THE PSYCHOLOGICAL IMPACT OF COVID-19 ON PEOPLE DURING LOCKDOWN AND ITS EFFECT ON STUDENT EDUCATION: AN ENSEMBLE LEARNING APPROACH

[1]Ranjini Mupra P, [2]Dr B Ashok, and [3]Dr Bindulal T S

[1]Research Scholar, Department of Computer and Information Science, Annamalai University

[2]Assistant Professor, Department of Computer and Information Science, Annamalai University

[3]Assistant Professor, Department of Computer Science, Government College, Nedumangaud, Kerala

Abstract

The COVID-19 pandemic has had a profound impact on various aspects of life, particularly on mental health and education. This study explores the psychological effects of COVID-19 lockdowns on individuals and the educational disruptions faced by students. This research proposes an advanced ensemble learning approach to analyze two datasets: the "Psychological Impact of COVID-19 on People" and the "Effect of COVID-19 on Student Education." The goal is to accurately predict psychological outcomes such as depression, anxiety, and stress, as well as the educational impact on students. The proposed methodology involves several key steps: data preprocessing, feature scaling, handling class imbalances, feature selection using Recursive Feature Elimination with Cross-Validation (RFECV), and model building using ensemble learning techniques.

Initially, categorical variables such as gender and student status are encoded numerically, and missing values are imputed using the mean of the respective columns. The features are then standardized to ensure uniformity. RFECV is applied to select the most significant features, enhancing the model's performance by eliminating less important features. To handle class imbalances in the data, we utilize the Synthetic Minority Over-sampling Technique (SMOTE), dynamically adjusting the k_neighbors parameter to ensure valid configurations. This step is crucial for creating a balanced dataset that enhances model performance. The model employs three primary classifiers: Random Forest, Support Vector Machine (SVM), and Gradient Boosting. Each classifier undergoes hyperparameter tuning using GridSearchCV, optimizing for accuracy. The best models from each classifier are then combined using a Stacking Classifier, which leverages the strengths of all three classifiers to improve prediction accuracy. The model is trained and evaluated on two datasets. For the psychological impact dataset, the targets are depression, anxiety, and stress. For the student education dataset, the outcome variable is predicted. The ensemble model's performance is assessed using metrics such as precision, recall, F1-score, and accuracy. The proposed model obtained an accuracy of 98.33% in predicting the psychological impact of COVID-19 on people, and 98.46% in predicting the impact of COVID-19 on student education. The results demonstrate the significant improvements of the ensemble model in prediction accuracy, achieving robust performance metrics.

Keywords: COVID-19, Psychological Impact, Lockdown, Student Education, Ensemble Learning, RFECV, SMOTE, GridSearchCV

1. Introduction

The COVID-19 pandemic has brought about an unprecedented global crisis, affecting virtually every aspect of life. The rapid spread of the virus led to the implementation of stringent lockdown measures worldwide, aiming to curb transmission and protect public health. While these measures were essential for controlling the outbreak, they also triggered significant disruptions in daily routines, social interactions, and economic activities. Among the most profoundly affected areas were mental health and education, particularly for students who faced abrupt shifts to remote learning environments [1].

1.1 Psychological Impact of Lockdowns

Lockdowns, though essential for controlling the pandemic, have had profound psychological impacts. The abrupt halt to normalcy, coupled with prolonged social isolation, economic uncertainty, and the constant fear of infection, has led to increased levels of anxiety, depression, and stress across different demographic groups. The mental health toll is particularly severe among vulnerable populations, including those with preexisting mental health conditions, individuals facing financial hardships, and those living alone. The disruption of social support systems and the reduced access to mental health services during lockdown periods have further exacerbated these issues [2].

1.2 Impact on Student Education

The educational sector has experienced significant upheaval due to the pandemic. The closure of schools and universities forced a rapid shift to online learning, creating a myriad of challenges for students and educators. The sudden transition to remote education exposed disparities in access to technology and internet connectivity, with students from disadvantaged backgrounds being disproportionately affected. Moreover, the lack of face-to-face interaction, reduced engagement, and the challenges of adapting to new learning environments have adversely impacted students' academic performance and overall well-being [3].

1.3 Need for Comprehensive Analysis

Understanding the dual impact of the pandemic on mental health and education requires a comprehensive analytical approach. Traditional methods may fall short in capturing the complex, interrelated effects of these unprecedented times. Advanced machine learning techniques, particularly ensemble learning methods, offer a powerful solution by integrating diverse datasets and uncovering intricate patterns that single models might miss. Ensemble learning combines the strengths of multiple machine learning models, improving prediction accuracy and providing robust insights into the multifaceted impacts of the pandemic.

1.4 Objectives of the Study

This study aims to explore the psychological impact of COVID-19 lockdowns on the general population and the specific effects on student education. By employing an ensemble learning approach, we seek to:

1. Quantify the levels of psychological distress, including anxiety, depression, and stress, during lockdown periods.
2. Assess the impact of psychological distress on academic performance and educational outcomes.
3. Identify high-risk groups and the factors contributing to increased vulnerability.
4. Provide actionable insights to inform targeted interventions and policy decisions.

1.5 Significance of the Study

The significance of this study lies in its potential to guide policymakers, educators, and healthcare providers in addressing the dual challenges posed by the pandemic. By offering a detailed analysis of the psychological and educational impacts of COVID-19, this research can help shape effective strategies to mitigate these effects and promote resilience in the face of ongoing and future crises. As the world continues to navigate the repercussions of the pandemic, understanding and addressing its psychological and educational impacts remain crucial for fostering a healthier and more equitable society.

2. Related works

Muhammad A S et al. [4] provides a comprehensive overview of the emergence and spread of COVID-19. The authors delve into the origins of the SARS-CoV-2 virus, tracing its initial outbreak in Wuhan, China, and discussing its rapid global dissemination. The article thoroughly examines the modes of transmission, highlighting the virus's highly contagious nature. It also provides an in-depth analysis of the characteristics of human coronaviruses, comparing SARS-CoV-2 with other coronaviruses like SARS-CoV and MERS-CoV. Pramod Soni [5] explores the environmental impact of the lockdown measures implemented during the COVID-19 pandemic in India. The author provides a detailed analysis of how different phases of lockdown affected atmospheric conditions, focusing on changes in air quality and pollution levels. The study leverages a variety of environmental data to demonstrate significant improvements in air quality, attributed to reduced industrial activities, vehicular traffic, and other anthropogenic emissions during the lockdown periods. Xiong J et al. [6] presents a thorough examination of the pandemic's

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

psychological effects on the general populace. The authors systematically review numerous studies to elucidate the widespread mental health challenges induced by the COVID-19 crisis, including heightened levels of anxiety, depression, stress, and other psychological disorders. They explore various demographic factors, such as age, gender, and pre-existing health conditions, that influence the extent of mental health impacts. The article also discusses the potential long-term psychological repercussions and emphasizes the urgent need for public health interventions and mental health support services.

Aras Bozkurt et al. [7] offers a comprehensive examination of how the COVID-19 pandemic has reshaped the educational landscape. Through a meta-narrative review, the authors synthesize a wide range of studies to analyze the multifaceted effects of the pandemic on education systems worldwide. The article discusses the rapid transition to online learning, highlighting both the challenges and opportunities that arose from this shift. It examines issues such as digital divide, accessibility, and the varying levels of preparedness among institutions and educators. Bakul J Parekh et al. [8] provides a poignant examination of the pandemic's effects on the mental and emotional well-being of children in India. The authors highlight the significant disruptions to daily life, education, and social interactions that have profoundly impacted children's psychological health. The study details how prolonged confinement, lack of peer interaction, and the stress experienced by parents have contributed to increased anxiety, depression, and behavioral issues among children. Additionally, the article underscores the exacerbation of existing socio-economic disparities, with vulnerable children facing greater challenges due to limited access to digital resources and mental health support. Thi Kieu Khanh Ho et al. [9] presents an innovative method for enhancing the accuracy of COVID-19 detection through the use of chest X-ray images. The authors propose a feature-level ensemble technique that combines multiple deep learning models to improve diagnostic performance. By integrating features extracted from different convolutional neural networks (CNNs), the ensemble model achieves a higher level of precision in distinguishing COVID-19 cases from other pulmonary conditions. The study's results demonstrate significant improvements in accuracy, sensitivity, and specificity compared to individual models.

Parisa Gifani et al. [10] presents a sophisticated approach for COVID-19 detection using CT scans. The authors employ an ensemble method that leverages transfer learning with deep convolutional neural networks (CNNs) to enhance the accuracy and reliability of COVID-19 diagnosis. By combining multiple pre-trained CNN models, the study effectively harnesses the strengths of each model to improve diagnostic performance. The results indicate that the ensemble model outperforms individual CNN models in terms of accuracy, sensitivity, and specificity. Aqeel M et al. [11] delivers a critical examination of the mental health impacts of varying lockdown measures on students. The authors conducted a follow-up study to compare the levels of mental wellbeing, anxiety, depression, and quality of life among students during full and partial lockdowns over a five-month period. The study highlights significant differences in mental health outcomes between the two lockdown phases, with students experiencing higher levels of anxiety and depression

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

during the full lockdown compared to the partial lockdown. The research underscores the importance of considering the intensity and duration of lockdown measures on students' mental health. Meenakshi Shukla et al. [12] offers a detailed exploration of the emotional and psychological impact of the COVID-19 pandemic and lockdown measures on young people in India. The study presents novel data highlighting the increased levels of worry, anxiety, and emotional distress experienced by youth during various phases of the lockdown. The authors employed comprehensive survey methodologies to capture the nuanced changes in emotions and concerns as the lockdown progressed. The findings underscore the significant toll that prolonged social isolation and disruption of daily routines have on young individuals' mental health. This research contributes valuable insights into the specific challenges faced by young people in India, emphasizing the need for targeted mental health interventions and support systems to help them cope with the ongoing and future impacts of the pandemic.

3. The Proposed Model

The COVID-19 pandemic has significantly impacted mental health and education worldwide, necessitating robust models to understand and mitigate these effects. The proposed model aims to predict the psychological outcomes and educational disruptions caused by the pandemic using ensemble learning techniques. The methodology includes data preprocessing, feature scaling, and handling class imbalances. The feature selection process is conducted using Recursive Feature Elimination with Cross-Validation (RFECV), which iteratively removes the least important features to enhance model performance by retaining only the most significant ones. By leveraging Random Forest, Support Vector Machine (SVM), and Gradient Boosting classifiers, and combining them into a Stacking Classifier, the model enhances prediction accuracy. The inclusion of SMOTE for handling class imbalances ensures that the model can effectively learn from imbalanced data, providing a comprehensive and reliable analysis of COVID-19's impact. This approach not only improves the accuracy of predictions but also offers valuable insights for policymakers and educators to address the challenges posed by the pandemic. Figure 1 depicts the working process of the proposed ensemble method.

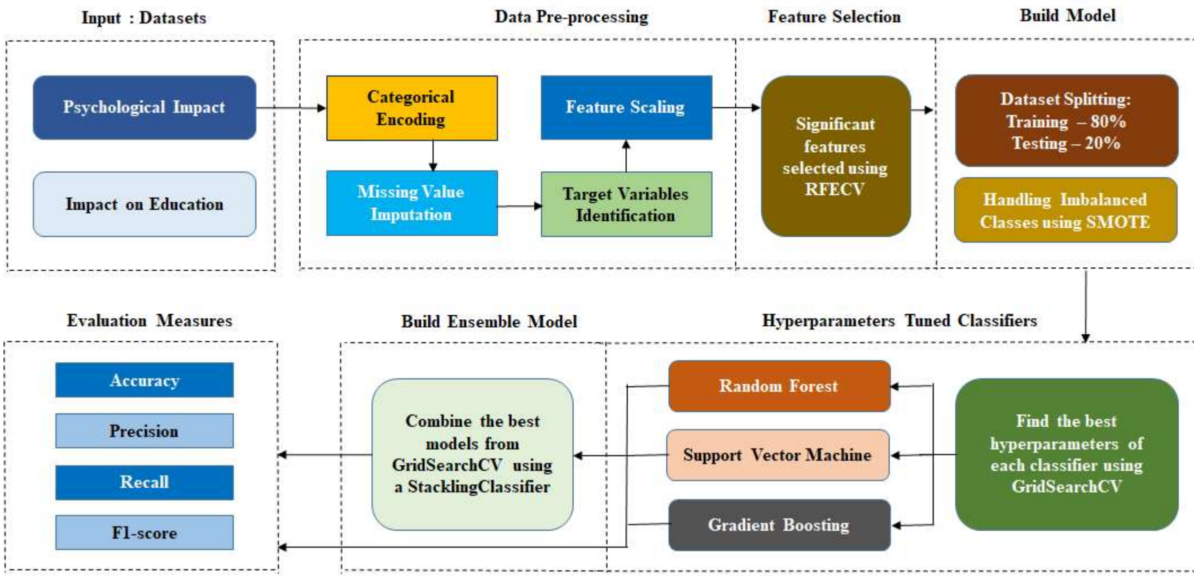


Figure 1: Block diagram of the Proposed Model

3.1 Data Pre-processing

Data pre-processing is a crucial step in the development of predictive models. This step involves transforming raw data into a format that is suitable for machine learning algorithms, ensuring that the model can learn effectively from the data. In the proposed ensemble model, data pre-processing includes handling missing values, encoding categorical variables, and standardizing numerical features.

3.1.1 Handling Missing Values

Handling missing values is an essential step in data pre-processing, as missing data can lead to biased estimates, reduced statistical power, and incorrect conclusions. There are several techniques for handling missing data, and one commonly used method is imputation. Imputation involves replacing missing values with substituted ones based on statistical techniques. Mean imputation involves replacing missing values in a dataset with the mean value of the non-missing values in the same column. This method is simple and can be effective when the proportion of missing data is relatively small.

$$Imputed\ value = \frac{\sum_{i=1}^{n} x_i}{n} \quad (1)$$

where:

x_i are the observed values in the column, and n is the number of non-missing values.

3.1.2 Encoding Categorical Variables

Encoding categorical variables is a vital step in data pre-processing, transforming non-numeric data into a numerical format that machine learning algorithms can utilize effectively. One common method for encoding categorical variables is label encoding, which assigns a unique integer to each category. For instance, if the categorical variable C has k distinct categories $c_1, c_2, \ldots, c_k$, label encoding maps each category c_i to an integer i, where $i \in \{0,1,\ldots,k-1\}$.

3.1.3 Standardizing Numerical Features

Standardizing numerical features is a key step in data pre-processing that ensures numerical variables contribute equally to a model's performance. Standardization transforms the features to have a mean of zero and a standard deviation of one. This process adjusts the scale of the data, making it easier for machine learning algorithms to converge and perform optimally. Standardization can be expressed as,

$$z_i = \frac{x_i - \mu}{\sigma} \tag{2}$$

where:

z_i is the standardized value of the feature, x_i is the original value of the feature, μ is the mean of the feature values, and σ is the standard deviation of the feature values.

3.2 Feature Selection

Feature selection is a critical step in building the proposed model to predict the psychological outcomes and educational disruptions caused by the COVID-19 pandemic. By identifying and retaining the most relevant features, feature selection helps to improve the model's performance and interpretability. It reduces the dimensionality of the data, which in turn minimizes the risk of overfitting, enhances the model's generalization capability, and reduces computational complexity. In the context of this study, employing Recursive Feature Elimination with Cross-Validation (RFECV) ensures that only the most significant features are used for training, thereby enhancing the robustness and accuracy of the predictions. RFECV recursively removes the least important features and builds the model with the remaining features. This process is repeated until the optimal number of features is reached, which maximizes the model’s performance based on cross-validation. The steps involved in this process are explained below,

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

Initialize the Estimator:

- Choose a machine learning estimator (e.g., Random Forest, SVM) that assigns weights to features based on their importance.

Rank Features:

- Train the estimator on the initial set of features.
- Rank the features based on their importance scores.
  Let X be the feature matrix and y be the target vector. The estimator f is trained on X to predict y: $f(X) \rightarrow y$
- The feature importances or weights www are determined by the estimator:
  $w = importance\ (f, X)$ (3)

Recursive Elimination:

- Remove the least important features.
- Retrain the estimator on the reduced set of features.
- Repeat this process until only a specified minimum number of features is left.
- At each iteration t, the least important feature X_{least} is removed:

  $X^{(t+1)} = X^t \backslash X_{least}$ (4)

  The model is retrained on the reduced feature set $X^{(t+1)}$:

  $f(X^{(t+1)}) \rightarrow y$ (5)

Cross-Validation:

- The performance of the model is evaluated using cross-validation at each iteration:
  $CV_t = \frac{1}{k} \sum_{i=1}^{k} Performance(f, X^{(t)}, y)$ (6)
  where k is the number of cross-validation folds.

Select Optimal Features:

- The optimal set of features X_{opt} is the one that maximizes the cross-validation performance:
  $X_{opt} = argmax_{X^{(t)}} CV_t$ (7)
  where $X^{(t)}$ is the set of features at iteration t.
- The selected optimal features are fed into the splitting process for training and testing purposes of the model.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

3.3 Build the Model

Model building for the proposed approach involves two critical steps, beginning with data splitting and addressing class imbalances. The dataset is first divided into training and testing sets to evaluate the model's performance on unseen data. Typically, an 80-20 split is used, where 80% of the data is reserved for training and 20% for testing. Given the nature of the datasets, class imbalances are common, particularly when dealing with psychological outcomes such as depression, anxiety, and stress. To tackle this, the Synthetic Minority Over-sampling Technique (SMOTE) is employed. These processes are explained in detail below.

3.3.1 Dataset Splitting

Dataset splitting is a crucial step in building a machine learning model, ensuring that the model's performance is evaluated on unseen data to prevent overfitting and assess its generalizability. In the proposed model, the dataset is split into training and testing sets. This involves partitioning the data into two subsets: one used to train the model and the other to test its performance. Let X be the feature matrix with m samples and n features, and y be the target vector with m samples. The dataset splitting process can be represented as:

Dataset Splitting : Algorithm 1
Given: (X,y) Let N be the total number of samples Let t be the test size ratio (e.g., 0.2) $N_{test} = \lceil t \times N \rceil$ $N_{train} = N - N_{test}$ $(X_{train}, X_{test}) = (X[: N_{train}], X[N_{train} :])$ $(y_{train}, y_{test}) = (y[: N_{train}], y[N_{train} :])$

Where $\lceil \cdot \rceil$ denotes the ceiling function, which ensures N_{test} is rounded up to the nearest integer. This process ensures that the model is trained on one subset of data and validated on another, providing an unbiased evaluation of its performance. This step is crucial for developing robust models that generalize well to unseen data, reducing the risk of overfitting and ensuring reliable predictions.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

3.3.2 Handling Imbalanced Classes

Handling imbalanced classes is critical in developing effective machine learning models, especially when predicting rare events or conditions, such as diagnosing diseases. One popular technique for addressing class imbalance is the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE works by generating synthetic samples for the minority class to create a more balanced dataset. The processes involved in this step are explained below,

Identify the Minority Class: Determine which class has fewer samples (the minority class) and which has more (the majority class).

For Each Minority Class Sample:

- Find its k-nearest neighbors (typically k=5).
- Randomly select one or more of these neighbors.

Generate Synthetic Samples: For each selected neighbor, create a synthetic sample by interpolating between the minority sample and the neighbor. The interpolation is done as follows:

$$x_{new} = x_i + \lambda \cdot (x_j - x_i) \quad (8)$$

where:

x_{new} is the generated synthetic sample, x_i is a minority class sample, x_j is one of the k-nearest neighbors of x_i, and λ is a random number between 0 and 1. After addressing class imbalances, the training data is used to build the model.

3.4 Classification Process

The classification process in the proposed model involves utilizing multiple machine learning algorithms to predict the psychological impact of COVID-19 on individuals and the educational disruptions faced by students. Initially, the dataset undergoes preprocessing steps, including the imputation of missing values, encoding of categorical variables, and standardization of numerical features. Recursive Feature Elimination with Cross-Validation (RFECV) is applied to select the most significant features. The model employs three primary classifiers: Random Forest, Support Vector Machine, and Gradient Boosting. The detailed description of each classification process is given below.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

3.4.1 Random Forest (RF)

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Process:

- **Bootstrapping**: Random Forest generates multiple subsets of the training data using bootstrapping.
- **Tree Construction**: For each subset, a decision tree is constructed. During this process:
    - A random subset of features is selected at each split point.
    - The best split is chosen based on a criterion like Gini impurity or entropy.
- **Aggregation**: The final prediction is made by aggregating the predictions of all the trees. Given T decision trees, the Random Forest prediction for an input x is:
  $$\hat{y} = majority\_vote\{h_t(x)\}_{t=1}^{T} \quad (9)$$
  where $h_t(x)$ is the prediction of the t^{th} decision tree.

3.4.2 Support Vector Machine (SVM)

SVM is a supervised learning model that finds the optimal hyperplane that maximizes the margin between different classes in a high-dimensional space.

Process:

- **Feature Transformation**: Data is transformed into a higher-dimensional space using a kernel function.
- **Optimal Hyperplane**: SVM finds the hyperplane that maximizes the margin between the nearest points of different classes.
- **Prediction**: New data points are classified based on which side of the hyperplane they fall. For a linear SVM, the decision function is:
  $$f(x) = w \cdot x + b \quad (10)$$
  where w is the weight vector, x is the input vector, and b is the bias term.

3.4.3 Gradient Boosting (GB)

Gradient Boosting is an ensemble technique that builds models sequentially, with each new model attempting to correct errors made by the previous models. It focuses on minimizing the loss function through gradient descent.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

Process:

- **Initial Model**: Start with an initial prediction, often the mean of the target variable.
- **Iterative Improvement**: For each iteration:
    - Compute the residuals of the current model.
    - Fit a new weak learner to these residuals.
    - Update the model by adding the predictions of the new learner, scaled by a learning rate.
- **Aggregation**: The final model is the weighted sum of all the weak learners.
  For each iteration m:
  $$F_m(x) = F_{m-1}(x) + \alpha h_m(x) \tag{11}$$
  Where $F_m(x)$ is the ensemble model at iteration m, α is the learning rate, and $h_m(x)$ is the new weak learner.

3.5 Hyperparameter Tuning with GridSearchCV

Hyperparameter tuning is a critical step in optimizing machine learning models to achieve the best performance. GridSearchCV is a popular method for hyperparameter tuning that performs an exhaustive search over a specified parameter grid. This involves training and evaluating the model for each combination of hyperparameters, and selecting the set that yields the best performance based on a chosen evaluation metric.

Process:

- **Parameter Grid**:
    - Specify a range of values for each hyperparameter that want to tune.
    - Let Θ represent the set of hyperparameters, and θ be a specific combination of hyperparameters.

    $$\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\} \tag{12}$$

- **Cross-Validation**:
    - For each combination of hyperparameters, the model is trained and evaluated using cross-validation, where the training data is split into multiple folds.
    - The dataset D is split into k folds.
    - For each combination θ, train the model on k−1 folds and validate it on the remaining fold.
    $$CV_\theta = \frac{1}{k}\sum_{i=1}^{k} Performance(\theta, D_{-i}, D_i) \tag{13}$$

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

where D_{-i} is the training set excluding the i[th] fold, and D_i is the validation fold.

- **Model Evaluation**:
  - The performance of each combination is evaluated based on a chosen metric.
- **Select the Best Model**:
  - The combination of hyperparameters that yields the best performance is selected as the optimal set.

$$\theta^* = argmax_{\theta \in \Theta} CV_\theta \tag{14}$$

The hyperparameters tuned using GridSearchCV for the Random Forest, SVM, and Gradient Boosting classifiers are given below.

Random Forest (RF)

- **Hyperparameters**:
  - Number of trees ($n_{estimators}$)
  - Maximum depth of trees (max_depth)
  - Minimum samples split (min_samples_split)
  - Minimum samples leaf (min_samples_leaf)

Support Vector Machine (SVM)

- **Hyperparameters**:
  - Kernel type (kernel)
  - Regularization parameter (C)
  - Kernel coefficient (γ) for RBF kernel

Gradient Boosting

- **Hyperparameters**:
  - Number of boosting stages ($n_{estimators}$)
  - Learning rate (learning_rate)
  - Maximum depth of trees (max_depth)

Hyperparameters-tuned classifiers are referred to as Optimized Random Forest (ORF), Optimized Support Vector Machine (OSVM), and Optimized Gradient Boosting (OGB).

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

3.6 Building Ensemble Model using Stacking Classifier

Stacking is an ensemble learning technique that combines multiple classification models via a meta-classifier. The idea is to train several base learners on the training dataset and then use their predictions as features for a second-level learning algorithm. This approach often improves predictive performance by leveraging the strengths of each individual model.

Step 1: Base Learners

The first step is to choose and train several base learners. These models can be different types or the same type with different hyperparameters.
Let $M_1, M_2, \ldots, M_n$ be the base learners. Each base learner M_i is trained on the training data (X, y):

$$M_i : X \to y, \qquad \forall_i \in \{1,2,\ldots,n\} \tag{15}$$

Step 2: Generating Meta-Features

After training the base learners, their predictions on the training data are used as features to train the meta-classifier. This process involves cross-validation to prevent overfitting.
For each base learner M_i, generate the predictions P_i on the training data: $P_i = M_i(X)$
The predictions from all base learners are concatenated to form the meta-features,

$$X_{meta} = [P_1, P_2, \ldots, P_n] \tag{16}$$

Step 3: Meta-Classifier

The meta-classifier M_{meta} is trained on the meta-features X_{meta} and the original target y:

$$M_{meta} : X_{meta} \to y \tag{17}$$

Step 4: Making Predictions

To make predictions on new data X_{test}:
Each base learner M_i makes predictions P_i^{test} on X_{test}: $P_i^{test} = M_i(X_{test})$.
The meta-features for X_{test} are formed by concatenating these predictions:

$$X_{meta}^{test} = [P_1^{test}, P_2^{test}, \ldots, P_n^{test}] \tag{18}$$

The meta-classifier M_{meta} makes the final prediction y_{pred}:

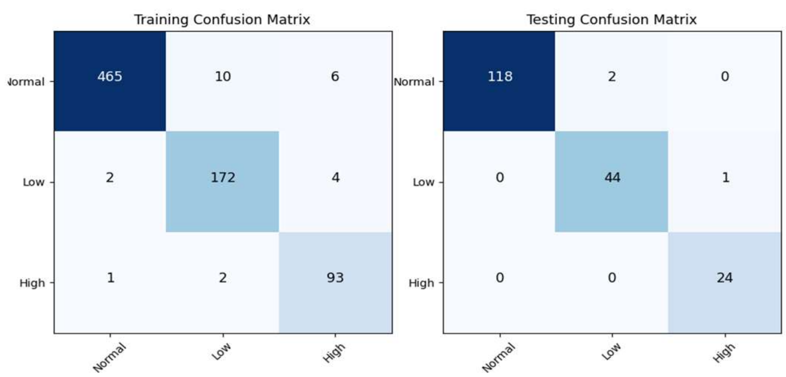
$$y_{pred} = M_{meta}(X_{meta}^{test}) \tag{19}$$

The meta-classifier synthesizes the predictions of the base learners, leading to improved accuracy and robustness. This ensemble approach is particularly useful in complex prediction tasks.

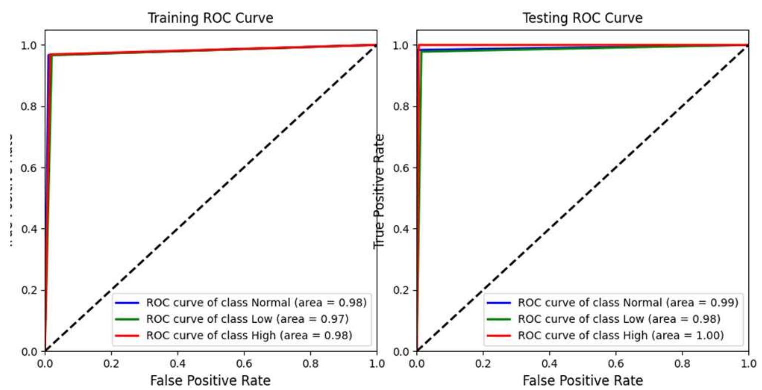
Vol. 21, No. 1, (2024)
ISSN: 1005-0930

4. Results and Discussion

The experimental validation of the proposed model was conducted using two publicly available datasets from the Kaggle Repository: "Social and Psychological Impact of COVID-19" [23] and "Impact of COVID-19 on Student Education" [24]. The Psychological dataset comprises 38 attributes, from which the RFECV model selected 10 key attributes: 'Participant', 'Openness', 'Restraint', 'Transcendence', 'Interpersonal', 'DASS_21', 'GHQ_12', 'SEC', 'Age', and 'Zest'. Similarly, the Student Education dataset contains 19 attributes, and the RFECV model identified 7 important attributes: 'Region', 'Age', 'Medium_OC', 'TimeSpent_SelfStudy', ''TimeSpent_SocialMedia', 'HealthIssue', and "TimeSpent_OC'.



(a) (d)



(b) (e)

Training Classification Report

	precision	recall	f1-score	support
Normal	0.99	0.97	0.98	481
Low	0.93	0.97	0.95	178
High	0.90	0.97	0.93	96
accuracy			0.97	755
macro avg	0.94	0.97	0.95	755
weighted avg	0.97	0.97	0.97	755

(c)

Testing Classification Report

	precision	recall	f1-score	support
Normal	1.00	0.98	0.99	120
Low	0.96	0.98	0.97	45
High	0.96	1.00	0.98	24
accuracy			0.98	189
macro avg	0.97	0.99	0.98	189
weighted avg	0.98	0.98	0.98	189

(f)

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

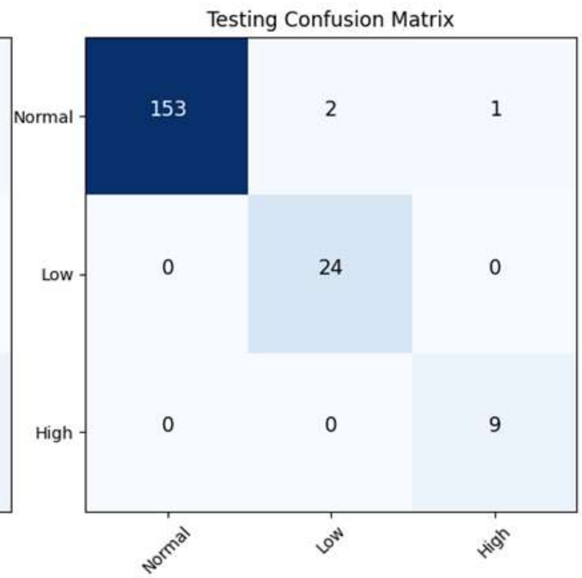
Figure 2: Results of Psychological – Depression target - Training Phase a) Confusion Matrix b) ROC Analysis c) Classification Report / Testing Phase d) Confusion Matrix e) ROC Analysis f) Classification Report

Figure 2 provides a detailed result analysis of the ensemble model on the Psychological – ‘Depression’ target with 80% of training data and 20% of testing data. The proposed model categorizes instances into 'Normal', 'Low', and 'High' classes, is exemplary across both training and testing datasets. The training confusion matrix shows high correct classification rates with minimal misclassifications, and the testing confusion matrix mirrors this accuracy. The classification reports indicate high precision, recall, and F1-scores for all classes, demonstrating that the model maintains a good balance between precision and recall. Specifically, the precision and recall scores are consistently above 90%, with F1-scores indicating strong overall performance. The ROC curves for both datasets exhibit high Area Under Curve (AUC) values, further affirming the model's capability to distinguish between the classes effectively.

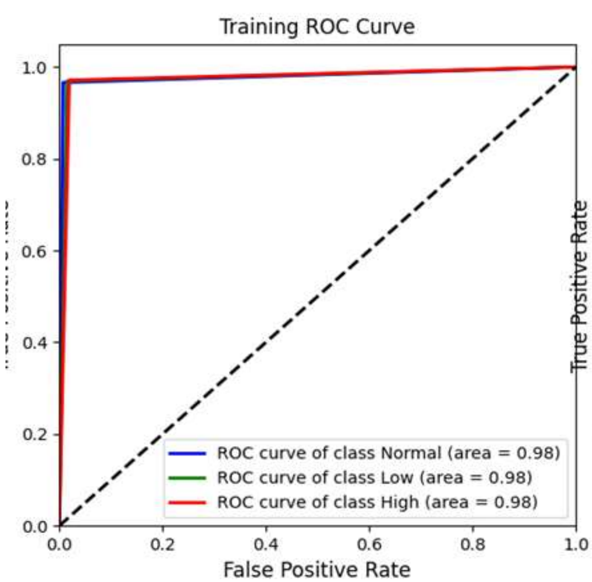
Figure 3 offers a detailed outcome analysis of the proposed approach on the Psychological – ‘Anxiety’ target with 80% of training data and 20% of testing data. Figure 3(a) depicts that the ensemble algorithm has recognized 601 samples into Normal class, 95 samples into Low class and 34 samples into High class on 80% of training data. Also, Figure 3(d) represents that the proposed model has categorized 153 samples into Normal class, 24 samples into Low class, and 9 samples into High class on 20% of testing data. Figures 3(b) and 3(e) illustrates the ROC examination of proposed technique under 80% of training and 20% of testing data. At last, Figures 3(c) and 3(f) demonstrates the classification report of the ensemble approach under 80% of training and 20% of testing data.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

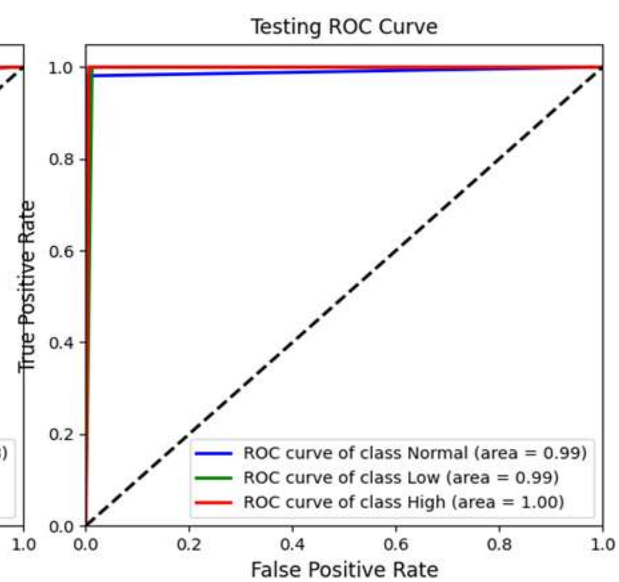
(a)



(d)



(b)



(e)

Training Classification Report

	precision	recall	f1-score	support
Normal	1.00	0.97	0.98	622
Low	0.90	0.97	0.94	98
High	0.71	0.97	0.82	35
accuracy			0.97	755
macro avg	0.87	0.97	0.91	755
weighted avg	0.97	0.97	0.97	755

(c)

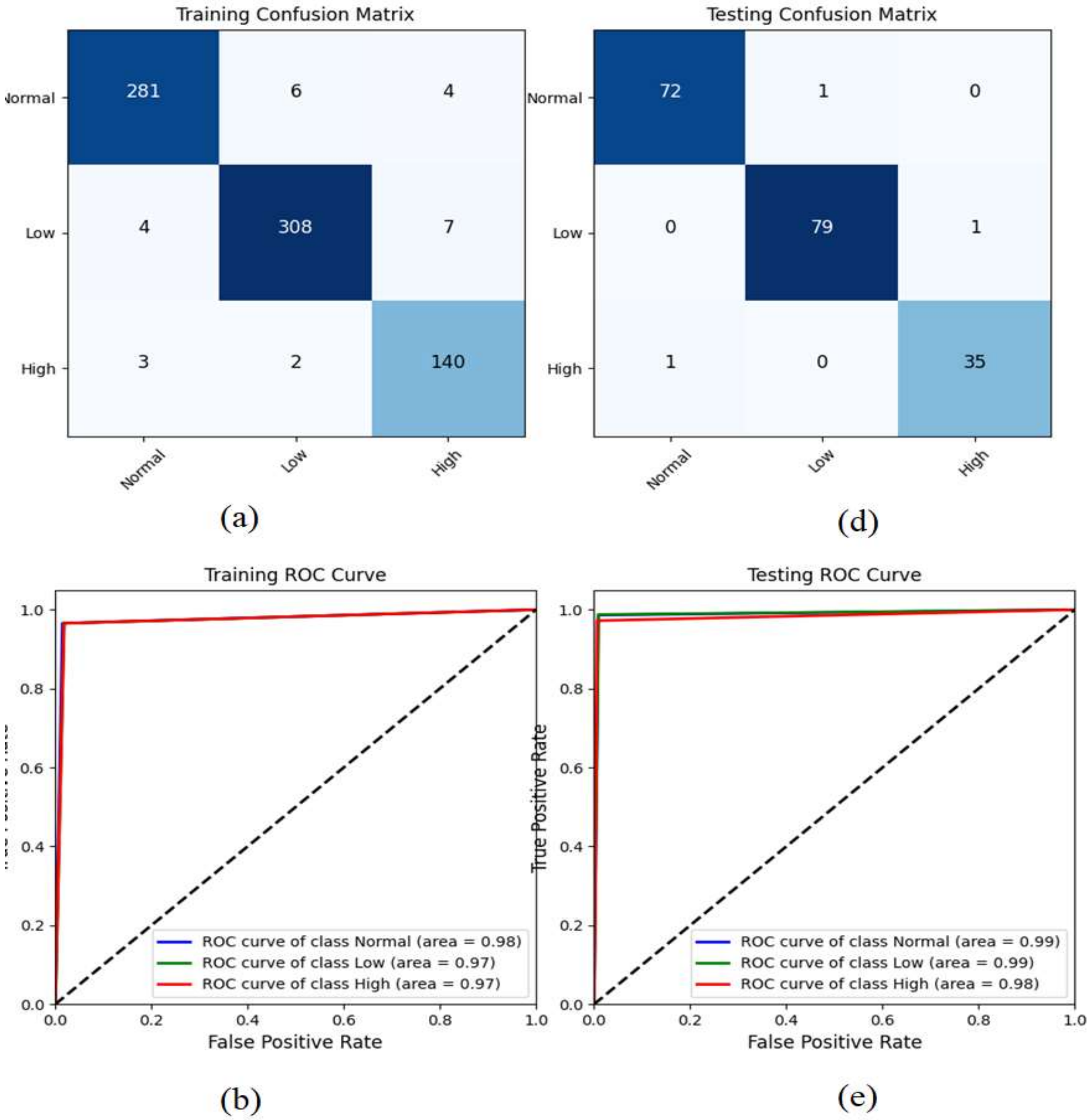
Testing Classification Report

	precision	recall	f1-score	support
Normal	1.00	0.98	0.99	156
Low	0.92	1.00	0.96	24
High	0.90	1.00	0.95	9
accuracy			0.98	189
macro avg	0.94	0.99	0.97	189
weighted avg	0.99	0.98	0.98	189

(f)

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

Figure 3: Results of Psychological – Anxiety target - Training Phase a) Confusion Matrix b) ROC Analysis c) Classification Report / Testing Phase d) Confusion Matrix e) ROC Analysis f) Classification



Training Classification Report

	precision	recall	f1-score	support
Normal	0.98	0.97	0.97	291
Low	0.97	0.97	0.97	319
High	0.93	0.97	0.95	145
accuracy			0.97	755
macro avg	0.96	0.97	0.96	755
weighted avg	0.97	0.97	0.97	755

(c)

Testing Classification Report

	precision	recall	f1-score	support
Normal	0.99	0.99	0.99	73
Low	0.99	0.99	0.99	80
High	0.97	0.97	0.97	36
accuracy			0.98	189
macro avg	0.98	0.98	0.98	189
weighted avg	0.98	0.98	0.98	189

(f)

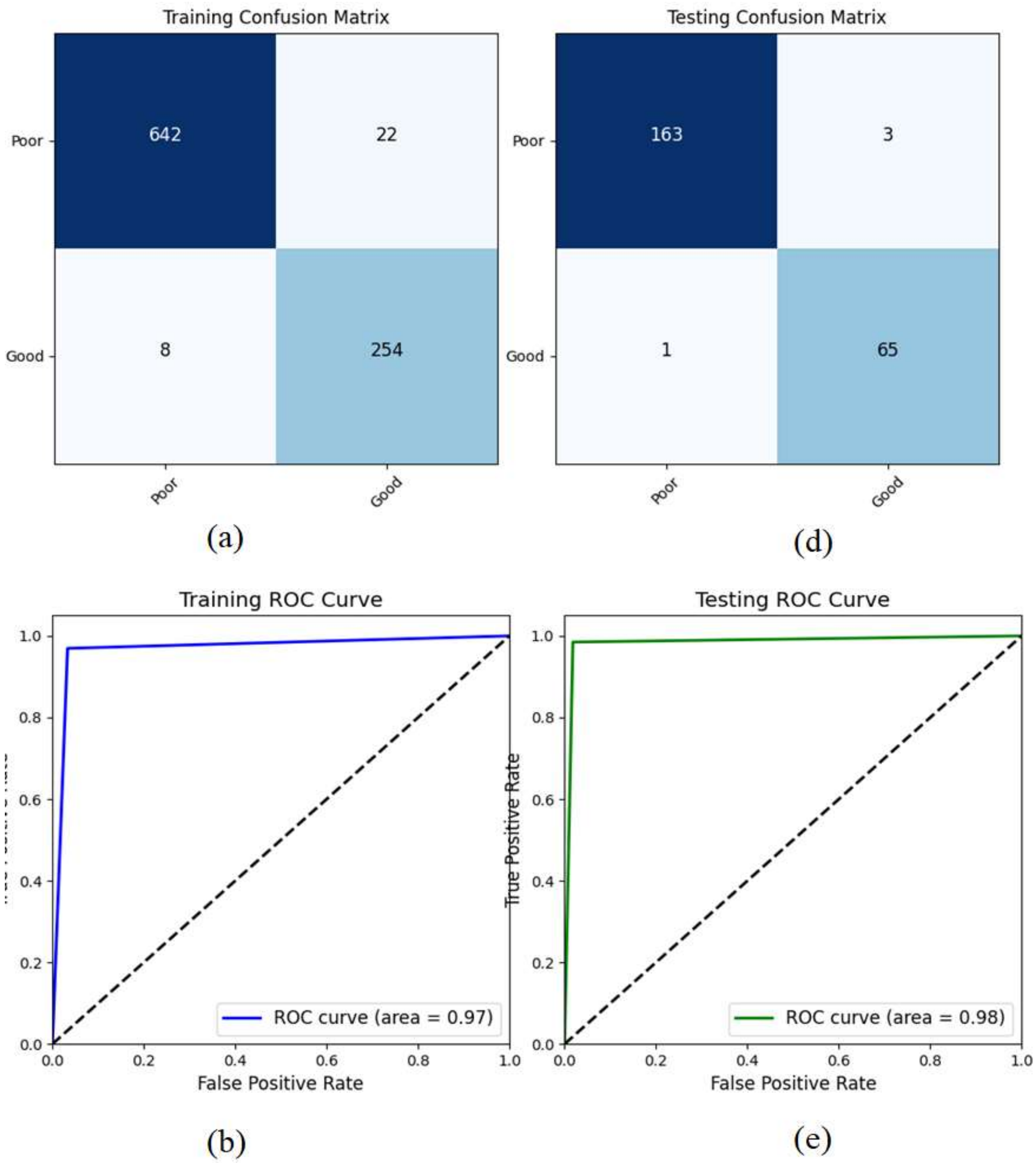
Report

Figure 4: Results of Psychological – Stress target - Training Phase a) Confusion Matrix b) ROC Analysis c) Classification Report / Testing Phase d) Confusion Matrix e) ROC Analysis f) Classification Report

The ensemble model's performance metrics reflect its effectiveness in categorizing instances into different stress levels in figure 4. The model achieved an impressive precision of 96%, indicating that among all instances predicted as a particular stress level, 96% were correctly classified. It also demonstrated a recall of 97%, meaning it accurately identified 97% of all instances belonging to a specific stress level. The F1-score, which combines precision and recall into a single metric, averaged at 96%, indicating a robust balance between correctly identifying instances and minimizing false positives and negatives. The Area Under Curve (AUC) for the ROC curve was 0.98, illustrating the model's strong overall ability to discriminate between different stress levels.

Figure 5 provides a comprehensive analysis of the proposed approach on the Education dataset with the 'Outcome' target, using 80% of the data for training and 20% for testing. Figure 5(a) shows that the ensemble algorithm classified 642 samples as 'Poor' and 22 samples as 'Good' in the training data. Similarly, Figure 5(d) indicates that the model identified 163 samples as 'Poor' and 3 samples as 'Good' in the testing data. Figures 5(b) and 5(e) present the ROC analysis for the proposed technique on the training and testing datasets, respectively. Finally, Figures 5(c) and 5(f) display the classification report for the ensemble approach on the training and testing datasets.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930



Training Classification Report

	precision	recall	f1-score	support
Poor	0.99	0.97	0.98	664
Good	0.92	0.97	0.94	262
accuracy			0.97	926
macro avg	0.95	0.97	0.96	926
weighted avg	0.97	0.97	0.97	926

(c)

Testing Classification Report

	precision	recall	f1-score	support
Poor	0.99	0.98	0.99	166
Good	0.96	0.98	0.97	66
accuracy			0.98	232
macro avg	0.97	0.98	0.98	232
weighted avg	0.98	0.98	0.98	232

(f)

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

Figure 5: Results of Education – Outcome target - Training Phase a) Confusion Matrix b) ROC Analysis c) Classification Report / Testing Phase d) Confusion Matrix e) ROC Analysis f) Classification Report

Table 1 reports an overall result analysis of the Ensemble model and Optimized Individual Techniques on the two applied datasets. On the Psychological dataset, the ORF, OSVM, and OGB models have attained accuracy of 91.25%, 93.18%, and 94.62%, precision of 91.02%, 93.05%, and 94.53%, recall of 91.69%, 93.22%, and 94.68%, F1-score of 91.62%, 93.70%, and 94.60%, ROC score of 91.79%, 93.89%, and 95.27% respectively. Similarly, the Ensemble Model has reported accuracy of 98.33%, precision of 96.56%, recall of 98.78%, F1-score of 97.67%, and ROC score of 98.56% on the Psychological dataset. Also, on the Education dataset, the ORF, OSVM, and OGB approaches have reached accuracy of 91.81%, 92.54%, and 94.55%, precision of 91.25%, 92.81%, and 94.20%, recall of 91.92%, 92.63%, and 94.67%, F1-score of 91.89%, 92.59%, and 94.58%, ROC score of 92.95%, 93.13%, and 94.58% respectively. Eventually, the Ensemble Model has reported accuracy of 98.46%, precision of 97.50%, recall of 98.00%, F1-score of 98.00%, and ROC score of 99.89% on the Education dataset.

Table 1: Result analysis of Ensemble and Individual techniques on two datasets

Metrics	**Models**			
	ORF	**OSVM**	**OGB**	**Ensemble Model**
Psychological_Impact Dataset – Average of three classes				
Accuracy	91.25	93.18	94.62	98.33
Precision	91.02	93.05	94.53	96.56
Recall	91.69	93.22	94.68	98.78
F1-Score	91.62	93.70	94.60	97.67
ROC Score	91.79	93.89	95.27	98.56
Education_Impact Dataset				
Accuracy	91.81	92.54	94.55	98.46
Precision	91.25	92.81	94.20	97.50
Recall	91.92	92.63	94.67	98.00
F1-Score	91.89	92.59	94.58	98.00
ROC Score	92.95	93.13	95.64	99.89

To showcase the improved performance of the proposed Ensemble model, a comprehensive comparison study is presented in Table 2. Figure 6 provides a brief accuracy inspection of the Ensemble model compared to existing models. The figure indicates that the Optimized Naïve Bayes model achieved the lowest accuracy of 87.15%. Meanwhile, the Optimized Logistic Regression and Optimized AdaBoost models achieved slightly higher accuracy values of 91.88%

and 92.14%, respectively. However, the proposed Ensemble model demonstrated superior performance with even higher accuracy.

Table 2: Comparative analysis of Ensemble technique with existing approaches

Methods	Accuracy	Precision	Recall	F1-Score	ROC score
Optimized Naïve Bayes	87.15	87.30	87.20	87.10	88.34
Optimized Logistic Regression	91.88	91.20	91.87	91.73	92.65
Optimized AdaBoost Algorithm	92.14	92.97	93.04	92.91	93.23
Ensemble Model	98.46	97.50	98.00	98.00	99.89

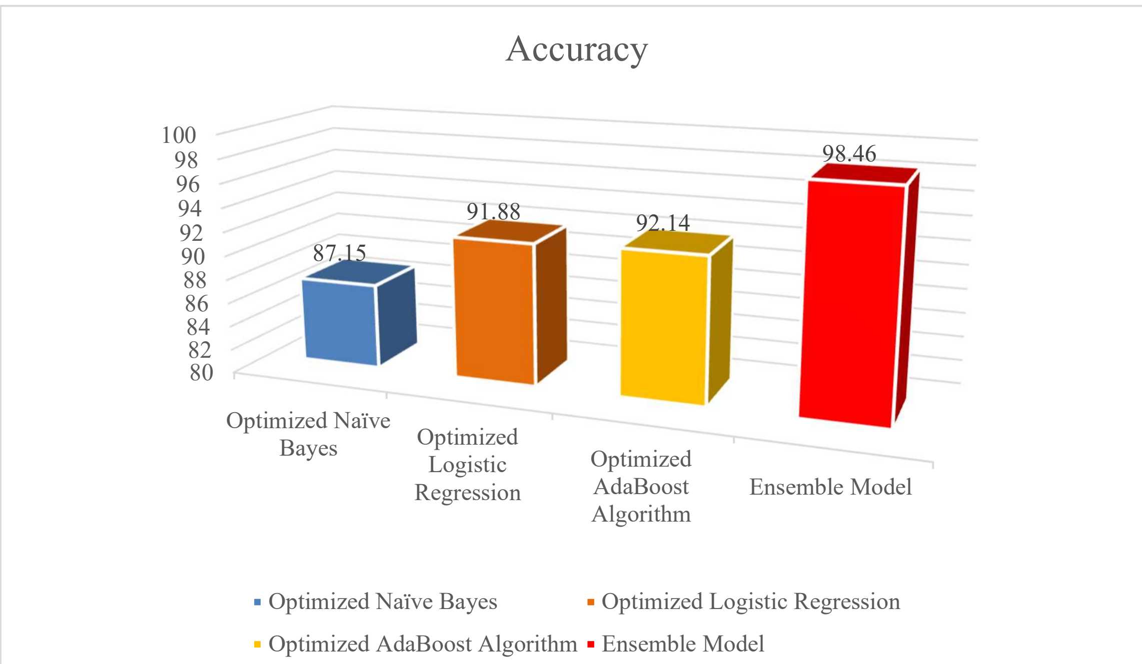


Figure 6: Accuracy analysis of Ensemble model with exiting methodologies

5. Conclusion

The COVID-19 pandemic has undoubtedly left a significant mark on various facets of life, with mental health and education being among the most affected areas. This study aimed to address the dual challenges of understanding the psychological impact of the pandemic on individuals and the educational disruptions faced by students. The methodology involved several critical steps to ensure the reliability and accuracy of the predictions. Data preprocessing was meticulously performed to handle missing values and encode categorical variables appropriately. Feature

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

selection through RFECV was employed to identify the most relevant attributes, thereby enhancing the model's performance. The inclusion of SMOTE addressed the issue of class imbalances, creating a balanced dataset that improved the learning process.

The ensemble model, comprising Random Forest, Support Vector Machine (SVM), and Gradient Boosting classifiers, proved to be highly effective. Each classifier was optimized using GridSearchCV for hyperparameter tuning, ensuring the best possible performance. The final Stacking Classifier, which combined the strengths of the individual models, achieved impressive accuracy rates of 98.33% for predicting the psychological impact of COVID-19 and 98.46% for predicting the impact on student education. The results from this study provide valuable insights into the extent of the pandemic's effects on mental health and education. The high accuracy of the predictions indicates that ensemble learning techniques can be instrumental in developing effective intervention strategies. Integrating natural language processing (NLP) techniques to analyze text data from social media and educational forums could provide deeper insights into the nuanced impacts of the pandemic in the future.

References

1. World Health Organization. Coronavirus disease (COVID-19) pandemic. Accessed September 1, 2021. https://www.who.int/emergencies/disease/novel-coronavirus-2019
2. Valeria Saladino, Davide Algeri, and Vincenzo Auriemma, “The Psychological and Social Impact of Covid-19: New Perspectives of Well-Being”, Frontiers in Psychology, Volume 11(1), 2020.
3. Kunal Chaturvedi, Dinesh Kumar Vishwakarma, and Nidhi Singh, “COVID-19 and its impact on education, social life and mental health of students: A survey”, Children and Youth Services Review, vol. 121, pp. 105866, 2021.
4. Muhammad A S, and Suliman K,Abeer K, “COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses”, Journal of Advanced Research, vol. 5(24), pp. 91-98, 2020.
5. Pramod Soni, Effects of COVID-19 lockdown phases in India: an atmospheric perspective”, Environ Dev Sustain, vol. 23(8), 12044–12055, 2021.
6. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, “Impact of COVID-19 pandemic on mental health in the general population: a systematic review”, Journal of *Affect Disord*, vol. 277, pp. 55–64, 2020.
7. Aras Bozkurt, Kadir Karakaya, Murat Turk, Özlem Karakaya, and Daniela Castellanos-Reyes, “The Impact of COVID-19 on Education: A Meta-Narrative Review”, Tech Trends, vol. 66(5), pp. 883-896, 2022.
8. Bakul J Parekh and Samir H Dalwai, “Psychosocial Impact of COVID-19 Pandemic on Children in India”, Indian Pediatrics, Volume 57(1), December 15, 2020.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

9. Thi Kieu Khanh Ho and Jeonghwan Gwak, "Feature-level ensemble approach for COVID-19 detection using chest X-ray images", PLoS One, vol. 17(7), pp. 430, 2022.
10. Parisa gifani, Ahmad Shalbaf, and Majid Vafaeezadeh, "Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans", International Journal of Computer Assisted Radiology and Surgery, vol. 16, pp. 115-123, 2021.
11. Aqeel M, Rehna T, Shuja KH, Abbas J., "Comparison of students' mental wellbeing, anxiety, depression, and quality of life during COVID-19's full and partial (smart) lockdowns: a follow-up study at a 5-month interval", *Frontier Psychiatry*, vol. 13, pp. 585, 2022.
12. Meenakshi Shukla, Rakesh Pandey, Tushar Singh, Laura Riddleston, Taryn Hutchinson, Veena Kumari, and Jennifer Y. F. Lau, "The Effect of COVID-19 and Related Lockdown Phases on Young Peoples' Worries and Emotions: Novel Data From India", Frontiers in Public Health, vol. 9, pp. 183, 2021.
13. Rahel Mulatie Anteneh, Anteneh Mengist Dessie, Melkalem Mamuye Azanaw, Denekew Tenaw Anley, Biruk Demissie Melese, Sefineh Fenta Feleke, Tiruayehu Getinet Abebe, and Achenef Asmamaw Muche, "The psychological impact of COVID-19 pandemic and associated factors among college and university students in Ethiopia: a systematic review and meta-analysis, 2022", Frontier Public Health, vol. 11, pp. 1136031, 2023.
14. Alvarez, F., Argente, D., and Lippi, F., "A simple planning problem for Covid-19 lockdown", Covid Economy, vol. 14, pp. 1–33, 2020.
15. Brooks, S., Webster, R. S., Woodland, L., Wessely, S., Greenberg, N., "The psychological impact of quarantine and how to reduce it: rapid review of the evidence", *Lancet,* vol. 395(10227), pp. 912-920, 2020.
16. Cai, H., Tu, B., Ma, J., Chen, L., Fu, L., Jiang, Y., "Psychological impact and coping strategies of frontline medical staff in Hunan between January and March 2020 during the outbreak of coronavirus disease 2019 (COVID19) in Hubei. China", Medical Science Monitor, vol. 26(1), pp. 171-186, 2020.
17. De Luca, R., and Calabrò, R. S., "How the COVID-19 Pandemic is Changing Mental Health Disease Management: The Growing Need of Telecounseling in Italy", *Innovative Clinical Neuroscience,* vol.17(6), pp. 7413339, 2020.
18. Batra K, Sharma M, Batra R, Singh TP, Schvaneveldt N., "Assessing the psychological impact of COVID-19 among college students: an evidence of 15 countries", *Healthcare*, vol. 9, pp. 222, 2021.
19. E.M. Aucejo, J. French, M.P. Ugalde Araya, B. Zafar, "The impact of COVID-19 on student experiences and expectations: Evidence from a survey", Journal of Public Economics, vol. 191, pp. 104271, 2020.
20. Showkat Ahmad Dar and Dr.Naseer Ahmad Lone, ""IMPACT OF COVID 19 ON EDUCATION IN INDIA", The Journal of Indian Art History Congress, vol. 26(2), pp. 47-55, 2021.

21. Mishra S, Sahoo S, Pandey S., "Research trends in online distance learning during the COVID-19 pandemic", *Distance Education, vol.* 42(4), pp. 494–519, 2021.
22. Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., ... & Agha, R., "The socio-economic implications of the coronavirus and COVID-19 pandemic: A review", *International Journal of Surgery, vol. 78*, pp. 185-193, 2020.
23. https://www.kaggle.com/datasets/mathchi/covid19-and-character-strengths/data
24. https://www.kaggle.com/code/swapnilnarwade/covid-19-students-survey/notebook