International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

An Integrated Learning Approach for Detecting Depression in University Students

R R Shantha Spandana[1], N Sukanya[2], A Uma Maheswara Reddy[3]

[1]Assistant Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
E-mail: shanthaspandana@gmail.com, ORC-ID: https://orcid.org/0009-0003-4236-1250
[2]P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
E-mail: sukanyadiya619@gmail.com , ORC-ID: https://orcid.org/0009-0007-2364-1715
[3]Assistant Professor, Department of MCA, Annamacharya PG College, Rajampet, Kadapa Dist. AP.
E-mail: maheshmca5535@gmail.com

Abstract: Depression among college students has become a major mental health issue that needs accurate and reliable computer screening methods. Conventional grading methods often have problems with uneven distribution of classes, duplicate features, and limited interpretability, which makes them less useful in the real world. This framework uses two different sets of data that are linked to depression: a Kaggle dataset on student depression and a large-scale mental health survey that was conducted during the COVID period and included demographic, academic, behavioral, and psychological questions. A lot of work goes into the preprocessing, like getting rid of nulls and duplicates, labeling, balancing the data with SMOTE, getting rid of features with RFECV and Stratified K-Fold validation, and making sure everything is normal with MinMax and Standard scaling. Random Forest, Gradient Boosting, Decision Tree, Gaussian Naive Bayes, Logistic Regression, Extra Trees, Support Vector Machine, XGBoost, LightGBM, CatBoost, SGD, LASSO, and MLP are some of the machine learning classifiers that are used. There are also hybrid ensemble strategies that combine MLP, SGD, and CatBoost. We check how well the model works by looking at its accuracy, precision, recall, F1-score, ROC-AUC, Cohen's kappa, and log loss. Voting and Stacking classifiers worked best for the sadness student dataset, getting a maximum accuracy of 98.0%. A Voting classifier worked best for the mental health dataset, getting a 99.3% accuracy. SHAP and LIME analysis make sure that things can be explained. Real-time prediction from user-provided inputs is possible with a Flask-based interface that uses SQLite for authentication. This shows that the sadness detection capability is strong, understandable, and deployable.

*“**Index Terms:** Depression Detection, Hybrid Machine Learning, Ensemble Models, University Students, Explainable AI, Mental Health Screening”.*

1. INTRODUCTION

Mental health disorders have become a big public health issue around the world. Depression is one of the most common and severe conditions in the world [1]. Young people in college are especially at risk because they are under a lot of pressure to do well in school, make new friends, deal with money problems, and not know what they want to do with their lives after college [2]. These stresses, along with developmental and environmental factors, make college students more likely than the general population to show signs of depression [3]. The problem is even worse in low- and middle-income countries, where it's harder to get mental health services, which raises the risks [4]. Because of this, early detection and help are needed to lessen the long-term effects of untreated sadness on health, social life, and schoolwork.

Even though people are becoming more aware of the problem, the ways that are currently used to spot sadness in students have major flaws. Self-reported surveys and clinical interviews are used a lot in traditional diagnostic methods, but they take a lot of time and are often affected by stigma, underreporting, and subjective bias [5]. Because of this, a lot of students who are affected don't get identified or wait too long to get help [6]. Recent research has looked into using data to help with mental health evaluations, but many of the current solutions have trouble finding the right mix between being able to predict performance and being reliable, clear, and scalable [7]. Aside from that, not much attention has been paid to checking whether performance gains seen are statistically significant and reliable across a wide range of student groups [8]. These holes make it harder for automated systems that can spot depression to be used in real academic situations.

The main goal of this study is to use survey-based mental health data to build a strong and reliable system for finding sadness early in college students. The suggested method aims to make predictions more accurate while also making sure that results are consistent and reliable. The study stresses clear decision-making and strict performance evaluation in addition to being able to make good predictions across multiple datasets. This study tries to show that the framework can be used with any group of students by carefully testing it on a variety of student groups. The study also wants to find out what causes student depression so that mental health workers and schools can better help students who are depressed [9].

This study is important because it could lead to tools for mental health screening in higher education that are scalable, reliable, and easy to understand. A reliable early detection system can help with timely interventions, lower student dropout rates, and general health and happiness of students. The suggested work also meets the practical and moral needs for using decision-support systems in sensitive healthcare settings by focusing on openness, understanding of uncertainty, and statistical validation. The main goal of this study is to connect academic research with real-life mental health uses. This will help people make better decisions and make universities healthier [10].

2. LITERATURE REVIEW

New developments in smart healthcare systems have shown that data-driven models are becoming more useful for diagnosing medical problems and judging mental health. Esha et al. [11] suggested a multi-view attention-based approach to classify disabilities related to lung cancer. They showed how combining different clinical features can help improve prediction. In the same way, Adnan Palash and Yousuf [12] looked into privacy-preserving learning for finding lung cancer using distributed data, focusing on training models together without sharing data centrally. These studies show good results in medical imaging areas, but they mostly look at physical health problems and don't look at how to test for mental health problems in sensitive groups like college students.

More and more, studies that focus on mental health have used smart models to look at sensitive psychological effects. Mumenin et al. [13] showed a hybrid method that can be interpreted to find suicidal thoughts in social media text. This shows how important it is for high-risk mental health applications to be able to be explained. At the same time, Akter et al. [14] and Ghosh et al. [15] created reliable deep learning systems for finding brain tumors and Alzheimer's disease, respectively. This shows how important it is for clinical decision-support systems to be reliable and strong. But these works are mostly about neurological or imaging-based diagnoses and don't talk enough about how to deal with doubt or how to generalize survey-based mental health data, which are inherently subjective and noisy.

People who study mental health are becoming more interested in uncertainty-aware and multimodal methods. Ahmed et al. [16] created a multimodal framework for classifying sadness that includes uncertainty approximation. This shows that confidence estimation can make predictions more reliable. Mumenin et al. [17] used standard surveys and machine learning to check for depression among college students, showing that it is possible to do mental health screenings for students automatically. Even though these studies are helpful, a lot of them use single-model methods or don't fully test their results on a lot of different datasets, which makes it harder to be sure that they are robust and can be used in the real world.

Many people agree that reliability and being able to explain things are very important for using intelligent systems in healthcare. Jia et al. [18] stressed that explainability is a key part of making sure safety in machine learning apps used in healthcare. Ahmad et al. [19] talked more about how important interpretable models are for helping doctors make decisions, and Rasheed et al. [20] gave a broad look at ethical, believable, and reliable machine learning in healthcare. These works lay the groundwork for mental health screening, but they often stay conceptual or domain-general and don't offer unified solutions that deal with accuracy, interpretability, uncertainty, and statistical confirmation all at the same time.

Overall, past study shows that intelligent healthcare modeling has come a long way, but there are still problems with making reliable, understandable, and statistically proven systems for finding depression in college students. A lot of studies only look at certain parts, like how accurate something is or how easy it is to explain. They don't look at reliability and uncertainty knowledge together. The current study was inspired by these problems and tries to fix them by suggesting a complete and reliable way to find depression early on. The focus is on making the

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

method strong, clear, and useful for real-life mental health situations in school settings.

3. MATERIALS AND METHODS

DDNet is meant to accurately and reliably find college students who are depressed by using two datasets that work well together: the Depression Student Dataset and a Mental Health Dataset from the COVID-19 era. The framework works with a structured process that includes importing data, cleaning it up, transforming it, optimizing features, supervised learning, and deploying it. SMOTE is used to fix class imbalance, and RFECV with Stratified K-Fold validation and, for the second dataset, RFE-based top-feature selection are used to improve the usefulness of features. A lot of different classifiers are trained and compared using a wide range of performance measures. These include tree-based [21], boosting [22], linear [23], probabilistic [24], and neural models [25]. By combining different ways of learning, a three-stage ensemble design makes the system even more robust. Some other methods that get better results are Standard Scaling with Voting for Mental Health Dataset, Stacking for Depression Student Dataset, and Min–Max scaling with Voting. LIME and SHAP make sure that models can be understood, and a Flask-based web interface with SQLite-backed authentication lets models be deployed in real time, allowing scalable early-screening applications.

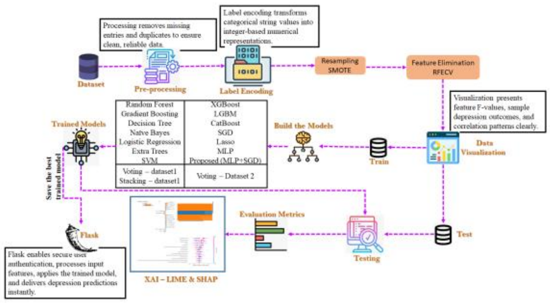

Fig.1 Proposed Architecture

Figure 1 shows the DDNet system architecture, which is made up of a structured pipeline with layers for data collection, preprocessing, model creation, evaluation, and deployment. At first, raw datasets on student mental health are cleaned, encoded, resampled using SMOTE, and their features are optimized using RFECV and RFE. To find the best model, several machine learning [26] and ensemble classifiers are learned and then compared. The finished hybrid ensemble is put together with XAI features and used through a Flask-based web interface to predict depression risks in real time and let users interact with it.

a) Dataset Collection:

The Depression Student Dataset, which we got from the Kaggle repository, has 502 records of college students with 10–18 demographic, academic, behavioral, and psychological characteristics, as shown in Fig. 2. There is a binary depression label (252 positive, 250 negative) that takes into account gender, age, academic pressure, sleep duration, financial stress, suicidal thoughts, and family background. Its structured survey style and nearly even class distribution make it a good choice for supervised depression prediction and comparing models.

	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

The Mental Health Dataset comes from a scientific data repository that has been reviewed by other scientists. It has 672 survey answers with 55 psychometric and sociodemographic attributes, as seen in Fig. 3. There are DASS-based depression signs, sleep-quality measures, lifestyle factors, and demographic information as variables. There are also two types of depression labels: 262 positive and 410 negative. Its multidimensional structure and modest class imbalance make it possible to do full psychological modeling and reliable testing of advanced grouping and classification methods.

	1. Express your consent to participate in the research and processing of anonymous data for scientific purposes.	2. Age range in years	3. Sex	4. Weight and height ratio/BMI score	5. Marital status	6. Education level	7. Occupation	8. Economic status	9. Residence area	10. Living status	...	11. During the past month, how many times, you feel too cold?	12. During the past month, how many times, you feel too hot?	13. During the past month, how many times, you had bad dreams?	14. During the past month, how many times, you have pain during sleep?	15. During the past month, how many times, you have trouble in sleeping because of any other reasons?	16. During the past month, how often have you taken medicine to help you sleep?	17. During the past month, how many times you did not sleep due to any program or other important case?	18. During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?	19. During the past month, how would you rate your sleep quality overall?
0	Agree	15-20	Male	Overweight	Married	Higher secondary	Unemployed	High	Urban	Without family	...	Three or more in week (3)	Less than once a week (1)	Not during last month (0)	Not during last month (0)	Not during last month (0)	Not during last month (0)	Not during last month (2)	Once or twice a week (2)	Fairly bad (2)
1	Agree	Above 30	Male	Normal	Married	Illiterate	Unemployed	Low	Rural	With family	...	Less than once a week (1)	Less than once a week (1)	Once or twice a week (2)	Three or more in week (3)	Three or more in week (3)	Not during last month (0)	Not during last month (2)	Not during last month (3)	Fairly good (1)
2	Agree	Above 30	Female	Normal	Unmarried	Primary	Unemployed	Middle	Rural	With family	...	Not during last month (2)	Three or more in week (3)	Less than once a week (1)	Not during last month (0)	Less than once a week (1)	Not during last month (0)	Not during last month (2)	Not during last month (3)	Fairly bad (2)
3	Agree	Above 30	Female	Normal	Married	Graduate/above	Student	Middle	Rural	With family	...	Less than once a week (1)	Not during last month (2)	Less than once a week (1)	Less than once a week (1)	Once or twice a week (2)	Not during last month (0)	Not during last month (2)	Not during last month (3)	Fairly good (1)
4	Agree	Above 30	Male	Normal	Married	Graduate/above	Student	Middle	Rural	With family	...	Not during last month (3)	Once or twice a week (2)	Less than once a week (1)	Less than once a week (1)	Once or twice a week (2)	Not during last month (0)	Less than once a week (1)	Not during last month (3)	Fairly bad (2)

5 rows × 55 columns

Fig.3 Mental Health Dataset

b) Pre-Processing:

To improve the quality of the data, make sure that the analysis is consistent, and make sure that features are relevant, a structured data preprocessing system was put in place. These steps set up a solid base for accurate classification and depression recognition that can be used in many situations.

1. Data Preprocessing: Both the Depression Student Dataset and the Mental Health Dataset went through a lot of data preparation to make sure that the structure was consistent and the data was reliable for analysis. Missing values and duplicate records were found and deleted in a planned way to stop unfair learning and extraneous effects on model parameters. Standardizing attribute formats ensures that all demographic, behavioral, and psychometric traits are shown in the same way. These improvements made the data more reliable, cut down on noise, and set a solid base for feature tuning and classification steps that came later.

2. Label Encoding: In order to make both types of data work with machine learning algorithms, categorical attributes were turned into numerical ones. This change kept the semantic differences between categories while turning qualitative answers into structured numerical inputs. Encoding made sure that psychometric answers, lifestyle indicators, and demographic factors could be processed correctly in statistical and optimization-based models. By allowing mathematical processes and distance calculations, label encoding made computing much faster and easier while also making pattern extraction more accurate.

3. Resampling the Data: To fix the class imbalance seen in the datasets, a synthetic oversampling technique was used to make sure that minority classes were better represented. Based on similarities in the feature space, artificial samples were made to even out the distributions of the classes without just duplicating the data. This method lessened the impact of biased decision boundaries and made it easier to spot cases of sadness. Balanced data helped with better recall, fewer wrong classifications of vulnerable students, and better model robustness and generalization across different population groups.

4. Feature Elimination: A recursive elimination approach and stratified cross-validation were used for feature selection to find the most discriminative predictors. Iteratively, irrelevant and redundant traits were taken away based on how much they helped with performance. This left a smaller but more useful set of features. For the Mental Health Dataset, the most important characteristics were tweaked even more to make the dataset more dimensional. By focusing on statistically significant factors, this process made computations faster, reduced the risk of overfitting, made the data easier to understand, and increased the accuracy of classification.

5. Data Visualization: Exploratory and statistical visualization methods were used to look at the relationships between attributes, the significance of features, and the distribution of classes. Analysis of variance scores measured how well each predictor could tell the difference between groups, and correlation matrices showed how factors were connected. Distributional plots showed that the resampling methods worked. These visual insights helped with choosing the right model, made things easier to understand, and gave real-world evidence for feature optimization and ensemble building in the depression detection framework.

c) Training and Testing:

An 80:20 split ratio was used to separate both datasets into training and testing groups so that model development and validation could be done in a planned way. The training subset was used to learn model parameters and set decision boundaries. The testing subset, on the other hand, was only used for evaluating performance on its own. This structured partitioning reduces overfitting, stops data leaks, and lets an objective evaluation of generalization ability across student records that haven't been seen.

d) Algorithms:

Random Forest makes predictions more accurate by using bootstrapped samples to build multiple decision trees and then adding up the results of these trees through majority vote. This ensemble method lowers the variation, stops overfitting, and makes generalization better. Its ability to describe interactions that don't follow a straight line helps with accurate depression classification [27].

$$Gini = 1 - \sum_{i=1}^{C} (Pi)^2 \qquad (1)$$

Through sequential learning, where each weak learner fixes mistakes made in earlier rounds, Gradient Boosting makes accuracy better. It improves decision boundaries and lowers bias by minimizing loss functions over and over again. This method for boosting improves the ability to tell the difference between things and the accuracy of predictions.

By dividing the feature space over and over again into homogeneous groups, Decision Tree makes classification easy to understand. Its hierarchical layout shows how relationships and interactions don't work in a straight line. Even though it's pretty simple, it provides analytical clarity and can be used as a benchmark for success.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

$$I\ (i) = 1 - \sum_{i=1}^{k} p_i^2 \qquad (2)$$

Gaussian Naïve Bayes uses probabilistic thinking with the idea of conditional independence. It uses Gaussian functions to model feature distributions and figures out posterior probabilities for classification [28]. It can be used as a lightweight baseline learner because it is fast on computers and stable in adaptation.

A logistic function is used to estimate the chances of each class in logistic regression, which describes decisions that have clear boundaries. Parameter optimization makes steady convergence and feature contributions that can be understood possible. It offers dependable baseline performance and works well with ensemble frameworks.

$$P(y = 1\ |X) = \frac{1}{1 + e^{-(W^T x + b)}} \qquad (3)$$

Randomized split selection during tree building in Extra Trees makes the system more robust. It lowers variance and improves generalization by making trees more diverse and adding their results together. This method catches complex patterns quickly and easily while keeping the ability to handle large amounts of data.

Support Vector Machine creates the best dividing hyperplanes by making the space between classes as big as possible. Modeling nonlinear relationships in spaces with more dimensions is possible with kernel changes. Its margin-based optimization makes it better at generalization and less likely to overfit high-dimensional data [29].

$$minimize\ \frac{1}{2}\ ||W||^2 + C \sum_{i=1}^{n} \xi_i \qquad (4)$$

Regularized gradient boosting is done by XGBoost with sequential tree building and shrinkage control. Structured optimization and complexity costs make predictions more accurate while keeping them from fitting too well. Its effective learning approach makes it possible to classify depression in a way that is scalable and accurate.

$$\hat{y}_i = \sigma\left(\sum_{k=1}^{K} f_k(x_i)\right), f_k \in F \qquad (5)$$

A leaf-wise tree growth method with depth constraints is used by LGBM to speed up convergence and improve accuracy. The processing load is lower with histogram-based learning. This system handles features with a lot of dimensions well, makes it easier to scale, and has good generalization performance.

CatBoost improves the accuracy of classification by using ordered boosting and more advanced category handling. It makes the model more stable and less likely to overfit by reducing forecast shift and gradient bias. Its structured learning mechanism helps it work consistently even when the features are spread out in different ways.

SGD uses incremental gradient updates to find the best model parameters over and over again. This method makes it possible to use less computing power and make the feature areas bigger. Its regularized optimization helps keep the trade-offs between bias and variance in check and keeps the accuracy of its predictions stable.

LASSO uses L1 regularization to make sure that model values are sparse. It does integrated feature selection and lessens overfitting by shrinking less important features toward zero. This makes it easier to understand and makes it easier to generalize.

MLP learns complex feature representations by using nonlinear activation functions and hidden layers that are linked to each other. Iteratively, backpropagation-based tuning makes network weights better. Its deep architecture records complex patterns of behavior, which makes it more accurate at classifying [30] and flexible.

$$\hat{y} = f\big(W^L f\big(W^{L-1} \cdots f(W^1 X + b^1) + b^{(L-1)}\big) + b^L\big) \qquad (6)$$

Through hierarchical aggregation, the three-stage group brings together different base learners. As the process goes through steps, the initial predictions get better and better until they reach meta-level fusion. This organized integration makes complementary learning better, lowers model bias, and makes the system more robust overall.

This mixed group uses a mix of linear, probabilistic, and neural learners along with a collection of voting choices. Normalized feature scaling makes sure that each model makes an equal input. Through different types of learning, consensus-based prediction lowers variance and improves classification stability.

The stacking structure takes the outputs of base learners and feeds them into a meta-level classifier. This stacked learning method improves the ability to tell the difference between things and makes predictions more accurate. By using the benefits of complementary modeling, hierarchical integration makes generalization stronger.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

This boosting-based ensemble takes results from more advanced gradient boosting models and adds them together to make the consensus more accurate. Scaling and choosing features makes data more consistent and useful. The diversified boosting integration makes the model more stable, lowers its volatility, and makes it more accurate at making predictions.

e) Integration of XAI & Flask Framework:

Adding explainable artificial intelligence (XAI) to the framework for finding depression makes it more open and trustworthy by letting users understand model results both locally and globally. A model-agnostic method is used to make instance-level explanations by figuring out the most important factors that affect each guess. This gives a clear picture of why decisions are made. Overall feature importance and behavior across multiple samples are summed up in complementary global explanations. This helps us get a full picture of model thinking and reliability in sensitive mental health applications.

For easy use, the explainability module is combined with a Flask-based structure that allows safe, real-time communication. The interface lets people fill out surveys, get predictions, and see details that are easy to understand all at the same time. This seamless integration makes it easier to use in real life, giving stakeholders access to accurate predictions along with useful explanations. This boosts trust, responsibility, and the use of intelligent mental health screening systems.

4. EXPERIMENTAL RESULTS

Accuracy: How well a test can tell the difference between sick and healthy people is called its accuracy. To get an idea of how accurate a test is, we should figure out what percentage of cases are true positives and true negatives. In terms of math, this can be written as

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \quad (7)$$

Precision: Precision is the percentage of correctly classified cases or samples compared to those that were correctly classified as positives. So, here is the method to figure out the precision:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} \quad (8)$$

Recall: In machine learning, recall is a metric that shows how well a model can find all the important instances of a certain class. It shows how well a model captures instances of a certain class. It is calculated by dividing the number of correctly predicted positive observations by the total number of real positives.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (9)$$

F1-Score: The F1 score is a way to rate the correctness of a machine learning model. It takes a model's accuracy and recall scores and adds them together. The accuracy metric counts how many times, across the whole dataset, a model made a correct guess.

$$F1\ Score = 2 * \frac{Recall \text{ X } Precision}{Recall + Precision} * 100 \quad (10)$$

AUC-ROC Curve: The AUC-ROC Curve shows how well a classification problem is solved at different benchmark levels. The True Positive Rate is plotted against the False Positive Rate by ROC. AUC measures how well the model can tell the difference between classes; a higher AUC means the model works better.

$$AUC = \sum_{i=1}^{n-1} (FPR_{i+1} - FPR_i) \cdot \frac{TPR_{i+1} + TPR_i}{2} \quad (11)$$

Cohen Kappa: A statistical test called Cohen's Kappa (º) is used to find out how much two raters (or judges, viewers, etc.) agree on how to group things into categories. It works really well when choices are arbitrary and the categories are nominal, which means they don't naturally fit in a certain order.

$$Kappa(k) = \frac{P_o - P_e}{1 - P_e} \quad (12)$$

Table.1 Performance Evaluation - Depression Student Dataset

ML Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
RF	0.842	0.842	0.842	0.842	0.960
GB	0.901	0.901	0.901	0.901	0.976
DT	0.842	0.841	0.842	0.841	0.836
NB	0.950	0.951	0.950	0.950	0.991
LR	0.941	0.941	0.941	0.940	0.992
ETs	0.871	0.872	0.871	0.871	0.957
SVM	0.931	0.931	0.931	0.931	0.988

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

XGB	0.901	0.901	0.901	0.901	0.978
LGBM	0.891	0.891	0.891	0.891	0.974
CatBoost	0.881	0.881	0.881	0.881	0.969
SGD	0.941	0.941	0.941	0.941	0.990
LASSO	0.950	0.952	0.950	0.950	0.990
MLP	0.911	0.923	0.911	0.909	0.992
Proposed	0.931	0.934	0.931	0.930	0.963
Voting Classifier	0.980	0.981	0.980	0.980	0.997
Stacking Classifier	0.980	0.981	0.980	0.980	0.997

The results shown in Table.1 show that the Stacking Classifier performs better overall than other models. It gets the most consistent results across all evaluation measures and shows strong generalization ability for depression detection.

Table.2 Performance Evaluation - Mental Health Dataset

ML Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
RF	0.982	0.982	0.982	0.982	0.992
GB	0.982	0.982	0.982	0.982	0.995
DT	0.970	0.970	0.970	0.970	0.982
NB	0.866	0.867	0.866	0.866	0.924
LR	0.738	0.738	0.738	0.738	0.789
ETs	0.982	0.982	0.982	0.982	0.979
SVM	0.738	0.738	0.738	0.738	0.787
XGB	0.982	0.982	0.982	0.982	0.996
LGBM	0.982	0.982	0.982	0.982	0.993
CatBoost	0.982	0.982	0.982	0.982	0.994
SGD	0.744	0.745	0.744	0.743	0.781
LASSO	0.738	0.738	0.738	0.738	0.789
MLP	0.970	0.970	0.970	0.970	0.982
Proposed	0.970	0.970	0.970	0.970	0.987
Voting Classifier	0.993	0.993	0.993	0.993	0.999

Table.2 shows that the Voting Classifier has the best total performance of all the models that were tested. It has better predictive stability and a strong ability to generalize for tasks that involve classifying mental health risks.

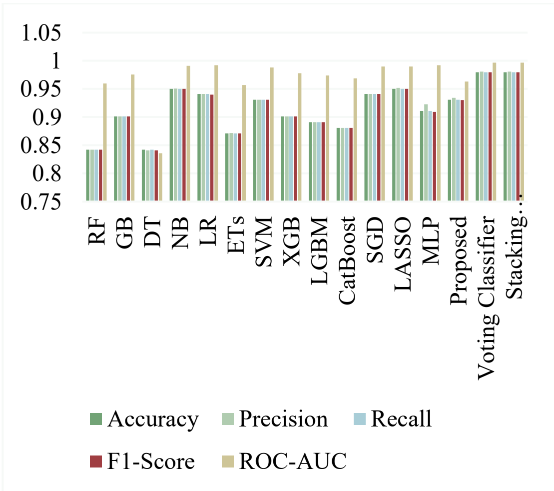

Fig.4 Comparison Graph – Depression Student Dataset

As shown in Fig.4, the comparison graph shows that the Stacking Classifier performs better than the others. It constantly gets higher scores for accuracy, precision, recall, F1-Score, and ROC-AUC, which highlights how reliable and well it classifies things.

Graph.2 Comparison Graph

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

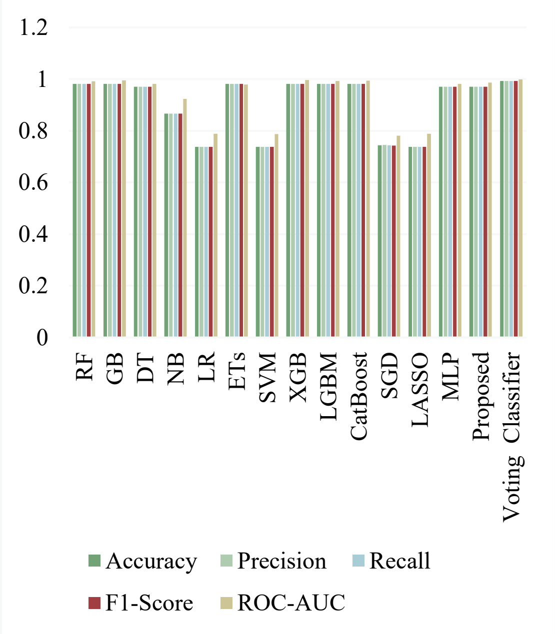

Fig.5 Comparison Graph – Mental Health Dataset

Figure 5 shows a comparison graph that shows how well the Voting Classifier did compared to the others. It consistently had high values for accuracy, precision, recall, F1-Score, and ROC-AUC, which shows how strong it is and how well it can tell the difference between mental health risks.

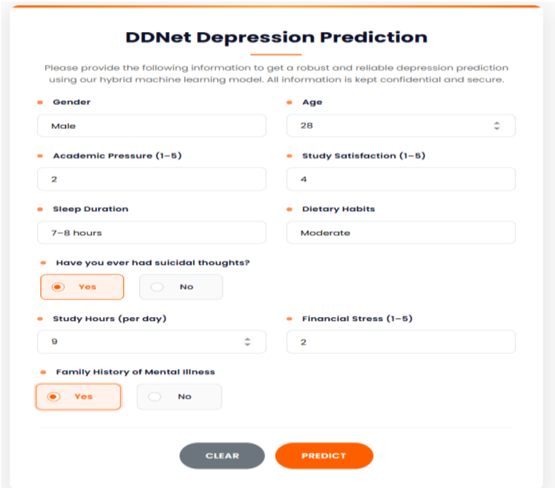

Fig.6 Enter the Input Data

Figure 6 shows the DDNet Depression Prediction interface. It has a structured input panel where users can enter information about their demographics, academics, lifestyle, and mental health. This helps the system make accurate and personalized estimates about their risk of depression.

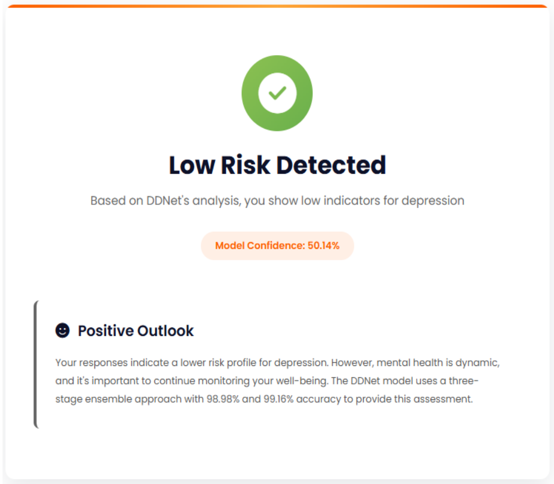

Fig.7 Prediction Result

The DDNet output interface shows a Low-Risk depression result in Fig.7, along with model confidence, positive outlook advice, and suggestions for ongoing mental health tracking. This gives users clear and actionable predictive insights.

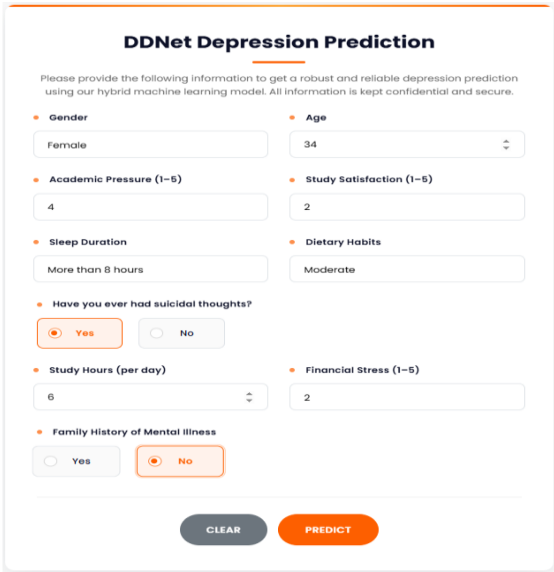

Fig.8 Enter the Input Data

Figure 8 shows that the DDNet input interface has an easy-to-use panel for entering grouped information, such as demographic, academic, lifestyle, and mental health indicators. This makes it easier to get correct data for assessing the risk of depression.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

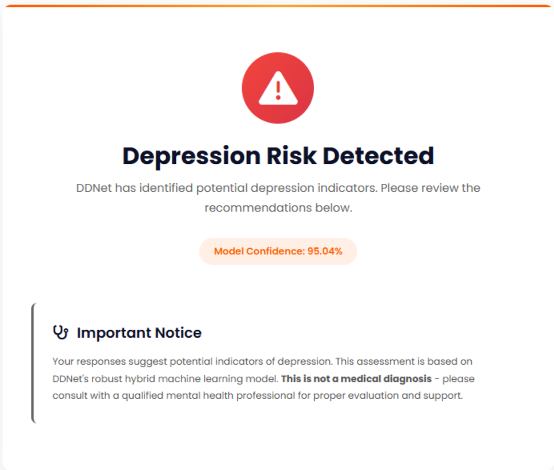

Fig.9 Prediction Result

The DDNet output interface shows a high-risk depression result in Fig.9, showing strong model confidence, highlighting key risk indicators, and making clear suggestions for next steps and professional mental health advice.

5. CONCLUSION

The DDNet framework that was created is very good at finding patterns of sadness in college students by using strong hybrid and ensemble learning methods. Full preprocessing, feature optimization, and class balance made it much easier to tell the difference between the two sets of data. Using the student sadness dataset for experiments confirmed that ensemble-based Voting and Stacking classifiers worked the best, getting 98.0% accuracy and doing better than individual classifiers like Random Forest, SVM, and MLP. This shows why putting together different kinds of students can help with generalization and steadiness. The Voting Classifier did very well on the mental health survey dataset, with an accuracy of 99.3%. It also had consistently high precision, recall, F1-score, and ROC-AUC values, showing that it could reliably classify even complicated, high-dimensional data. When explainable AI methods like SHAP and LIME were used together, it became easier to see how each feature contributed, which increased trust and made the predictions easier to understand. Also, using trained models with a Flask-based interface made it possible to predict sadness in real time based on inputs from users, proving that it was useful. Overall, the results show that using hybrid ensemble learning along with explainability and deployable interfaces makes depression identification accurate, reliable, and easy to understand in school settings.

In the future, improvements could include adding longitudinal and real-time behavioral data to the framework. This would allow for constant tracking of student mental health trends. Passive data collection and early danger alerts may be made easier by integrating with mobile and wearable platforms. Adapting to different languages and cultures can help make things more generalizable across a wide range of academic groups. Adding adaptive learning methods could make it possible for models to be improved on the fly as new data comes in. Closer integration with university counseling systems may also make it easier to provide timely help, make personalized support suggestions, and make mental health policies based on data in higher education settings.

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] Naresh, M., Gudditti., M., Viswanath, & SunilKumarReddy, M.T. (2014). Distributed Utility-Based Energy Efficient Cooperative Medium Access Control in MANETS.

[7] Islam, M. R., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine

learning techniques. Health Information Science and Systems, 6(1), 1–12.

[8] Choudhury, A. A., Khan, M. R. H., Nahim, N. Z., Tulon, S. R., Islam, S., & Chakrabarty, A. (2019, June). Predicting depression in Bangladeshi undergraduates using machine learning. In Proceedings of the IEEE Region 10 Symposium (TENSYMP) (pp. 789–794).

[9] Gao, L., Xie, Y., Jia, C., & Wang, W. (2020). Prevalence of depression among Chinese university students: A systematic review and meta-analysis. Scientific Reports, 10(1), 15897.

[10] Gollust, S. E., Eisenberg, D., & Golberstein, E. (2008). Prevalence and correlates of self-injury among university students. Journal of American College Health, 56(5), 491–498.

[11] Esha, J. F., Islam, T., Pranto, M. A. M., Borno, A. S., Faruqui, N., Yousuf, M. A., Azad, A., Al-Moisheer, A. S., Alotaibi, N., Alyami, S. A., & Moni, M. A. (2024). Multi-view soft attention-based model for the classification of lung cancer-associated disabilities. Diagnostics, 14(20), 2282.

[12] Palash, M. I. A., & Yousuf, M. A. (2024, May). A federated learning-based model for the detection of lung cancer from CT scan images. In Proceedings of the 6th International Conference on Electrical Engineering, Information and Communication Technology (ICEEICT) (pp. 741–745).

[13] Mumenin, N., Basar, M. R., Hossain, A. B. M. K., Hossain, M. A., Hasan, M. L., & Hossain, M. N. (2023, December). Suicidal ideation detection from social media texts using an interpretable hybrid model. In Proceedings of the 6th International Conference on Electrical, Information and Communication Technology (EICT) (pp. 1–6).

[14] Akter, A., Nosheen, N., Ahmed, S., Hossain, M., Yousuf, M. A., Almoyad, M. A. A., Hasan, K. F., & Moni, M. A. (2024). Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor. Expert Systems with Applications, 238, 122347.

[15] Ghosh, T., Palash, M. I. A., Yousuf, M. A., Hamid, M. A., Monowar, M. M., & Alassafi, M. O. (2023). A robust distributed deep learning approach to detect Alzheimer’s disease from MRI images. Mathematics, 11(12), 2633.

[16] Kumar, K., Udaya Suriya Rajkumar, D., Viswanath, G., & Mahalakshmi, J. (2024). A Hybrid Particle Swarm Optimization and C4.5 for Network Intrusion Detection and Prevention System. International Journal of Computing, 23(1), 109-115. https://doi.org/10.47839/ijc.23.1.3442

[17] Mumenin, N., Kabir Hossain, A. B. M., Hossain, M. A., Debnath, P. P., Nusrat Della, M., Hasan Rashed, M. M., Hossen, A., Basar, M. R., & Hossain, M. S. (2024). Screening depression among university students utilizing GHQ-12 and machine learning. Heliyon, 10(17), e37182.

[18] Jia, Y., McDermid, J., Lawton, T., & Habli, I. (2022). The role of explainability in assuring safety of machine learning in healthcare. IEEE Transactions on Emerging Topics in Computing, 10(4), 1746–1760.

[19] Ahmad, M. A., Teredesai, A., & Eckert, C. (2018, June). Interpretable machine learning in healthcare. In Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI) (pp. 559–560).

[20] M, Chaithanya., Viswanath, G., Dunna, Nikitha Rao,, G, Prathyusha. (2023). A Real Time Online Food Ording Application Based Django Restfull Framework Juni Khyat journal, 13(9), 154-162.

[21] Abdullah, A. A., Hassan, M. M., & Mustafa, Y. T. (2022). A review on Bayesian deep learning in healthcare: Applications and challenges. IEEE Access, 10, 36538–36562.

[22] Al-Turjman, F., Zahmatkesh, H., & Mostarda, L. (2019). Quantifying uncertainty in Internet of Medical Things and big-data services using intelligence and deep learning. IEEE Access, 7, 115749–115759.

[23] Chua, M., Kim, D., Choi, J., Lee, N. G., Deshpande, V., Schwab, J., Lev, M. H., Gonzalez, R. G., Gee, M. S., & Do, S. (2022). Tackling prediction uncertainty in machine learning for healthcare. Nature Biomedical Engineering, 7(6), 711–718.

[24] Islam, M. R., Daria, S., Das, R., & Hasan, M. R. (2021). A nationwide dataset on the mental health of the Bangladeshi population due to the COVID-19 pandemic. Data in Brief, 38, 107347.

[25] Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., Demyttenaere, K., Ebert, D. D., Green, J. G., Hasking, P., & Murray, E. (2018). WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. Journal of Abnormal Psychology, 127(7), 623.

[26] Karyotaki, E., Cuijpers, P., Albor, Y., Alonso, J., Auerbach, R. P., Bantjes, J., Bruffaerts, R., Ebert,

D. D., Hasking, P., Kiekens, G., Lee, S., McLafferty, M., Mak, A., Mortier, P., Sampson, N. A., Stein, D. J., Vilagut, G., & Kessler, R. C. (2020). 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 in Psychology, 11, 1759.

[27] Evans, T. M., Bira, L., Gastelum, J. B., Weiss, L. T., & Vanderford, N. L. (2018). Evidence for a mental health crisis in graduate education. Nature Biotechnology, 36(3), 282–284.

[28] Viswanath, G., Abirami, V., & Prathyusha, G. (2024). Hybrid Feature Extraction With Machine Learning To Identify Network Attacks. International Journal Of Hrm And Organizational Behavior, 12(3), 217-228.

[29] Zhang, L., Zhao, S., Yang, Z., Zheng, H., & Lei, M. (2024). 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, 24(1), 581.

[30] Sutter, B., Chiong, R., Budhi, G. S., & Dhakal, S. (2021, July). Predicting psychological distress from ecological factors: A machine learning approach. In Proceedings of the 34th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 341–352). Springer.