IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, May 2026
Impact Factor: 8.2

Depression Risk Prediction among Tech Employees using AdaBoost Tree

Dr. N. Sree Divya, Dr. U. Chaitanya, Chenemoni Vaishnavi, Pendyala Jaya Charan
Assistant Professor, Department of Information Technology
Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, India
nsreedivyait@mgit.ac.in, cvaishnavi_it221214@mgit.ac.in
uchaitanyait@mgit.ac.in, pjayacharan_it221247@mgit.ac.in

Abstract: *Depression is increasingly becoming an issue among those people working in the technology industry. The majority of people working in such an environment tend to work for some time and face numerous challenges associated with the high pressure of tasks, strict deadlines, which consequently may lead to a certain amount of stress affecting the mental state of these employees. Therefore, it is crucial to diagnose depression at risk in order to give companies a possibility to provide their employees with appropriate support. This paper applies a machine learning approach to detect the probability of depression at risk among employees of the technology industry. The paper analyzes a survey among technology professionals using the algorithm of AdaBoost. The analyzed dataset consists of the personal data about employees of the technology sector including the infor-mation about their age, job and overall health condition. The pre-processing stage of data was performed in order to prepare the dataset to build a prediction model. The performance of the algorithm AdaBoost was then compared to other algorithms aimed at forecasting depression such as Logistic Regression, Random Forest and Support Vector Machine. According to the findings, AdaBoost outperforms other approaches applied to the analysis of the specified data. Consequently, machine learning algorithms may be successfully used to forecast diseases*

Keywords: Depression Prediction, AdaBoost, Machine Learning, Mental Health, Ensemble Learning, Workplace Analytics

I. INTRODUCTION

There have been various developments in the work culture due to the emergence of the technology industry. Working overtime, meeting deadlines, and performing well in work are common expectations. Such a scenario may lead to stress, anx-iety, and even depression. Depression not only brings damage to the health of the individual but could also have negative impacts on the productivity of the firm. The common approach in determining depression among employees is through psy-chological tests. But there is another approach where machine learn- ing algorithms may detect early signs of depression. In this research, the AdaBoost will be used as the algorithmic tool in predicting de-pression.

II. RELATED WORK AND LITERATURE SURVEY

Machine learning methods have proven to be highly effec-tive in the analysis of data in the healthcare sector, especially for depression prediction. The use of traditional classifiers such as Logistic Regression, Support Vector Machine, and Decision Trees is common due to the simplicity and efficiency of these methods in data classification. But with more complex data, more sophisticated machine learning approaches, like en-semble and deep learning approaches, are gaining popularity. Patel and Gupta (2022) have used a machine learning ap-proach using AdaBoost and Random Forest methods for predicting mental illnesses. Their study was aimed at making the prediction system more understandable by using fea-ture importance. Although their prediction system had good accuracy, there was no real-time testing of the model.

The method Gradient Boosted Trees by Das Mehta (2021) applied for prediction purposes on health datasets was successful as far as capturing the non-linearity and analyzing large sets of data. Nevertheless, this model required high com-putational power and lacked transparency. Hassan Lee (2024) suggested a new explana- ble boosting model through the use of SHAP (SHapley Ad- ditive Explanations). Even though this model could explain feature importance, it required additional computational power. In their research on stacked ensemble models, Kumar et al. (2020) considered the use of multiple techniques to achieve prediction accuracy. While it led to better results, the model became more complicated, making its implementation challenging. Roy Bansal (2022) discussed the use of trans- fer learning approaches through tree-based ensemble models. Although these approaches proved to be valuable, they required large datasets and accurate parame-ters.

Selvaraj and Mohandoss (2020) compared various machine learning models such as Logistic Regression, SVM, and Deci-sion Trees for mental health prediction. Their study concluded that ensemble methods generally outperform individual clas-sifiers in terms of accuracy.

Nguyen and Park (2022) have suggested the development of attention-based deep learning models for mental health studies. Such models could successfully identify complex patterns yet had the need for large amounts of data and were not easy to explain.

Matsuda and Cho (2024) offered adaptive boosting, where model parameters were adjusted automatically for better accuracy; however, it increased computation.

The paper by Dasgupta and Mehra (2021) discussed deci-sion tree ensembles, stressing the benefits from the collaboration of different learners. Still, overfitting occurred.

In their comparative study, Zhang and Kumar (2023) stated that AdaBoost was more accurate because of its ability to fix errors.

Ethical issues, particularly fairness, transparency, and pro-tection of personal data, were emphasized by Banerjee and Das (2024).

Lastly, Wang and Luo (2023) presented the decision tree mod- els that had an optimized version characterized by automatic feature selection, which contributed to improving accu-racy without increasing complexity. Sharma and Verma (2023) suggested the use of a hybrid machine learning algorithm that combines both Logistic Regres- sion and Random Forest. Their method ensured high accuracy by taking advantage of interpretability and robustness. Li and Chen (2022) conducted research based on deep neural networks in the context of depression detection from behav- ioral and survey data. They managed to achieve excellent accuracy results, but the require-ment was a large amount of data and lack of interpretability. Garcia and Martinez (2021) created a comparative frame- work for evaluating multiple machine learning algorithms. The au-thors found out that ensemble methods performed better than standalone classifiers in terms of both accuracy and robust-ness; however, the disadvantage included increased compu-tational complexity and training time. In summary, according to the reviewed literature, ensemble techniques offer better performance than other models in the context of predicting depression; however, some challenges exist..

III. EXISTING SYSTEMS

Existing approaches for depression prediction primarily rely on traditional machine learning techniques applied to survey-based or clinical datasets. Logistic Regression is one of the most commonly used methods due to its simplicity and interpretability in binary classification problems. However, it often struggles to capture complex nonlinear relationships present in mental health data

Another approach for predicting depression that has also been widely used is support vector machines (SVM). SVM works well in classification tasks, but it requires proper tuning of the parameters to achieve optimal results, and it does not scale well with the size of the dataset. Random Forest technique uses several decision trees to create a more accurate prediction. It is suitable for working with non-linear data, but it may lead to overfitting and lack of

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, May 2026
Impact Factor: 8.2

interpretability. The recent advances in the field include ensemble methods such as AdaBoost algorithm, which works by repeatedly focusing on training samples that have been misclassified earlier. AdaBoost outperforms other approaches in terms of accuracy due to the use of boosting mechanism. However, it suffers from sensitivity to outliers and noises in the data. Neural networks and other deep learning methods have also been tried to predict depression. These models require extensive computing resources and large datasets and are thus far not efficient enough to be applied effectively. To conclude, current systems have contributed greatly to the development of depression prediction technology; however, they suffer from certain dis-advantages such as poor data, insufficient interpretability, and lack of real-time adaptation. These limitations make it essential

IV. COMPARATIVE ANALYSIS AND STUDY

The comparison of various machine learning methods used for depression prediction is provided in the below table (Table I). Various methods have been compared in terms of the type of dataset used, accuracy, and constraints associated with them. The following table illustrates the performance of various machine learning techniques used for depression prediction. One can clearly observe that methods like AdaBoosting and Random Forest give higher accuracy because of their capabil-ity to use multiple weak learners in building the final model. Methods such as deep learning give higher accuracy; however, these methods require extensive training datasets as well as computational powers. Methods such as Logistic Regression and Support Vector Machines (SVM) are simple and easy to interpret, though they fail to deliver good performance with complex datasets. From the above discussion, it can be seen that the method of AdaBoosting gives better results in terms of accuracy because of its error correcting nature.

V. DATASET DESCRIPTION

This dataset contains structured survey data that was taken from employees in the tech industry. This dataset contains variables related to demographics, work, and psychological factors.

TABLE I: FEATURE DESCRIPTION

S.No	Feature	Description
1	Age	Age in years
2	Gender	Male/Female/Other
3	Working Hours	Weekly working hours
4	Sleep Duration	Hours of sleep per day
5	Stress Level	Scale (1–10)
6	Job Satisfaction	Low/Medium/High
7	Work-Life Balance	Rating scale
8	Mental History	Yes/No
9	Depression Risk	Target (0/1)

This dataset is split into 80for testing purposes. This dataset contains some attributes used for predicting the risk of de-pression among employees. Age is defined as the age of the individual, in years, while Gender refers to whether an employee is either male, female, or another gender. Working Hours refers to the number of working hours per week, while Sleep Duration is the duration of sleeping hours per day.Stress Level is a measurement ranging from 1 to 10, which describes the stress level of the employee. Job Satisfaction is measured using low, medium, and high to indicate how satisfied an employee feels with his/her work. Work Life Balance is provided with a rating scale measuring how good an employee can balance professional and personal life. The Mental History variable describes whether the employee has ever been affected by any mental illness (Yes or No). The response variable Depression Risk is shown as 0 or 1, where 0 represents no risk of depression and 1 represents depression risk. The dataset is split into training (80testing (20develop a model, while the testing dataset will be used to validate the model's performance..

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, May 2026
Impact Factor: 8.2

VI. PROPOSED SYSTEM

A. Data Preprocessing

Data preprocessing is an integral part of the proposed system since the performance of the model will depend on the quality of the data entered. Raw data can have missing values, inconsistent format, and categorical variables which cannot be used directly for training. Thus, several steps of data preprocessing are performed prior to model training. The data preprocessing procedures include:

Handling Missing Values: In case there are missing or null values in the data, appropriate methods are performed to ensure that the integrity of data is maintained.

Categorical Variable Encoding: Machine learning algo-rithms use numeric input only. Thus, categorical variables, such as gender, country, and company size, are converted to numeric input by means of labeling or other encoding techniques.

Normalizing Features :Numerical features of the dataset can have a high value range. For ensuring the effective learning of the model, feature normalization is performed.

Splitting into Training/Testing Subsets :To perform the training process properly and to assess its efficiency, data splitting is carried out.

B. AdaBoost Model

The proposed solution utilizes the Adaptive Boosting (Ad-aBoost) algorithm, which is a boosting method that aims at increasing classification accuracy. Instead of having a single learner, Adaboost uses multiple weak learners to create a strong one. Initially, all instances of the training set have equal weights. The weak learner is trained on the entire training set. At each stage, the performance of the weak learner is evaluated. The weights of the instances that are misclassified by the model are increased, while those that are correctly classified are decreased. As a result, the next learner pays more attention to the difficult examples. The final prediction of the AdaBoost classifier can be described by the following formula:

$$F(x) = \sum_{m=1} \alpha m h m(x)$$

where hm(x) denotes the mth weak learner, and αm rep-resents its corresponding weight. the weight of this learner, depending on its performance. At the final step, all weak learners vote using their respective weights to make the final classification decision. Through this process, AdaBoost sig-nificantly reduces variance and bias, thus leading to improved prediction performance. Therefore, it is a suitable tool for depression risk prediction models.

VII. SYSTEM FLOWCHART

The system flowchart shows the overall working process of the depression risk prediction system. It gives a visual represen- tation of the entire process that takes place during the data flow from its collection until the prediction is made. Every stage in the process chart is a crucial one for the development and eval- uation of the machine learning model.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 6, Issue 3, May 2026

Impact Factor: 8.2

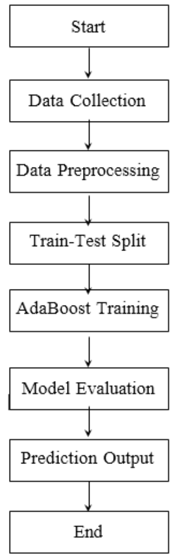

Fig. 1. Workflow of Depression Risk Prediction System

Firstly, there is the data collection stage, which collects all the data required for further actions. Secondly, the data preproces- sing stage cleans and transforms the data. Thirdly, the data is se- parated into two groups – the train set and the test set. The Al- daBoost model gets trained on the training dataset, while the model's performance is evaluated based on the test data. Finally, the prediction result is produced, showing the risk of depress- ion.

VIII. PERFORMANCE EVALUATION

Evaluation metrics used:

A. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

B. Precision

$$Precision = \frac{TP}{TP + FP}$$

C. Recall

$$Recall = \frac{TP}{TP + FN}$$

D. F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The AdaBoost algorithm showed better balanced perfor-mance on all evaluation measures. The system flowchart illus- trates the workflow process of the depression risk predication system. It starts from data gathering, in which the relevant data for employees are obtained. The gathered data are then sent to the data preprocessing phase, in which cleaning and transformation take place to render the data ready for analysis. After the preprocessing phase, the dataset is partitioned into training and testing datasets through the train-test splitting technique. The AdaBoost algorithm is then trained through the training dataset to learn the patterns and relationships in the dataset. Finally, the model is tested on the testing dataset, and the system produces the output prediction that determi-nes the risk level of the depression in an individual.

Performance evaluation of the proposed model was done through various measurements such as accuracy, precision, recall, and F1 score. Accuracy represents the overall correct-ness of prediction; precision denotes the percentage of cor- rect predictions of positive instances from the total predic- tion of positives, while recall represents the total correct prediction of positives out of the actual positives. Overall, adaboost proved to exhibit better performance through all per- formance metrics.

IX. PERFORMANCE EVALUATION

The accuracy of depression prediction algorithms is as-sessed based on common classification metrics. Such metrics play a crucial role in comparing different machine learning algo- rithms and evaluating their efficiency in recognizing people at risk.

A. Accuracy

Accuracy is the most frequently applied metric, measuring the overall percentage of correct predictions made by the algorithm. It can be calculated as the ratio of correctly pred-icted examples to the total number of examples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

B. Precision

Precision reflects the percentage of correct predictions of positive outcomes among all positive predictions made by the algorithm. It helps in assessing the reliability of positive predictions made by the model.

$$\text{Precision} = \frac{TP}{TP + FP}$$

C. Recall

Recall or sensitivity is the fraction of all actual positives that have been correctly identified by the model.

$$\text{Recall} = \frac{TP}{TP + FN}$$

D. F1-Score

F1 Score is the harmonic mean of Precision and Recall. It is useful when the class distribution is unbalanced.

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

E. Discussion of Metrics

When it comes to predicting depression, the use of a single measure like accuracy might be inadequate, especially if the dataset is unbalanced. Measures such as Precision and Re-call become vital to ensure that all the people at risk are identified. The F1-score measure is an amalgamation of both measures and has been extensively utilized in the evalu- ation of mental health prediction algorithms. In summary, the above measures play an indispensable role in comparing various machine learning techniques, including Logistic Regression, Random Forest, Support Vector Ma- chines, and AdaBoost..

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, May 2026
Impact Factor: 8.2

X. CHALLENGES AND RESEARCH GAPS

Even though there have been tremendous strides in machine learning techniques in predicting depression, there are still a number of difficulties and research gaps, especially when it comes to employees working within the technological sector.

A. Data Availability and Quality

The first difficulty is the lack of quality and accurate data. Information about mental health is highly confidential, and organizations tend to be reticent about sharing such data for security reasons. Consequently, researchers tend to use small or self-reporting surveys as their datasets, which can potentially limit the accuracy of the models and their validity.

B. Class Imbalance

The data sets for depression problems tend to be imbal-anced, where the data of non-depression is much more than that of depression. This creates biased data sets in favor of the majority class and thus makes it difficult for the system to recognize high-risk individuals. Despite the use of methods like resampling and cost-sensitive learning, it is not an easy task to balance the data set.

C. Lack of Domain-Specific Models

Most researches have been conducted on generic data, not targeted at a particular industry. The conditions in the tech industry are special in that workers tend to be subjected to stressful conditions such as working overtime and having tight deadlines.

D. Model Interpretability

However, many machine learning models, including en-semble learning and deep learning techniques, are black box models. Although they are highly accurate, their inability to be interpreted poses challenges in the comprehension of how decisions were made, hence limiting their implementation in organizations and the healthcare sector.

E. Privacy and Ethical Concerns

Handling sensitive information like mental health data re-quires great care to prevent misuse. Maintaining confidentiality and privacy of data and ensuring ethical use of data are among the challenges in using machine learning models in the workplace.

F. Real-Time Prediction Limitations

Current solutions to mental health problems rely on static data obtained from surveys. They fail to detect changes in the patient's mental state due to lack of dynamic predictions and monitoring.

G. Integration with Organizational Systems

An additional problem is the failure of prediction models to integrate into organizational decision-making systems. Most studies are still in the experimental phase and not imple-mented in real organizational settings. Overcoming this divide is an unresolved matter.

H. Research Gaps

From the above-mentioned challenges, the following re-search gaps may be identified: • The lack of large-scale and high-quality data sets related to tech- nology workers
• The need for explainable and interpretable machine learn-ing models
• The lack of attention towards real-time and continuous moni- toring of mental well-being
• Insufficient incorporation of machine learning models into organizational infrastructure

• Ethical considerations for using AI in an appropriate manner

It is important to resolve the above-mentioned challenges and research gaps to create effective depression prediction sys-tems.

XI. APPLICATIONS

Depression Prediction using machine learning has many applications in practical life scenarios, especially in organizations and healthcare sectors.

• Wellness Tracking for Employees: Organizational man-agement will be able to track the wellness of their employ-ees and take necessary steps to protect their health.

• Detection of Early Depressive Symptoms: Through this system, it would be possible to detect any mental prob- lems in the initial stages and treat accordingly.

• Human Resource Analytics: The Human resource mana-gement can analyze the state of wellness of the employees and devise strategies to improve the productivity of employ- ees and increase their satisfaction level.

• Healthcare Support System: This system, if integrated with the healthcare sector, will be helpful for diagnosing patients and devising their treatment plans.

• Suggestions for the Patient: Systems will recommend personalized suggestions to patients in terms of stress management techniques and lifestyle changes.

XII. ETHICAL CONSIDERATIONS

The implementation of machine learning in mental health prediction gives rise to certain ethical issues that need to be properly addressed. First, given that the data on depression is extremely sensitive, there are serious concerns about its privacy and confidentiality. Any kind of unauthorized acces-sibility or usage of such data will have serious repercussions. Another significant issue that emerges is algorithmic bias. If the algorithm itself includes any kind of bias, then the pre-dictions that emerge may be biased or incorrect. Thus, people might receive the wrong assessment regarding their depression risk, which could negatively affect decision-mak- ing at the organizational level. Moreover, there are ethical issues related to transparency and explainability. Many ML algorithms used by organiza- tions are black boxes. This is why it is vital that all the predic- tions can be adequately explained. It is also crucial to take into account the other ethical issues. For instance, systems should comply with principles of informed consent and accountability. Machine learning sys- tems should not replace professional medical diagnosis.

XIII. DISCUSSION

The review of the current machine learning methods used for depression prediction reveals that there is a rising trend in applying data-driven approaches in diagnosing men- tal disor-ders. Numerous machine learning algorithms such as Logistic Regression, Support Vector Machines (SVM), Random Forest, and AdaBoost have already been applied to predict depressive disorders. Each algorithm possesses spe- cific strengths and weaknesses which depend largely on the properties of the data.

The traditional models like Logistic Regression are very simplistic, understandable, and computationally inexpensive. They perform well if there is a linear relationship between the feature variables and the tar- get variables. But, their efficiency is compromised if there is any nonlin- earity in the dataset. The Support Vector Machine model has proven its efficacy in classification tasks, but its performance depends on the tuning of parameters.

It has been observed that the ensemble method, which includes Random Forest and AdaBoost, demonstrates better perfor- mance than single classifiers. In addition, the Random Forest algorithm increases the efficiency of predictions be-cause mul- tiple decision trees are used. Moreover, variance reduction is one of the properties of this algorithm, making it less vulnerable to overfitting. The algorithm AdaBoost works on boosting meth- odology where learning takes place iteratively, and emphasis is given to misclassification cases. Based on numerous research studies, it has been seen that AdaBoost provides better accu- racy in depression prediction cases.

Deep learning methods have also been investigated for predicting depression, especially in cases where the data set is extensive and complicated. These machine learning models can detect complex correlations between variables, yielding high accuracy rates. But these models need a lot of computing power and huge amounts of data, which is not always feasible. Moreover, they tend to act like black boxes, making them uninterpretable, which is a major limitation, especially in fields such as mental health

The other important observation found in most of the studies was their dependence on the quality of data. The data used for building the model for predicting depression mostly comes from surveys, which are prone to subjective biases. In addition, missing values and imbalances in the datasets can pose further problems in making accurate predictions.

In terms of practical applications, machine learning algo-rithms that predict depression have great potential applica-tion in workplace settings, particularly within the technology industry. Companies can use these algorithms to keep track of employee well-being, detect any possible mental problems, and adopt preventative actions. Nonetheless, when applied in practice, such algorithms must take into account important ethical concerns.

On the whole, there is a trade-off among accuracy, interpretabil- ity, and computation. Although modern methods like AdaBoost and deep learning algorithms achieve greater accuracy, simple methods have greater interpretability and simplicity. Thus, the choice of an optimal method relies on the particular needs of the problem at hand, which include access to data and com- putational power.

In summary, machine learning algorithms have demon-strated potential as an effective method of identifying indi-viduals at risk of depression. Nonetheless, additional research is needed in order to overcome various limitations associated with machine learning models. These include concerns relating to data accuracy as well as model interpretation.

XIV. FUTURE WORK

AWhile substantial strides have been made in using machine learning approaches for predicting depression, there are still many opportunities available for making improvements and conducting more research in this area. One such opportunity is the incorporation of data from real-time data sources, such as wearable devices, apps on mobile phones, and workplace monitoring. This will allow obtaining continuous data about patients' sleep and activity levels, and help make more ac-curate predictions about their mental state. Another aspect of future research in this field is the use of more sophisticated deep learning methods. Such methods include RNNs, LSTMs, and transformer networks. These models have proven partic-ularly effective at capturing sequential patterns in data, and as such can be very useful for analyzing temporal patterns in mental well-being.

Enhancing model interpretability is another essential area of research that needs to be addressed. The implementation of explainable artificial intelligence approaches like SHAP (SHapley Additive Explanations) and LIME (Local Interpreta-ble Model-Agnostic Explanations) will enable the creation of predictions that are clear and reliable. This is especially impor-tant in mental health contexts, where predictions must be un-derstood by healthcare professionals and other stakeholders. Class imbalance and data quality issues continue to be ma-jor problems. Further research should explore the develop-ment of improved preprocessing algorithms, such as sampling techniques and noise reduction processes.

The creation of hybrid approaches based on combining several machine learn- ing and deep learning strategies is another direction worth exploring. By combin- ing the ad-vantages of various methodologies, such solutions may help increase efficiency. Validation across domains and industries is necessary as well since the models should not be restricted to a particular set of data or conditions. It is vital to test these approaches within various organizations to increase their applicability in the real world. What is more, integration of these systems into organizational and healthcare platforms may further improve their usability. Creating intuitive interfaces and decision support systems will be helpful. It is important for future research to pay attention to ethical concerns as well. The issues of data privacy, responsibility, and other ethical concerns should be addressed. The need to create a unified framework for implementing machine learning models for mental health prediction and treatment cannot be overestimated. Overall, the future improvements should aim at increasing accuracy, interpretabil- ity, scalability, and ethical compliance of the solutions under consideration.

XV. CONCLUSION

This paper provided an extensive review of the various techniques utilized in predicting depression risks using ma-chine learning, with emphasis on technology workers. Several models such as Logistic Regression, Support Vec- tor Ma-chines, Random Forest, AdaBoost, and Deep Learn- ing were explored, taking into account their effectiveness, strengths, and weaknesses. In this study, it is evident that the ensemble mod-els like AdaBoost prove to be more effective since they work on im- proving the accuracy rate through iterative learning of mis- classification errors. They can handle the complex and nonlin- ear relationships existing within mental health data. Even though the other models are simpler, they play a critical role in the model due to their simplicity and interpretabil- ity. The review indicated several problems that need to be ad-dressed such as data quality, class imbalance, lack of data sets for specific domains, interpretability, among others. Moreover, ethical considerations such as privacy and fair- ness need to be taken into account when developing these models. Lastly, the paper addressed some of the potential applica- tions of depression prediction models within workplace set- tings..

Some other challenges were also mentioned in the survey, in- cluding poor data quality, class imbalance, scarcity of do-main-specific datasets, and lack of model interpretability. Ethical considerations like data privacy, fairness, and trans-parency also play an important role in deploying such sys-tems.

Moreover, the potential uses of such predictive systems in the workplace setting were also explained. Depression predi-ction systems could prove beneficial in enhancing the well-being of employees and organizational efficiency. The neces-sity of implementing such systems in practical settings was also emphasized.

Finally, machine learning algorithms provide a good op-portunity to detect depression risk at an early stage. It will help in developing preventive measures and interventions. It is imperative to maintain an appropriate trade-off between ac-curacy and interpretability in such systems. In the future, researchers must focus on designing practical depression pre-diction systems.

REFERENCES

[1] Patel, S. & Gupta, A. (2022). Explainable AdaBoost & Random Forest. IEEE Transactions on Affective Computing.
[2] Das, P. & Mehta, A. (2021). Gradient Boosted Trees for Prediction. Springer – Neural Computing and Applications.
[3] Hassan, M. & Lee, J. (2024). Explainable Boosting with SHAP. IEEE Journal of Biomedical and Health Informatics.
[4] Kumar, R., Sharma, V. & Singh, P. (2020). Stacked Ensemble Models. IEEE Access.
[5] Roy, S. & Bansal, P. (2022). Transfer Learning with Tree Ensembles. Springer – Cognitive Neurodynamics.
[6] Selvaraj, A. & Mohandoss, L. (2020). ML Models for Mental Health Prediction. Springer Journal.
[7] Nguyen, T. & Park, S. (2022). Attention-based Models. IEEE TNNLS.
[8] Matsuda, N. & Cho, Y. (2024). Adaptive Boosting for Personalization. IEEE Access.
[9] Dasgupta, R. & Mehra, L. (2021). Decision Tree Ensemble Methods. IEEE Access.
[10] Zhang, Y. & Kumar, S. (2023). Tree Models vs AdaBoost Comparison. Springer – Applied AI Reviews.
[11] Banerjee, A. & Das, R. (2024). Ethical Issues in AI for HR. Springer – AI & Ethics.
[12] Wang, J. & Luo, T. (2023). Optimized Decision Trees with Feature Selection. IEEE Access.