The Original

AN EXPLAINABLE AND ROBUST AI FRAMEWORK FOR MENTAL HEALTH SCREENING IN HIGHER EDUCATION

Puvvada Nagesh,[1] Malladi Venkata Laxmi Haripriya Priya[2]

[1]Assistant Professor, Departtment of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., Andhra Pradesh - 522502,India.
Email: pnagesh.qa@gmail.com
[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,Vaddeswaram,522502.
Guntur Dist., Andhra Pradesh,India. Email: venkatalaxmiharipriya.malladi9@gmail.com

Abstract: University student depression is a serious mental health problem, which requires accurate, interpretable and practical computational screening methods. The paper presents a full machine learning system to identify depression, with two different datasets of depression: a student depression dataset on Kaggle and a massive mental health survey in the COVID-19 pandemic. The files contain details on the demographics, academics, behavior, and mental health of people. Quite a bit of preprocessing is performed, including removing null and duplicate data, encoding labels, correcting class asymmetry using SMOTE, selecting features with RFECV and Stratified K-Fold validation, and normalizing with MinMax and Standard scaling. We consider a variety of ML models, including RF, Gradient Boosting, DT, Naive Bayes, LR, Extra Trees, SVM, XGBoost, LightGBM, CatBoost, SGD, LASSO, and MLP. Ensemble Voting and Stacking classifiers are also investigated by us. Performance is measured by standard evaluation measures and the findings indicate that the accuracy is 98.0% and 99.3% in the two sets. SHAP and LIME ensure that models are comprehensible. It can predict real-time sadness with a web application created in Flask and authenticated with SQLite as it proves to be powerful and effective.

“**Keywords:** Depression detection, ensemble learning, student mental health, feature selection, explainable artificial intelligence, SMOTE, Flask-based prediction system, machine learning classification”.

I. INTRODUCTION

One of the most prevalent and crippling mental illnesses, depression has been a global public health problem [1]. University students are susceptible to stress due to academic pressure, social and emotional shifts, financial issues, and lack of direction [2]. These pressures and also the developmental and environmental variables increase the rates of depression in students compared to the general population [3]. Poor mental health care and chronic stigma in low- and middle-income countries are likely to enhance the chances of undiagnosed and untreated depression [4]. Academics and social life: Early detection and support could prevent long-term health, social and health problems. Even though more individuals are becoming aware of student mental health issues, depression can hardly be detected. Conventional methods of diagnostics are based on self-reported questionnaires and interviews with clinicians, which are time-consuming, resource-intensive, and subject to stigmatization-induced bias and underreporting [5]. In turn, not all ailing pupils will be subjected to therapy or delay the consultations with professional help [6]. Subsequently, other studies investigated the data and machine learning of mental health assessment.Current systems have a hard time achieving accuracy,

dependability, projected occurrences, transparency, and scalability. This restricts their academic use [7]. There was no great focus on statistical significance and consistency of performance gains in diverse student populations [8]. People find it hard to use automated ways of diagnosing depression due to loopholes. The present research is based on a mental health survey data to come up with a plausible model in early depression diagnosis among university students. The proposed approach enhances accuracy, clarity, consistency, and reliability of forecasts. To demonstrate the generalizability of the framework and present digestible information to allow mental health professionals and schools to make decisions, this research evaluates the framework in several student groups [9]. This study has the potential to result in scalable, reliable and intelligible mental health screening systems in higher education. This would enable intervention in time and improvement of the student well-being as well as ethical healthcare decision-support norms [10].

II. LITERATURE REVIEW

The recent developments in smart healthcare systems demonstrated that medical conditions could be diagnosed with the help of data-driven algorithms and that mental conditions could be evaluated. Esha et al. [11] also came up with a multi-view attention-based paradigm of lung cancer disability detection and discovered that a combination of clinical markers is more predictive of impairment. Palash and Yousuf [12] explored privacy-sensitive distributed learning to detect lung cancer whereby data is not centralized to train the model. The studies are useful in the diagnostics of physical health, but the majority of them are medical imaging studies and cannot be extended to the mental health analysis and at-risk groups, such as university students.

Mental health is the use of intelligent models in sensitive psychological outcomes. Mumenin et al. [13] constructed a hybrid explainable suicidal ideation identification model on social media as a way to highlight explainability in high-risk mental health models. Akter and Ghosh [14, 15] designed deep learning-based models that detect brain tumors and Alzheimer disease and showed that reliability and robustness are important attributes of clinical decision-support systems. These works concentrate on neurological or image-aided diagnoses and do not comment on the vagueness, subjectivity, and generalization of survey-based data on mental health.

Uncertainty-conscience and multimodal learning are widely used in depression studies. Ahmed et al. [16] introduced a multimodal system of depression classification that approximates uncertainty in order to demonstrate that trust in predictions is enhanced by confidence. Mumenin et al. [17] used machine learning and standardized questionnaires in order to screen the university students on depression, which demonstrated that automated mental health assessment could be done. The studies have been worthwhile, but most have used one model or not testing it, and it is difficult to believe that the conclusions made are strong and can be applied to the real world.

It is agreed by many that healthcare machine learning systems ought to be transparent and reliable. Jia et al. [18] emphasized explainability as a safety aspect, and Ahmad et al. [19] as a clinical decision-making one. Rasheed et al. [20] examined ethical, interpretable, and trustworthy healthcare machine learning applications. These research works include the most vital issues, yet they hardly include models of mental health screening.

Nevertheless, reliable, interpretable and statistically acceptable methods of identifying depression among college students are yet to be established. Accuracy or explainability are considered by many studies without taking robustness or uncertainty into account. Irrespective of these constraints, this paper suggests a holistic paradigm of early depression detection in institutions of learning.

III. Materials And Methods

The method will be able to recognize depression among university students automatically in two sets of surveys. It has the ability to deal with high dimensional, heterogeneous, unevenly distributed data and also capable to extrapolate to groups. SMOTE is based on standardized preprocessing and splitting to equalize classes. Recursive Feature Elimination is a cross-validation-stabilized and stratified, feature selection algorithm that eliminates discriminative features. We discuss numerous supervised models and arrive at the conclusion that the complimentary learners and Voting and Stacking are simpler and enhance model stability and reliability. Speed can be better by scaling features and organizing the data in a way that will allow it to be fast. Explainable AI methods render feature contributions more transparent. An open-source, secure Flask interface provides real-time mental health screening, is scalable, reliable, and open.

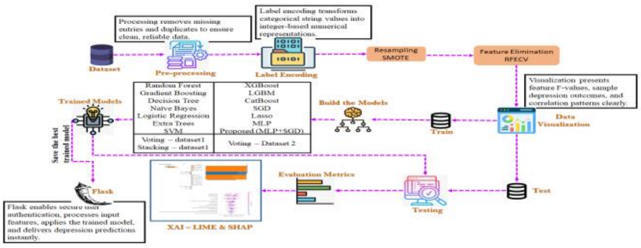


Fig. 1. Proposed Architecture

This system design demonstrates an entire machine learning pipeline of smart data processing and prediction. First, data is ingested, preprocessed and labeled. The data is then resampled by SMOTE and RFECV is employed to drop features. We train and evaluate numerous machine learning models on traditional performance metrics. The system includes system testing, presentation of results, and explainable AI such as LIME and SHAP. A Flask-based interface is provided that allows you to securely access the system and deploy predictions in real time.

A) Dataset Collection

The Student Depression Dataset was obtained in an open-access publicly available source of data and is frequently used in mental health studies. The variables in the dataset are 55 and collected through structured survey responses. These characteristics are demographic, academic, and lifestyle factors, as well as psychiatric signs of depression. The goal variable indicates the level of depression of the kids. This dataset is also quite dimensional and the number of classes is not even and this is similar to the distribution of mental health in the real world. It is an appropriate option to test powerful and generalizable depression detection models since it possesses numerous different features.

	Gender	Age	Academic Pressure	Study Satisfaction	Sleep Duration	Dietary Habits	Have you ever had suicidal thoughts ?	Study Hours	Financial Stress	Family History of Mental Illness	Depression
0	Male	28	2.0	4.0	7-8 hours	Moderate	Yes	9	2	Yes	No
1	Male	28	4.0	5.0	5-6 hours	Healthy	Yes	7	1	Yes	No
2	Male	25	1.0	3.0	5-6 hours	Unhealthy	Yes	10	4	No	Yes
3	Male	23	1.0	4.0	More than 8 hours	Unhealthy	Yes	7	2	Yes	No
4	Female	31	1.0	5.0	More than 8 hours	Healthy	Yes	4	2	Yes	No

Fig.2 Depression Student Dataset

Fig.3 Mental Health Dataset

B) Pre-Processing

The effectiveness of the proposed system in terms of resilience, accuracy, and applicability relies on the effectiveness of data preprocessing to ensure consistent data depression detection, data accuracy, balanced representation, informative feature selection, and unbiased evaluation.

Data Cleaning: The data set was verified, and the accuracy was checked before analysis. The missing values and duplications in records were eliminated to reduce biassed learning and wrong results. There were no null or duplicate data as well as in the data set. This implies that the projections of mental health are credible, legitimate and usable.

Categorical Data Encoding: Supervised learning models were fed with numbers that were the category characteristics on the demographic data and the questionnaire responses. This translation maintains response category differences and allows the user to perform mathematical operations on them which educates them on how behavioral indicators influence the results of depression.

Class Imbalance Handling: The mental health of the depressed and non-depressed were not balanced. Synthetic oversampling brought about equal representation of classes and enhanced minority class learning. This phase eases the detection and recall of depression, reduces biased forecasts,and offers credibility to the healthcare category issues.

Feature Significance Analysis: To identify the relevance of attributes we employed statistical analysis of balanced data using statistical significance tests. This strategy has been used to eliminate weak predictors by measuring the correlations between individual characteristics and the desired outcome. It is better to prioritize the most crucial features first, to enhance learning, consistency, and understanding of the situation in sensitive mental health prediction.

Feature Selection: Iterative feature reduction and cross-validation were used to identify the most accurate predictors.To reduce dimensionality and overfitting, we eliminated unnecessary characteristics. This step will make mental health data easier to understand, generalize, and describe, which is interrelated and subjective.

C) Training and Testing:

The enhanced feature set was divided into separate training and testing subsets that were not combined, allowing performance to be evaluated fairly. This separation ensures that learning and validation of the model are carried out on data distributions that are not related. This type of partitioning is needed to check the extent to which something can generalize and to prevent stated findings being excessively positive. The analysis remains realistic in the real deployment environment through maintaining the proportion of classes in splitting. This renders performance measurement to be reliable and repetitive.

D) Algorithms:

Random Forest: RF is another form of classifier, which incorporates many decision trees to give a more precise and robust judgment. It reduces overfitting risk, and generalizes more on high-dimensional complex data by introducing randomness to both data sampling and feature selection.

Gradient Boosting: The Gradient Boosting algorithm generates a strong predictive model by assembling weak learners sequentially, with each step devoted to correcting errors made in the previous step. This repeated cycle of prediction allows predictions to be more precise and enables you to characterize complex and nonlinear patterns of data.

Decision Tree: The DT provides an easy method of categorizing data by continuously dividing it into similar groups according to the demands of features. It is able to capture nonlinear interactions effectively and it is interpretable and is therefore useful in understanding the decision limits and baseline performance.

Gaussian Naive Bayes: Gaussian Naive Bayes is a kind of classifier which involves the use of probabilities to explain the distribution of features under the assumption of Gaussian. Its simplicity allows it to learn easily and to be stable in high dimensional space, contributing to the efficiency, robustness and diversity of ensemble learning.

Logistic Regression: Logistic Regression makes a linear guess of the likelihoods of all classes based on the linear decision boundary. This renders it highly generalizable and comprehensible. It provides precise predictions and acts as a reliable baseline classifier and thus enables binary classification tasks to be operated with ease and consistency.

Extra Trees: Extra Trees introduces more randomization to the tree building process to achieve more diversity in the ensembles. It is good to use in cases where there are noisy and high-dimensional data and it can be used to make good predictions.

Support Vector Machine: The SVM constructs the most optimal separation boundary by maximizing the margins of classes towards a transformed feature space. It is effective with high-dimensional data distributions that are overlapping and are useful in achieving excellent generalization and strong classification.

XGBoost: XGBoost is an effective boosting system; it employs gradient based learning and regularization to enhance accuracy of predictions. It is able to capture complex relationships between features without overfitting, resulting in highly accurate and robust classification performance.

LightGBM: LightGBM is a gradient boosting model which employs a histogram-based learning to make it faster and even scalable. It can deal well with large and high-dimensional data, finding a solution fast and being able to make correct predictions with minimal effort.

CatBoost: CatBoost is a gradient boosting algorithm that is capable of utilizing a diverse set of features and maintaining prediction bias at a low level. Its gradual learning methodology renders it more capable of generalizing and resilience, which is excellent in structured data with intricate interactions of features.

Stochastic Gradient Descent: A method of learning based on optimization to update model parameters by bits is called Stochastic Gradient Descent. It can work fast and expand, which is why it is simple to adapt and learn to work in environments with a significant amount of data that varies rapidly.

LASSO: LASSO regularizes a model to reduce complexity by promoting the sparsity of learnt parameters. This allows easier comprehension, reduces overfitting and is simpler to generalize as learning is concentrated on the most useful features.

Multilayer Perceptron: Multilayer perceptron is a form of feedforward neural network that is able to learn nonlinear interactions of a complicated nature through layered representations. It has a high modeling capacity, which enhances the degree of predictions, and features of complex relationships are readily captured.

Proposed Ensemble Model: A combination method is a structured method that combines a large number of complementary classifiers in the ensemble model. It reduces prejudice and disparity by equalizing the capacities and deficiencies of all individuals. This renders the model more robust, precise and capable of generalizing highly on various kinds of data.

Min–Max Scaler: The Min-Max Scaler makes features take on a fixed range such that they equally contribute in the learning process. It is a preprocessing step that stabilizes the numbers, accelerates the convergence and allows distance-based and gradient-driven classifiers to interact.

Voting Classifier: Voting Classifier is based on consensus strategy to combine the prediction of a large number of base learners. By utilizing various kinds of classifiers, it reduces individual errors, increases the stability and generalizability of the system and the stability and generalizability of predictions.

Stacking Classifier: In the Stacking Classifier, a higher-level learner combines the output of a large number of base models. It is a hierarchical integration that relies on the advantages of each level to make superior and more precise categorization judgments and minimize bias and volatility.

e) Integration of XAI & Flask Framework:

The addition of XAI to the depression detection framework provides both local and global interpretability of model predictions, thus making it more transparent and trustworthy. We apply a model-agnostic approach to discover the most significant factors that influence individual predictions to generate instance-level explanations. This is the reason why it is easy to comprehend why a decision was taken. Complementary global explanations provide a summary of the behaviour of the features of many samples, and how important they are. This assists us in understanding entirely the functioning of the model and how dependable it is in delicate mental health uses.

It is constructed as an explainability module within a Flask-based framework enabling safe real-time interaction to ensure it can be applied to real life. The interface allows consumers to input survey responses, receive predictions and view explanations that are sensible simultaneously. This seamless integration facilitates its application in reality where stakeholders can receive the correct forecasts together with helpful explanations. This enhances confidence, accountability, and utilization of intelligent mental health screening mechanisms.

IV. Experimental Results

Accuracy: How well a test can distinguish between sick and healthy people is called the accuracy of a test. To determine the degree of accuracy of a test, we need to determine the percentage of the true positives and true negatives out of the cases we examined. In mathematics it can be expressed as:

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \,(1)$$

Precision: Precision is a measure of the proportion of correctly identified cases or samples out of the number of cases identified as positive. The equation to work out the accuracy is then:

$$\mathrm{Precision} = \frac{\text{True Positive}}{\text{True Positive } + \text{ False Positive}} \,(2)$$

Recall: In machine learning, recall is a metric of the ability of a model to locate each and every pertinent instance of a particular class. It is the ratio of the correctly predicted positives to the actual positives. This informs you of how a model is good at capturing incidences of a given class.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \,(3)$$

F1-Score: F1 score is an option to measure the performance of a machine learning model. It combines accuracy and recall scores of the model. The accuracy statistic is the number of times that a model made a valid prediction on the entire dataset.

$$\text{F1 Score} = 2 * \frac{\text{Recall X Precision}}{\text{Recall} + \text{Precision}} * 100(1)$$

AUC-ROC Curve: One method of testing the effectiveness of a classification problem is by testing the AUC-ROC Curve. ROC indicates True Positive and False Positive on the same graph. AUC is used to determine the ability of the model to distinguish the classes. The larger the AUC, the more the model is good at this.

$$\mathrm{AUC} = \sum_{i=1}^{n-1} (\mathrm{FPR}_{i+1} - \mathrm{FPR}_i) \cdot \frac{\mathrm{TPR}_{i+1} + \mathrm{TPR}_i}{2} \qquad (5)$$

Table. 1. Performance Evaluation Table - Dataset

ML Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	kappa	LogLoss
RandomForest	0.842	0.842	0.842	0.842	0.960	0.678	0.305
GradientBoosting	0.901	0.901	0.901	0.901	0.976	0.799	0.208
DecisionTree	0.842	0.841	0.842	0.841	0.836	0.676	5.472
NaiveBayes	0.950	0.951	0.950	0.950	0.991	0.899	0.248
LogisticRegression	0.941	0.941	0.941	0.940	0.992	0.879	0.135
ExtraTrees	0.871	0.872	0.871	0.871	0.957	0.739	0.312
SVM	0.931	0.931	0.931	0.931	0.988	0.859	0.136
XGBoost	0.901	0.901	0.901	0.901	0.978	0.799	0.180
LGBM	0.891	0.891	0.891	0.891	0.974	0.778	0.212
CatBoost	0.881	0.881	0.881	0.881	0.969	0.757	0.231
SGD	0.941	0.941	0.941	0.941	0.990	0.879	0.164
LASSO	0.950	0.952	0.950	0.950	0.990	0.899	0.133
MLP	0.911	0.923	0.911	0.909	0.992	0.814	0.231
Proposed	0.931	0.934	0.931	0.930	0.963	0.857	2.394
VotingClassifier	0.980	0.981	0.980	0.980	0.997	0.960	0.147
StackingClassifier	0.980	0.981	0.980	0.980	0.997	0.960	0.089

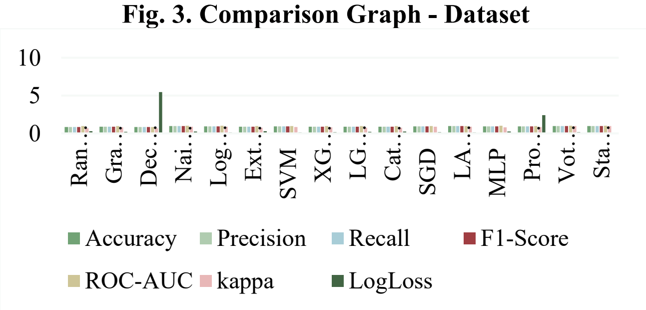
This performance analysis examines several machine learning models and compares them based on accuracy, precision, recall, F1-score, ROC-AUC, kappa, and log loss to demonstrate that the ensemble methods of Voting and Stacking classifiers are the most effective.

Table.2 Performance Evaluation Table - Depression Student Dataset

ML Model	Accuracy	Precision	Recall
RandomForest	0.982	0.982	0.982
GradientBoosting	0.982	0.982	0.982
DecisionTree	0.970	0.970	0.970
NaiveBayes	0.866	0.867	0.866

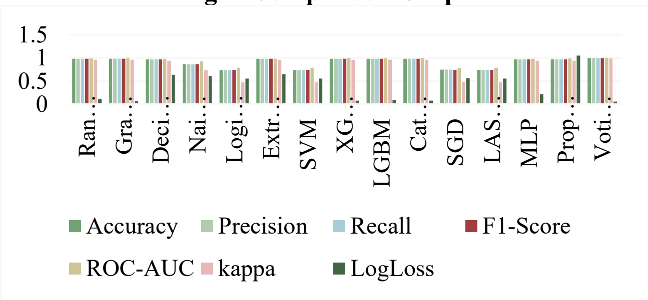
This performance appraisal table indicates the relationship between various machine learning models on various metrics. It demonstrates that ensemble and boosting methods are more accurate, have higher ROC-AUC, and kappa, and the VotingClassifier provides the highest overall performance.

Fig. 3. Comparison Graph - Dataset



This comparison graph indicates the performance of the various machine learning models in terms of accuracy, precision, recall, F1-score, ROC-AUC, kappa and log loss. It leaves no doubt that ensemble classifiers are more reliable than any other classifier.

Fig 4. Comparison Graph



This comparison graph depicts the performance of various machine learning models on various metrics. It demonstrates that ensemble and boosting algorithms achieve higher accuracy, ROC-AUC and kappa at much lower log loss than single classifiers.

Fig.5 Enter the input data

In this image, the DDNet Depression Prediction interface is shown as a user receives personal data, such as age, gender, academic pressure, sleep duration, stress, and mental health history to be predicted regarding the possibility of depression.

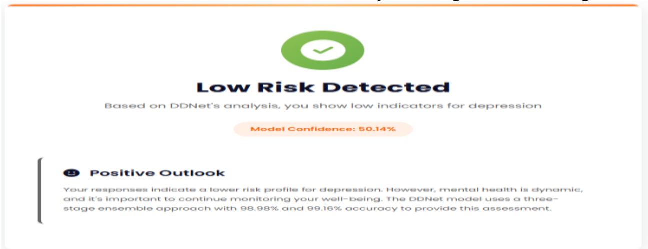


Fig.6 Predicted Result

This output display indicates that the probability of depression is low and there is a possibility of a 50.14% probability of the model being accurate. It also indicates that mental health still should be monitored with the help of the predictive analysis provided by DDNet.

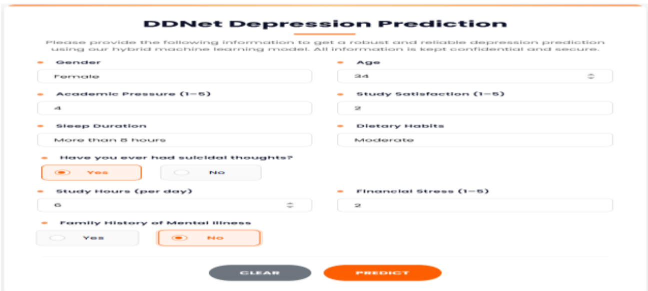


Fig. 7 Enter the input data

The following picture depicts the DDNet Depression Prediction input screen, which gathers data concerning the user, including their age, sex, academic stress, sleep duration, stressors, and mental health signals in an attempt to identify their susceptibility to depression.

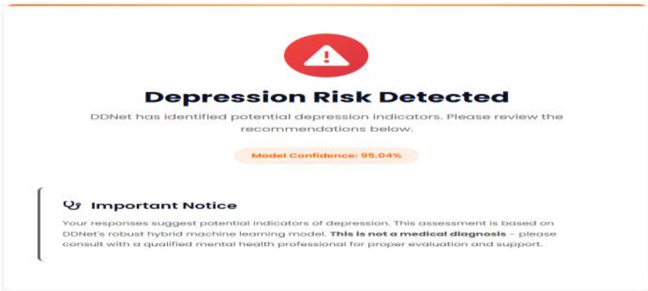


Fig. 8 Predicted Result

This output screen indicates that there is a detected risk of depression, the model confidence level is 95.04. It demonstrates that it can be depressive and implies that users should read the recommendation and visit mental health professionals to have an examination.

V. CONCLUSION

The DDNet system is decent at identifying depressed college students with the combination of hybrid and ensemble learning approaches. The capacity to distinguish between the datasets was enhanced by preprocessing, optimizing features, and class balancing. In the case of student data, the ensemble voting and stacking classifiers achieved 98.0% accuracy, and voting achieved 99.3% accuracy in the case of the mental health survey data. When SHAP was combined with LIME, it became clearer and easier to understand and when applied to Flask, it became possible to make predictions in real time. There is good, precise and comprehensible identification of depression in the framework. This renders it an effective option in reliable mental health screening and application in educational institutions.

Longitudinal and real-time behavioral data can be included in future studies to enable continuous surveillance, and mobile and wearable technologies can be used to passively obtain data and give immediate danger alerts. Generalizability and dynamic updates can be promoted with multilingual adaptability and adaptive learning, and the integration with institutional counseling systems can be used to promote quick interventions and data-based mental health policies in higher education.

REFERENCES

1. Ruan, Z., Yang, P., Huang, J., Yang, K., Lv, Y., & Zhang, Z. Q. (2024). Automatic depression detection among higher education students based on DeepFM. IEEE Transactions on Instrumentation and Measurement, 73, 1-10.
2. Nykoniuk, M., Basystiuk, O., Shakhovska, N., & Melnykova, N. (2025). Multimodal data fusion for depression detection approach. Computation, 13(1), 9.
3. Rony, J. H., Syeed, M. M., Khan, R. H., Fatema, K., Hossain, M. S., & Uddin, M. F. (2024, September). Predicting Depression Among University Students: a Comparative Assessment of ML & DL models using XAI. In 2024 23rd International Symposium on Communications and Information Technologies (ISCIT) (pp. 175-180). IEEE.
4. Taoussi, C., Lyaqini, S., Metrane, A., & Hafidi, I. (2025). Enhancing machine learning and deep learning models for depression detection: A focus on smote, roberta, and cnn-lstm. Informatica, 49(14).
5. Su, W. (2025). Depression prediction model based on deep learning and psychological feature extraction. Discover Artificial Intelligence, 5(1), 387.
6. B. J. Sadock, V. A. Sadock, and P. Ruiz, Comprehensive Textbook of Psychiatry, vol. 1. Lippincott Williams & Wilkins Philadelphia, 2000.
7. M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, ''Depression detection from social network data using machinelearning techniques,'' Health Inf. Sci. Syst., vol. 6, no. 1, pp. 1–12, Dec. 2018.
8. A.A.Choudhury,Md.R.H.Khan,N.Z.Nahim,S.R.Tulon,S.Islam,and A. Chakrabarty, ''Predicting depression in Bangladeshi undergraduates using machine learning,'' in Proc. IEEE Region 10 Symp. (TENSYMP), Jun. 2019, pp. 789–794.
9. L. Gao, Y. Xie, C. Jia, and W. Wang, ''Prevalence of depression among Chinese university students: A systematic review and meta-analysis,'' Sci. Rep., vol. 10, no. 1, p. 15897, Sep. 2020.
10. S.E.Gollust, D. Eisenberg, and E. Golberstein, ''Prevalence and correlates of self-injury among university students,'' J. Amer. College Health, vol. 56, no. 5, pp. 491–498, Mar. 2008.
11. J. F. Esha, T. Islam, M. A. M. Pranto, A. S. Borno, N. Faruqui, M. A. Yousuf, A. Azad, A. S. Al-Moisheer, N. Alotaibi, S. A. Alyami, and M. A. Moni, ''Multi-view soft attention-based model for the classification of lung cancer-associated disabilities,'' Diagnostics, vol. 14, no. 20, p. 2282, Oct. 2024.
12. M.I. Adnan Palash and M. A. Yousuf, ''A federated learning-based model for the detection of lung cancer from CT scan images,'' in Proc. 6th Int. Conf. Electr. Eng. Inf. Commun. Technol. (ICEEICT), May 2024, pp. 741–745.
13. N. Mumenin, M. R. Basar, A. B. M. K. Hossain, M. A. Hossain, M. L. Hasan, and M. N. Hossain, ''Suicidal ideation detection from social media texts using an interpretable hybrid model,'' in Proc. 6th Int. Conf. Electr. Inf. Commun. Technol. (EICT), Dec. 2023, pp. 1–6.

A. *Akter, N. Nosheen, S. Ahmed, M. Hossain, M. A. Yousuf, M. A. A. Almoyad, K. F. Hasan, and M. A. Moni, ''Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor,'' Exp. Syst. Appl., vol. 238, Mar. 2024, Art. no. 122347.*

14. T. Ghosh, M. I. A. Palash, M. A. Yousuf, M. A. Hamid, M. M. Monowar, and M. O. Alassafi, ''A robust distributed deep learning approach to detect Alzheimer's disease from MRI images,'' Mathematics, vol. 11, no. 12, p. 2633, Jun. 09, 2023.
15. S. Ahmed, M. Abu Yousuf, M. M. Monowar, A. Hamid, and M. O. Alassafi, ''Taking all the factors we need: A multimodal depression classification with uncertainty approximation,'' IEEE Access, vol. 11, pp. 99847–99861, 2023.
16. N. Mumenin, A. B. M. Kabir Hossain, M. A. Hossain, P. P. Debnath, M. Nusrat Della, M. M. Hasan Rashed, A. Hossen, M. R. Basar, and M. S. Hossain, ''Screening depression among university students utilizing GHQ-12 and machine learning,'' Heliyon, vol. 10, no. 17, Sep. 2024, Art. no. e37182.
17. Y. Jia, J. McDermid, T. Lawton, and I. Habli, ''The role of explainability in assuring safety of machine learning in healthcare,'' IEEE Trans. Emerg. Topics Comput., vol. 10, no. 4, pp. 1746–1760, Oct. 2022.
18. M. A. Ahmad, A. Teredesai, and C. Eckert, ''Interpretable machine learning in healthcare,'' in Proc. IEEE Int. Conf. Healthcare Informat. (ICHI), Jun. 2018, pp. 559–560.
19. K. Rasheed, A. Qayyum, M. Ghaly, A. Al-Fuqaha, A. Razi, and J. Qadir, ''Explainable, trustworthy, and ethical machine learning for healthcare: A survey,'' Comput. Biol. Med., vol. 149, Oct. 2022, Art. no. 106043.

A. A. Abdullah, M. M. Hassan, and Y. T. Mustafa, ''A review on Bayesian deep learning in healthcare: Applications and challenges,'' IEEE Access, vol. 10, pp. 36538–36562, 2022.

20. F.Al-urjman,H.Zahmatkesh,andL.Mostarda,''Quantifyinguncertainty in Internet of Medical Things and big-data services using intelligence and deep learning,'' IEEE Access, vol. 7, pp. 115749–115759, 2019.
21. M.Chua,D.Kim,J.Choi,N.G.Lee,V.Deshpande,J.Schwab,M.H.Lev, R. G. Gonzalez, M. S. Gee, and S. Do, ''Tackling prediction uncertainty in machine learning for healthcare,'' Nature Biomed. Eng., vol. 7, no. 6, pp. 711–718, Dec. 2022.
22. M. R. Islam, S. Daria, R. Das, and M. R. Hasan, ''A nationwide dataset on the mental health of the Bangladeshi population due to the COVID-19 pandemic,'' Data Brief, vol. 38, Oct. 2021, Art. no. 107347.
23. R. P. Auerbach, P. Mortier, R. Bruffaerts, J. Alonso, C. Benjet, P. Cuijpers, K. Demyttenaere, D. D. Ebert, J. G. Green, P. Hasking, and E. Murray, ''Who world mental health surveys international college student project: Prevalence and distribution of mental disorders,'' J. abnormal Psychol., vol. 127, no. 7, p. 623, 2018.
24. E. Karyotaki, P. Cuijpers, Y. Albor, J. Alonso, R. P. Auerbach, J. Bantjes, R. Bruffaerts, D. D. Ebert, P. Hasking, G. Kiekens, S. Lee, M. McLafferty, A. Mak, P. Mortier, N. A. Sampson, D. J. Stein, G. Vilagut, and R. C. Kessler, ''Sources of stress and their associations with mental disorders among college students: Results of the world health organization world mental health surveys international college Student initiative,'' Frontiers Psychol., vol. 11, p. 1759, Jul. 2020.
25. T. M. Evans, L. Bira, J. B. Gastelum, L. T. Weiss, and N. L. Vanderford, ''Evidence for a mental health crisis in graduate education,'' Nature Biotechnol., vol. 36, no. 3, pp. 282–284, Mar. 2018.
26. N. Meda, S. Pardini, P. Rigobello, F. Visioli, and C. Novara, ''Frequency and machine learning predictors of severe depressive symptoms and suicidal ideation among university students,'' Epidemiol. Psychiatric Sci., vol. 32, p. 42, Jan. 2023.
27. L. Zhang, S. Zhao, Z. Yang, H. Zheng, and M. Lei, ''An artificial intelligence tool to assess the risk of severe mental distress among college students in terms of demographics, eating habits, lifestyles, and sport habits: An externally validated study using machine learning,'' BMC Psychiatry, vol. 24, no. 1, p. 581, Aug. 2024.

B. *Sutter, R. Chiong, G. S. Budhi, and S. Dhakal, ''Predicting psycholog ical distress from ecological factors: A machine learning approach,'' in Proc. 34th Int. Conf. Ind., Eng. Appl. Appl. Intell. Syst., Kuala Lumpur, Malaysia. Springer, Jul. 2021, pp. 341–352.*