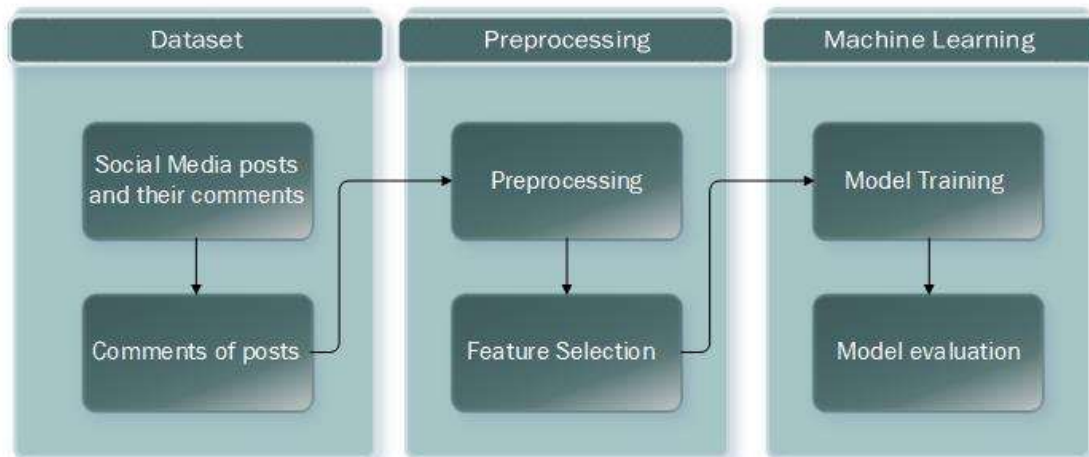

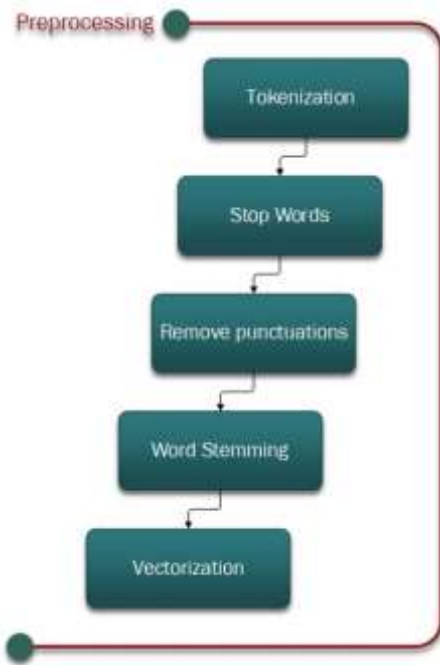Fig. 1. Graphical Overview of Proposed Model.

Fig. 2.   Steps Involved in Preprocessing Phase.

As test data consist of thousands of comments and those comments consist of millions of word stream, so despite of useful features, this word stream also contain many useless features that increase the processing time and create distortion in the data. Hence, it is necessary to remove these useless features and select the only features that are significantly useful. The Random Forest technique is used for this purpose. It is a subsist of 4-12 hundred decision trees. Each of the decision tree is made over a random extraction of analysis from the dataset and a random selection of features. The decision trees are de-correlated and less prostrate to overfitting hence, improving the purity of node which allow the decrement of impurity from all over the tree and selecting the best feature of the data. An automatic feature selection is used and 1000 best features from the data are selected and passed to machine learning model for the classification.

D.   Machine Learning Model

In this research work, XGBoost technique is used in order to classify the comments on the posts. XGBoost technique was proposed in 2016, it is a boosting technique to increase the performance and decrease the processing time. In the proposed model it is used to classify the comments on the posts into four classes i.e. r/Schizophrenia, r/Autism, r/OCD and r/PTSD. Furthermore the accuracy of the model is calculated through F-measure, which is a standard statistic calculation for machine learning classifiers. F-measure is the average recall and precision. Recall states the sensitivity of the machine learning model. It defines the ratio of accurately predicted positive observations to the total no. of the observations presented in the actual class. For example, a good schizophrenia classifier should predict maximum comments from the Schizophrenia subreddit.

$$Recall = \frac{TP}{(TP + FN)} \tag{2}$$

Similarly, Precision is the total ratio of accurately predicted positive observations to the total number of positive predictive observations. For example, when a classifier labels a comment as from the Schizophrenia subreddit then its prediction must be correct.

$$Precision = \frac{TP}{(TP + TP)} \tag{3}$$

The F-measure is usually preferred over accuracy. It is a harmonic mean of Precision and Recall.

$$F1\ Score = 2 * \frac{Recall*Precision}{(Recall + Precision)} \tag{4}$$

The accuracy is calculated as a whole performance of the model.

$$Accuracy = \frac{(TP + TN)}{TP+TN+FP+FN} \tag{5}$$

IV.   RESULTS AND DISCUSSION

In this research work, the dataset has generated about 32330 comments on 4000 posts on clinical subreddit. In each experiment, K-fold (i.e. K=10) cross validation is performed to analyze the effectiveness of the proposed methodology. Four widely used performance measure, namely Accuracy, Precision, Recall, F-measure is used in each experiment.

A.   Response to RQ-1

In order to respond RQ-1, first, an experiment is performed to investigate the effectiveness of the proposed methodology employing the XGBoost classifier. Firstly, the training and cross-validation score in terms of accuracy is shown in Fig. 3. The class wise (in terms of OCD, Autism, PTSD, and Schizophrenia) performance of XGBoost in terms of Precision, Recall, and F-measure is shown in Table I, while the average performance in terms of Accuracy is shown in Fig. 3. Besides, the training Log Loss (>0.69) indicate the low risk in prediction and better performance of XGBoost classifier.
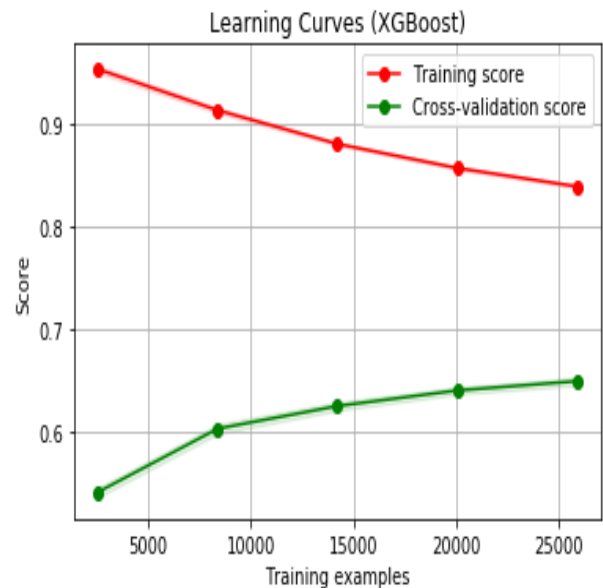

Fig. 3.   Learning curve of XGBoost.

Log-loss is also a measurement of the accuracy that includes the concept of probability assurance. As classification accuracy solely is not enough to analyze the strength of the prediction so, log loss is also a significant measure. Basically, it is a Cross entropy concerning the true labels and predicted probabilities. It is calculated as:

$$Log\ Loss - \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} y_{i,j} log(p_{ij}) \qquad (6)$$

Where N is the number of samples and M is the number of classes, y_{ij} is true label and p_{ij} is predicted probability. In this case, there are four classes. The accuracy achieved is 68% and the log-loss result is shown in Fig. 4.

The results of Fig. 3, Fig. 4, and Table I indicate the efficacy of the proposed methodology to identify the mental illness disease using posts and related comments. The main consequences of the experimental results are as follows:

a) The accuracy of training model gradually decreases when the number of training examples is increased, such as training accuracy with 10000 training examples is higher than 15000 examples with minor differences.

b) Cross-validation accuracy gradually increases with the increase in the number of examples. Such as, cross-validation accuracy with 10000 training examples is higher than 5000 examples with minor differences.

c) Class wise classification performance of XGBoost in terms of F-measure remains better. For example, in terms of F-measure, the better performance XGBoost for OCD (0.58), Autism (0.72), PTSD (0.63) and Schizophrenia (0.70) indicate the efficacy of the proposed methodology for the identification mental disorders.

B. Response to RQ-2.

In order to respond RQ-2, the performance of the proposed methodology (i.e. XGBoost) is compared with other ML classifiers. According to literature, it has been identified that for the text categorization approach SVM and NB are most accurate and commonly employed classifiers for predicting mental illness. Thus, the performance of XGBoost is compared with SVM and NB. The average performance of NB and SVM is shown in Fig. 5 and Fig. 6, respectively. The results of Fig. 5 and Fig. 6 indicates the performance (Training Score) of SVM which remains better than the performance of NB. For example, the performance of NB (i.e. Accuracy = 0.69) and SVM (i.e. Accuracy = 0.74) with 25000 training examples. However, its performance is less than the performance of XGBoost (i.e. Accuracy = 0.82).

Class wise performance of NB and SVM is shown in Table II and Table III.

The results of Table II and Table III indicates that the class wise performance of NB, and class wise performance of SVM. Such as, the performance of SVM (i.e. F-measure = 0.71) is better than NB (i.e. F-measure = 0.70). However, the performance of SVM is less than the performance of XGBoost (i.e. F-measure = 0.72). Table IV shows comparative assessment of XGBoost, NB and SVM in terms of F-measure.

The result in Table IV indicates the best performance of the proposed model XGBoost in terms of F-measures. The results show that the performance of XGBoost for the classification of OCD comments (F-measure = 0.58), Autism (F-measure = 0.72), PTSD (F-measure = 0.63) and Schizophrenia (F-measure = 0.70) remain better than the performance of NB and SVM.
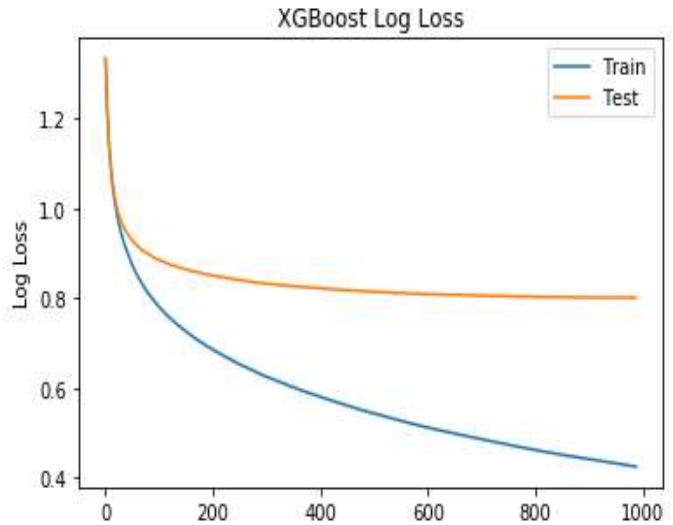

Fig. 4.   Log-Loss of XGBoost.

TABLE I.   PERFORMANCE EVALUATION OF XGBOOST

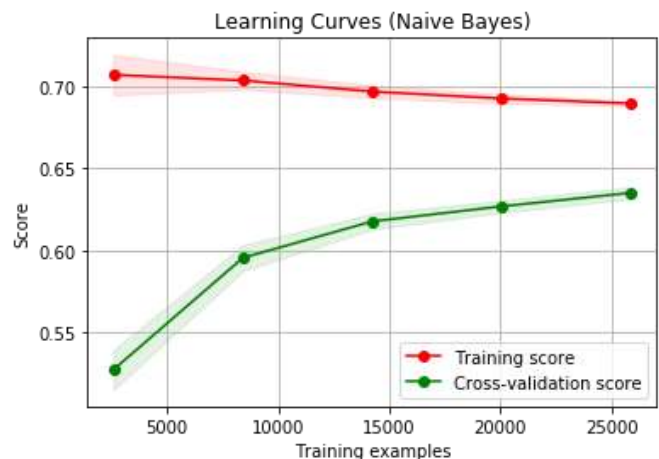| Class	Performance Measures		
	Precision	Recall	F-measures
OCD	0.87	0.44	0.58
Autism	0.69	0.76	0.72
PTSD	0.76	0.54	0.63
Schizophrenia	0.60	0.84	0.70

Fig. 5.   Learning Curve of NB.
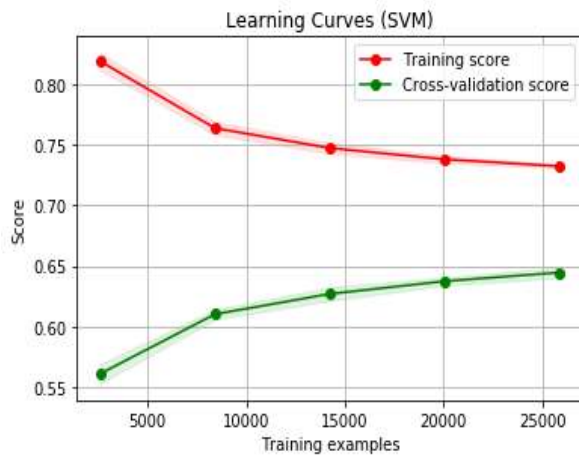
Fig. 6. Learning Curve of SVM.

TABLE II. PERFORMANCE EVALUATION OF NB

Class	Performance Measures		
	Precision	Recall	F-measures
OCD	0.93	0.31	0.47
Autism	0.65	0.76	0.70
PTSD	0.70	0.49	0.57
Schizophrenia	0.52	0.74	0.61

TABLE III. PERFORMANCE EVALUATION OF SVM

Class	Performance Measures		
	Precision	Recall	F-measures
OCD	0.84	0.43	0.57
Autism	0.67	0.75	0.71
PTSD	0.69	0.52	0.60
Schizophrenia	0.55	0.74	0.63

TABLE IV. COMPARISON OF PROPOSED APPROACH WITH STATE OF THE ART CLASSIFIERS

Mental Illness Class	Classifiers Performance		
	NB	SVM	XGBoost
OCD	0.47	0.57	**0.58**
Autism	0.70	0.71	**0.72**
PTSD	0.57	0.60	**0.63**
Schizophrenia	0.61	0.63	**0.70**

V. THREATS TO VALIDITY

In this research work, some treats of using large number of posts and relevant comments are considered that can affect the prediction techniques. The first threat is related to use the limited number of datasets. In this paper, only one dataset is used that is constructed through 32330 comments on 4000 posts. The inclusion of new posts and comments might affect the efficacy the proposed methodology. The second threat is related to comparison of the proposed methodology with the classifiers namely SVM and NB. The finding might be altered when Random Forrest and other widely known outperformed classifiers are reported. Finally, an effectiveness of the proposed methodology is reported by considering the posts and related comments of only four mental illness diseases. The reported effectiveness of proposed methodology might be altered in the existence of posts of more diseases (i.e. class labels in this case).

VI. CONCLUSION

In this paper, a novel ML methodology is proposed to classify the patient's mental illness on the basis of their posts (along with their relevant comments) shared on the well-known social media platform "Reddit". The proposed methodology is exploit by leveraging the capabilities of widely-used classifier namely "XGBoost" for accurate classification of data into four mental disorder classes (Schizophrenia, Autism, OCD and PTSD). The experimental results indicate the effectiveness of the proposed methodology in terms of classification of the mental illness. Though, several mental illness diseases are reported, however, four mental disorder classes are considered, namely schizophrenia, OCD, PSTD, and ASD. A 1000 posts besides related comments are gathered from each of the clinical sub-reddit. Each experiment is performed with k-fold cross validation and used widely known performance measures. The main consequences of the proposed study are as follows: 1) the class-wise performance of XGBosst indicate the effectiveness of the proposed methodology in terms of identifying the mental illness with respect to posts and related comments, Such as the highest F-measure (i.e. 076 for the Autism class and 0.70 for Schizophrenia class); 2) as compared to NB, the XGBoost has made 18.96%, 2.7%, 9.5% and 12.8% improvement in the classification decision for OCD, Autism, PTSD and Schizophrenia, respectively; 3) similarly, as compared to SVM, the XGBoost has made 1.7%, 1.3%, 4.7% and 10% improvement in the classification decision for OCD, Autism, PTSD and Schizophrenia respectively; and 4) the performance of XGBoost remain more significantly better than NB as compared to SVM. In the future work, more classes (i.e. posts of other mental illnesses) can be included to compute the efficacy of the proposed methodology.

REFERENCES

[1] T. B. Üstün, "The global burden of mental disorders," Am. J. Public Health, vol. 89, no. 9, pp. 1315–1318, 1999.

[2] R. C. Kessler et al., "Screening for serious mental illness in the general population," Arch. Gen. Psychiatry, vol. 60, no. 2, pp. 184–189, 2003.

[3] [3] C. Chang et al., "All-cause mortality among people with serious mental illness (SMI), substance use disorders , and depressive disorders in southeast London : a cohort study," 2010.

[4] M. Ernst, J. L. Gowin, C. Gaillard, R. T. Philips, and C. Grillon, "brain sciences Sketching the Power of Machine Learning to Decrypt a Neural Systems Model of Behavior," pp. 1–17, 2019.

[5] M. Srividya, S. Mohanavalli, and N. Bhalaji, "Behavioral Modeling for Mental Health using Machine Learning Algorithms," 2018.

[6] R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, and H. Wang, "Depression detection from social network data using machine learning techniques," Heal. Inf. Sci. Syst., vol. 6, no. 1, pp. 1–12, 2018.

[7] S. F. Behaviors, A. A. Augustine, S. Vazire, D. Ph, N. Holtzman, and S. Gaddis, "Manifestations of Personality in Online Social Networks :," vol. 14, no. 9, 2011.

[8] A. G. Reece and C. M. Danforth, "Instagram photos reveal predictive markers of depression," EPJ Data Sci., 2017.

[9]   H. Zhuang et al., "Multimodal classification of drug-naïve first-episode schizophrenia combining anatomical, diffusion and resting state functional resonance imaging," Neurosci. Lett., vol. 705, no. April, pp. 87–93, 2019.

[10]   B. Torabi Moghadam et al., "Analyzing DNA methylation patterns in subjects diagnosed with schizophrenia using machine learning methods," J. Psychiatr. Res., vol. 114, no. March, pp. 41–47, 2019.

[11]   M. Shim, H. J. Hwang, D. W. Kim, S. H. Lee, and C. H. Im, "Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features," Schizophr. Res., vol. 176, no. 2–3, pp. 314–319, 2016.

[12]   H. G. Schnack, "Improving individual predictions: Machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases)," Schizophr. Res., 2017.

[13]   K. I. Karstoft, A. Statnikov, S. B. Andersen, T. Madsen, and I. R. Galatzer-Levy, "Early identification of posttraumatic stress following military deployment: Application of machine learning methods to a prospective study of Danish soldiers," J. Affect. Disord., vol. 184, pp. 170–175, 2015.

[14]   K. I. Karstoft et al., "Bridging a translational gap: Using machine learning to improve the prediction of PTSD," BMC Psychiatry, vol. 15, no. 1, pp. 1–7, 2015.

[15]   G. N. Saxe, S. Ma, J. Ren, and C. Aliferis, "Machine learning methods to predict child posttraumatic stress: A proof of concept study," BMC Psychiatry, vol. 17, no. 1, pp. 1–13, 2017.

[16]   J. W. Kanen, "Computational modelling reveals contrasting effects on reinforcement learning and cognitive flexibility in stimulant use disorder and obsessive-compulsive disorder : remediating effects of dopaminergic D2 / 3 receptor agents," 2019.

[17]   N. Reggente et al., "Multivariate resting-state functional connectivity predicts response to cognitive behavioral therapy in obsessive–compulsive disorder," Proc. Natl. Acad. Sci., vol. 115, no. 9, pp. 2222–2227, 2018.

[18]   F. Lenhard et al., "Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach," Int. J. Methods Psychiatr. Res., vol. 27, no. 1, pp. 1–11, 2018.

[19]   J. F. Cohn et al., "Automated Affect Detection in Deep Brain Stimulation for Obsessive-Compulsive Disorder," pp. 40–44, 2018.

[20]   S. Levy, M. Duda, N. Haber, and D. P. Wall, "Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism," Mol. Autism, vol. 8, no. 1, pp. 1–17, 2017.

[21]   Q. Tariq, J. Daniels, J. N. Schwartz, P. Washington, H. Kalantarian, and D. P. Wall, "Mobile detection of autism through machine learning on home video: A development and prospective validation study," PLoS Med., vol. 15, no. 11, pp. 1–20, 2018.

[22]   A. Crippa et al., "Use of Machine Learning to Identify Children with Autism and Their Motor Abnormalities," J. Autism Dev. Disord., vol. 45, no. 7, pp. 2146–2156, 2015.

[23]   A. Mahmood and P. Srinivasan, "Twitter bots and gender detection using Tf-idf notebook for PAN at CLEF 2019," CEUR Workshop Proc., vol. 2380, 2019.