International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

An Efficient Novel Approach for Early Detection of Mental Health Disorders Through Distributed Machine Learning Paradigms from Public Societal Communication

T. Jayasri Devi[1], Adapa Gopi[2]

Submitted: 18/09/2023 **Revised**: 17/11/2023 **Accepted**: 29/11/2023

Abstract: The main factor contributing to impairment worldwide is mental illness. In underdeveloped countries, approximately 2 million people commit suicide annually, and 85% of those who have mental health issues go untreated. According to a study, using social media excessively might cause depression and anxiety. This research study focuses on social media, specifically Facebook, Twitter, and Instagram, and mental health. Using PRISMA criteria on PubMed and Google Scholar, a search of the literature was conducted from January 2010 to June 2022 to find studies addressing the relationship between social media sites and mental health. Social media can provide users with a sense of community, but excessive and rising use of it, especially among the weak, is associated with depression and other mental health issues. The World Health Organization (WHO) estimates that anxiety affects one in every thirteen individuals worldwide. According to the WHO, anxiety disorders are the most common kind of mental illness in the world, with specific phobias, major depressive disorder, and social phobia coming in first and second, respectively. The World Health Organization (WHO) estimates that anxiety affects one in every thirteen individuals worldwide. According to the WHO, anxiety disorders are the most common kind of mental illness in the world, with specific phobias, major depressive disorder, and social phobia coming in first and second, respectively. Depression can be brought on by a variety of events, such as the death of a loved one, the loss of a job, a divorce, and other traumatic situations. These emotions are normal when we are worried. Everyone has experienced sorrow. Contrarily, depression and depression are not the same thing. Depression is a psychological ailment that requires pharmacological treatment. This research blends personal interests in random forest, stacking, and boosting algorithms with mental health. Artificial intelligence's branch of machine learning is frequently used to identify illnesses. It also gives doctors a platform to evaluate vast amounts of patient data and come up with the best course of action based on the patient's medical condition.

Keywords: *Machine learning, World Health Organization, mental health, societal*

1. Introduction

College campuses are frequently concerned about student mental health issues. Demand for services from college counselling centres has increased dramatically, some experts have declared a mental health crisis in higher education. Some experts have declared a mental health crisis in higher education [1]. Evidence suggests that engineering students experience anxiety and depression at considerably higher rates than the overall population [2]. Even though data do not support the concept that engineering students are more prone than non-engineering students to suffer from disorders such as depression, understanding mental health is a big issue for this group owing to several aspects of the engineering curriculum [3].

The continually poor retention rates for bachelor's engineering degrees make investigating mental health among engineering students extremely essential. Several research [4] have discovered correlations between poor student mental health, retention, and academic success. Furthermore, research has indicated that the current engineering curriculum generates cultures of stress and guilt, which may lead to engineering students' poor mental health. Improving engineering students' overall mental health might be a critical strategy for graduating more engineers.

Poor mental health, when paired with microaggressions directed at students of color, women, and first-generation college students, is likely to result in considerably worse retention rates and academic performance for members of disadvantaged groups in engineering. As a result, researching how mental health differs among engineering student groups might hold the answer to graduating more diverse engineering cohorts [5].

In contrast to past studies on the mental health of engineering students, which concentrated on a single school or a small number of mental health issues, the data utilized in this study came from several locations all throughout the United States. This study compares engineering students to the general population in eight different mental health variables using well-known population-scale mental health instruments. Finally, it investigates whether and how

[1] *1Research Scholar, Dept.of.CSE, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India*
ORCID ID 0000-0002-9173-176X
[2] *2Associate Professor, Dept.of.CSE, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India.*
ORCID ID : 0000-0003-2161-5592

inequalities in mental health influence various demographic groups in engineering education. The data shown here was collected in early 2020, before it was realized that COVID-19 was prominent among institutions that had abandoned their physical campuses in favor of online learning. As a result, our study highlights the mental health issues and inequalities that are frequent in "normal" engineering schools. There have been several reports of pandemic data. Universities are increasingly being asked to handle student mental health concerns. Depression is growing increasingly widespread among college students, according to a nationwide poll of undergraduate students, who reported feeling dread or concern to the point that they were unable to function. As a result, it is not surprising that psychological stress has a significant impact on student attrition [6].

Poor mental health, when paired with microaggressions directed at students of color, women, and first-generation college students, is likely to result in considerably worse retention rates and academic performance for members of disadvantaged groups in engineering. As a result, researching how mental health differs among engineering student groups might hold the answer to graduating more diverse engineering cohorts. While it is still in its infancy, engineering-specific mental health research has advanced significantly in recent years. Interventions targeted at enhancing the mental health of engineering graduate and undergraduate students have been adopted in recent years [7]. Early research in this field investigated the relationship between mental health and participation in service-learning activities. More than 44 percent of the 582 engineering students who self-identified as men in a 2008 poll exhibited indicators of sadness. According to multiple recent polls, engineering students reported mental health difficulties at significantly greater rates than the general population.

Even though engineering students are less likely than other students to seek treatment for mental health disorders, they are less likely to seek therapy than other college students [8]. Furthermore, research demonstrates that engineering programs in general generate stress and humiliation. These characteristics may indicate that engineering students have distinct mental health needs than other categories of students. Finally, this work analyses the creative union of distributed machine learning paradigms with information from social communication to develop an effective strategy for the early detection of mental health diseases. This strategy aims to progress the field of mental health by providing scalable, timely, and exact insights on people's general mental health using technology and the analysis of massive volumes of internet communication [9].

2. Related Works

As a result of recent societal developments, the prevalence of mental health problems and psychological ailments has skyrocketed. The World Health Organization (WHO) defines "mental health" as the ability to manage life's pressures to the best of one's abilities while continuing to function properly and successfully at work and contribute to society [10].A person's way of life, including stress at work, a bad financial situation, family problems, interpersonal issues, violence, and environmental factors, is most likely the underlying cause of aspects that have an influence on mental health. These situations may have a role in mental health problems such as depression, anxiety, stress, and other psychological illnesses that influence quality of life. Mental disease affects around 450 million individuals worldwide, accounting for 13% of all ailments. According to the WHO, one out of every four persons will have a mental illness at some time in their life [11]. The WHO launched a policy in 2018 to address the physical concerns of persons suffering from major mental diseases. A person suffering from a significant mental illness, such as schizophrenia, bipolar disorder (BD), psychotic disorder, or depression, frequently dies sooner than the general population. Furthermore, it is estimated that 350 million individuals worldwide suffer from depression, which can lead to suicide thoughts and attempts [12]. The early detection and treatment of mental health problems is critical. People suffering from mental illnesses can benefit from early identification, precise diagnosis, and effective treatment [13].

Mental illness may have major consequences for the persons affected, their families, and society. Face-to-face interviews, self-reporting, or the distribution of questionnaires are commonly utilized in conventional mental health detection procedures. Traditional processes, on the other hand, are usually laborious and time-consuming [14]. Thus, in the past, wearable sensors and telephones were employed in experiments to identify mental illness and enhance healthcare. These tools, however, are mostly employed by persons who have been diagnosed with mental illness and have been continuously monitored over time [15]. Research recently presented a novel way for diagnosing mental health disorders in online social networks (OSNs), and OSNs have grown in popularity in recent years, providing users with a new means of contact and information sharing. Millions of people use OSNs on a daily basis all around the world. OSN users can submit different types of data (such as text, images, videos, and audios) about their everyday activities to express their feelings and opinions. They may also engage with their pals by leaving comments on other people's blogs. As a result, this new area of research is linked to big data research and the rise of online service networks (OSNs) like Facebook, YouTube, twitter, Instagram, and Sina Weibo [16]. OSNs like Twitter, Facebook, and Sina Weibo are used as data sources for online studies and crowdsourcing by researchers from the West and the East. Psychological stress, unhappiness, mental disease, and suicidal thoughts were some of the mental health issues identified in OSNs. Understanding data sets, data analysis techniques, feature extraction strategies, classifier

The present state of mental health detection in OSNs influences performance (accuracy and efficiency), problems, restrictions, and future work. The goal of this systematic review is to conduct critical assessment research on the process of diagnosing mental health disorders using OSN data. For analyzing data from user-posted texts on OSNs, two typically used methodologies are dictionary-based and machine-learning methods. Nonetheless, both techniques have downsides. As a result, academics are currently exploring other approaches to improve the efficacy and performance of the analysis. Overfitting, model interpretation, and generalization are all common training concerns with classical machine learning. Consequently, the researcher resorted to deep learning techniques, which have shown to be a useful tool in recent years. This is done so that machine learning, particularly in the context of health data, can perform more difficult jobs [17].

Before researchers obtained access to the datasets utilized in this study in 2018 and 2020, they had been the focus of several investigations. Fasmer et al. utilized a similarity graph approach to demonstrate that patients with depression and patients with schizophrenia have lower time series regularity than controls. Enrique et al. set a starting point for categorization jobs (depressed vs. non-depressed days). They tested a variety of machine learning classification algorithms, including naive Bayes, nearest neighbors, random forest, support vector machine (SVM) in the linear kernel, radial basis function kernel (RBF), Gaussian process, decision tree, AdaBoost, quadratic discriminant analysis (QDA), and neural network (ANN).The best results were obtained by combining naive Bayes with QDA, although at the price of precision (0.543), yielding a precision of 0.65 and a Gaussian process recall of 0.733. In August 2020, Jakob et al. did a deep learning investigation on the data following class balancing using SMOTE. A Matthews correlation coefficient (MCC) of 0.65 and an accuracy of 84 percent were used to distinguish between the depressed condition groups. According to the study's findings, machine learning algorithms performed well in distinguishing between depressed patients and healthy controls using actigraphy data, however deep learning with limited amounts of data resulted in overfitting [18-20].

The feature engineering approach used in this work seeks to reveal the finest attributes that define symptom clusters, as previously addressed by Liddle, to find the best representation of mental diseases. Psychomotor and disorganization were designed and conveyed through the characteristics employed, in line with Liddle's five primary clusters. These clusters are recognized as critical components that are particularly relevant to depression and schizophrenia diseases. They include symptoms that might be assessed using actigraphy recordings. The feature engineering approach used in this work seeks to reveal the finest attributes that define symptom clusters, as previously addressed by Liddle, to find the best representation of mental diseases. Psychomotor and disorganization were designed and conveyed through the characteristics employed, in line with Liddle's five primary clusters. These clusters are recognized as critical components that are particularly relevant to depression and schizophrenia diseases. They include symptoms that might be assessed using actigraphy recordings [21-23]. Several MDD studies have included continuous assessments of daily physical activity using activity monitors as a more quantitative strategy. These studies demonstrated a variety of behavioral alterations, including decreased activity levels throughout the day, sleep issues, and disturbance of the circadian rhythm, as well as their improvement during therapeutic therapy. In this context, we recently measured locomotor activity, i.e., spontaneous physical activity in daily life, in patients with MDD for more than one week and discovered that patients with depression exhibited more intermittent behavioral patterns characterized by lower mean activity levels associated with occasional bursts of locomotor activity compared to healthy subjects. Several MDD studies have included continuous assessments of daily physical activity using activity monitors as a more quantitative strategy. These studies demonstrated a variety of behavioral alterations, including decreased activity levels throughout the day, sleep issues, and disturbance of the circadian rhythm, as well as their improvement during therapeutic therapy. In this context, we recently measured locomotor activity, i.e., spontaneous physical activity in daily life, in patients with MDD for more than one week and discovered that patients with depression exhibited more intermittent behavioral patterns characterized by lower mean activity levels associated with occasional bursts of locomotor activity compared to healthy subjects [24-28].

3. Problem Statement:

The application of machine learning in mental health has the potential to improve the accuracy and speed of diagnosis, reduce the burden on healthcare providers, and enhance our understanding of mental health disorders. The idea behind this approach is to use machine learning algorithms and distribute them across public communication platforms to detect early signs of mental health disorders. The goal is to increase the efficiency of early detection by leveraging the vast amounts of data generated through public communication. This can potentially help individuals get the support they need before their condition becomes severe. However, the effectiveness of this approach depends on the quality of the data and the accuracy of the machine learning algorithms used.

4. Methodology:

Stacking Algorithm:

Stacking is one of the most well-liked and successful ensemble techniques in machine learning. It is comparable to voting ensembles in that it distributes weights to machine learning algorithms, despite having two layers of models ground models and meta models. As a result, stacking outperforms all other ensemble strategies in machine learning. In this article, stacking ensemble approaches will be discussed. The fundamental principle of these strategies will be discussed first, followed by mathematics and many operational procedures. We will write code for these algorithms to utilize them on data. The key lessons from this plan will be highlighted in the end. Stacking is a popular and effective ensemble strategy in machine learning. Despite having two layers of models ground models and meta models it acts similarly to voting ensembles in that it distributes weights to machine learning algorithms. Stacking is the greatest solution since no other ensemble strategy exceeds it in machine learning. Sure! Stacking is a type of ensemble learning that combines the predictions of numerous base models to generate a more accurate and resilient meta model. To create the final prediction, the meta model learns how to weight the predictions of the basic models. Here's a stacking algorithm:

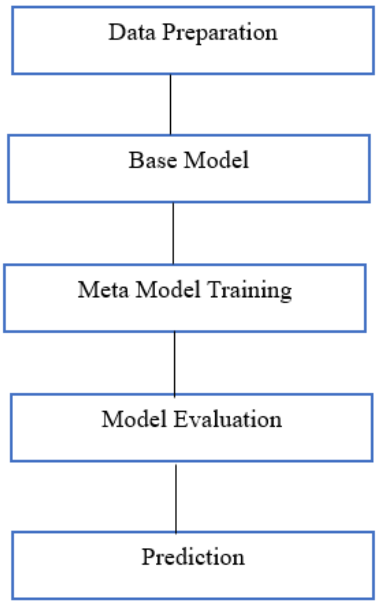

Fig. 1 Stacking Algorithm

In figure (1), Data Preparation: Collect information on mental health issues from social media networks. Clean, normalize, and tokenize the text before preprocessing the data. Divide the dataset into two parts: training and hold-out validation. The validation set will be used to train the meta model, while the training set will be utilized to train the base models. **Base Models Training:** As foundation models, select a variety of machine learning techniques. Support Vector Machines (SVM), Naive Bayes, Random Forest, and Recurrent Neural Networks (RNN) are among examples. On the training set, train each base model independently. For each fundamental model ,1. Fit the model to the training data with the retrieved features and labels.2. Make predictions for the validation dataset.

Meta Model Training: Collect the basic models' predictions on the validation set. These projected values will be utilized as attributes in the meta model. Select a metamodel algorithm. As the meta model, a basic linear regression or a more complicated model, such as a neural network, is frequently utilized. On the validation set, train the meta model.

1. As input features, use the anticipated values from the basic models.

2. Teach the meta model how to weigh the predictions from the basic models in order to create the final prediction.

Model Evaluation: Once trained, utilize the meta model to make predictions on the test set or fresh data. Analyze the stacked model's performance using several measures like accuracy, precision, recall, F1-score, and AUC-ROC. **Prediction:** When predicting fresh data, run it through each of the underlying models to generate their own forecasts. Use the base model predictions as input features for the trained meta model. The meta model then integrates the base models' predictions to get the final forecast for the new data. The goal of stacking is to combine the capabilities of several base models, which may excel in different elements of the data, utilizing a meta model to improve forecast performance. Each basic model learns from the training data and predicts on the validation data set. These predictions are then used to train the meta model, which learns how to integrate the predictions of the underlying models optimally. When compared to separate base models, the final stacked model is intended to generalize better and perform better. The main advantage of stacking is its capacity to capture complex patterns in data and give more robust predictions by limiting the influence of individual model biases and limits. However, careful tweaking and cross-validation are required to avoid overfitting and optimize the ensemble model's hyperparameters.

Base Models Predictions: The base model i produces a prediction indicated as _i,j for each base model i (where i = 1 to N) and each data point j (where j = 1 to the number of data points in the validation set).The predictions of the base models are integrated into a matrix, where each row corresponds to a data point and each column corresponds to a forecast of a base model.

Base Models Predictions Matrix:

$$\begin{pmatrix} y_{1,1} & \cdots & y_{N,1} \\ \vdots & \ddots & \vdots \\ y_{1,M} & \cdots & y_{N,M} \end{pmatrix}$$

Here, M is the number of data points in the validation set.

Meta Model Training: The meta model uses the basic models' predictions as input characteristics and learns how to combine them to create the final prediction. Let us represent the projected values from the base models as a matrix _base, where each row corresponds to a data point and each column refers to a prediction from a base model.:

$$\hat{Y}\_base: \begin{pmatrix} y_{1,1} & \cdots & y_{N,1} \\ \vdots & \ddots & \vdots \\ y_{1,M} & \cdots & y_{N,M} \end{pmatrix}$$

The validation set is used to train the meta model, with the true target labels (y) serving as the training target. The meta model learns weights or coefficients (denoted as w) for each prediction made by the base model in order to integrate them and generate the final prediction.:

Meta Model Equation:

$$\boldsymbol{y_{final}} = \sum \boldsymbol{w_i} * \boldsymbol{y_i} \quad \boldsymbol{for\ all\ i=1\ to\ N} \qquad \text{Eq (1)}$$

The weight (or coefficient) allocated to the prediction of base model i is represented by w_i. During the training phase, the meta model learns these weights.

Prediction: Each base model i creates its prediction (_i) while making predictions on fresh data points. These forecasts are then sent into the trained meta model, which combines them with the learnt weights to produce the final prediction (_final) for the new data point.t. Training the meta model often entails minimizing a loss function (e.g., mean squared error or cross-entropy) using techniques such as gradient descent or other optimization algorithms to determine the appropriate weights (w_i).). It's worth noting that the stacking approach can also handle multi-class classification and regression issues. The fundamental concept stays the same: the predictions of the base models are integrated using a meta model to increase the ensemble's overall predictive performance.

5. Boosting

Boosting is a popular machine learning ensemble approach that combines many weak learners to generate a strong learner. It is a type of supervised learning that is often used in classification and regression applications. The core principle of boosting is to train a sequence of models repeatedly, with each subsequent model focused on fixing errors made by previous models. To begin, the original training dataset is used to train a weak learner. A weak learner is a basic model that slightly outperforms random guessing. After training the original model, boosting assigns more weights to misclassified examples or those with higher prediction errors. We are taught to place a higher weight on misclassified situations when we follow models. Boosting, in this sense, focuses on difficult-to-predict circumstances and aims to correct mistakes made by previous models.

Decision trees, also known as decision stumps, are shallow trees with only a few tiers that are commonly utilized by inexperienced learners. Other models, like linear models or neural networks, can, nevertheless, be employed as weak learners. Boosting strategies differ in how weights are assigned to instances, weights are updated, and weak learner predictions are combined. Some popular boosting algorithms are AdaBoost (adaptive boosting), Gradient Boosting Machines (GBM), XGBoost, LightGBM, and CatBoost. Another ensemble learning approach that may be used to treat mental health illnesses in social media data is boosting. Boosting creates a powerful learner by merging numerous weak learners (typically basic models). Boosting may be used in the context of mental health illnesses on social media in the following ways:

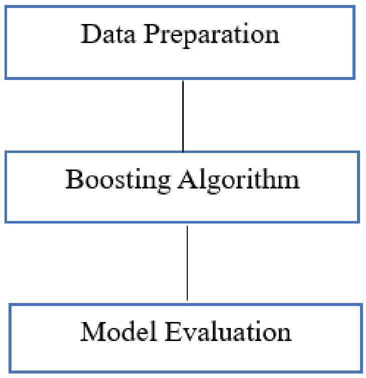

Fig. 2 Boosting Algorithm

In figure (2), Data Preparation: Collect useful data on mental health issues from social media networks. Clean, normalize, and tokenize the text before preprocessing the data. Extract important textual elements that can aid in the identification of mental health-related information. To evaluate the model, divide the dataset into two parts: training and validation (or test).

Boosting Algorithm: As a weak learner, use a basic model or a rudimentary learning method. In boosting, decision trees are frequently employed as weak learners. Set the weights for the training data: At the start, provide identical weights to all of the samples in the training set. These weights are used to control each sample's influence throughout the training process. Repeat the following procedures for a predetermined number of boosting rounds (or until a stopping requirement is met):

1. Use the existing weights to train the weak learner on the training data.
2. Assess the performance of the weak learner on the training data.

3. Determine the error of the slow learner. The mistake is weighted depending on the sample weights, giving misclassified samples additional weight.
4. Determine the contribution of the weak learner to the final forecast. This contribution is defined by an error rate-dependent weight (alpha). The lesser the contribution, the greater the mistake.
5. Adjust the sample weights based on the errors produced by the inexperienced learner. Misclassified samples are given more weight, whereas correctly classified samples are given less weight. To generate the final boosted model, combine the predictions of all weak learners using their individual weights (alpha).

Model Evaluation: Make predictions on the validation (or test) set using the boosted model. Analyze the boosted model's performance using conventional evaluation measures like accuracy, precision, recall, F1-score, and AUC-ROC. Boosting works by iteratively teaching weak learners on data, with each learner focused on the prior learner's faults. During training, the algorithm provides misclassified samples with extra weight to make them more significant in the learning process. The final boosted model combines all weak learners' predictions, each weighted by its alpha value, to generate a strong learner who benefits from the combined knowledge of the weak learners. Boosting efficiently decreases bias and variance, resulting in increased generalization and test set performance. It's worth mentioning that common boosting algorithms for implementing the boosting approach include AdaBoost (Adaptive Boosting) and Gradient Boosting Machines (GBM). These methods differ somewhat in how they update sample weights and calculate contributions, but the essential notion stays the same. Boosting may be a powerful strategy for dealing with uneven data or when the underlying models are simple yet complimentary in nature.

Boosting Algorithm (AdaBoost):

Data Preparation: Collect information on mental health issues from social media networks. Clean, normalize, and tokenize the text before preprocessing the data. Extract important textual elements that can aid in the identification of mental health-related information.

To evaluate the model, divide the dataset into two parts: training and validation (or test).

Initialize Weights: Set the weight of each sample in the training set to zero. At first, all samples have equal weights, indicated by 'w_i', where 'i' is the sample index in the training set.

$$w_i = \frac{1}{N} \quad for\ all\ i = 1\ to\ N \qquad \text{Eq(2)}$$

(where N is the number of samples in the training set)

Boosting Rounds: Repeat the following procedures for a predetermined number of boosting cycles (T):

Train Weak Learner: Use the current sample weights ('w_i') to train a weak learner (e.g., a decision tree with minimal depth) on the training data. On the weighted training set, the weak learner attempts to minimize classification error.

TrainWeakLearner(Data, Labels, w_i) Weak Learner_t

Evaluate Weak Learner: Evaluate the performance of the weak learner on the training data:

$$Error_t = (Week\ Learner(t), Data, Label, w(i)) \qquad \text{Eq(3)}$$

Where 'Compute Error' calculates the weighted error of the weak learner.

Calculate Contribution: Calculate the weak learner's contribution to the final prediction ('alpha_t') based on its weighted error ('Error_t'). A larger mistake leads to a smaller contribution:

$$\alpha_t = 0.5 * \log \frac{1 - Error_t}{Error_t} \qquad \text{Eq(4)}$$

Update Sample Weights: Based on the performance of the poor learner, update the sample weights ('w_i'). Increase the weight of incorrectly categorised samples while decreasing the weight of correctly recognised ones:

$$w_i = w_i * \exp\left(-\propto_t * y_i * ht_{xi}\right) \qquad \text{Eq(5)}$$

Where 'y_i' is the i-th sample's true label, 'h_t(x_i)' is the weak learner's prediction for the i-th sample, and 'exp' is the exponential function.

Normalize Sample Weights: Normalize the sample weights so that they sum to 1:

$$\boldsymbol{w_i = \frac{w_i}{sum_w}} \qquad \text{Eq(6)}$$

Combine Weak Learners: Combine the predictions of all weak learners using their respective weights ('alpha_t') to create the final boosted model.

$$\boldsymbol{final_{Prediction} = \sum \propto_t * WeakLearner_{tx\ for\ all\ t=1\ to\ T}} \qquad \text{Eq(7)}$$

Model Evaluation: Make predictions on the validation (or test) set using the boosted model.

Analyze the boosted model's performance using conventional evaluation measures like accuracy, precision, recall, F1-score, and AUC-ROC. Data Preparation: This stage entails gathering and prepping data, as well as dividing it into training and validation sets. Initialize Weights: Each sample is given a starting weight before boosting. Initially, all weights are equal, giving each sample equal value in the training process. **Boosting Rounds:** The method executes a sequence of boosting rounds (T rounds) repeatedly. It trains a weak learner on the training data using the current sample weights in each round. **Train Weak Learner:** On the weighted training set, a weak learner, often a basic model (e.g., decision tree), is trained to minimize the classification error.

Evaluate Weak Learner: The performance of the weak learner is assessed using training data and sample weights. **Calculate Contribution:** The performance of the weak learner is assessed using training data and sample weights. **Update Sample Weights:** The sample weights are adjusted based on the poor learner's performance. Misclassified samples are given more weight, whereas correctly classified samples are given less weight. **Normalize Sample Weights:** The sample weights are normalized to total to one, guaranteeing that they retain probabilities. **Combine Weak Learners:** The final boosted model is built by combining the predictions of all weak learners based on their weights. **Model Evaluation:** The performance of the boosted model is assessed using the validation (or test) set. During training, boosting successfully concentrates on misclassified data, enhancing the model's performance and making it more resilient. The final model is a hybrid of numerous weak learners, each specializing in different areas of the data, resulting in a more accurate and stronger predictor.

One of the most important advantages of boosting is its ability to improve model performance by combining the talents of several poor learners. It typically leads to better accuracy when compared to using a single model. Boosting algorithms are well-known for their ability to deal with complex data patterns and perform well over a wide range of datasets.

However, boosting may be computationally costly and may demand careful parameter optimization. It is also prone to overfitting if not properly regularized. With appropriate parameter selection and regularization methodologies, boosting algorithms can be valuable tools for machine learning applications.

6. Random Forest:

Random forest is a popular machine learning ensemble learning technique that combines several decision trees to create predictions. It is well-known for its ability to complete tough tasks and produce reliable results in a variety of industries. Here's a quick rundown of Random Forest: Collective Learning: Random Forest belongs to the ensemble learning algorithm family, which integrates several distinct models to create a more robust and accurate model. Random Forest's models are all decision trees. Trees of Decision: Decision trees are simple but powerful machine learning models that learn an if-else set of rules to predict outcomes. Each decision tree splits the data based on qualities and builds a tree-like structure to arrive at the final forecast. A solitary decision tree, on the other hand, is prone to overfitting and may fail to generalize to new data adequately. Construction of Random Forest: Random Forest overcomes the limits of a single decision tree by constructing an ensemble of many decision trees. Each Random Forest decision tree is trained using a randomly selected subset of the training data and a randomly selected subset of features. This randomization adds variation to individual trees, lowering the danger of overfitting and improving overall model performance. To make predictions, Random Forest employs a voting process. For classification problems, each decision tree in the Random Forest guesses the class of a new instance separately, with majority vote determining the final prediction. The final prediction in regression tasks is the average of the predictions produced by the different trees.

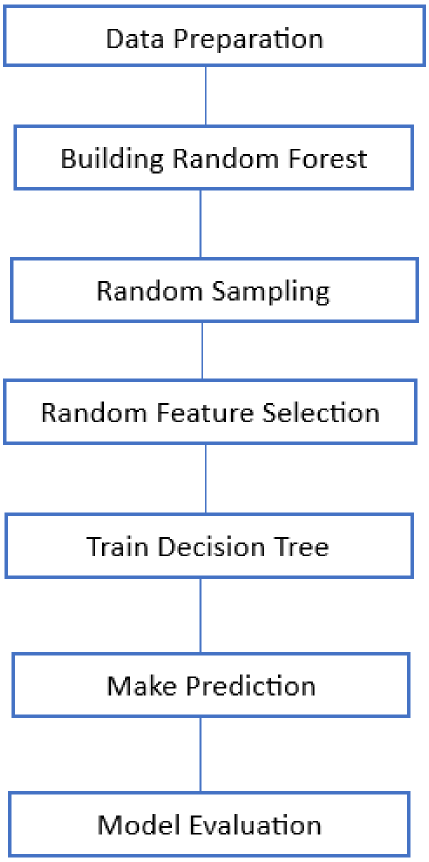

Fig. 3 Random Forest

Random forest provides a measure of feature importance that shows how much each feature contributes to the model. It computes the average impurity reduction (e.g., Gini index or entropy) caused by a characteristic over all decision points. There are trees in the woods. The relevance of features can help determine the most significant traits for the work at hand. Random forest has the following advantages: - Random Forest is very accurate and works well on a wide range of tasks, including classification and regression. It successfully handles huge datasets with high dimensionality. Because of the ensemble of trees and the random feature selection, Random Forest is resistant to overfitting. It gives a measure of feature relevance, which aids in feature selection and issue comprehension. Random forest is a versatile and effective algorithm that is utilized in a variety of applications such as finance, healthcare, and picture classification. It is a popular choice in machine learning due to its capacity to handle difficult problems, avoid overfitting, and offer solid predictions.

Random Forest is an ensemble learning approach that may be applied to social media data to treat mental health conditions. It is based on the idea of building several decision trees and combining their predictions to increase accuracy and resilience. Here's how the Random Forest algorithm works in the context of social media mental health illnesses. **In figure (3), Data Preparation:** Collect useful data on mental health issues from social media networks. Clean, normalize, and tokenize the text before preprocessing the data. Extract important textual elements that can aid in the identification of mental health-related information. To evaluate the model, divide the dataset into two parts: training and validation (or test). **Building Random Forest:** Specify the number of decision trees (n_estimators) that will comprise the Random Forest. Each tree will be trained using a portion of the training data. For each decision tree (t ranging from 1 to n_estimators).

Random Sampling: Sample a chunk of the training data at random (with replacement). This subset is known as the current tree's training set. This subset is generally the same size as the original training set, but with certain data points repeated and others deleted, introducing variation in the training data for each tree. **Random Feature Selection:** Choose a subset of features at random (a subset of the total features) for training the decision tree. This technique brings variation and unpredictability into the attributes examined for each tree. **Train Decision Tree:** Train the decision tree using the randomly picked features and the randomly sampled training data. The CART (Classification and Regression Trees) technique is commonly used to train the decision tree.

Making Predictions: To generate a set of predictions, run each new data point through each decision tree in the Random Forest. **Classification:** If the job involves categorization (for example, predicting whether a post is connected to a mental health issue), the final prediction is made by majority vote. As the final forecast, the class with the most votes from the decision trees is chosen. **Regression:** If the job involves regression (for example, estimating the severity of a mental health issue), the final prediction might be the average of all the decision trees' predictions. **Model Evaluation:** Make predictions on the validation (or test) set using Random Forest. Analyze Random Forest's performance using conventional assessment measures like as accuracy, precision, recall, F1-score, and AUC-ROC.

Random Forest creates an ensemble of decision trees, each trained on a random subset of the training data and characteristics. This diversity and randomization in data sampling and feature selection prevents overfitting and improves the model's generalization capabilities. Combined forecasts from all decision trees are used to make the final prediction. The class with the most votes is chosen in classification problems, whereas the average of the predictions is utilized in regression tasks. Random Forest is well-known for its capacity to cope with high-dimensional data, deal with noise, and produce feature significance ratings. It is an effective algorithm for social media mental health analysis because it can capture complicated interactions between information and increase prediction accuracy. The Random Forest algorithm is comprised of various mathematical equations linked to the creation of decision trees, the combination of their predictions, and the assessment of their value. Let us go through the essential mathematical principles and equations that are employed in the Random Forest technique.

Data Preparation: Before delving into the mathematical formulae, the training data and labels for the algorithm must be available.

Building Random Forest: Random Sampling (Bootstrapping): In the Random Forest, a random subset of the training data is produced with a replacement for each decision tree 't'. This is known as bootstrapping. Let's call 'N' the total number of samples in the training set and 'N sample' the size of the randomly sampled subset. The random sampling equation is as follows:

N_sample = N (with replacement)
Eq(8)

Random Feature Selection:

In the Random Forest, a random subset of characteristics is chosen for each decision tree 't'. This increases the diversity and unpredictability of the features utilised to train the decision tree. Let's call 'M' the total number of features and 'M selected' the size of the randomly selected subset of features. The following is the random feature selection equation:

M_selected << M
Eq(9)

Train Decision Tree: As weak learners, Random Forest employs decision trees. The CART (Classification and Regression Trees) technique is commonly used to train each decision tree. The CART method and its mathematical calculations are beyond the scope of this description, but the basic idea is to partition the data recursively depending on the selected characteristics to form a tree structure that predicts the target variable (classification or regression).

Making Predictions:

For each new data point 'X_new'
Eq(10)

Classification:

In the case of classification problems, the Random Forest uses majority voting to integrate the predictions from all decision trees. The class label with the most votes from the decision trees is the final prediction. The majority voting equation can be represented mathematically as:

FinalPrediction = mode(Predictions_1, Predictions_2, ..., Predictions_n_estimators)
Eq(11)

Where'mode' reflects the most common class label in the collection of predictions from all decision trees.

Regression:

In the case of regression issues, the Random Forest averages and combines the forecasts of all decision trees. The averaging equation can be expressed numerically as follows:

FinalPrediction = (Prediction_1+Prediction_2 +..+Prediction_n_estimators)/n_estimators Eq(12)

Where 'Prediction_t' is the t-th decision tree's prediction and 'n_estimators' is the total number of decision trees in the Random Forest.

Model Evaluation: Finally, the Random Forest is assessed using common evaluation measures like as accuracy, precision, recall, F1-score, and AUC-ROC on the validation (or test) set. It should be noted that the Random Forest algorithm is an ensemble approach that generates several decision trees with unpredictability in the input and feature selection. When these decision trees are combined, they increase prediction performance and generalisation to new data. Because of its durability and capacity to handle complicated datasets, the Random Forest technique is frequently utilised in a variety of machine learning applications.

Density:

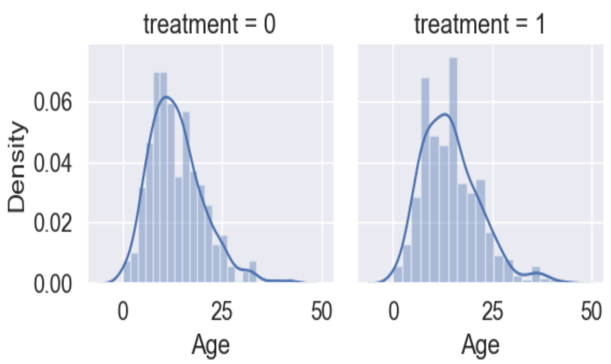

Fig. 4 Density

Feature Importance

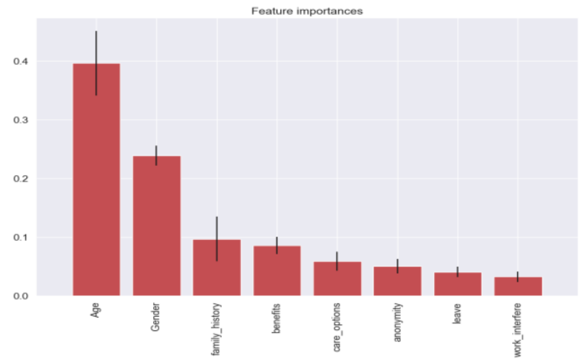

Fig. 5 Feature Importance

These graphics make it easy to examine how different factors impact the model's predictions. A distinct visualisation strategy may be used to demonstrate the importance of a specific characteristic, depending on the algorithm and library utilised.

Confusion Matrix

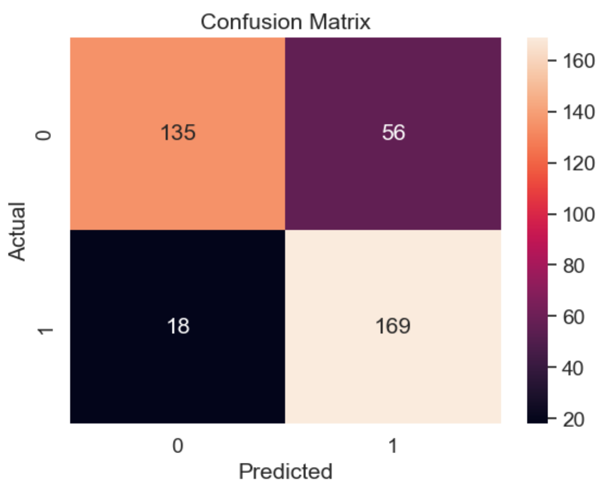

Fig. 6 Confusion Matrix

An example of a confusion matrix is a square matrix, where the rows indicate expected labels, and the columns reflect actual labels. The diagonal elements of the matrix reflect the number of cases that were properly recognized, whereas the off-diagonal components represent the instances that were incorrectly categorized.

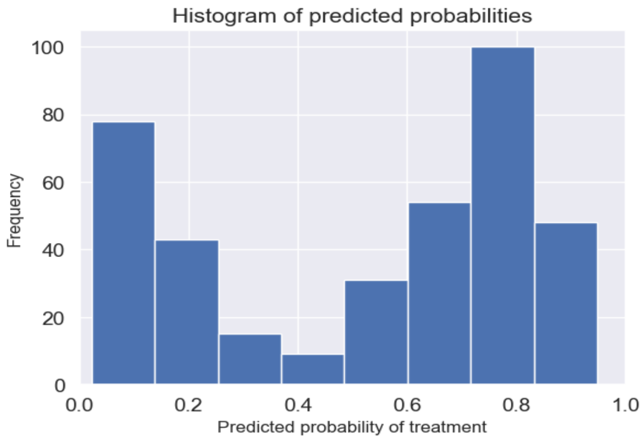

Fig. 7 Histogram of predicted probabilities

Histogram:

A histogram is a graphical depiction of a dataset's distribution. It depicts the frequency or count of observations that fall inside distinct ranges or bins of a continuous variable. Histograms are often used for exploratory data analysis and comprehending a feature's underlying distribution.

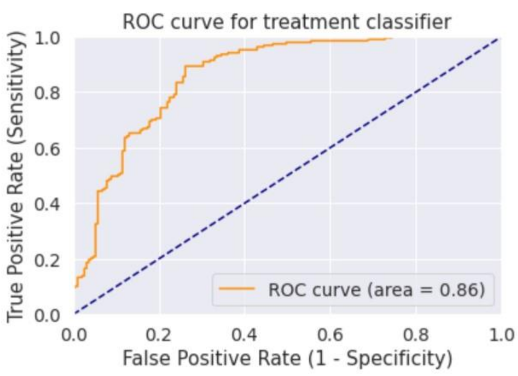

Fig. 8 ROC curve treatment classifier

ROC curve treatment classifier:

The receiver operating characteristic (ROC) curve visually represents the performance of a binary classifier at various classification levels. The ROC curve is created by graphing the true positive rate (TPR) vs. false positive rate (FPR) at various threshold levels.

Comparison results:

These ensemble learning approaches, which include Stacking, Boosting, and Random Forest, are critical in the investigation of mental health in societal communication. These algorithms can assist uncover mental health-related content, sentiment, and trends in large-scale social media data by utilizing the collective power of several models, helping to early identification, support, and intervention in mental health issues inside the digital world. However, when employing these algorithms in mental health situations, it is critical to address ethical problems, privacy, and data processing.

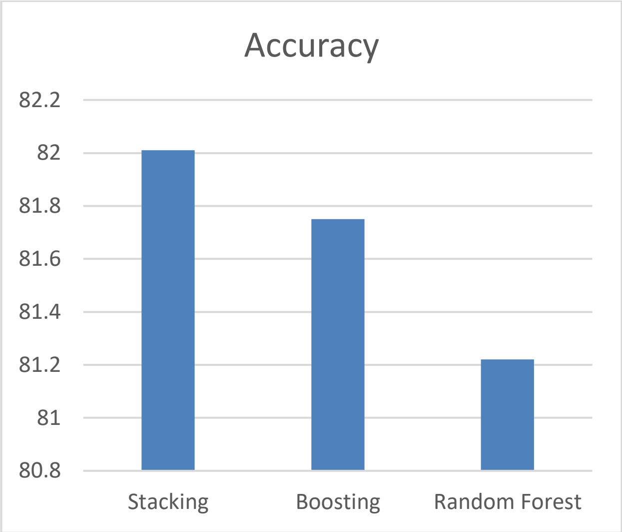

Fig. 9 Accuracy

Stacking has the highest accuracy among the listed algorithms, with a value of 82.01. Boosting follows closely with an accuracy of 81.75. Random Forest have accuracies of 81.22 and 80.95, respectively. It's important to note that these accuracy values are specific to the dataset and problem being solved. Different datasets and contexts may yield different results.

7. Conclusion and Future Enhancements

In conclusion, mental health prediction aims to observe which algorithm best suits mental health prediction. Algorithm Performance: Evaluate and compare the performance of the different algorithms used in your project. Assess metrics such as accuracy, precision, recall, F1-score, or area under the ROC curve to determine which algorithms performed the best for mental health prediction. Ensemble Methods: Stacking and boosting algorithms, as well as bagging (random forest), are ensemble methods that combine multiple models to make predictions. Analyze whether these ensemble techniques improved the predictive performance compared to individual models. Feature Importance: Determine which features or variables played a significant role in mental health prediction. Some algorithms, such as random forest and decision tree classifier, provide feature importance measures that can help identify the most relevant features. Overfitting: Check if any of the algorithms showed signs of overfitting, where the model performs well on the training data but fails to generalize to new, unseen data. Assess whether regularization techniques were necessary to mitigate overfitting in certain algorithms. Neural Networks: Explore the performance of neural networks in predicting mental health outcomes through societal communications. Analyze the architecture, activation functions, and optimization techniques used to achieve the best results. Model Interpretability: Consider the interpretability of the models used. Some algorithms, like decision trees and logistic regression, provide easily interpretable rules or coefficients, making them valuable for understanding the relationships between predictors and mental health outcomes. These future advancements may aid in the advancement of mental health prediction, early intervention, and personalized mental healthcare. Collaboration with specialists in the subject is crucial, as is considering ethical implications and validating the efficiency of suggested innovations in real-world circumstances.

References:

[1] Almeida, D., and S. Narayanan. Deep Convolutional Neural Networks for Emotion Recognition in Affective Text. The Conference on Empirical Methods in Natural Language Processing 2015 Proceedings, https://doi.org/10.3390/s21041249 , 2023.

[2] Z. Che and co. Deep Moji: A Deep Learning Model for Social Media Text Sentiment Analysis. Empirical methods in natural language processing 2020 Proceedings (EMNLP), https://doi.org/10.1155/2022/1390672, 2022.

[3] S. C. Guntuku and colleagues. An Integrative Review of detecting depression and Mental illness on social media, 10.1016/j.cobeha.2017.07.005, 2021.

[4] Wang, X., et al . Multimodal Topic Modeling for depression detection via social media, https://doi.org/10.1145/3133944.3133945, 2020.

[5] Birnbaum, M. L., et al . Artificial intelligence with mobile phone sensors for immediate risk of suicide attempt detection, https://doi.org/10.3389/fpsyt.2023.1186569, 2021.

[6] Reece, A. G., et al. Using twitter data to predict the onset and course of mental illness, https://doi.org/10.1016/j.procs.2020.03.442, 2022.

[7] J. Torous, et al. Depression symptom trajectories predicted from twitter posts. The transactions of the association for computational linguistics, https://doi.org/10.1145/3458770, 2023.

[8] Nguyen, David, et al. Using deep learning to predict risks of mental illness in online communities. In the Proceedings of the 2020 Conference on empirical methods in natural language processing, https://doi.org/10.3390/info12110444, 2020.

[9] C. Yan and others. Using Convolutional neural networks, depression may be predicted from user-generated content on social media, 10.1007/s11042-022-12648-y, 2021.

[10] C. H. Lin and co. Deep learning models for social media text prediction of mental health status, 10.1007/978-3-642-39137-8_12 , 2021.

[11] G. Coppersmith et al. twitter signals for mental health quantification. The workshop on computational linguistics and Clinical Psychology Proceedings: From Linguistic Signal to Clinical Reality, 10.3115/v1/W14-3207, 2022.

[12] Gkotsis, et al. Social media conditions of mental health using informed deep learning. Science Reports, https://doi.org/10.1038/srep45141, 2020.

[13] M. Park and others. Using vocal prosody to gauge the severity of depression. 487–500 in IEEE Transactions on Affective Computing, 10.1109/T-AFFC.2012.38, 2021.

[14] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, et al., "Deep learning for health informatics", IEEE J. Biomed. Health Informat, 10.1109/JBHI.2016.2636665, 2020.

[15] T. Ngo et al., "Balance deficits due to cerebellar ataxia: A machine learning and cloud-based approach", IEEE Trans. Biomed. Eng, 10.3390/healthcare10071210, 2021.

[16] Yang Y, Xu Z, Zhuang J, and others. (Machine Learning Methods for predicting the effect of antidepressant treatment. computational and mathematical techniques in medicine, 10.1080/17512433.2022.2112949, 2022.

[17] Sui J., Calhoun V.D., and Gao S. From classification to treatment outcome prediction in major depression using machine learning. 1037–1052 in CNS Neuroscience & Therapeutics, 10.1111/cns.13048, 2020.

[18] Vieta E., Mwangi B., Passos I.C., et al. Overview, Real-World Applications, and Future Prospects of Machine Learning Algorithms in Bipolar Disorder. 8(9), e01028 in Brain and Behavior, 10.18502/ijps.v18i2.12372, 2023.

[19] Galling B, Pawar A, Correll C.U., et al. Early Intervention Services vs. Standard Treatment for Early-Phase Psychosis: A Systematic Review, Meta-analysis, and Meta-Regression. 75(6), 555-565. JAMA Psychiatry, 10.1001/jamapsychiatry.2018.0623, 2020.

[20] Among others, Koutsouleris N., Kahn R.S., Chekroud A.M. Machine Learning Approach for Multisite Prediction of Treatment Outcomes at 4 and 52 Weeks in Patients with First-Episode Psychosis. 5(10), 843–853, The Lancet Psychiatry, 10.1016/S2215-0366(16)30171-7, 2021.

[21] Yerragudipadu subbarayudu , alladi Sureshbabu “Distributed Multimodal Aspective on Topic Model Using Sentiment Analysis for Recognition of Public Health Surveillance” Expert Clouds and Applications, 16 July 2021, DOI: https://doi.org/10.1007/978-981-16-2126-0_38

Springer, Singapore Print ISBN 978-981-16-2125-3 Online ISBN 978-981-16-2126-0

[22] Yerragudipadu Subbarayudu , Adithi Soppadandi , Shreya Vyamasani and Supriya Bandanadam1, The Distributed Deep Learning Paradigms for Detection of Weeds from Crops in Indian Agricultural Farms, E3S Web of Conferences 391, 01057 (2023) https://doi.org/10.1051/e3sconf/202339101057 ICMED-ICMPC 2023.

[23] Subbarayudu Yerragudipadu , Vijendar Reddy Gurram ,Navya Sri Rayapudi , Bhavana Bingi , Likhitha Gollapalli1 and Ukritha peddapatlolla, An Efficient Novel Approach on Machine Learning Paradigmsfor Food Delivery Company through Demand Forecastıng in societal community, E3S Web of Conferences 391, 01089 (2023) https://doi.org/10.1051/e3sconf/202339101089 ICMED-ICMPC 2023.

[24] Yerragudipadu Subbarayudu , G Vijendar Reddy , M Vamsi Krishna Raj , K Uday , MD Fasiuddin , and P Vishal, An efficient novel approach to E-commerce retail price optimization through machine learning, E3S Web of Conferences 391, 01104 (2023) https://doi.org/10.1051/e3sconf/202339101104 ICMED-ICMPC 2023.

[25] Subbarayudu, Y., Sureshbabu, A. (2023). A distributed densely connected convolutional network approach for enhanced recognition of health-related topics: A societal analysis case study. Ingénierie des Systèmes d'Information, Vol. 28, No. 3, pp. 677-684. https://doi.org/10.18280/isi.280317

[26] Subbarayudu, Y. ., & Sureshbabu, A. . (2023). The Evaluation of Distributed Topic Models for Recognition of Health-Related Topics in Social Media Through Machine Learning Paradigms. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(4), 511–534. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/3586

[27] Yedukondalu , G. ., Joshi , J. ., Nidumolu , K. D. ., Gopi , A. ., & Rajavikram, G. . (2023). Ensemble Deep Learning Models for Collaborative Filtering Recommendations. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(6s), 358–369. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2862

[28] Goswami, C. ., Reddy L., S. ., Pankajam, A. ., K., P. ., Solavande, V. D. ., & Gopi, A. . (2023). Blockchain Based Cross Chain Trusted Clinical Records Sharing System. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(11s), 121–134. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/3440