International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Analysis of Infliction of Mental Illness in Tech Industry

Himanshu Udupa G, Gourav H G, Shukruth J, P Prajeeth Shetty

Dr. M P Pushpalatha, Professor, Dept of CSE, JSS STU, Mysoressss

Himanshu Udupa G, Dept of CSE, JSS STU, Mysore
Gourav H G, Dept of CSE, JSS STU, Mysore
Shukruth J, Dept of CSE, Mysore
P Prajeeth Shetty, Dept of CSE, Mysore

---***---

Abstract - In today's environment, mental wellness is more important than ever. High stress, long hours, work pressure, building a reputation for oneself, and a work-life balance between personal and professional life are all symptoms of mental illness disorder. Mental health problems are highly prevalent among workers in the tech workplace. This research used a dataset called the mental health in tech survey, which gathered data from people all over the world. The goal of the proposed research is to determine the frequency of mental health disorders among technical employees versus non-technical employees in the workplace. We have explored what are the differences in the prevalence of mental disease and attitudes toward mental health by geographic location? What are the most powerful predictors of mental illness in the workplace, as well as specific attitudes regarding mental health? It is important detection and diagnosis of mental health conditions. This research illuminates the most effective techniques for improving employee well being and encouraging them to seek treatment when necessary

Key Words: *Mental Health, Tech Industry, Machine Learning, Occupational stress, Psychological well-being*

1. INTRODUCTION

Our emotional, psychological, and social well-being are all part of our mental health. It has the potential to influence our interactions with others, as well as our professional performance and physical wellness. The problem of mental health is receiving an increasing amount of attention these days. For persons with mental illnesses, having a positive attitude toward obtaining therapy is critical. Mental disease is strongly characterized in society as a shortcoming in a person, and most people are not comfortable admitting that they are suffering from mental illness. According to WHO high stress, long hours of labour, work pressure, and the desire to make a name for oneself characterize people in technical sectors.

The dataset for this research comes from Open Source Mental Illness (OSMI), an
organization committed to raising awareness and providing resources for mental health
in the tech community. The data includes survey responses from 2014, 2016, 2017, 2018, and 2019, offering a longitudinal perspective on the evolving mental health landscape in tech workplaces. The study also includes deep depression detection model with the help of deep learning based on data tracking activity of the individuals.

2. OBJECTIVES

The main objectives of this paper were to:

i. Identify Influential Factors: Study the previous versions to determine demographic and occupational factors such as family history of mental illness, gender, age, and self-employment status, that correlate with mental health disorders among tech industry employees.
ii. Data Collection and Analysis: To collect and process the OSMI Mental Health in Tech Survey datasets, ensuring all relevant variables are accurately formatted for model training.
iii. Model Development and Evaluation: To implement various machine learning models to predict the likelihood of a mental health disorder, depression and diagnosis and assess its performance.
iv. Insights and Future Research: To generate insights on significant predictors of mental health issues and propose additional variables for future analyses to deepen the understanding of mental health in the tech workplace.
v. Build a Potential Solution: Develop a NLP based chatbot to provide emotional support and assistance to individuals struggling with mental health issues. It can help individuals access mental health resources, offer guidance and support.

Additionally, we aim to address several critical questions through research and interviews, including:

- What steps should upper management take to improve mental health outcomes?
- Which racial groups are most affected by mental health disorders (MHD)?
- Which racial groups are more willing to discuss mental health openly?
- What types of jobs are more impacted by MHD?
- How do remote, hybrid, and non-remote work settings influence MHD?
- Which gender groups receive the least mental health support?
- What can the tech industry and employers do to enhance mental health support for employees?

3. LITERATURE SURVEY

Md Milon Uddin et al.(2022) a study on **Mental health Analysis in Tech Workplace** focused on five primary areas of information. Respondents' demographic and geographic information, such as age, gender, country, state, and family history of mental illness. Basic information regarding the work environment: for example, whether you are self employed or not, the number of employees you have, whether you work

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

remotely or not, whether you work for a tech firm or not and whether or not work interferes with your mental health.
Sandhya P et al.(2019) a work which is on **Prediction of Mental Disorder for employees in IT Industry** focuses on predicting mental health disorders in IT industry employees using machine learning techniques. It analyses a dataset of surveys from IT professionals to identify factors like age, gender, and work environment that affect mental health. Several machine learning algorithms were applied for analysis. The study emphasizes the importance of mental health check-ups, flexible work environments and employer support to improve employee well-being.
Rachna Narula et al.(2024) inspected the study on **Mental Health problem prediction of Tech Employees Using Machine Learning** on Open Source Mental Illness(OSMI) dataset by training various machine learning algorithms. A comparison of various types of ML algorithms for mental health prediction is offered in this research report. The outcomes show that, with an accuracy of 94.42 percent, the Bagging model performed better than competing methods. The potential for improved diagnostic accuracy and customized treatment strategies is highlighted in these findings, which add to the expanding body of research on machine learning-based mental health prediction.
Kiran Polimetla et al.(2022) proposed a project in **Mental Health in the Tech Industry: Insights from Surveys and NLP Analysis** that examines the mental health challenges faced by professionals in the technology sector, using data from the Open-Source Mental Illness (OSMI) surveys conducted between 2014 and 2019. By applying both quantitative analysis and Natural Language Processing (NLP) techniques, the research identifies trends in mental health disorders, attitudes, and support systems across various demographic groups and job roles.

4. METHODOLOGY

4.1 **Data Collection** - This study utilizes data collected through annual surveys conducted by Open Source Mental Illness (OSMI) from 2014 to 2021 (excluding 2015). These surveys were distributed online and targeted individuals working in the tech industry. Participants were asked to respond to questions about their mental health, workplace conditions, and attitudes toward mental health support and resources within their companies.
The study also includes additional data collected from Simula.no for depression detection. This data was monitored with an actigraph watch worn at the right wrist. The dataset contains the following: Two folders, whereas one contains the data for the controls and one for the condition group. For each patient we provide a csv file containing the actigraph data collected over time.

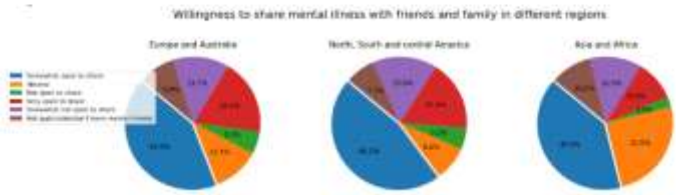
4.2 **Data Survey** - The survey consisted of understanding dataset, Likert scale items and open-ended responses. Key areas covered in the survey include:

1. **Demographics**: Age, gender, geographic location, and race.
2. **Job Role**: Participants' roles in the tech industry, including developers, DevOps, and other job categories.
3. **Mental Health**: Presence of diagnosed mental health conditions, such as anxiety or depression.
4. **Workplace Environment**: Attitudes and policies toward mental health in the
workplace.
5. **Willingness to Discuss**: Participants' willingness to discuss mental health issues with their employers and colleagues.
6. **Access to Resources**: Availability and usage of mental health benefits and
resources provided by employers.
7. **Work Setup**: Types of work arrangements, including remote, hybrid, or in-office work environments.

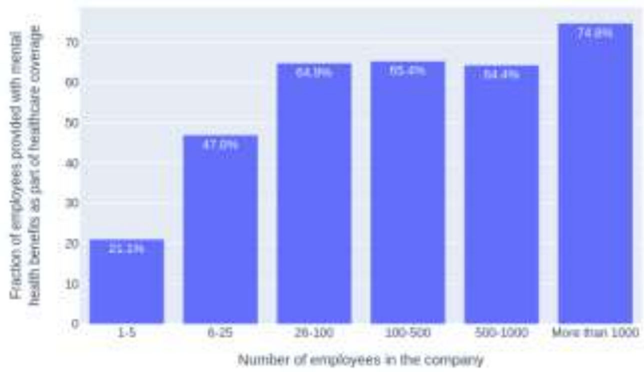
The depression detection dataset contained columns such as: timestamp (one minute intervals), date (date of measurement), activity (activity measurement from the actigraph watch). In addition, there was a MADRS scores(in scores.csv). It contains the following columns: patient id, days (number of days of measurements), gender (1 or 2 for female or male), age, inpatient (1: inpatient, 2: outpatient), madrs1 (MADRS score when measurement started), madrs2 (MADRS when measurement stopped) etc.

4.3 **Data Analysis** - For machine learning to apply, we first removed unnecessary fields like comments and timestamp. Dataset was cleaned making it appropriate for further analysis. Many of the attributes had empty values as input so default values were assigned to it. For Integer it is 0, float is 0.0 and string is NaN. Now for gender attribute we have make it all in standard format by replacing any unknown inputs to standard input.
The next step is Data Encoding. Every column present in the dataset was categorical data. The categorical values were encoded manually using numbers starting from 0. Later the dataset was scaled and fitted.
Data Visualization helped in visualizing the data and draw insights from data

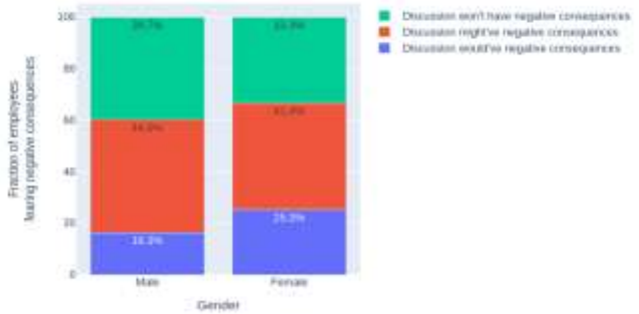
- Willingness of employees to share mental illness with different regions



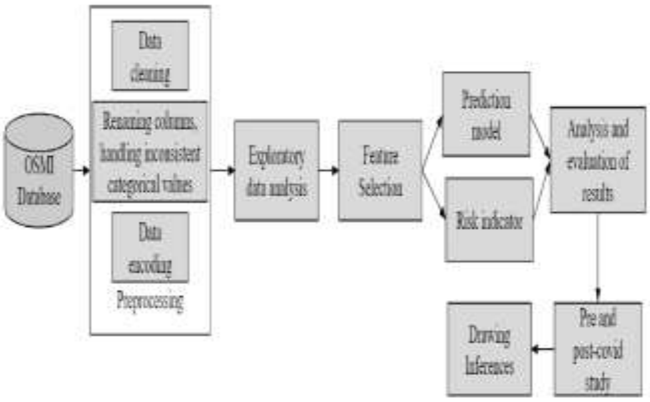
- Fraction of employees in companies of different sizes provided with mental health benefits as a part of health coverage



- Fear of negative consequences in male and female employees on discussing mental health issues with their employers



• Machine Learning techniques were applied and compared that suits best for dataset. The dataset was split into training and testing set with 70:30 ratio. Below figure provides a thorough representation of the machine learning model's underlying workflow, covering the full procedure from start to finish. It explains how the algorithm efficiently processes the orderly succession of train inputs and their related outputs. Additionally, it illustrates how the trained model is then applied to evaluate fresh data, using a set of independent variables. In our research, a logistic regression was proven to be 80% accurate.



Models we tested as follows:

K-Nearest Neighbors Classifier(KNN) - The KNN classifies input data into coherent clusters based on the similarity to previously trained data. The class that the input data shares the most distant neighbors. In our study, the KNN classifier was able to obtain an accuracy of 89.89%.

Logistic Regression – The Logistic Regression is used to forecast either a categorical or a discrete dependent variable using a set of independent variables. Because it represents the link between input and output variables, a logistic regression yields probabilistic values in the range of 0 to 1. These measurements suggest that there is a very high likelihood that specific outcomes, such as 0 or 1, yes or no, and true or false will occur. In our research, a logistic regression was proven to be 91.41% accurate.

Decision Trees - It employs a tree like structure to express the statistical likelihood or sequence of events, action, and result. The decision tree classifier methods separate the attributes at each node in the tree to determine whether splitting is optimal in each class. Our investigation discovered an accuracy rating of 85.35% when using the Decision Tree Classifier.

Random Forest - Random Forest classifier is a method used to address classification and regression issues. It merges several decision trees to form an amalgamation known as "the forest". In our test, the Random Forest classifier succeeded in obtaining an accuracy of 91.41%.

XGBoost - By using a training sample selected at random with replacement, xgboost creates subgroups from the data. They use each set of data to train their decision trees. On the XGBoost Classifier, we get 93.43 percent accuracy by using the Bagging Classifier.

ADABoost – Adaptive Boosting a powerful ensemble machine learning algorithm that combines multiple weak learners (classifiers) to create a strong, accurate predictive model, particularly effective for classification tasks. In our study we obtained accuracy of 88.38%.

Gradient Boost – Gradient Boosting combines multiple "weak" models (like decision trees) to create a strong predictive model by iteratively focusing on correcting errors from previous models, ultimately minimizing a loss function. We obtained accuracy of 93.93% best among all the algorithms.

In order to understand and study the employees' risk of progressing to a mental health issue, a risk indicator was built. The risk indicator was modelled using various clustering techniques where the employees were fractionated into 3 clusters representing different levels of risk of the employees. The following clustering models were implemented: K means clustering, K means++, Partition around Medoids clustering etc. The following metrics were used to evaluate the performance of the above clustering models: Silhouette score, Calinski Harabasz score, Sum of squared errors

The deep depression detection is built using deep neural network model has an input layer, then two convolutional layers (sliding window of 30 minutes, sliding step 10 minutes), then one max-pooling layer (to pick up the most responsed outputs), then one convolutional layer, and a finally max-pooling layer. Finally a dense layer.

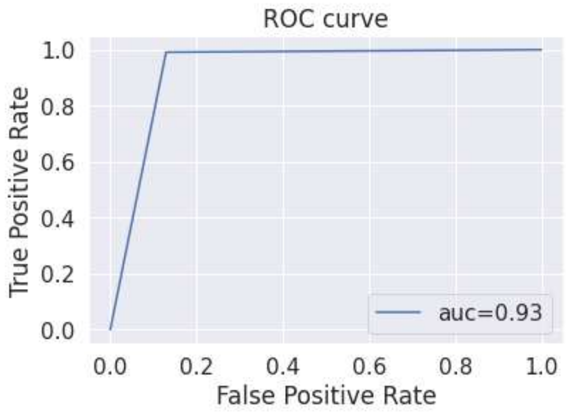
```
>>> model.summary()
Model: "model_7"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_12 (InputLayer)        [(None, 65407, 1)]        0
_________________________________________________________________
conv1d_33 (Conv1D)           (None, 6538, 300)         9300
_________________________________________________________________
conv1d_34 (Conv1D)           (None, 651, 300)          2700300
_________________________________________________________________
max_pooling1d_1 (MaxPooling1 (None, 325, 300)          0
_________________________________________________________________
conv1d_35 (Conv1D)           (None, 30, 300)           2700300
_________________________________________________________________
global_max_pooling1d_4 (Glob (None, 300)               0
_________________________________________________________________
dense_8 (Dense)              (None, 2)                 602
=================================================================
Total params: 5,410,502
Trainable params: 5,410,502
Non-trainable params: 0
```

5. RESULTS AND INFERENCE

• While applying the above machine learning algorithms on the dataset we evaluated these methods using metrics such as Mean absolute error, Mean squared error, Root mean square error, accuracy, precision, recall and f1-score.

Model	Accuracy	Root Mean Squared Error
KNN	89.8989	0.3178
Logistic Regression	91.4141	0.293
Decision Tree	85.3535	0.3827
Random Forest	91.4141	0.293
ADA Boost	88.3838	0.3408
XGBoost	93.4343	0.2562
Gradient Boost Classifier	93.9393	0.2461

- Here from the table we can see that Gradient Boost Classifier has highest accuracy of 93.93% and least root mean square error of 0.2461.



- ROC curve of Gradient Boosting Classifier
- Participants who battle with mental illness but refuse to seek treatment had the most concerns/fears (of negative workplace consequences) about disclosing a mental illness, as well as the least availability of benefits and support resources.
- Supervisors might be trained to spot mental problems and take the initiative to reassure staff to get treatment as needed to avoid negative consequences, as they are likely to be the first to notice any issues.
- Offering benefits to cover mental health treatments encourages people to seek help.
- There are fewer negative effects in the job when leave allowance and anonymity protection are provided.

6. INSIGHT GENERATION AND FUTURE RESEARCH

Data visualization and machine learning models collectively helped in insight generation leading to deeper understanding of problem. The insight generation mainly revolved around multiple questions and attributes from dataset.

The key areas of the study included:

1. **Demographics**: Age, gender, geographic location and race.
2. **Job Role**: Participants' roles in the tech industry, including developers, DevOps, and other job categories.
3. **Mental Health**: Presence of diagnosed mental health conditions, such as anxiety or depression.
4. **Workplace Environment**: Attitudes and policies toward mental health in the workplace.
5. **Willingness to Discuss**: Participants' willingness to discuss mental health issues with their employers and colleagues.
6. **Access to Resources**: Availability and usage of mental health benefits and resources provided by employers.
7. **Work Setup**: Types of work arrangements, including remote, hybrid, or in-office work environments.

Implementation of Descriptive statistics and Correlation Analysis combined with the results of machine learning addressed specific research questions related to demographic groups and relationship between employee and employer.

The main objective of this research is to answer an important question :

To what extent can we predict tech workers' mental health condition using machine learning techniques?

The above question can be answered through 2 sub questions, as follows:

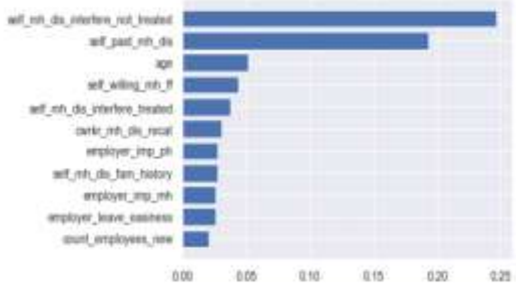
Q1. *Which machine learning technique provides good performance in predicting tech workers' mental health conditions?* And

Q2. *What are the contributing factors in predicting tech workers' mental health conditions?*

The first sub question can be answered from the previous study. That is gradient boosting algorithm had a better performance and showed considerably higher scores compared to other machine learning algorithms with f1-score more than 0.93 while others had score of around 0.7-0.8.

Acquiring predictions from a machine learning model and evaluating its performance is usually the primary goal of performing data science techniques. Asides from the main question, a feature importance analysis can be performed to investigate the contribution of each feature to the models that previously had been developed. The most influential factors in predicting a target attribute can be identified with feature importance.

Feature importance analysis is carried out using Gradient Boost as the model that shows the highest f1 score among the other models.



The x-axis represents the total gain in models across all splits when a feature is included in the model. y-axis represents attributes that are essential in predicting the mental disorder of tech workers. These attributes are : *"If you have a mental health disorder, how often do you feel that it interferes with your work when not being treated effectively?", "Have you had a mental health disorder in the past?", "age" etc.*

Two factors that show the highest contribution to predicting someone's mental health condition are the question *"If you have a mental health disorder, how often do you feel that it interferes with your work when not being treated effectively?"* with 0.025 total gain and *"Have you had a mental health disorder in the past?"* with 0.19 total gain.

The above features covers questions across all the areas including from personal conditions, related to employer and co-workers. Of the top 10 features, five features are related to an individual's condition, four features are related to the employer, and one feature is related to co-workers.

The features that focus on individual's condition are:

- *If you have a mental health disorder, how often do you feel that it interferes with your work when NOT being treated effectively (i.e., when you are experiencing symptoms)?*

- Have you had a mental health disorder in the past?
- Do you have a family history of mental illness?
- How willing would you be to share with friends and family that you have a mental illness?
- If you have a mental health disorder, how often do you feel that it interferes with your work when being treated effectively?

The features that target relation between employer and employee are:

- Overall, how much importance does your employer place on physical health?
- Overall, how much importance does your employer place on mental health?
- If a mental health issue prompted you to request a medical leave from work, how easy or difficult would it be to ask for that leave?
- Have you ever discussed your mental health with your employer? And how many people are there in your organization?

The question related to co-worker is: “*If they knew you suffered from a mental health disorder, how do you think that your team members/co-workers would react?*”

Apart from the above research question we were able to draw various other useful insights through statistical analysis. The two major methods used are:

Descriptive Statistics: Used to summarize demographic information and the prevalence of mental health conditions across different demographic groups.

Correlation Analysis: To explore relationships between workplace factors (e.g., job role, work setup) and mental health outcomes.

Using a combination of descriptive statistics and trend analysis, we aimed to answer the following research questions:

1. Which racial group is most affected by mental health disorders and were willing to discuss their disorder openly?
2. Which job roles are more impacted by mental health disorders?
3. How does remote, hybrid and non-remote work affect mental health?
4. Which gender group reports receiving the least mental health support?

The above questions can lead to solution for What can upper management do to address mental health disorders more effectively? And What can the tech industry and employers do to improve mental health resources and support?

By addressing these questions, the methodology aims to provide actionable insights into the mental health challenges faced by tech industry professionals and offering solutions for improving mental health support.

1. African American individuals report higher rates of mental health disorders compared to other racial groups. Despite this, there is no corresponding increase in mental health resources or support specifically targeted for this group. This disparity underscores the need for more focused mental health resources and support initiatives for African American employees to address their unique challenges.

African American respondents were the most willing to discuss mental health issues (83%), followed by Multiracial (73%), White (70%), and Asian respondents (45%).

2. Mental health conditions varied significantly based on job roles:

High Impact Roles: Developers and DevOps professionals reported the highest levels of mental health disorders, with nearly 45% reporting anxiety and burnout. These roles were identified as more demanding, with higher workloads and pressure.

Management Roles: While managers reported mental health issues, their prevalence was lower (around 30%), likely due to greater access to mental health support.

Developers and DevOps professionals are the most impacted by mental health disorders, with higher reported rates of anxiety, burnout and stress compared to other roles. This highlights the urgent need for tailored mental health resources and support systems specifically designed for these high-pressure job categories.

3. The type of work setting had a noticeable effect on mental health:

- **Non-Remote Workers**: Individuals who had never worked remotely reported higher levels of stress and burnout, with about 40% experiencing mental health disorders.
- **Remote Workers**: Remote workers, on average, reported slightly better mental health outcomes, citing improved work-life balance as a contributing factor.
- **Hybrid Workers**: Hybrid workers fell between these two groups in terms of mental health conditions, with 35% reporting mental health issues.

4. Gender played a critical role in the availability and quality of mental health support:

- **Women and Non-Binary Individuals**: Women and non-binary respondents consistently reported receiving poorer mental health support compared to men. Over 35% of women and 32% of non-binary respondents experienced either unsupportive or inadequate responses to mental health concerns.
- **Men**: While men were less likely to report mental health disorders, they also encountered fewer issues in accessing mental health resources. Over 27% of male experienced either unsupportive or inadequate responses to mental health concerns.

From the above generated insights we can recommend certain measures that can be undertaken in an IT company that makes sure to improve the mental health of workers’ from tech background.

Action Items for upper management: Survey responses highlighted several key action items for upper management to improve mental health support:

1. **Management Training**: Training programs for upper management on how to handle mental health discussions and responses, particularly for female and non-binary employees are needed.
2. **Resource Allocation**: Increased mental health resources should be allocated to African American employees, as they reported higher proportions of mental health disorders but lower access to support.
3. **Open Communication**: Encouraging open discussions about mental health across all levels of the organization, particularly in high-impact roles like developers and DevOps, could reduce stigma and improve mental health outcomes.

Recommendations to improve Mental health in Tech Industry:

- Promote better balance between work and personal life.
- Encourage open discussion on mental health and raise awareness among the employees.

- Offering more mental health holidays, enhanced healthcare options, and additional mental health resources are crucial for employee well-being.
- Educating the employees on mental health benefits and resources currently available.
- Take measures to curb racial abuse and improve the gender neutrality so that people from various race and gender can have a good place of work.

7. FUTURE WORK

A number of other future possibilities are to be taken into account in order to enhance the effectiveness of Machine Learning Model on Binary Classification. In the first place, features that address work-life balance, stress levels, social support or workplace policies will be used to predict the mental disorder that can help us understand the aspects that affect the mental health of IT employees better. A chat-bot will be designed to provide emotional support and assistance to individuals struggling with mental health issues. It can help individuals access mental health resources offer guidance and support. With the integration of Language translation this chatbot will be very efficient as it will be able to break the language barriers.

8. CONCLUSION

There are many suggestions that employers and employees could keep in mind. Employers need to keep track of number of their employees having mental disorder. Employers should allow flexible work environment with flexible work scheduling and break timings. They should allow employees to choose flexible place of work. They should give day-to-day feedback and guidance for nurturing employees' health. This type of model could be used to detect metal health progress among employees and also could lead to policy changes. Employees could talk to colleagues and their managers about their problem freely. Hence upper management could help them to get correct aid with beneficiaries like work from home, flexible timings, more leaves, many more. Employees should know health benefits provided by their organization participate in any wellness programs.

9. REFERENCES

1. World Health Organization, Promoting mental health: concepts, emerging evidence, practice: a report of The World Health Organization, Department of Mental Health and Substance Abuse in collaboration with the Victorian Health Promotion Foundation and the University of Melbourne. 2005.
2. Awal, G. K., & Rao, K. (2021). Can machine learning predict an employee's mental health? In *International conference on information, communication and computing technology*.
3. Teles, A., Rodrigues, I., Viana, D., Silva, F., Endler, M., & Rabêlo, R. (2019, June). Mobile mental health: A review of applications for depression assistance. In *2019 IEEE 32nd International Symposium on Computer Based Medical Systems (CBMS).*
4. B. Saha, T. Nguyen, D. Phung and S. Venkatesh, "A Framework for Classifying Online Mental Health-Related Communities With an Interest in Depression," in *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1008-1015, July 2016.