A Machine Learning Perspective on Emotional Dichotomy during the Pandemic

Dr.A.Sharada[1], Manasa Jonnalagadda[2]

[1]*Professor, Computer Science & Engineering, G Narayanamma Institute of Technology & Science, Telangana, India*
[2] *M.Tech Student, Computer Science & Engineering, G Narayanamma Institute of Technology & Science, Telangana, India*

---***---

Abstract - *Mental health is a stabilizing force of an individual's emotional well-being, and any distress can cause imbalances in one's conventional routine and plethora of mental disorders. Mental health concerns usually took a backseat during the pandemic and impacts seamless functioning for teachers and students in educational environment. Depression is a mental health condition manifesting constant elevation or lowering of person's mood and little interest in everyday activities causing substantial impairment in everyday life. Depression in particular is influenced by complex array of factors including everyday stress, academic strain, compounded negative emotions and panic due to COVID-19 outbreak. Research conducted in healthcare domain in par with Artificial Intelligence provides various methods for detection and diagnosis of depression. However, minimal research is conducted predicting depression based on individual's situation and their environment in early stages. The objective of this study is to propose a context aware model for teachers and students for predicting risk of depression in educational framework and pandemic. The datasets are created through structured self-reporting questionnaires and potential variables for depression risk are identified with Regression analysis. Related context information is extracted in relevance with each potential variable and Convolutional Neural Networks is applied for depression risk prediction. Subsequently, accuracy of the proposed model for teachers and students is evaluated with performance metrics and comparative analysis of Multiple Regression and Convolutional Neural Networks.*

Key Words: **Mental Health, Depression Risk, Convolutional Neural Networks, Multiple Regression, Machine Learning.**

1. INTRODUCTION

COVID-19 is a global humanitarian cataclysm that has left the world in shambles over the recent years. In India, it has enforced rapid transition in education, IT, healthcare, and other sectors, to digitize and implement various strategies for their seamless functioning [10]. Specifically, schools and colleges were forced to run emergency online learning/classes causing prolonged social isolation and increasing academic stressors on both teachers and students. It also had major impact on everyone's life, disturbing individual's conventional activities along with their physical and mental health. Mass fear and uncertainty has reflected disparaging effect in holistic well-being of a person steering strong emotions like stress, anxiety, anger, depression, and other complex array of factors. Work stress, difficult financial situation, family issues, personal and professional problems, changes due to the COVID-19 and other psychological and environmental parameters originating from an individual's way of life contribute to distress and mental health disorders.

According to World Health Organization (WHO), depression ranks high among common mental debilities. Depression is a mental health condition manifesting constant elevation or lowering of person's frame of mind and loss of interest in daily activities causing substantial impairment in everyday life. Given the current shifts in the educational landscape over the past years, depression has become increasingly common in teachers and students in India. It is an emotional dichotomy found in various strata of the society and in different age groups. Parameters like complete burnout, extreme work strain due to academic and curricular responsibilities in teachers; and academic stressors, peer and societal pressure in students could afflict the individual's ecosystem. Thus, a souring need rises to support the emotional well-being of teachers and students by predicting depression risk in preliminary stages to potentially reduce the escalation of the illness and in turn improve their quality of life.

Machine Learning and Deep Learning based mental health explorations [11] have attracted lot of attention to predict mental disorders using multimodal data like text, images, and videos. Approaches like Deep Neural Networks (DNN) and Regression has opened a new frontier to address early screening, detection, prediction, and diagnosis of various disorders by tracking compound emotional parameters associated with the mental health challenges. The statistical and computational methods extended by Machine Learning assist in constructing robust automated prediction and detection of depressive symptoms with the ability to learn and train from data. Multimodal data relying on frequent measurements of depression status procured from various sources have been implemented with deep learning models for early recognition of depression symptoms in the individuals. However, minimal research exists for classifying and predicting individual's emotional state based on their situation and circumstances which is context aware. The proposed system emphasizes on predicting risk of depression in teachers and students based on potential variables affecting the depression risk in pandemic and educational environment along with their context information. Real time data is collected for teachers and students via structured questionnaires and is further synthesized with Conditional Tabular Generative Adversarial Networks (CTGAN) model. Regression analysis is employed to identify potential variables influencing depression risk/prone predictor (DPP) along with their context information. Convolution Neural Networks (CNN) is then designed to compute the values of the potential variables using their related context information followed by calculation of DPP to predict risk of depression for both teachers and students. The objectives of the proposed work are:

- The study is to identify depression risk based on workforce, emotional well-being, academic stressors, professional growth, and other factors affecting during the pandemic in teachers and students.
- To study and identify the context information related to the potential factors prompting depression risk.
- To study the depression risk occurring in both teachers and students during pandemic by taking real-time data.

The rest of the paper is organized as follows. Section 2 focuses on outlining the previous studies conducted on mental health problems and depression detection with using various Machine Learning models and Deep Learning models. The proposed model for depression risk prediction in teachers and students is addressed in Section 3 by utilizing Multiple Regression and CNN. Section 4 presents the results and performance evaluation followed by conclusions and future extensions in Section 5.

2. RELATED WORK

2.1 Mental Health, Depression, and its Impact in Recent Times

Emotional well-being is one of the most vital concerns in health care sector with accentuated indication of its influence worldwide. WHO defined mental health as “complete cognitive, emotional and social behavior of the individual.” Depression is conceded by WHO as the single largest contributor to health-related constrictions like disturbed quality of life, compounded negative emotions, cognitive and psychological imbalances, and even suicidal deaths. Irrespective of age, people with depression tend to hide their interwoven emotions and suffer in silence due to social stigma in the society leading to prolonged delay in finding necessary help needed. Drastic changes over the past years due to COVID-19 pandemic has fueled intricate health, societal, economic, and educational changes around the world in teachers and students. [6], [17]. On a broader spectrum, identifying mental imbalances make use of traditional methods (face-to-face interviews and self-reporting questionnaires) which are interminable and labour-intensive. Additionally, clinical diagnosis of depression is difficult because the illness itself manifests in diverse ways and diagnosis is highly dependent on specialists’ expertise. Thus, integration of traditional methods and technology such as wearable sensors provide a way for periodic monitoring and treating the patients with mental health problems. Nonetheless, development of automated detection and prediction with objective evaluation was highly desirable to complement traditional methods and medical diagnosis.

2.2 Research Contributions to Mental Health Predictions

Incorporating health care and technology is beneficial in discovery and prediction of various physical and mental health conditions making it possible to provide proper treatment and care needed for the individuals. The advancement of Machine Learning algorithms and Deep Learning techniques have been bound to offer new applications for learning and identifying mental health symptoms, building models for prediction and detection for disease progression and improving performance of the developed models. Various statistical, computational and reinforcement techniques have been effective in enabling automated detection and prediction of depression. Growth in availability of data alongside enhancements to computing power has directed to a rise in research and applications deep learning techniques.

Machine Learning algorithms like Random Forest [9], Naive Bayes [13] and Support Vector Machine (SVM) [7], [28] are commonly used techniques in many previous studies to detect symptoms of major depressive disorder using datasets encompassing behavioral patterns of patients. Alongside of these techniques, Logistic Regression, K-Nearest Neighbor and Neural networks are used for extracting predictive features and identifying general anxiety and depressive disorder. In study conducted in the mental health domain, Nemesure et al [19] aimed to detect Major Depressive Disorder and Generalized Anxiety Disorder by collecting samples from undergraduate students through general health screening and psychiatric assessment. Use of XBoost classifier became the first known proposed system for predicting the two disorders by enabling the model to learn and train from electronic health records data. Although the system showed potential its predictive validity by detecting unknown psychiatric diagnosis, the general health screening process may not have covered all cases within the population.

Similarly, large partial datasets of health survey data through PhQ-9 and SF-25 scales were taken as input to an Ensemble binary classifier [25] for identifying depression. The classifier was to be stable and robust while predicting depressed cases based on correlation between quality-of-life scale and depressive symptoms. However, reliability and sensitivity of the developed systems are to be assessed on added datasets.

Data driven approaches were introduced for modeling a system for diagnosis of depressive disorder during the pandemic. Supervised and unsupervised learning alongside of clinical data saved time and cost in data collection which is essential while designing a data driven model [4]. Predictive modeling for inferring depression became popular with multimodal data. Large datasets of audio and video data acquired from Reddit posts with Deep learning models such as Recurrent Neural Networks (RNN) [1] is presented for prediction and diagnosis of suicide risks. Similarly, multimodal data collected through question answer mechanism was utilized to mood disorders and identify the individual's mental state of mind. Autoencoders and Long Short Term Memory were employed for this purpose. First model extracted features from facial expressions and speech response whereas the latter model analyzed the temporal information of the stimulated responses. Autoencoders and Long Short Term Memory are also utilized to prevail over misdiagnosis of bipolar disorder as unipolar disorder [24]. Small datasets acquired through audio and video data are extracted from feature selection methods like Regression and Support Vector Machine was implemented for classifying depression [21]. Apart from this, Regression analysis and CNN are combined to automatically assess depression and foresee depression severity [29] from human behavior. The model created two sequence descriptors with the help of gaze directions and facial action primitives. The results of this model attained from training the system on AVEC 2016 DAIC-WOZ database, achieved significant progress compared to previous state-of-art in terms of estimating depression severity.

From all these previous studies it can be observed that Machine Learning and Deep Learning techniques are integrated, or a single technique is employed for prediction and diagnosis. Often, detection and diagnosis were determined by considering symptoms of mental health disorders. The data streamlined for depressive disorder prediction was clinical data, patients' voice or visual data. Acquiring clinical data for analysis is a challenging task as hospitals and clinics do not disclose patients' records due to privacy and data sensitivity. Availability of resources, specifically equipment needed for attaining audio and video data becomes difficult if they are not within the affordable range. As such, it is necessary to consider the diverse situations along with depressive symptoms for prediction depression risk in pilot stages and provide the help needed for the individuals to cope.

3. DEPRESSION RISK PREDICTION

Depression is an ongoing problem, and it can last for weeks, months, or years. The proposed solution determined depression risk or DPP in teachers and students during COVID-19 in educational environment using Machine Learning algorithm and Deep Learning models. It is important to recognize and address the harmful effects of dejected emotional state of the individual to avoid long term complications in academic and personal life. The proposed Depression Prone/Risk Predictor is developed in four steps as shown in Fig-1.

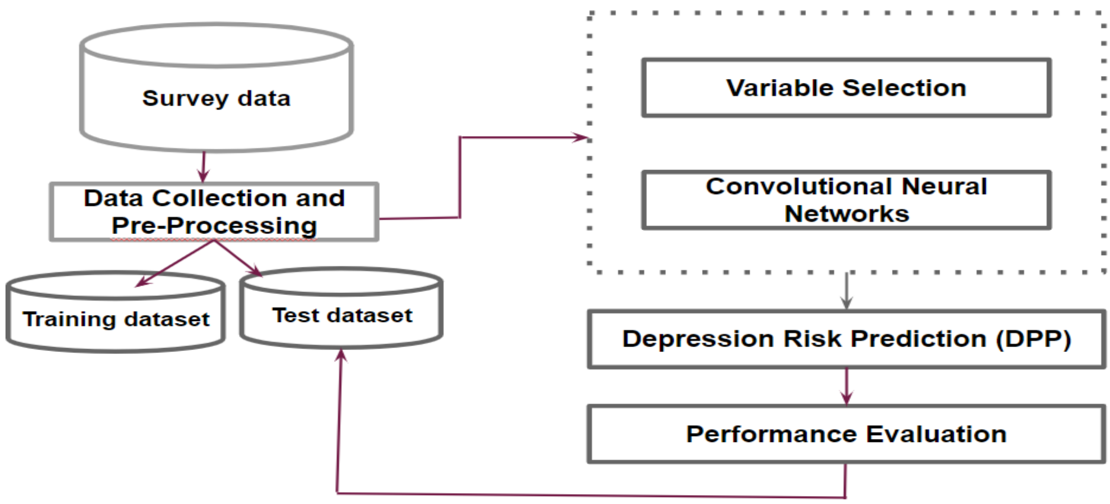

Fig-1: Architecture for Depression Risk Predictor

3.1 Data Collection and Pre-Processing

Clinical datasets or data collected through health surveys are generally used as depression datasets. However, due to the scarcity of open sourced datasets for depression in teachers and students, datasets on basis of global pandemic and academic factors have been created. In the first of proposed system, primary data is collected through structured questionnaires for both teachers and students. The self-reporting questionnaire are designed separately for teachers and students in reference to standard depression scales such as Patients Health Questionnaire (PHQ-9), Beck's Depression Inventory (BDI), Canada Mental Health survey during the pandemic to assess general mental health for teachers [26] and USA Mental Health Survey during pandemic to analyze depression of students [2]. The questionnaires contain 50 questions categorized into 8 sections on the basis of factors influencing the depression risk. Both the questionnaires are validated by the Psychiatrist and real time data is collected by conducting a survey through google forms in various educational institutes across Hyderabad city. The survey is conducted during the time of stringent lockdowns by keep in view emotional state of the individual for past one month. The sampled data is further synthesized using CTGAN. The model is well defined for formulating data for categorical data in tabular format based on frequency of values in each tabular column. Teachers and students Datasets are preprocessed and are split into 80% training data and 20% testing data.

3.1.1 Teachers' Dataset

Teachers' 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. For example, for the variable 'Balancing personal and professional life' values are ordered as: not at all exhausting, mildly exhausting, moderately exhausting, and severely exhausting. This self-reporting questionnaire contains categorical data and is labeled based on CES-D. This labeled data is used to generate a dataset of 10,000 samples using CTGAN which are stored in a Comma Separated Values (CSV) file. The generated data is evaluated against the real time data to attain a similarity measure and to evaluate the model. An evaluation of 0.73187 was achieved indicating that the synthesized data was 73.18% like the real sample data. The sample dataset for teachers in CSV file is shown in Table-1.

Table-1: Teachers' Dataset

S.No	Attribute/Variable	Name	Value
1	x1	Workload	1: Never 2: Rarely 3: Occasionally 4: Always
2	x2	Class Preparation & Online Teaching	1: Never 2: Rarely 3: Occasionally 4: Always
3	x3	Working Hours	1: Never 2: Rarely 3: Occasionally 4: Always
….	….	…...	….
49	x49	Helplessness	1: Not at all 2: Several days 3: More than half the days 4: Nearly everyday
50	x50	Loneliness	1: Not at all 2: Several days 3: More than half the days 4: Nearly everyday

3.1.2 Students' Dataset

Students' dataset contains variables influencing depression risk such as academic workload, challenges faced in adapting to online learning, career planning, balance of personal and professional life, effects of stress on health, sleep and eating habits, social interactions and effects of pandemic forced on the individual. For example, for the variable 'Career planning during pandemic' values are ordered as: not at all satisfying, mildly satisfying, moderately satisfying, and severely satisfying.

This self-reporting questionnaire contains categorical data and is labeled based on CES-D. This labeled data is used to generate a dataset of 10,000 samples using CTGAN which are stored in a CSV file. The generated data is evaluated against the real time data to attain a similarity measure and to evaluate the model. An evaluation indicating that the synthesized data was 76.89% like the real time dataset was achieved. Table-2 shows sample dataset for students in CSV file.

Table-2: Students' Dataset

S.No	Attribute/Variable	Name	Value
1	x1	Overburdened with Workload	1: Nil 2: Low 3: Moderate 4: High
2	x2	No of sitting hours	1: Nil 2: Low 3: Moderate 4: High
3	x3	Effectiveness of Online Learning	1: High 2: Moderate 3: Low 4: Nil
....	....	......	....
49	x49	Helplessness	1: Not at all 2: Several days 3: More than half the days 4: Nearly everyday
50	x50	Loneliness due to Social Isolation	1: Not at all 2: Several days 3: More than half the days 4: Nearly everyday

3.2 Multiple Regression with Backward Elimination

In the second step, Multiple Regression with Backward Elimination is implemented for identifying potential variables significantly affecting depression risk in teachers and students. In this approach, a significance level α is fixed separately for teachers and students. It indicated the probability of rejecting the true null hypothesis. Multiple Regression model is fit with all the features and input features with highest p-value are identified. If the feature with highest p-values is greater than α, the feature is eliminated from the dataset. The model is fit again with this new dataset and the process is repeated until highest p-value from all the remaining features in the dataset is less than the significance level thus eliminating less important or noteworthy features.

3.2.1 Potential Variables Identified for Teachers

For teacher's dataset, a significance level is 0.1 is chosen and out of 50 input features ten potential variables of depression risk have been identified. Table-3 demonstrates the result of regression analysis with p-value <=0.1. 'Std Error' represents the standard error, and 't-value' is a t-test statistical value indicating the statistics of the influence of variables.

Table-3 Potential Variables for Teachers' Dataset

Variable	Description	Coeff (estimate)	Std Error	t-value	P > |t|
x4	Academic & other curricular responsibilities	-0.012423	0.008	-1.500	0.134
x5	Method of Instructions for teaching	0.020720	0.008	2.663	0.008
x10	Managing students' behavior	0.017433	0.008	2.155	0.031
x14	Professional growth during Pandemic	0.013284	0.008	1.698	0.090
x17	Adapting to online teaching platform	0.015574	0.008	1.880	0.060
x27	Managing time to achieve personal & professional goals	0.014560	0.009	1.631	0.103
x33	Nutrition: Overeating or poor Appetite	0.011402	0.008	1.406	0.130
x38	Concern about one's health & safety protocols	0.018196	0.009	1.996	0.046
x41	Restlessness	-0.012795	0.009	-1.352	0.127
x48	Losing interest in work	0.015779	0.008	1.943	0.052

In addition to this regression equation is formed. The probability of data on depression risk is generated as output which is the y-intercept. Each of the selected potential variables for teachers is multiplied by the estimated value, and then the multiplied results are all added up to obtain the probability information on depression risk in teachers.

$$DPP = (2.4697) + (0.0124*x4) + (0.0207*x5) + (0.0174*x10) + (0.0132*x14) + (0.0155*x17) + (0.0145*x27) + (0.0114*x33) + (0.0181*x38) + (-0.0127*x41) + (0.0157*x48) \quad (1)$$

3.2.2. Potential Variables Identified for Students

For student's dataset, a significance level=0.01 is selected and out of 50 input features seven potential variables of depression risk have been identified. Table-4 displays the result of Multiple Regression with p-value <=0.01. 'Std Error' represents the standard error, and 't-value' is the t-test statistical value indicating the statistics of the influence of variables.

Table-4 Potential Variables for Students' Dataset

Variable	Description	Coeff (estimate)	Std Error	t-value	P > |t|
x6	Preparedness of the topic	0.027871	0.011	2.566	0.010
x15	Support from friends/classmates	0.027745	0.012	2.869	0.018
x20	Job offers during pandemic	0.034288	0.012	2.375	0.004
x30	Effects of stress on physical and mental health	0.038418	0.012	3.144	0.002
x37	Nervousness and Anxiety	0.031912	0.012	2.685	0.007
x40	Worry about financial health	0.023554	0.010	2.354	0.019
x33	Nutrition: Overeating or poor Appetite	0.011402	0.008	1.406	0.130
x41	Restlessness	0.032338	0.012	2.649	0.008

Regression equation is also attained for predicting depression risk for students in the last step of the proposed system.

DPP = (0.5209)+(0.0278*x6)+(0.0277*x15)+(0.0342*x20)+(0.0384*x30)+(0.0319*x37)+(0.0235*x40)+(0.0323*x41) (2)

3.3 Identifying Context Information

In the penultimate step of the proposed solution, related context information for each of the potential variable is identified by implementing Multiple Regression with Backward Elimination described in the previous step. The potential variables obtained in the previous step are used as the dependent variable and the relevant context information is obtained in high relation with the potential depression variables.

3.3.1 Context Information for Teachers

A significance level of 0.1 and relevant context is identified for each of the 10 potential variables. For example, the potential variable x48: 'Losing interest in work' is associated with context information variables: x2: Class preparation and teaching time, x7: Meeting classroom expectations, x31: Effects of stress on physical and mental health and x38: Concern about one's health and safety protocols. The results of this step are presented in Table-5.

Table-5 Context Information for depression Risk in Teachers

Potential Variable	Related Context Information
x4	x2, x16, x19, x23, x28, x32, x39, x47
x5	x7, x11, x21, x22, x43, x49
x10	x1, x26, x40
x14	x31, x37
x17	x25, x28, x36
x27	x21, x26, x37
x33	x29, x40, x47
x38	x7, x16, x23, x24, x26, x43
x41	x8, x19, x25, x44, x50
x48	x2, x7, x31, x38

3.3.2 Context Information for Students

A significance level of 0.01 and relevant context is identified for each of the 7 potential variables. For example, the potential variable x20: 'Job offers during Pandemic' is associated with 3 context information variables: x17: Career planning, x38: Concern about one's health and safety protocols and x43: Feeling down and hopeless. Table-6 projects results of this step.

Table-6 Context Information for depression Risk in Students

Potential Variable	Related Context Information
x6	x12, x21, x22, x23, x42, x44, x45, x46, x48
x15	x5, x16, x19, x28, x39, x44, x48
x20	x17, x38, x43
x30	x2, x25, x39, x44, x48
x37	x2, x9, x12, x23, x25, x32, x38, x39, x42, x45, x46, x48
x40	x2, x5, x12, x13, x16, x17, x19, x22, x25, x27, x38, x39, x43, x44, x45, x46, x49
x41	x1, x9, x22, x42, x46, x48

3.4 CNN & Prediction of Depression Risk

CNN models are primarily developed for dealing with image classification problems in which the model learns to represent two-dimensional data referred to as feature learning. The same procedure can be harnessed on one-dimensional categorical data in which the model extracts features from sequence of observations. In the last step of the proposed alternative, CNN model is implemented on categorical data with The Rectified Linear Unit (ReLu) as the activation function and Adam optimizer. Once the CNN model is constructed, the model is trained and compiled to better fit the system and then evaluated against the test data to predict the output values. Flatten and Dropout are also added to the model to improve its performance.

3.4.1 Prediction of Depression Risk for Teachers

CNN model is designed with context variables as input and the value of potential variable as the output. 10 potential variables are identified for depression risk and each one is used for a CNN. 10 CNNs are learned separately to restrict the number of hidden layers and reduce the time for learning. The output of each CNN is substituted in the regression equation (1) presented to predict the degree of depression. Fig-2 shows the context prediction process for the DPP.

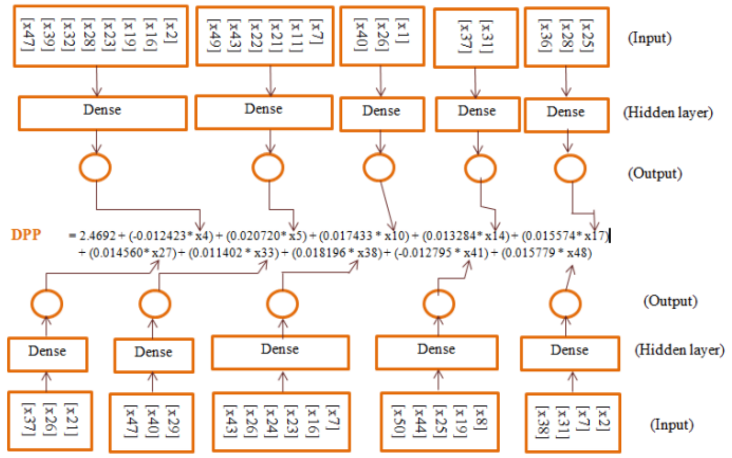

Fig-2: Proposed CNN process for Depression Risk Prediction in Teachers

3.4.2 Prediction of Depression Risk for Students

Each of the 7 potential variables identified for depression risk are used with CNN model. 7 CNNs in total are learned separately to restrict the number of hidden layers and cut back the time for learning. For the connection of the learned CNNs, the regression equation is applied. The output of each CNN is substituted in the regression equation (2) to predict the degree of depression and Fig-3 shows the context prediction process for the DPP.

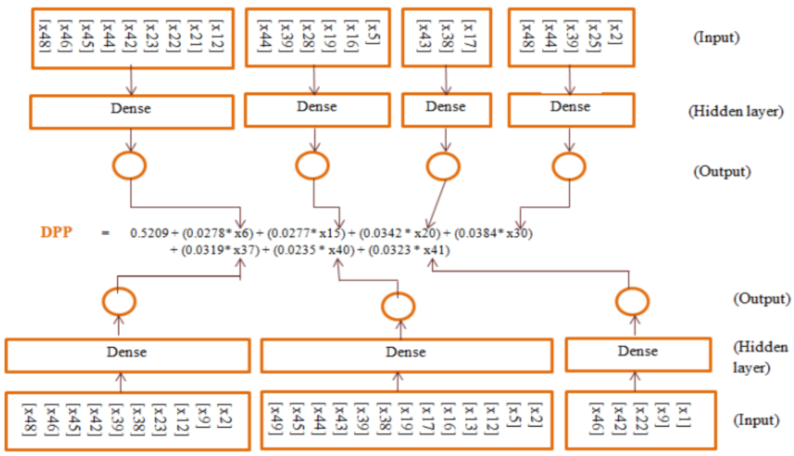

Fig.-3: Proposed CNN process for Depression Risk Prediction in Teachers

4. RESULTS AND PERFORMANCE MEASURES

Machine Learning algorithms and Deep Learning techniques have been formulated for the prediction of depression risk in teachers and students. The proposed CNN model uses dynamic context information for depression risk prediction and individual learning promotes learning in a brief time. The variables identified through CNN are substituted in regression equation for predicting depression risk. The predicted depression risk for students and teachers has been categorized into 4 stages: no risk, mild risk, moderate risk, and severe risk on the basis of CES-D [20] of National Mental Health Center. The performance of teachers' and students' module is evaluated in terms of goodness-of-fit, accuracy and residual error. A comparative analysis with General CNN, Multiple Regression and the proposed CNN model are executed to predict the depression risk independently for both teachers and students.

From the available performance metrics, for Multiple regression RSE and Multiple R^2 score is used for evaluating the results obtained. Residual Standard Error, RSE is defined as "the difference between predicted value in the regression model and the actual data." Multiple R^2 is the "variance rate of dependent variables" which can be explained by the regression model. In the General CNN, the performance measures chosen are Loss function and Accuracy. Loss function is used to assess the "difference between the data obtained by learning and real data. Larger the value is, the more the learned data is inconsistent with the real data." Mean Absolute Percentage Error (MAPE) and Accuracy are the metrics for evaluating the performance of proposed CNN model. MAPE can be considered as a loss function to define the error termed by the model evaluation. Using MAPE accuracy of the model can be estimated in terms of the differences between the actual and estimated values.

4.1 Results and Performance Evaluation for Teachers

Teachers' module was implemented with dataset containing 10,000 samples and 50 input variables. The target variable, DPP was predicted by variable selection reducing the input features from 50 to 10. Related context information for each of these 10 potential variables is identified using regression analysis. CNN is modeled for each potential variable to find its value with related context information as input to the deep learning model. The DPP was predicted by computing its value in the regression equation (1) along with these 10 potential variables and their context information.

Table-7 Comparative Analysis and Performance Evaluation for Teachers

Depression Risk Prediction for Teachers		
Multiple Regression	**RSE**	0.7305
	Multiple R^2	0.0094
CNN	**Loss**	1.4742
	Accuracy	0.3890
Proposed CNN Model	**MAPE**	0.3047
	Accuracy	0.6952

From Table-7, it is observed that the proposed CNN module displayed a best performance. In regression analysis to predict teachers' depression risk, a RSE was measured to be about 0.7 indicating a larger difference between predicted values and actual values. In addition, proposed model is observed to have a lesser value of loss and higher accuracy than general CNN. This comparative analysis shows that the proposed CNN model has better performance with accuracy of 0.69 than multiple regression analysis and general CNN in terms of depression risk for teachers.

4.2 Results and Performance Evaluation for Students

Students' module was implemented with dataset containing 10,000 samples and 50 input variables. The target variable, DPP was predicted by variable selection reducing the input features from 50 to 7 variables. These 7 potential features along with their related context information are fit to the proposed CNN and the value of DPP was predicted using the regression equation (2).

Table-8 Comparative Analysis and Performance Evaluation for Students

Depression Risk Prediction for Students		
Multiple Regression	**RSE**	1.0476
	Multiple R^2	0.0408
CNN	**Loss**	1.2519
	Accuracy	0.3751
Proposed CNN Model	**MAPE**	0.4386
	Accuracy	0.6113

From Table-8, it is observed that the proposed CNN model displayed a best performance. In regression analysis to predict students' depression risk, RSE was measured to be about 1.04 indicating a larger discrepancy between predicted value and actual value. The proposed model is observed to have a lesser value of loss and higher accuracy than general CNN. This comparative analysis shows that the proposed CNN model has better performance with accuracy of 0.61 than multiple regression analysis and general CNN in terms of depression risk for students.

5. CONCLUSIONS AND FUTURE ENHANCEMENTS

The proposed system for predicting depression risk in teachers and students was developed using Multiple Regression for feature selection and CNN for prediction. A questionnaire based data was collected independently for both teachers and students. For teacher's module 10 potential variables and for students' module 7 potential variables are selected and related context information is identified followed by depression risk prediction. An accuracy of 69.52% for teachers and 61.13% for students is achieved for depression risk prediction. It is observed that depression risk prediction for teachers' module has better performance and results in comparison with depression risk prediction for students' module. The real time data collection is challenging as perception and ideology about mental health differs in individuals. Time is a limiting factor for conducting survey for wide range of respondents. In addition, synthesized data was used along with real time data to measure depression risk accuracy. The proposed system can be extended further significantly increasing the model accuracy by accumulating more real time data and implementing different Machine Learning techniques. The model can be enhanced to predict stress levels, anxiety, and other mental health conditions. Lastly, the model can be enhanced by designing a front end system with end to end integration to focus on individual predictions.

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] Basheti IA, Mhaidat QN, Mhaidat HN Prevalence of anxiety and depression during COVID-19 pandemic among healthcare students in Jordan and its effect on their learning process: A national survey. PLoS One, 2019.

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

[4] 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.

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

[6] 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. 2021, Neurol.; 268:1561–1569.

[7] De Choudhury M, Counts S, Horvitz E social media as a measurement tool of depression in populations, 2013.

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

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

[10] 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.

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

[12] 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 (5):667-673.

[13] Hassan AU, Hussain J, Hussain M, Sadiq M, Lee S Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. 2017, IEEE, New York.

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

[15] 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

[16] Lin, E. et al. A deep learning approach for predicting antidepressant response in major depression using clinical and genetic biomarkers. 2018, Front. Psychiatry 9, 290.

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

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

[19] 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, yr 2021.

[20] Radloff LS The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. 1977 Applied Psychological Measurement;1(3):385-401.

[21] S. M. Alghowinem, T. Gedeon, R. Goecke, J. Cohn and G. Parker, Interpretation of Depression Detection Models via Feature SelectionMethods, in 2020, IEEE Transactions on Affective Computing.

[22] S. Song, L. Shen and M. Valstar, Human Behaviour-Based Automatic Depression Analysis Using Hand-Crafted Statistics and Deep Learned Spectral Features, 13th IEEE International Conference on Automatic Face & Gesture Recognition, 2018.

[23] Stough, Laura & Baker, Lynn. Identifying Depression: In Students with Mental Retardation. Teaching Exceptional Children, 1991.

[24] Su, M. H., Wu, C. H., Huang, K. Y. & Yang, T. H Cell-coupled long short-term memory with l-skip fusion mechanism for mood disorder detection through elicited audiovisual features. 2019, IEEE Trans. Neural Netw. Learn.

[25] Tao, X., Chi, O., Delaney, P.J. et al. Detecting depression using an ensemble classifier based on Quality of Life scales. Brain Inf. 2021.

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

[27] Tran, T. & Kavuluru, R Predicting mental conditions based on history of present illness in psychiatric notes with deep neural networks. J. Biomed. Inform. 75, S138–S14, 2017.

[28] Tsugawa S, Kikuchi Y, Kishino F, Nakajima K, Itoh Y, Ohsaki H Recognizing depression from twitter activity. pp. 3187–3196, 2015.

[29] 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.

BIOGRAPHIES

Dr.A.sharada, Professor in CSE dept, obtained Ph.D from JNTUH in the area of Speech Recognition has been working in G.Narayanamma Institute of Technology & Science for last 20 years. She is working on various R&D projects in the areas of Voice Assistants for Elderly and Patient care, Mental Health Analysis, Assistive Technology. She published around 25 papers in International Journals like Springer and ACM and in International conferences. She has developed content for Gradiance portal of Stanford university. To her credit, she has a **license** for the research work done in finalizing ***Telugu Phone set and Phoneme chart applicable to Speech Recognition research*** under Creative Commons Unported License. Received **Best Teacher Award** from Institute for Exploring Advances in Engineering in Mar, 2018. Dr.Sharada has adopted many books for various universities and reviewed books on UML, C, Data Structures, C++ and JAVA for popular publishers like Tata McGrawHill, Pearson, Oxford University Press etc.

Manasa Jonnalagadda received B.Tech degree in 2018 from G. Narayanamma Institute of Technology & Science and worked at Infosys, India for 2 years. She is currently a final year M.Tech student of G. Narayanamma Institute of Technology & Science and working as a software programmer at Thomson Reuters. Her primary research interest is Machine Learning and Deep Learning Applications in Healthcare domain.