A Novel Approach In Twitter Based Mental Disorder Prediction

S.Sinduja[1], Priyadharshini J[2],Pavithra A[3], Pragathi S[4], Vigneshwari S[5]
[1]Assistant professor, Dept of Computer science and engineering
[2, 3, 4, 5]Dept of Computer science and engineering
[1, 2, 3, 4, 5] Vivekanandha College of engineering for women

Abstract- *Personal and ubiquitous sensing technologies such as smartphones have allowed the continuous collection of data in an unobtrusive manner. Machine learning methods have been applied to continuous sensor data to predict user contextual information such as location, mood, physical activity, etc. Recently, there has been a growing interest in leveraging ubiquitous sensing technologies for mental health care applications, thus, allowing the automatic continuous monitoring of different mental conditions such as depression, anxiety, stress, and so on. This paper surveys recent research works in mental health monitoring systems (MHMS) using sensor data and machine learning. We focused on research works about mental disorders/conditions such as: depression, anxiety, bipolar disorder, stress, etc. We propose a classification taxonomy to guide the review of related works and present the overall phases of MHMS. Moreover, research challenges in the field and future opportunities are also discussed.. The health related tweets are collected from the twitter using R tool and twitter Api. The objective of this study is to discover the major disease and cause of death all over the world through the people tweets.The TM-EM (Expectation-Maximization) is the proposed approach used to minimizing the prediction error on topic distributions. The proposed system also suggests the doctors for affected user.*

Keywords- twitter Api, tweets, Expectation-Maximization, R tool.

I. INTRODUCTION

Mental health problems are common worldwide including changes in mood, personality, inability to cope with daily problems or stress, withdrawal from friends and activities, and so on. In 2010 mental health problems were the leading causes of years lived with disability (YLDs) worldwide with depressive and anxiety disorders among the most frequent disorders [1]. Polanczyk et al. estimated a worldwide prevalence of mental disorders in children and adolescents of 13.4% [2]. In the last years, their prevalence has increased even further [3], [4]. Mental disorders can have a serious impact for the patients but also for their families, friends and society, since it is difficult to cope with the implications of someone close having a mental illness.

Dealing with a mental disorder can be physically, economically and emotionally demanding. Work impairment is one of the adverse consequences of mental illness [5] and is also the leading cause for hospital admissions [6]. According to the World Mental Health Survey Consortium, the proportion of respondents who received treatment for emotional or substance-use problems is much larger in developed than in less-developed countries [7]. Nonetheless, the unmet need for treatment of mental disorders is a major problem in both, developed and less-developed countries but being larger in the latter.

Advances in automated data processing, machine learning and natural language processing (NLP) present the possibility of utilizing these massive data sources for public health monitoring and surveillance, as long as researchers are able to address the methodological challenges unique to this media. Numerous studies have been published recently in this realm, including studies on pharma covigilance, identifying smoking cessation patterns, identifying user social circles with common experiences, monitoring malpractice, and tracking infectious dis-ease spread.

The use of social media for health monitoring and surveillance indeed has many drawbacks and difficulties, particularly if done automatically. For example, traditional NLP methods that are applied to longer texts have proven to be inadequate when applied to short texts, such as those found in Twitter.Something seemingly simple, such as searching and collecting relevant postings, has also proven to be quite challenging, given the amount of data and the diverse styles and wording used by people to refer to the topic of interest in colloquial terms inherent to this type of media.

In this paper, we survey novel research works about mental health monitoring systems (MHMS) that make use of sensors and mainly machine learning. We focused on research works about mental disorders/conditions such as: depression, anxiety disorders, bipolar disorder, stress, epilepsy, etc. The goal of this work is to survey relevant work that illustrates the current use of technology (multimodal sensing and machine learning) for automatic and adaptive mental health monitoring.

II. RELATED WORKS

SibulelaMgudlwa and TikoIyamu [1] study can be used to guide integration of social media with healthcare big data by health facilities in the communities. The study contributes to healthcare workers' awareness on how social media can possibly be used to improve the services that they provide to the needy. Also, the study will benefit information systems and technologies and academic domains, particularly from the health services' perspective.Objectives of the study were to examine and gain a better understanding of the complexities that are associated with the use of social media and healthcare big data, through influencing factors, and to develop a framework that can be used to improve health-related services to the patients.

ChonlateeKhorakhun and Saleem N. Bhatti [2] examined the provision of alerts for a remote health monitoring (RMA) application by leveraging an online social media platform (OSMP). Using Facebook as our example platform, we find there are many facilities and features that such an OSMP can offer for RMA alerts. The use of an OSMP allows us to implement communication between actors in a career network which includes the patient and medical professionals. The OSMP also allows alerts to be delivered as Facebook notifications. We have found that the latency of delivery of such alerts is perfectly acceptable over WLAN and 3G, with alerts typically delivered in a few seconds. Our proofof-concept implementation of an alert mechanism shows the feasibility of using OSMPs for alerts.

Michael j. Paul et al [3] describes topics pertaining to the session, "Social Media Mining for Public Health Monitoring and Surveillance,". In addition to summarizing the content of the session, this paper also surveys recent research on using social media data to study public health. The survey is organized into sections describing recent progress in public health problems, computational methods, and social implications. The goal of this session was to create a single venue for cross-disciplinary researchers to present research on social media mining for public health monitoring and surveillance. The session provided a forum to share new research in a variety of important public health areas, including the detection of disease outbreaks and awareness; pharmacovigilance, including interactions withnatural products and dietary supplement

ShaftabAhmed ; M YasinAkhtar Raja [4] designed a knowledge based systems; cyber domain information interchange, data management, security and ubiquitous cloud for telemedicine to enable virtual hospital services have been discussed. The wireless sensor networks for disaster management or emergency assistance with the help of infrastructure-less communication and role of social networking has been presented. The design and development initiatives and projects at hand in these areas have been briefly described.

III. PROPOSED METHODOLOGY

The proposed system is designed to identify the disease affected by people all over the world. This work is done by analysis the real time twitter data. The framework of the proposed system is shown in figure 1.

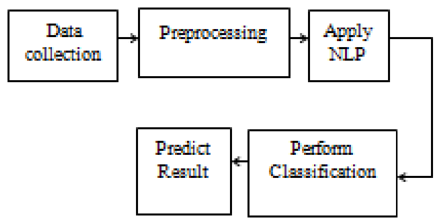

Figure 1. Proposed Framework

Data collection

The data collection is initial process in health monitoring system. The dataset used in this research is twitter health related tweets. These tweets are collected using twitter API and R tool. After api key authentication process the twitter allows to extract the required tweets through R tool.

Preprocessing

The collected tweets contains stop words, special characters, other language tweets etc., these unwanted characters has to be removed before applying clustering algorithm. The stops words such as the, as, you, we etc., has to be removed in order to obtain a better accuracy in classification process.

Natural language processing (NLP)

Usually tweets contain many words that aren't necessary to understand the general idea of the text. These high frequency words like 'a', 'the', and 'of' are called stop words and can be outright ignored in many situations. This is the approach fields like InformationRetrieval (IR) take to reduce the dimensionality of their term spaces and improve performance. It is less useful in the field of text mining, however, because these words can often lend information and clarify semantics.

Stemming (or lemmatization) is also commonly used in IR. To group similar words into one, we reduce words to their stem, or root form. For example, "walking", "walk", "walked", and "walker" will all be reduced to the root word "walk". Although it can be argued that this effect is not as harsh as stop word lists, it can still be harmful to the semantics of the text.

The Expectation-Maximization algorithm

The Expectation-Maximization algorithm is used in maximum likelihood estimation where the problem involves two sets of random variables of which one, X, is observable and other, Z, is hidden. In simpler words algorithm works in following two steps:

E-step

Estimates the expectation of the missing value i.e. unlabeled class information. This step corresponds to performing classification of each unlabeled document. Probability distribution is calculated using current parameter.

M-step

Maximizes the likelihood of the model parameter using the previously computed expectation of the missing values as if were the true ones.

Step 1-

Given: X - Labeled data
Z- Missing values
θ - Unknown parameter
L(θ;X,Z) = p(X,Z|θ) - likelihood function (probability).
L(θ|X) ϵ {α p(X| θ): α >0} , p(X| θ) * p(Z|θ)

Step 2:

With the given variables X, Z and θ, the maximum likelihood estimation of the unknown parameters is calculated by the marginal likelihood of the observed data. The value obtained is not tractable.

Finding maximum likelihood

$$L(\theta;X) = p(X|\theta) = \sum_z p\ (X,Z|\theta)$$

Step 3- calculate expected value of log likelihood function.

$$Q(\theta|\theta^{\wedge}(t)) = \sum_{Z} x\theta^{\wedge}(t)\ [\log L(\theta;X,Z)]$$

Step 4- find the parameters that maximizes this quantity

$$\theta^{\wedge}(t+1) = \arg\max Q(\theta|\theta^{\wedge}(t))$$

The observed data points X, may be discrete which are finite or taken from countably infinite set or it may be continuous which are taken from uncountably infinite set.

The EM algorithm can be applied to other data models when one of the parameters Z or θ is known. This shows the iterative nature of algorithm that continues to predict the values until converges to a particular and specific value.

The EM provides the classification output such as major disease affected by human and their country. The system also provides the result of each disease details with their region. The following section describes the experimental result.

IV. EXPERIMENTAL RESULT

The proposed system is developed in java language. The netbeans IDE is utilized for front end design. NetBeans is an integrated development environment (IDE) for Java. NetBeans allows applications to be developed from a set of modular software components called modules. MYSQL is used for database access.MYSQL is an open source relational database management system (RDBMS).The experiment evaluation of the research is shown in following results.

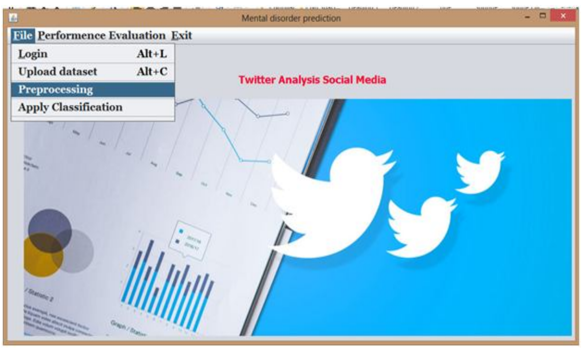

Figure 3. Implementation screen shot

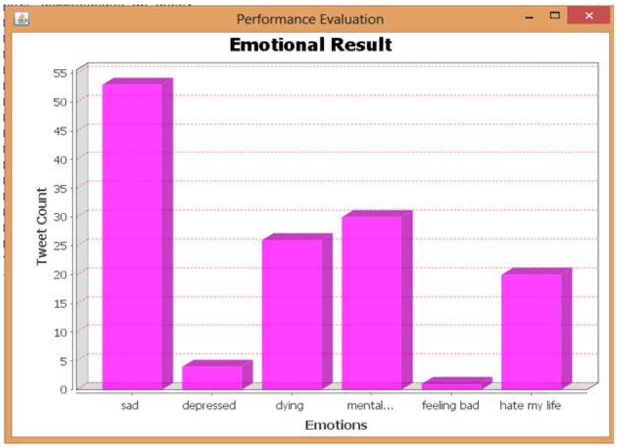

Figure 2. Emotional Result

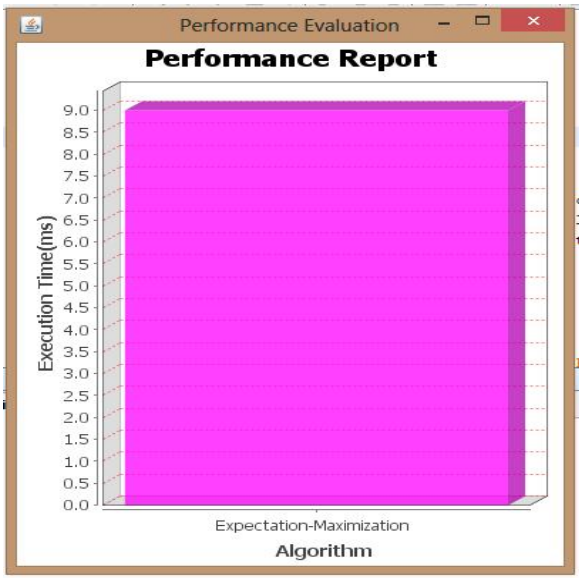

Figure 2. Execution time report

Figure 2 gives the emotional tweets results. The figure 3 gives the implementation screen developed in java language.Figure 4 gives the execution time which is 9 ms.

V. CONCLUSION

In this paper, we surveyed state-of-the-art research works on mental state monitoring with a primary focus on those which use sensors to gather behavioral data and EM to analyze these data. The system identified key characteristics among the reviewed literature and proposed classification taxonomy that we believe, will help new researches in this field to understand the overall structure of such systems. The proposed study presented some of the research challenges of MHMS and future opportunities to advance the field. Based on the surveyed literature, the application of multimodal sensing technologies along with machine learning methods represents a great opportunity in the advancement of providing mental health care technology tools for treatment.

REFERENCES

[1] Khorakhun, Chonlatee, and Saleem N. Bhatti. "Alerts for remote health monitoring using online social media platforms." In *e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on*, pp. 177-181. IEEE, 2013.

[2] Mgudlwa, Sibulela, and TikoIyamu. "Integration of social media with healthcare big data for improved service delivery." *South African Journal of Information Management* 20, no. 1 (2018): 1-8.

[3] Paul, Michael J., AbeedSarker, John S. Brownstein, AzadehNikfarjam, Matthew Scotch, Karen L. Smith, and Graciela Gonzalez. "Social media mining for public health monitoring and surveillance." In *Biocomputing 2016: Proceedings of the Pacific Symposium*, pp. 468-479. 2016.

[4] Whiteford H.A., Degenhardt L., Rehm J., Baxter A.J., Ferrari A.J., Erskine H.E., Charlson F.J., Norman R.E., Flaxman A.D., Johns N., et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010 Lancet, 382 (9904) (2013), pp. 1575-1586.

[5] Polanczyk G.V., Salum G.A., Sugaya L.S., Caye A., Rohde L.A. Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents J. Child Psychol. Psychiatry, 56 (3) (2015), pp. 345-365.

[6] Twenge J.M.Time period and birth cohort differences in depressive symptoms in the us, 1982–2013.

[7] Olfson M., Druss B.G., Marcus S.C. Trends in mental health care among children and adolescents New England J. Med., 372 (21) (2015), pp. 2029-2038.

[8] Moore R.C., Depp C.A., Wetherell J.L., Lenze E.J.Ecolog ical momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults.

[9] Firth J., Torous J., Nicholas J., Carney R., Rosenbaum S., Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? a meta-analysis of randomized controlled trials J. Affect. Disord., 218 (2017), pp. 15-22.

[10] Torous J., Friedman R., Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions JMIR mHealthuHealth, 2 (1) (2014), Article e2