Datasets Creation for Depression Risk Prediction in Educational Environment during Pandemic

Dr.A.Sharada and Manasa Jonnalagadda

*Professor, G Narayanamma Institute of Technology & Science(For Women),
CSE Hyderabad, Telangana
sharada@gnits.ac.in
M.Tech, G Narayanamma Institute of Technology & Science(For Women),
CSE Hyderabad, Telangana
jonnalagaddamanasa7@gmail.com*

Abstract

Emotional well-being of any individual is crucial and predominant to handle everyday stress and work productively. Mental health usually takes a setback during the pandemic, and it is vital to monitor individuals suffering to improve their quality of life. Datasets attained over the years through strategic medical screening and diagnosis of mental health illnesses forms a value asset for understanding the conditions better and help patients cope from it. Datasets play a vital role in enabling Machine Learning and Deep Learning applications to predict and classify results accurately. Specifically, as the explorations of Machine Learning is growing swiftly in the field of health care sector it is essential to utilize high-quality datasets for precise screening, detection, and diagnosis of mental health conditions. Multi modal data such as textual, audio and video datasets designed through National health surveys, clinical data, electronic health records as part of patients screening and treatment process, are frequently used for detection of prediction of depressive disorder. Nevertheless, less research and datasets exist on predicting depression risk in educational environment by considering factors like academic stressors, depression symptoms and panic due to Covid-19. The current study focuses on creating high value datasets with respect to the mentioned factors for depression risk prediction in teachers and students.

Keywords: Datasets, Self-reporting Questionnaires, Mental Health, Depression Risk, Machine Learning.

1.   INTRODUCTION

Mental Health Illness is one of the most serious and prevalent public health problems. In this respect, educational establishments play a critical and strategic role in imparting knowledge sharing, preparing teachers and students for social challenges and acting as a principle environment for their physical and mental well-being. The World Health Organization (WHO) defined mental health as "the condition of a person who can handle his/her stress in life according to his/her ability but is still able to work normally and productively as well as contribute to the society." Many previous studies conducted on mental health issues reveal that emotional well-being of the individual takes a major seatback during the epidemic. COVID-19 pandemic created a global emergency at an alarming rate forcing rapid transmission of functioning across various industries. Rigorous lockdowns, loss of work and shelter, bad financial strain, prolonged social isolation, closure of educational establishments and more such disruptions directed a huge impact on every aspect of human life around the world since 2019. Specifically, educational establishments implemented emergency strategies to introducing virtual learning platforms, altered teaching, and learning styles, accentuated academic and work life stress and uncertainty due to COVID-19. Subsequently, these factors have become a driving force to poor mental health problems, little interest in day-to-day activities, accentuated stress, anxiety and compounded negative emotions.

In clinical terms depression is a mood disorder with constant feeling of emptiness, loneliness, and sadness. Logistics have revealed [22] that an estimate of around 53 million have suffered with depressive mood disorder due to the pandemic and such cases are only ever increasing. It is chronic suffering that can affect children, adolescents, and adults. It contributes to physical, emotional, economic, and social downfall of the individual and if neglected for a long time it may also lead to suicidal deaths. So, it is essential to recognize and diagnose this emotional dichotomy at the earliest and get the needed treatment. Additionally, it is also vital to identify or accurately predict diverse situations negatively affecting the holistic well-being of the individual and manage them.

The advancements in Machine Learning and Deep Learning approaches paired with mental health domain has mainstreamed for screening, predicting, and diagnosing of plethora of mental health illnesses. Extensive research exists on diagnosis and early intervention of depression prediction in elderly and young adults by analyzing longitudinal clinical data obtained from National health surveys and health centers. Many techniques utilize either structured, unstructured, or semi structured data for predicting depression in varying age groups. Factors like age, cognitive functioning, smoking, alcohol addiction, health conditions and so on are commonly considered depression factors in elderly. Similarly, history of substance abuse, sexual orientation, personal relationships, racism and so on are the factors frequently used in depression prediction in young adults. It is observed that the datasets used are often clinical data and electronic health records containing sensitive patient information and is difficult to gain access to as they are not open sourced. It is also noted that obtaining reliable datasets is extremely important for accurate predictions. It is in this context a system is proposed to predict depression risk in teachers and students by employing Multiple Regression and Deep Neural Networks. The datasets used for teachers and students in

the proposed system are created through self-reporting questionnaires in view with the academic stresses and social landscape in educational environment during the pandemic.

The rest of the paper is structured as follows. Section 2 focuses on delineating the previous studies conducted on depression detection with using various datasets and Machine Learning and Deep Learning models. The datasets created for the proposed systems for depression risk prediction in teachers and students addressed in Section 3. Section 4 pertains conclusions and future improvements in creating the datasets for depression risk prediction.

2.   RELATED WORK

Mental health and Depression are one of the most crucial and relevant aspects to address in healthcare. On a broader bandwidth, diagnosis and treatment of depression and mental health illnesses make use of traditional and conventional methods like in person interviews, self-assessing questionnaires, cognitive and behavioral therapies. Depression is particularly challenging to diagnose medically because clinical and self-assessment reports are vastly dependent on subject experts and diagnosis methods; the disorder manifests in varying forms for different people irrespective of age; it is rather time- consuming and labour-intensive.

Accentuated rise in research and applications of Machine learning and deep learning techniques is observed as there is growth in data available and improving computing power. Electronic health records, clinical data collected via National health surveys, personal interviews among others are used to assist in screening, predicting and treatment of depression and other mental health disorders by tracking compounded behavioral patterns of the individuals.

Electronic Health Records (EHR) datasets [21] collected from undergraduate students are utilized in identifying and diagnosis of MDD and GAD with a novel machine learning pipeline. The datasets encompass of features like demographic information, unbalanced diet, blood pressure and other health conditions, substance abuse, housing status and so on. The disorders are assessed by screening the biometric markers and patient characteristics in the questionnaire.

National Health and Nutrition Examination Survey (NHANES) data [27] is used to classify depressive and non-depressive individuals by employing Ensemble binary classifier. National Center for Health Statistics (NCHS) designed the survey to gather large health-related information through PHQ-9 and SF-25 scales from US domestic population. These rich large partial datasets are protected by the federal laws and are for health care research purposes only.

Data driven models [5] are introduced for designing and collecting datasets via multiple questionnaires to identify high risk of depressive disorder. The questionnaires encompassed of demographic information, physical and mental health questions assessing depression indirectly and questions related to depression symptoms formulated via Beck Depression Inventory (BDI) [2] General Anxiety Disorder 7 (GAD-7) [23], Resilience Appraisals Scale (RAS) [16], PHQ-9. Data is collected from college students who have undergone health screening at KAIST clinic Papalardo

center over a period.

Social media is a people's platforms and significant rise in social media encroachments has leveraged the medical health experts to use individuals' posts to diagnose physical and mental illnesses. XBoost classifier is developed for classifying patients' emotional disorders by taking user posts from "Reddit".[17] Detection of suicide risks [1] in patients was enabled by using Reddit posts collected and analyzed textual/sentence data over a year. The features mainly focused while accumulating the data are substance abuse and addictive disorders, depressive disorders and conduct disorders.

Depression is also predicted by systematically analyzing Web of Science (WoS) core data collection designed by mapping knowledge gathered on depression research from 2004 to 2019. The data is visualized in the form of valid records of text files as retrieved after immense research on articles related to depression and major depressive disorder [25].

From the above mentioned studies, it can be observed that multi modal datasets are utilized with various Machine Learning techniques for predicting and diagnosing a plethora of illnesses and emotional conditions in the health care domain. The datasets mainstreamed for depression prediction and mood prediction are either medical data, survey data collected as part of closed research, patients voice or video data. The main challenge for obtaining reliable and high quality datasets is that patients' data is highly sensitive and are henceforth not available in open-source. The other challenge is also the affordability and availability of resources for attaining audio and video data. Electronic health records, clinical data collected via National health surveys, personal interviews among others are used to assist in screening, predicting and treatment of depression and other mental health disorders by tracking compounded behavioral patterns of the individuals. However, less research exists in predicting depression risk in teacher and students by considering academic stressors and pandemic factors in the educational environment. As such, there is a need for designing suitable datasets for students and teachers by including diverse situations effecting the individual along with depression symptoms.

3.   DATASETS CREATION FOR DEPRESSION RISK PREDICTION

Datasets have always been an integral part of machine learning for smooth and fast training of the models for attaining precise results. For depression predictions, analysis, and diagnosis, it is important to generate large-scale, well-balanced datasets. Clinical data generated through simulation treatments and diagnosis of patients; survey data collected by various health organizations in different forms are commonly utilized for implementing deep learning models. However, it is difficult to gain access to these health records as they are sensitive and private.

Furthermore, due to scarcity of rich open- source datasets for depression in teachers and students considering the academic stressors and the current outbreak, datasets have been created. In this study, the datasets for both teachers and students are primed separately by designing a questionnaire for depression risk in reference to the standard depression scales and surveys conducted by healthcare for mental health analysis in recent times. The respondents were asked to answer the questions keeping in view their

mental state for past one month. Datasets for both teachers and students are corroborated by the Psychologist to achieve accurate predictions for the same.

3.1   Teachers' Dataset

The pandemic has impacted the mental state of teachers with high stress levels since the beginning of the global crisis [3] [7]. Academic stress, adapting to online teaching, isolated lockdowns has accentuated teachers stress levels which was often accompanied by anxiety and depression symptoms. Not many studies were conducted during Covid-19 to scale the symptoms of stress, anxiety, and depression among teachers and so it is important to do and make appropriate predictions with deep learning and machine learning. Teacher's questionnaire is designed in reference to the standard self-reporting screening scales such as: Beck's Depression Inventory (BDI) which is a 21-item, 'self-report rating inventory that measures characteristic attitudes and symptoms of depression'; The Self-Efficacy Scale consisting of 10 items for measuring 'self-efficacy of the individual in correlation with emotions and work satisfaction'; Canada Mental Health Survey during Pandemic to 'assess the general mental health of teachers by with the survey focused on questions related to panic due to the outbreak, depression symptoms and burnout due to academic stress'. The questionnaire contains 50 close ended questions categorized into eight sections based on factors influencing depression risk outlined as below:

a)   Section I & II:

Section I gathers demographic information and Section II focuses on the academic stress that teachers undergo in terms of workload, managing curricular and classroom responsibilities and result oriented teaching as shown in Table 1.

Table 1: Section II in Teachers' Questionnaire

	Always	Occasionally	Rarely	Never
Burdened with work				
Time spent on class preparation & online teaching				
Continuous shifts in work				
Juggling between teaching & other curricular responsibilities				
Learning new teaching style				
Result oriented teaching				

b)   Section III:

This section emphasizes on emotional well-being of teachers regarding classroom expectations, providing support to the teachers and monitoring them as presented in Table 2.

Table 2: Section III of Teacehers' Questionnaire

	Not at all	Mildly	Moder ately	Sever ely
Meeting classroom expectation				
Providing emotional support for students				
Maintaining positive & energetic attitude				
Managing student behavior & multiple responsibilities				
Students focus & response while teaching				
Students Regularity				
Relation with students				

c)　　Section IV:

As indicated in Table 3, this section attempts to identify the stress induced due to workload, professional growth during the pandemic and work satisfaction. Self-efficacy in professional life that assesses how much people believe they can achieve their goals, manage academic work and classroom ambience despite difficulties.

Table 3: Section IV in Teachers' Questionnaire

	Not at all	Mildl y	Mode rately	Seve rely
Overall professional growth from past one year				
Upholding professional standards				
Colleagues Support for sharing workload				
Learning and adjusting to technology				
Job contentment during current situations				
Delivering values that have to be imparted in students' lifetime				
Focus on applicatory skills & conceptual knowledge				

d)　　Section V & VI:

These sections as shown in Table 4 attempt to measure professional relationship with management, financial aid and emotional burnout of teachers caused due to personal and work- related activities in everyday life.

Table 4: Section V & Section VI in Teachers' Questionnaire

	High	Moderate	Low	Nil
Relation with management				
Support from your department & college board				
Performance recognition, Promotions				
Financial Incentives				

	Not at all	Mildly	Moderately	Severely
Balancing Personal & Professional Life				
Meeting professional & personal expectations				
Manage time achieving personal & professional goals				
Managing student behavior & multiple responsibilities				
Spending quality time with family				
Being motivated & confidant to complete daily tasks				
Constant support from family and friends				

Table 5: Section VII in Teachers' Questionnaire

	Never	Sometimes	Frequently	Always
Health. Stress in professional, personal life has negative effects on my health.				
Sleep. You're getting sufficient sleep every night.				
Nutrition. You're eating regularly to improve concentration and performance.				
Hydration. You're drinking water while working and staying hydrated throughout the day.				
Movement. You're setting aside time to be physically active.				
Exhaustion. You're feeling energetic or a little tired while working throughout the day.				

e)   Section VII:
This section outlines on assessing depression symptoms like sleep pattern, energy levels, diet, physical activities, and overall health condition of teachers as given in Table 5.

f)   Section VIII:
The last section attempts to project questions allied to effects of pandemic to assess the strain and mental tensions instigated in the individual that affects their motivation and concentration in performing any task meritoriously as in Table 6.

Table 6: Section VIII in Teachers'/Students' Questionnaire

	Nearly everyday	**More than half days**	**Several days**	**Not at all**
Feeling nervous or anxious				
Worry about its effects on your health and safety				
Worry about its effects on your family's health and safety				
Worry about its effects on job & financial health				
Restlessness				
Irritability				
Hopelessness				
Feeling frustrated				
Feeling afraid as if something awful might happen				
Losing interest in doing things				
Guilt & grief due to death of loved ones				
Trouble concentrating on anything				
Feeling bad about not being able to help your family or community				
Loneliness				

3.2   Students Dataset
Students' questionnaire is designed in reference to the standard self-reporting screening scales such as: Beck's Depression Inventory (BDI) which is a 21-item, The Self-Efficacy Scale consists of 10 items, Academic Stress Scale comprises of 40 items developed in 'Order to adopt to the Indian conditions to assess the efficiency of behavioral programs in managing academic stress in improving academic performance'; USA Mental Health Survey during pandemic to analyze depression of students with the survey focused on questions related to panic due to the outbreak, depression symptoms and unparalleled challenges faced by the students during COVID-19. This study presents a structured questionnaire arranged into eight sections described below:

a)   Section I & II:

Demographic information is collected in Section I & Section II outlines on how online learning and digital classes effect in accordance with academics and on gaining knowledge, shown in Table 7.

Table 7: Section II in Students' Questionnaire

	High	Moderate	Low	Nil
Overburdened with classwork				
No. of sitting hours during online classes and daily classwork				
Online learning with regards to theory courses				
Online learning with regards to practical courses				
Online learning with regards to the assessment and exams				
Overall subject knowledge gained during online classes				

b)   Section III:

The focus of the section as presented in Table 8 is on the academic stress faced by the students in the educational environment in terms of personal adequacy, academic learning, and motivation to learn.

Table 8: Section III in Students' Questionnaire

	Not at all	Mil dly	Mode rately	Sever ely
Preparedness of topic				
Maintaining positive and energetic attitude				
Managing daily classwork				
Focus and response during online classes				
Regularity of attending online classes				
Relation with teachers				
Doubts clarification during online classes				

c)   Section IV & V:

These sections emphasized on the availability of resources for attending virtual classes and assess how lack of any of the resources given in table could be frustrating. It also attempts to analyze effects of pandemic in career planning and job opportunities as projected in Table 9.

eTable 9: Section IV and Section V in Students' Questionnaire

	Not at all	Mildly	Moderately	Severely
Irregular internet				
Unavailability of physical resources(laptop, desktop)				
Non Conducive environment at home				
Lack of daily schedule				
Little help from your classmates/ friends to complete your work				
Lack of access to study materials required for course preparation				

	Not at all	Mildly	Moderately	Severely
Overall career planning from past one year				
Effectiveness of training and development programs provided by college				
Internships and/or Job opportunities provided by college during current situations				
Job offers during pandemic compared to last year				
Growth in technical skills and communication skills				
Support from faculty for career growth				
Motivation for training and development				

d)   Section VI:

This section attempts to analyze the self-efficiency and life skills of students to achieve their goals in both academic and personal life despite any difficulties and challenges as reflected in Table 10.

Table 10: Section VI in Students' Questionnaire

	Not at all	Mildly	Moderately	Severely
Balancing academics and Personal Life				
Meeting personal & academic expectations				
Manage time in achieving personal and academic goals				
Spending quality time with family				
Being motivated and confident to complete daily tasks				
Support from family and friends				

e)   Section VII:

This section presented in Table 11 highlights on assessing the depression symptoms effect the individual's daily schedule such as emotional exhaustion, physical activities, sleep pattern and impact of stress on overall health.

Table 11: Section VII in Students' Questionnaire

	Never	Some times	Freque ntly	Alw ays
Health. Stress in professional, personal life has negative effects on my health.				
Sleep. You're getting sufficient sleep every night.				
Nutrition. You're eating regularly to improve concentration and performance.				
Hydration. You're drinking water while working and staying hydrated throughout the day.				
Movement. You're setting aside time to be physically active.				
Emotional Exhaustion. You're feeling energetic or a little tired while working throughout the day.				
Personal Goals. You're motivated and are able to meet personal goals.				

f)   Section VIII:

The last section attempts to project questions allied to effects of pandemic to assess the strain and mental tensions instigated in the individual that affects their motivation and concentration in performing any job diligently same as teachers given in Table 6.

Teachers' and students' questionnaires are comprised of 50 questions each with multiple choice questions applied on 4-point Likert's scale. In both the datasets, the input variables have categorical data, and the target variable is depression risk. As all the questions in the google forms are mandatory to answer, the datasets created are structured eliminating the possibility of NULL values. A total of 163 samples for teachers and 255 samples for students are collected from educational institutes in Hyderabad which are then synthesized with CTGAN to generated large datasets and are stored in Comma Separated Values (CSV) files.

4.   CONCLUSIONS

Reliable and quality datasets are vital for predicting and classifying depression risk accurately. In this study, datasets for teachers and students are collected by conducting a survey in various educational institutes of Hyderabad distributed through google forms. Each questionnaire contains 50 questions prepared based on the emotional state of individuals leading to depression risk. A 4-point Likert's scale is applied on each

categorical variable. The dataset contains variables impelling depression risk such as daily workload, challenges faced in adapting to new teaching styles, professional growth, balance of personal and professional life, effects of stress on health, sleep and eating habits and effects of pandemic imposed on the individual. This self-reporting questionnaires contain categorical data and are labelled based on CES-D. Subsequently, they are applied with Multiple Regression for feature selection and deep learning model to predict risk of depression in teachers and students by classifying depression risk into no risk, mild risk, moderate risk, and severe risk. This approach of creating datasets has saved time, money and was not labour-intensive. It is observed that many teachers experience stress and emotional burnout due to difficulty in managing class and students during online classes, financial strain, low job satisfaction and physical health conditions. Similarly, students experience emotional exhaustion, high pressure, and stress due to academic strain, less self- confidence and adequacy while performing personal and academic tasks and social isolation. It is critical to predict risk of depression in early stages to potentially avoid prolonged mental imbalances, reduce the escalation of the disorder and improve quality of life. The questionnaires can however be improved to also include psychiatric stressors and bio-markers effecting the individual which are also important for analyzing depression risk.

REFERENCES

[1]   Alambo, Amanuel & Gaur, Manas & Lokala, Usha & Kursuncu, Ugur & Thirunarayan, Krishnaprasad & Gyrard, Amelie & Sheth, Amit & Welton, Randon & Pathak, Jyotishman. Question Answering for Suicide Risk Assessment Using Reddit, 2019.

[2]   Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. An inventory for measuring depression. Arch. general psychiatry 4, 561–571 (1961).

[3]   Cachón-Zagalaz, J., Sánchez-Zafra, M., Sanabrias-Moreno, D., González-Valero, G., Lara-Sánchez, A. J., and Zagalaz-Sánchez, M. L. (2020). Systematic review of the literature about the effects of the COVID-19 pandemic on the lives of school children. Front. Psychol. 11:2457.

[4]   Chintalapudi N, Battineni G, Amenta F. Sentimental Analysis of COVID-19 Tweets Using Deep Learning Models. Infectious Disease Reports, 2021.

[5]   Choi, B., Shim, G., Jeong, B. et al. Data- driven analysis using multiple self-report questionnaires to identify college students at high risk of depressive disorder. Sci Rep 10, 7867, 2020.

[6]   Coppersmith, Glen & Leary, Ryan & Crutchley, Patrick & Fine, Alex. Natural Language Processing of Social Media as Screening for Suicide Risk. Biomedical Informatics Insights, 2018.

[7]   Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020a). COVID-19 and learning loss—disparities grow and students need help. McKinsey & Company,

December, 8.

[8]   Di Carlo D.T., Montemurro N., Petrella G., Siciliano G., Ceravolo R., Perrini P. Exploring the clinical association between neurological symptoms and COVID-19 pandemic outbreak: A systematic review of current literature. J. Neurol. 2021;268:1561– 1569.

[9]   De Choudhury M, Counts S, Horvitz E (2013) Social media as a measurement tool of depression in populations.

[10] Du, J. et al. Extracting psychiatric stressors for suicide from social media using deep learning. BMC Med. Inform. Decis. Mak. 18, 43, 2018.

[11] Fatima I, Mukhtar H, Ahmad HF, Rajpoot K (2018) Analysis of user- generated content from online social communities to characterise and predict depression degree. J Inform Sci 44(5):683– 695.

[12] Geraci, J. et al. Applying deep neural networks to unstructured text notes in electronic medical records for phenotyping youth depression. Evid. Based Ment. Health 20, 83–87, 2017.

[13] Gkotsis, G. et al. Characterisation of mental health conditions in social media using Informed Deep Learning. Sci. Rep. 7, 45141, 2017.

[14] Harapan H, Itoh N, Yufika A, Winardi W, Keam S, Te H, Megawati D, Hayati Z, Wagner AL, Mudatsir M. Coronavirus disease 2019 (COVID-19): A literature review. J Infect Public Health. 2020 May;13(5):667-673. Epub 2020 Apr 8. PMID: 32340833; PMCID: PMC7142680.

[15] J. Gaun, Artificial intelligence in healthcare and medicine: Promises, ethical challenges and governance, Chin. Med. Sci. J., vol 34, no. 2, pp. 76-83, Jun. 2019.

[16] Johnson, J., Gooding, P. A., Wood, A. M. & Tarrier, N. Resilience as positive coping appraisals: Testing the schematic appraisals model of suicide (sams). Behav. Res. Ther. 48, 179–186 (2010).

[17] kamal, Mohsin & Khan, Saif Ur Rehman & Hussain, Shahid & Nasir, Anam & Aslam, Khurram & Tariq, Subhan & Ullah, Mian. (2020). Predicting Mental Illness using Social Media Posts and Comments. International Journal of Advanced Computer Science and Applications. 11. 607-613.

[18] Kuo ES, Vander Stoep A, Herting JR, Grupp K, McCauley E. How to identify students for school-based depression intervention: can school record review be substituted for universal depression screening?. J Child Adolesc Psychiatric Nurs.2013;26(1):42-52

[19] Lin, H. et al. Detecting stress based on social interactions in social networks. IEEE Trans. Knowl. Data En. 29, 1820–1833, 2017.

[20] Minkos M.L., Gelbar N.W. Considerations for educators in supporting student learning in the midst of COVID-19. Psychol. Sch. 2021;58:416–426.

[21] Nemesure, M.D., Heinz, M.V., Huang, R. et al. Predictive modeling of depression

and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. Sci Rep 11, 1980 (2021).

[22] Prevalence of anxiety and depression following COVID-19 among healthcare students in Jordan and its effect on their learning process: a national survey.

[23] Spitzer, R. L., Kroenke, K & Löwe, B. A brief measure for assessing generalized anxiety disorder: the gad-7. Arch. internal medicine 166, 1092–1097 (2006).

[24] Teacher Mental Health Survey, Canadian Teachers Federation, Canada, 2020.

[25] Wang Hui, Tian Xuemei, Wang Xianrui, Wang Yun., Evolution and Emerging Trends in Depression Research From 2004 to 2019: A Literature Visualization Analysis, Frontiers in Psychiatry, volume 12. 2021