IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 4, 2020, accepted February 24, 2020, date of current version March 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2977887

Realizing a Stacking Generalization Model to Improve the Prediction Accuracy of Major Depressive Disorder in Adults

NIVEDHITHA MAHENDRAN[1], P. M. DURAI RAJ VINCENT[1], (Member, IEEE), KATHIRAVAN SRINIVASAN[1], (Senior Member, IEEE), VISHAL SHARMA[2], (Member, IEEE), AND DUSHANTHA NALIN K. JAYAKODY[3,4], (Senior Member, IEEE)

[1]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore 632014, India
[2]Department of Information Security Engineering, Soonchunhyang University, Asan 31538, South Korea
[3]Centre for Telecommunication Research, Faculty of Engineering, Sri Lanka Technological Campus, Padukka 11500, Sri Lanka
[4]School of Computer Science and Robotics, National Research Tomsk Polytechnic University, 634050 Tomsk, Russia

Corresponding author: Dushantha Nalin K. Jayakody (nalin@tpu.ru)

This work was supported in part by the Scheme for Promotion of Academic and Research Collaboration (SPARC), Ministry of Human Resource Development, India, under Grant SPARC/2018-2019/P145/SL, in part by the framework of the Competitiveness Enhancement Program of the National Research Tomsk Polytechnic University, Russia, in part by the Russian Foundation for Basic Research (RFBR) according to the research project under Grant 19-37-90037/19 and Grant 19-37-90105/19, in part by the International Cooperation Project of Sri Lanka Technological Campus, Sri Lanka, and in part by Tomsk Polytechnic University, Russia, under Grant RRSG/19/5008.

ABSTRACT Major depressive disorder (MDD) is a persistent psychiatric mood disorder that is prevalent from a few weeks to a few months, even for years in the worst cases. It causes sadness, hopelessness in the individuals; sometimes, it forces them to hurt themselves. In severe cases, MDD can even lead to the death of the individual. It is challenging to diagnose MDD as it co-occurs with many other disorders (Co-Morbid) and many other reasons such as mobility, lack of motivation, and cost. The way to diagnose MDD is usually high ended that is challenging for the regular clinicians to diagnose. Therefore, to make their work more comfortable, and to predict MDD at the early stages, we have developed an ensemble-based machine learning model. The data collected has been cleaned with a preprocessing technique, and feature selection are performed using wrapper based methods; moreover, in the final step, a stacking based ensemble learning model is implemented to classify the MDD patients. Furthermore, KNN Imputation is implemented for preprocessing, Random Forest-Based Backward Elimination for feature selection and multi-layer perceptron, SVM and Random Forest as low-level learners in stacking generalization model. The results show that the prediction accuracy of the stacking generalization model is superior to the individual classifiers.

INDEX TERMS K-nearest neighbors, major depressive disorder, multilayer perceptron, random forest, random forest-based feature elimination, stacking generalization and support vector machine.

I. INTRODUCTION

A. MOTIVATION

Depression is the most persistent forms of psychopathology [1], which affect one's way of thinking, behavior, and feel. It causes prolonged feelings of hopelessness and sadness [2]. Depression is different from occasional mood swings or being upset, which are for a few days to years. The reason for a person to suffer from depression is debatable, and the available treatments are ineffective for many patients [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Jihwan P. Choi.

In one of the clinical trials "STAR*D," it is recorded that only one-fourth of the patients affected with depression showed improvement because of its covert heterogeneous nature [4]. According to WHO, it is estimated that 57 million people who are 18% of the global estimate from India are victims of depression [5]. Global statistics show that around 4.4% of the total population are affected by depression, and it is found common in females [6]. Based on the DSM-5 scale, there are nine symptoms of depression among, which, if the individual shows a minimum of five symptoms from nine, and then he/she would be considered depressed. Also, in those five, at least one should be losing interest or depressed mood [7].

When there is a strong, intense and long-lasting feeling of depression, then it is declared as Major Depressive Disorder (MDD) [2]. DSM (Diagnostic and Statistical Manual of Mental Disorders) is a reference manual, which provides a common standard for classifying mental disorders based on the self-reported ratings.

There are many factors for MDD; significant factors are stress, genetics, biochemical reactions, and hormonal imbalance [2]. The three parts of the brain that play an essential role in MDD are Hippocampus, Amygdala, and prefrontal cortex [8]. The hippocampus is placed in the center of the brain. The task of the hippocampus is to store memory and to regulate a hormone called cortisol. It is found that during phases of depression, the hippocampus secretes cortisol in excess, which affects the production of neurons (brain cells) and causes memory problems [3], [8]. The amygdala is the center for generating emotional responses such as pleasure and fear. High levels of cortisol make amygdala enlarged and hyperactive, which causes insomnia and irregularities in the secretion of hormone and many other chemicals [3], [8]. The Prefrontal cortex, which is located in the frontal part of the brain, is responsible for forming memories and decision making. An excess amount of cortisol shrinks the prefrontal cortex [3], [8].

MDD is challenging to diagnose as it is usually comorbid (co-occurs with another disorder) such as in older adulthood; it occurs with coronary heart disease and type 2 diabetes. MDD in early adulthood increases the risk of heart disease [9]. In many cases, depression co-occurs with an anxiety disorder and leads to substance abuse, smoking, and alcohol consumption in excess [10]. According to NHMS (2015-2016), about one in 20 individuals, i.e., 5-25% who are above 18 years of age has been affected by MDD at least once in India. Globally, 322 million were suffering from depression as of 2015 report [5].

According to the recent report by WHO, carried out by the National Care of Medical Health (NCMH), India is mentioned as the most depressed country in the whole world, followed by China and the USA. There are scaring numbers in India with Depression, Anxiety, Schizophrenia and Bipolar, which go unreported. The report also shows that 6.5% of the Indian population is suffering from some mental disorder. It clearly shows that there is a considerable gap between the therapist or professionals and the patients. The numbers are so depressing in India and help is a rare sight, as the suicide rate in India is 10.9% of every hundred thousand people [5].

Also, the disturbing fact is that among the depressed people, 80% who are diagnosed with some mental disorder do not turn up for treatment. It is also predicted that it would increase by 20% in 2020. If these mental disorders are left undiagnosed, they will impair both physical and mental health, especially in adults. The numbers go up day by day and have become a taboo topic in this millennial. Though there is a considerable number of cases in MDD reported and unreported, the research in this area is very scarce. These scary figures motivated us to develop a model, which can be helpful for clinicians or mental health care professionals to diagnose early and start the treatment or therapy.

Individuals suffering from MDD might have certain features in common, mostly feeling low, almost all the time, which would last weeks or years in the worst case when compared with a healthy individual [11].

In general, to treat individuals affected by MDD, it is essential to be diagnosed [10]. There are many constraints like motivation, cost, and mobility which makes a depressed individual get professional help tedious. The methods handled to predict the outcome are usually mechanistic ways, such as brain wave imaging, genetic which uses high ended equipment unavailable to the clinicians who are trying to make treatment decisions [12]. So we have chosen a better alternative which is to develop a Machine learning model with features of a self-rated report made from demographic attributes and symptom scores. MDD can be effectively diagnosed when we look for a pattern in the data instead of concentrating on individual attributes which can be done efficiently by using machine learning techniques. Machine learning techniques can predict the outcomes strongly by identifying the best combination of variables.

Machine learning is a process of constructing a model that learns from data and makes predictions without being programmed explicitly [13]. The process of prediction using machine learning involves mostly three stages [14]. It involves the acquisition of data, or preprocessing, where the data are cleaned and made ready for further processing, feature selection where the dimension of features is reduced by selecting only the necessary features. The final stage is choosing a suitable model for prediction. There are three most widely used machine learning models [15] they are supervised learning, unsupervised learning, and reinforcement learning, among which a suitable model can be chosen. Machine learning is the most attractive because it makes minimum or no assumptions about the data also has a more exceptional ability to solve complex problems [16].

Along with Machine Learning, nowadays, Deep Learning also has gained popularity among researchers in the field of health care [17]. It includes advanced algorithms such as Deep Convolutional Neural Networks (DCNN), Deep Belief Network (DBN), and many others [18]. With these upcoming and advanced techniques, the classification and prediction of a disorder can be done with higher efficiency.

In this study, we have implemented the Stacking ensemble model along with Random Forest-based Backward Elimination as a Feature Selection method. There are not many ensembles-based models in the existing works. The Stacking generalization concept has not been explored well in the previous works. The Stacking ensemble method decreases the variance and also improves the predictive power of the model. Hence, we have used three completely unrelated algorithms to form a base and meta-learners to form a stacking ensemble, which was used in the classification of those who are suffering from MDD.

B. PROBLEM STATEMENT

This work aims to develop a robust machine learning model to analyze the data collected, preprocess the data for missing values as missing values can make the prediction less efficient. After cleaning, the essential features of the feature set are selected to make sure only the features that affect the target variable directly are used and then to implement an ensemble-based learning method to predict MDD with better accuracy. The individual implementation of classifiers always have issues such as over-fitting, non-robust, difficulty in selecting the optimal parameters. On the other hand, when they are combined by averaging or majority voting, the problems between them are compensated; hence stacking ensemble has been implemented by combining all the base learners.

C. OUR CONTRIBUTION

The data collected is cleaned using KNN imputation, then selected necessary features using novel Random forward, based Backward Elimination (RF-BE). Further, the stacking ensemble model combining Random Forest, Multilayer Perceptron, and Support Vector Machine has been implemented. The remaining paper is structured as a theoretical background, related works, materials and methods, methodology, results and discussions and conclusion.

II. THEORETICAL BACKGROUND

A. DATA PREPROCESSING

Preprocessing is the initial and most significant stage of the machine learning technique. If there are missing values in the data or redundant and irrelevant features, it would significantly impact the predicted outcomes. Hence it is essential to clean the data before further processing [16]. There are many techniques to treat the data with missing values and redundant data, such as deletion methods, which include pairwise deletion and list wise deletion, then there are imputation methods such as mean, regression imputation, and many others [19]. In this paper, we have used KNNImputation methods to clean the data collected.

B. FEATURE SELECTION

Feature selection is the process of selecting the best subset of features that have a significant impact on the predicted outcomes. For prediction/classification to be efficient and reduce the computation time, feature selection is the most needed one [20]. Feature selection methods are usually categorized into three types [14] – Wrapper methods, Filter methods, and embedded methods. Figure 1 shows the diagrammatic representation of all three methods of feature selection. The three types of feature selection are:

1) WRAPPER METHOD

They are a form of the black-box method [14], where we feed them with features and do not know the internal structure or functionality. It is done using random searching, in which cross-validation is done to predict the advantage of adding a feature or removing it from the subset. The wrapper method wraps the implementation of the chosen model, along with the search technique [20]. They sometimes lead to over-fitting and time-consuming [19].

2) FILTER METHOD

Unlike the wrapper method, the filter method is independent of the learning model [14]. The features are selected based on their performance scores on the statistical measures [19]. The filter method ranks every feature, based on their relevance score. Filter methods are faster than other feature selection methods, but they are not computationally effective as they lack interaction with the learning model [20].

3) EMBEDDED METHOD

In embedded methods, the training of a machine learning model itself will have a feature selection step as a part of it (decision tree, LASSO) [14]. Although it is more effective than filter and wrapper methods, it is highly dependent on the learning model, thus it is difficult to use with any other learning model [20].

C. MACHINE LEARNING MODELS

Based on the approach used, the machine learning models are classified into Supervised, Unsupervised and Reinforcement learning and a special type called ensemble learning [19].

1) SUPERVISED LEARNING

Supervised learning is for performing prediction for labeled class data [18]. The model is trained and analyzed for essential features and then tested with unlabeled class data [13]. The model gains knowledge in training and applies the knowledge in the testing phase with real-world data [15]. Based on the data on hand, it is of 2 types: Classification and Regression, where the former is for discrete binary data and later is for continuous data [13].

2) UNSUPERVISED LEARNING

Unlike supervised learning, Unsupervised does not involve any training, and the data is unlabeled [14]. Unsupervised learning identifies patterns in the given unlabeled data. Here the model is introduced with data from the real world and made to gain knowledge on their own and find hidden meaning from the data [15]. Widely used unsupervised algorithms are Clustering and Association rule mining [20].

3) REINFORCEMENT LEARNING

Reinforcement learning models are mostly goal-oriented models [20] based on the feedback given by the environment or an entity in the environment, which is external [15]. For example, when a dog (entity) fetches a stick, it is rewarded with a biscuit (Positive feedback) and if it fails, then no biscuit (Negative feedback) and in gaming (environment).

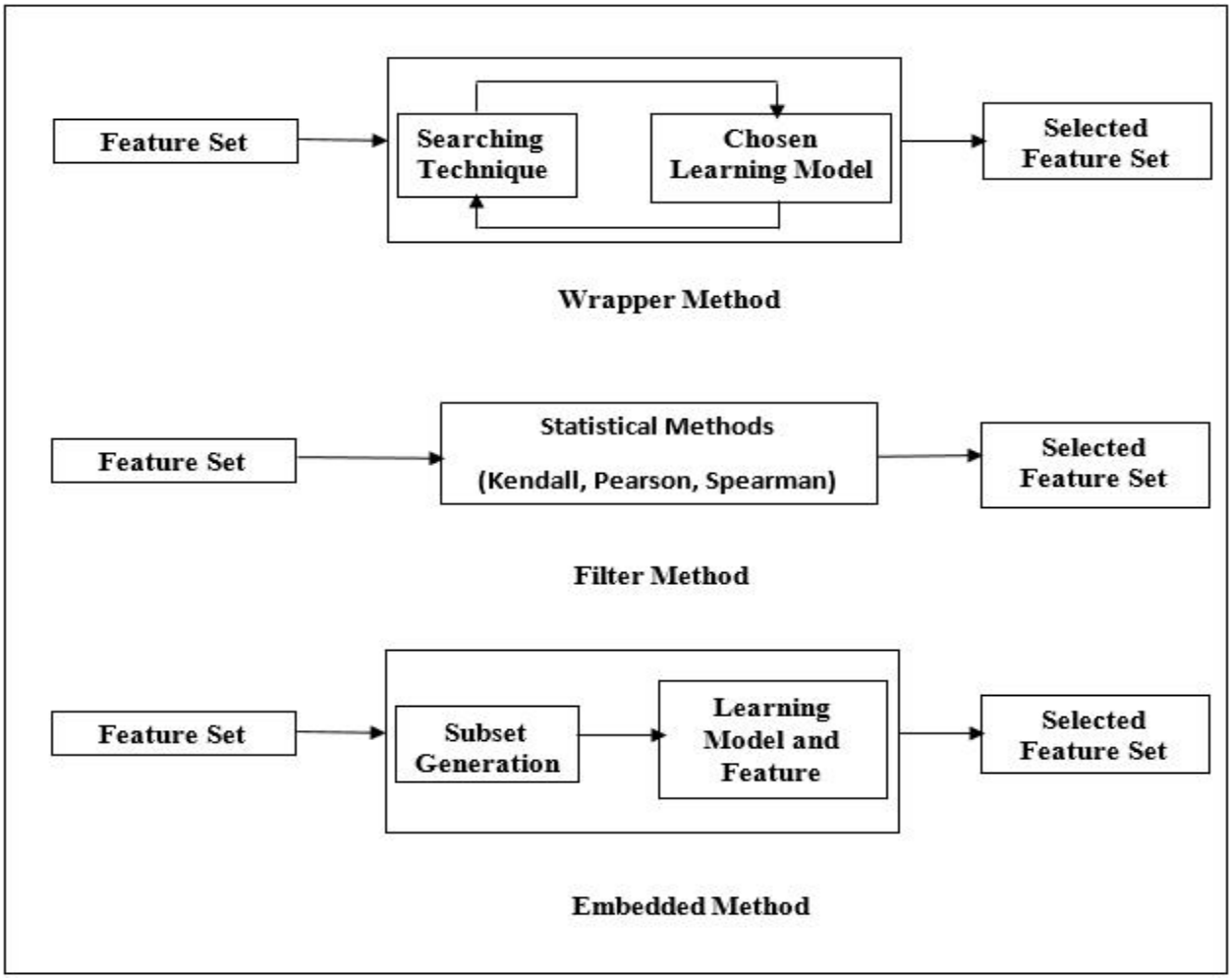

FIGURE 1. Feature Selection Methods.

D. ENSEMBLE LEARNING

Ensemble learning consists of meta-algorithms, which combine several learning models into one model, which improves the prediction accuracy, decreases bias, and variance [22]. The ensemble learning is stronger when compared to the individual learner algorithms because of its generalization ability [23]. Based on the bias, variance, and prediction accuracy, it is classified into three types: Bagging, Boosting, and Stacking [22].

1) BAGGING

Bagging is also called bootstrap aggregation, which is a simple yet effective ensemble algorithm [22]. Bagging offers diversity by creating replicas of the training data of the entire training data. It trains different base learner classifiers by sampling with replacement [23]. Bagging decreases the variance and suited for a limited size dataset.

2) BAGGING

Bagging is also called bootstrap aggregation, which is a simple yet effective ensemble algorithm [22]. Bagging offers diversity by creating replicas of the training data of the entire training data. It trains different base learner classifiers by sampling with replacement [23]. Bagging decreases the variance and suited for a limited size dataset.

3) BOOSTING

Boosting is an iterative approach combining various weak learners, which results in low training error. Unlike bagging, where the training data are bootstrapped replicas that have a possibility of having the same instances in each training set, boosting the training dataset for each classifier is the instances that are misclassified by the previous classifier [22]. All the instances are initially given uniform weights, and the first base learner is implemented with weighted cases. The error is calculated, and the weights of misclassified instances are increased, and correctly classified instances weights are decreased. Boosting decreases bias in the learning process [21].

4) STACKING

Stacking tries to achieve greater prediction accuracy by implementing different lower level learners and then combine them using a high-level meta-base learner [23].

The meta-level base learner combines all the predictions of multiple individual lower-level base learners and forms a stacking generalization model. The stacking model increases the prediction accuracy, but difficult to analyze theoretically [23].

III. RELATED WORKS

The authors [24] have tried to implement an SVM-FoBa (support vector machine-based forward and backward searching strategy) as a wrapper method to find the parameters that distinguish bipolar disorder from major depressive disorder (MDD). They have implemented other benchmark feature selection techniques and compared them with the developmental SVM-FoBa model and found that the SVM-FoBa performs better than them. They have also evaluated the selected features by doing weight analyses. After applying the proposed feature selection method, the classification accuracy of the classifiers improved significantly.

The approach of wrapper based feature selection SVM-RFE (Support vector machine-based Recursive Feature Elimination) handled by the authors selects the best subset for analyzing agro-industrial products [25]. They also implemented RF-RFE (Random Forest-Based Recursive Feature Elimination) along with SVM-RFE. They found that RF-RFE performed better than SVM-RFE and compared the results with KWS (Kruskal Wallis Statistic Method) and showed that RF-RFE selects the feature subset in a better manner.

The authors [26] had aimed to develop a credit scoring model using an ensemble machine learning algorithm. They have proposed a novel heterogeneous model combining bagging and stacking ensemble approaches known as bstacking model. To evaluate the efficiency of this model, they have implemented many other methods such as homogeneous and heterogeneous ensemble models, individual base learners. Their approach differs in various ways from all other approaches in trainable Fuser, pool generation, and base learners selection. They have employed four base learners XGBoost, GPC, RF, and SVM and trained them as bagging ensembles, and then they developed a bstack stacking approach by combing these bagging trained models using a Meta learner. After implementing bstack and testing them against the benchmark models, they found it outperforms all the other benchmark models.

The authors [27] have implemented all three ensemble models bagging, boosting, and stacking to classify Arabic tweets into predetermined categories. They have implemented a decision tree, Naïve Bayes, and Sequential Minimal Optimization as individual learners. Then they have developed ensemble models with the same classifiers and found that the ensemble models outperform individual learners. Also, among the ensemble models, stacking takes more time to execute as it has two layers, i.e., meta layer and base layer. Boosting takes less time than stacking and bagging takes less time than boosting and stacking.

A cascaded classifier is a form of ensemble method [28]. The authors have implemented a cascaded classifier and three stacking approaches for classifying the characteristics of a pulmonary nodule. The cascaded method consists of two levels, and stacking was applied at both the levels. Each level has base learners such as SVPS, SVM, KNN, and RF and stacked together. The results show that the cascaded classifier ensemble approach performs better than the individual classifiers and other ensemble approaches.

SVM is the most widely used successful machine learning classifier; however, when it comes to optimized features with C value in predicting petroleum reservoir characteristics, SVM does not perform well [29]. To overcome this issue, the author has proposed a stacking based ensemble method of SVM. They have also developed a traditional bagging technique with SVM and random forest. They have built a stacking ensemble of SVM with different C values and implemented SVM individually also. They have compared the results of the bagging ensemble, stacking ensemble of SVM, and individual implementation, and they found that the stacking ensemble of SVM outperforms the other two applications.

The structure of protein folds is always almost similar such instances are termed as homology [30]. To overcome homology and predict the structural class of protein along with their varying homology, the authors have developed an ensemble classifier. The proposed ensemble stacking method consists of four base learners RF, SVM, Logistic Regression, and Instance-Based. They have implemented three ensemble techniques Voting, Multi-scheme, and stacking, among which stacking produces better results.

The authors [31] have developed three ensemble models based on bagging ensemble technique M5' ensemble, CART ensemble, and hybrid M5' and CART ensemble to find the parameters of ground peak time-domain. They have tested the developed model against the statistical error parameters. They have also compared the developed model with other existing machine learning models, and the results show that the M5' and hybrid of the M5' and CART produce accurate results than other techniques. They have further tested the robustness of both the models and showed that the hybrid model is more efficient than the M5' ensemble.

The authors [32] have proposed an ensemble machine learning model for the clinicians to use in predicting the transfer of patients from regular ward to pediatric ICU care when it is likely. They have developed two boosting ensemble models gradient tree boosting and adaptive boosting. They have also developed a third model by averaging the boosting methods to form another ensemble model. They have proved that the data-driven ensemble approach performs effectively than other traditional models.

Table 1 shows the related works and methodologies used by them. From the table, we can see that the dataset used by the existing works is mostly private and not available as open source. Thus, we have used an in-housed dataset, which was collected in real-time through a standard questionnaire.

TABLE 1. Related works.

Author Name	Feature Selection/ Ensemble Method	Methodology	Accuracy	Dataset
Jie, N. F, Zhu, M. H., Ma, and et al	Feature Selection	SVM-FoBa	92.1%	Closed
Granitto, P. M., Furlanello and et al	Feature Selection	SVM-RFE	91%	Closed
Xia, Y., Liu, C., Da, B, and et al.	Ensemble Learning	Bstacking (bagging and stacking), XGBoost, GPC, RF, and SVM.	88%	Closed
Abdelaal, H. M., Elmahdy, and et al.	Ensemble Learning	Boosting, bagging, Stacking (Naïve Bayes, Decision Tree, SMO)	87.4%	Closed
Kaya, A	Ensemble Learning	Cascaded Method (SVPS, SVM, KNN, and RF)	88.80%	Closed
Anifowose, F., Labadin, J., and Abdulraheem, A	Ensemble Learning	Stacking ensemble (SVM)	89%	Closed
Kedarisetti, K. D., Kurgan, L., and Dick, S.	Ensemble Learning	Stacking (RF, SVM, LR)	73%	Closed
Hamze-Ziabari, S. M., and Bakhshpoori, T	Ensemble Learning	CART, Hybrid M5	84.6%	Closed
Rubin, J., Potes, C., Xu-Wilson and et al	Ensemble Learning	Gradient boosting, Adaptive Boosting.	92%	Closed

IV. METHODOLOGY

The whole process consists of 3 stages, namely, Preprocessing, Feature Selection and Classification. Figure 2 shows the diagrammatic representation of the whole process.

A. MATERIALS AND METHODS

The dataset we have used for this study is collected from a psychiatric department in a hospital that has around 3040 records and 22 features, including the demographic features and symptom scores. The dataset has been split into equal halves of training and testing data, each having 1520 records and 12 features after feature selection. On the selected training set, the model is trained, and outcomes are classified. Then, with the testing data, the trained model was evaluated for accuracy. Table 3 shows the feature descriptions.

B. PREPROCESSING

The collected dataset has missing values; hence to clean them, we have used the KNN Imputation technique. The KNN Imputation algorithm finds the distance between closest records and identifies the K most intimate record and calculates weights according to widely used distance measures

TABLE 2. Notations used in the pseudo-code.

Notation	Description
V	List of all features
α[i]	List of p-values of the features
X	Largest p-value
T[i]	List of features after eliminating the largest p-value
S	Results after training the classifier with selected features
W_i	Weights of the features
R	Rank of features based on their weights
M	Minimum Weight

such as Minkowski, Euclidean, and Manhattan distance measures [33], [34]. In R Studio, for implementing KNN Imputation, we have used knnImputation method, which is under the package ''DMwR.'' The parameters we have considered in the knnImputation method are, 'k' (k = 4), which is the number of neighbors and the distance measure. We have used the Manhattan distance measure (absolute difference between 2 records) used with KNN Imputation to clean the data. The Manhattan distance measure for computing distance between the two records is [33],

$$MD\,(X, Y) = \sum_{k=1}^{n} |X_i - Y_i| \tag{1}$$

C. FEATURE SELECTION

In the depression data set we have used, there are about 22 features and 2,300 records. The dataset has features that are dependent on each other. Dependent features directly affect the accuracy of the classification model because of the feature interaction [35]. To reduce the feature interaction and remove redundant variables, we have implemented a novel feature selection algorithm Random Forest-based Backward Elimination (RF-BE).

In R Studio, we have used the ''caret'' package for implementing the feature selection method.

The backward elimination is a searching technique that takes the dataset as a whole with all the predictors and eliminates the least performing predictor one by one [36]. The random forest is a combination of decision trees, and we have combined backward elimination and random forest and developed a wrapper-based feature selection technique for obtaining better results.

D. ENSEMBLE CLASSIFIER MODEL

The feature selection technique RF-BE selects the best possible features that are important in predicting depression, which is a binary classification problem. The next step is to apply a machine learning technique to classify depressed and not depressed records. In this study, we have implemented a stacking ensemble classifier method consisting of Multilayer Perceptron (MLP), SVM, and a random forest as low-level base learners and multilayer perceptron as a meta-level algorithm.

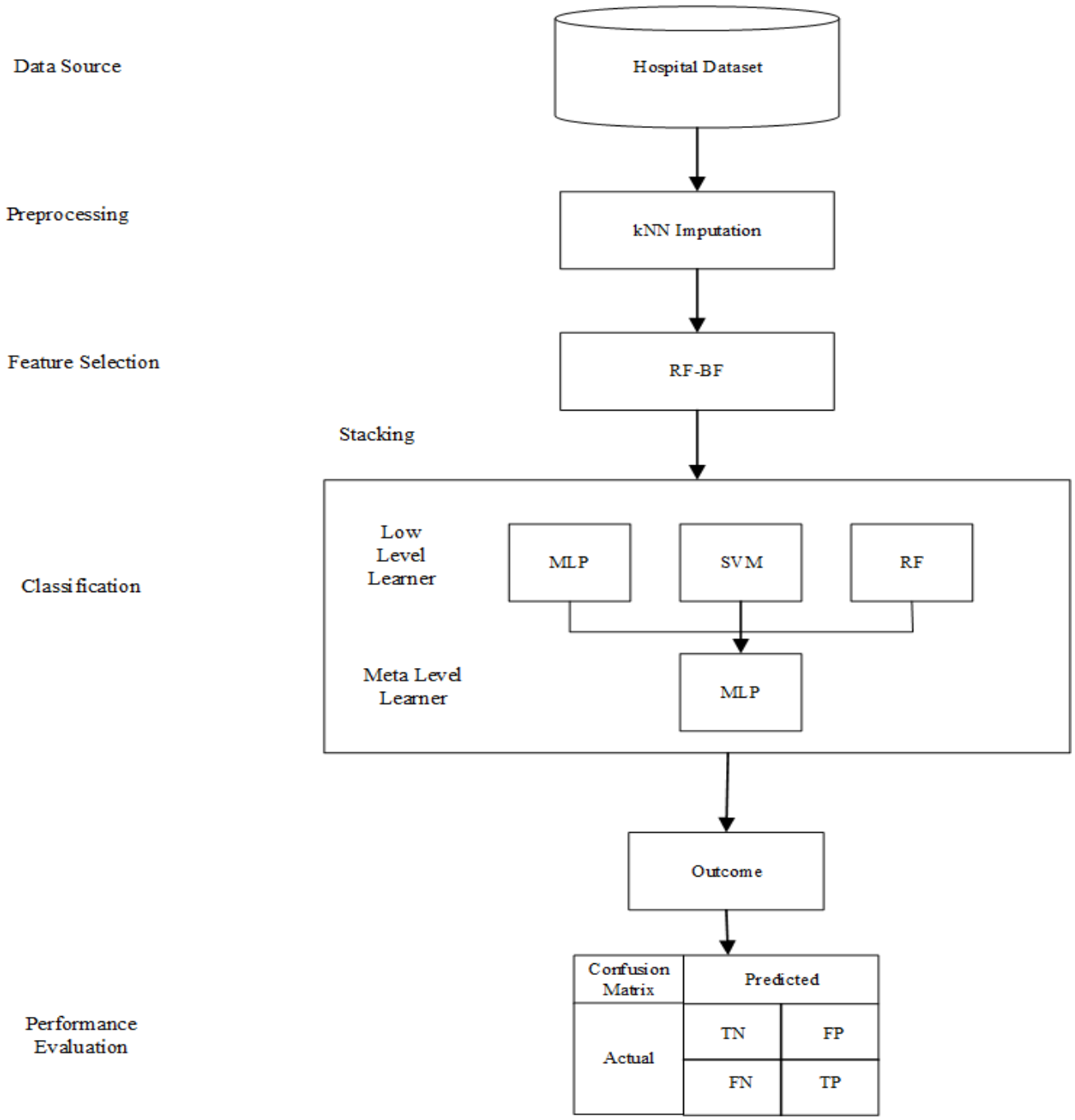

FIGURE 2. Process Flow Diagram.

The three base learners are first implemented separately and then formed as stacking ensemble with a multilayer perceptron as meta-level. A multilayered perceptron is a feed-forward artificial neural network [37]. ANN's functionalities and structure are similar to the functionalities and the structure of the brain. The neurons are the functional units which are connected to each other and function in parallel. These neurons are connected using a connecting link which has the information for solving problems [38], [39].

For implementing MLP in R Studio, we have used the MLP method, which is under the package ''RSNNS,'' with parameters, a number of hidden layers, learning rate (0.25), and learning function (Back Propagation).

A support vector machine (SVM) is a supervised machine learning algorithm for regression and classification problems. Each attribute is a point in n-dimensional space (n – the number of features) [38]. For performing linear classification, a hyperplane is formed, which separates the two classes

TABLE 3. Feature description.

Features	Description
Age	Age of the individual (Average age is 28)
Gender	Male/Female
Falling Asleep	The time is taken for an individual to sleep once he/she goes to bed.
Waking up too early	Waking up at the correct time or early or too late.
Sleeping too much	Sleeping for more hours than required. For example, more than 7-8 hours.
Feeling Sad	Feeling sad for half of the time or none or nearly all the time.
Feeling Irritable	Feeling irritable half of the time or none or nearly all the time.
Feeling Anxious or Tense	Feeling anxious or tense half of the time or none or nearly all the time.
The response of Your Mood to Good or Desired Events	Mood lights up when the desired event happens or normal, or it never brightens up any time.
The mood in Relation to the Time of Day	Mood changes based on time or it is bad/worse always.
The Quality of Your Mood	Mood-based on the day that something happened made the individual sad, or it is sad for no reason.
Decreased Appetite	Eating very less or rarely eat in the 24 hours of the day.
Increased Appetite	Eating more than required, such as eating both during mealtime and between the meals.
Decreased Weight (Within the Last Two Weeks)	Losing more weight in a particular period of time.
Increased Weight (Within the Last Two Weeks)	Gaining more weight in a particular period of time.
Concentration/Decision Making	Struggling to focus on something or failing to make decisions.
View of My Future	Having optimistic/pessimistic thoughts about the future.
Thoughts of Death or Suicide	Thinking about harming oneself or attempting suicide always or not at all.
Capacity for Pleasure or Enjoyment	Normal feeling of enjoyment in any pleasurable actions or feeling extremely irritable in any pleasurable actions.
Feeling restless	Difficulty in focusing on anything and always pacing around or none at all.
Bodily symptoms	Having symptoms like faster heartbeats, visions blurred, chest pain, sweat, shivering, and shaking or none at all.
Panic/Phobic symptoms	Having panic attacks consistently or none at all
Outcome	Depressed or Not Depressed.

well enough. For non-linear classification, kernel methods like Gaussian kernel, Laplace kernel, Polynomial kernel are used [40], [41].

For implementing SVM in R Studio, we have used the SVM method under the package, ''e1071'' with parameters, type (classification), and kernel (linear).

Random forest is a learning algorithm that is a combination of several decision trees. It predicts by averaging the decision tree predictions [42]. The tree is built with a root node that has the feature that contributes significantly by separating the sample classes. Then the tree is divided into many branches until an outcome (decision node or lead) is attained [43].

It is challenging to obtain optimal parameters using a Multilayer perceptron; the robustness of SVM is not always good and random forest has overfitting issues [44]. To overcome these issues and to improve the prediction accuracy, we have developed a stacking ensemble model. In Stacking, these three models will compensate for the issues and offer better results.

For implementing the Random Forest in R Studio, we have used the randomForest method, which is under the package, ''randomForest,'' with parameters such as, ntree (500) and Importance (TRUE).

Pseudocode RF-BE

Input: List of all predictors: $B_0 = B_1, B_2 \ldots . B_n$

Step 1: Select all the features in the list $V = [1, 2, 3 \ldots . n]$

Step 2: Find the p-value of all the features in the list

$$\alpha\,[\mathrm{i}] = p - value\ of\ all\ the\ attributes$$

Step 3: Eliminate the feature which has the largest p-value For i in 0: n

$$If\,\alpha\,[\mathrm{i}] > p(thresholdp - value)$$
$$X\,[i] = \alpha[i]$$
$$T\,[i] = B\,[i] - X[i]$$

Step 4: The classifier with selected features is trained.

$$S = \text{RF-train}(T, V)$$

Step 5: Compute the weight

$$W_i = \sum_i S_i T_i$$

Step 6: Determine the rank of features based on their weights

$$R = (W_i)^2 \forall i$$

Step 7: Find the feature with the smallest rank

$$M = \min(R)$$

Step 8: Eliminate the feature with the smallest rank and update the list.

$$L = [V\,(M)\,, M]$$
$$V = V(1 : M - 1, M + 1 : length\,(V))$$

Output: Ranked Feature List.

V. RESULTS

The original dataset consists of 3040 records with 22 features. After preprocessing the data with the help of KNN Imputation, the missing values are handled. In the second step, feature selection is made by applying a novel wrapper-based feature selection method, RF-BE, which eliminates low

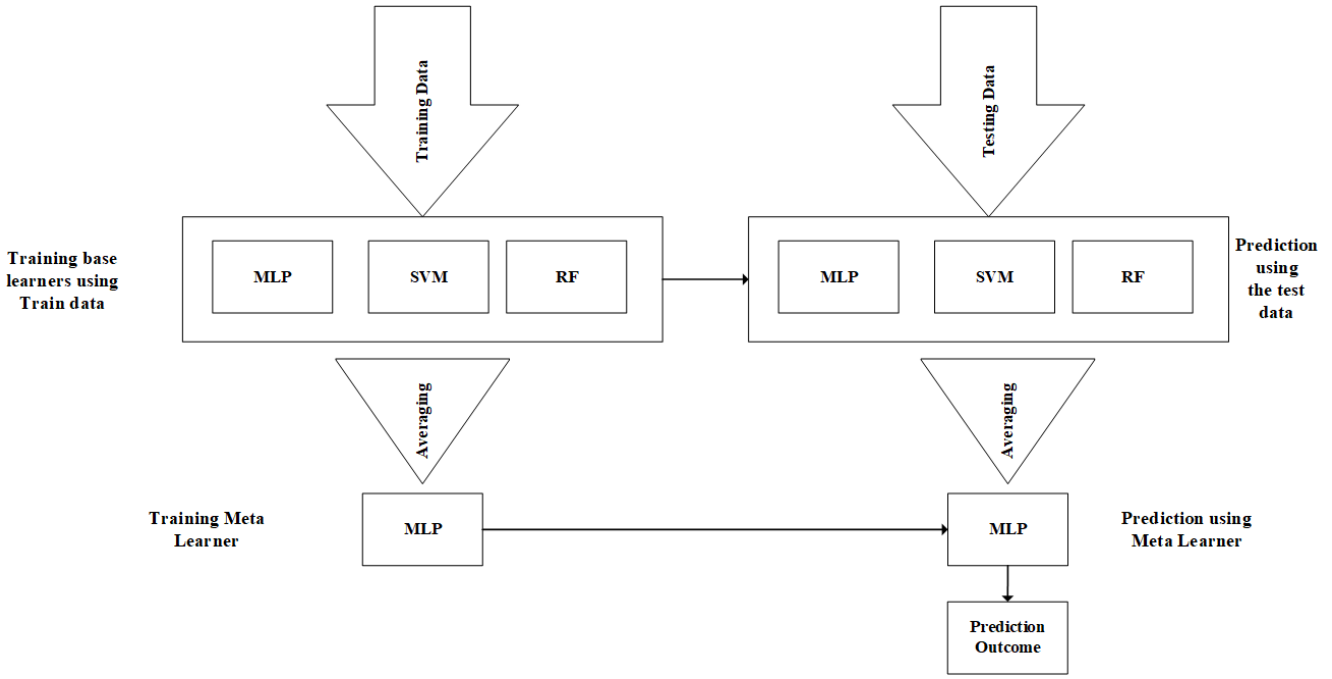

FIGURE 3. Stacking Generalization Models.

performing features in iteration and ranks them according to their evaluation score. Twelve features have been selected by the algorithm, which is independent and will not interact much in the way of affecting the target variable. After choosing the critical features, the classification is divided into two parts in the first part the base learners are trained, and the outcomes are predicted individually. In the second part, to overcome the issues with the individual implements of the base learners, we have developed a stacking generalization algorithm. In the stacking generalization algorithm, MLP, SVM and DT are trained as low-level base learners, and the predicted outcomes are inputted to the meta-level learner, which is MLP Figure 3 shows the process flow of the stacking generalization model. The steps involved in developing a stacking generalization model are as follows:

1) The individual base learners (MLP, SVM, and DT) are trained on the training data.
2) Using the base layer of the model prediction is done for both training and testing data.
3) The top layer, which has the meta-learner (MLP), is trained on the predictions from the base layer that has been made on training data.
4) Using the base layer prediction done on testing data, the top layer prediction is made.

A. PERFORMANCE EVALUATION

For the evaluation of the developed model, we have used the confusion matrix. The confusion matrix is a table that is used to evaluate the performance of a classification model, which was tested on test data, and the true values are known. The basic terms of confusion matrix were True Positive (TP) when the model predicted positive, and the actual is also positive, True Negative (TN) when the model predicted negative, and the actual is also negative, False Positive (FP) when the model predicted positive, but the actual was negative, False Negative (FN) when the model predicted negative but the actual is positive [45]. Using the confusion matrix, various rates can be computed, and in this study, we have used accuracy, sensitivity, specificity, precision, and measure for evaluating our model. The confusion matrix parameters can be explained as follows,

Accuracy – It is the percentage of cases classified correctly as depressed or not depressed.

Sensitivity – It is used to find cases that are identified correctly as depressed.

Specificity– It is used to find cases that are identified as without depression.

Precision– It is used to find out the percentage of cases that the classification model has found out as depressed and are actually depressed.

FMeasure – FMeasure is the harmonic mean of Recall and Precision. It helps in finding out the precision and robustness of the classification model.

The formula for performance metrics is given in Table 4 [45].

The dataset is split into training and testing sets, which are equal halves, each consisting of 1520 records. The prediction accuracy of MLP, SVM, and Random Forest are 96.38%, 95.06%, 96.90%, respectively, when implemented individually. Figure 4 shows the error rate plot for the random forest. Figure 5 shows the plotted neural network and its connection

TABLE 4. Performance metrics.

Performance Metrics	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Specificity	$\frac{TN}{Negative}$
Sensitivity	$\frac{TP}{TP+FP}$
Precision	$\frac{TP}{TP+FP}$
FMeasure	$\frac{2\times(Precision\times Recall)}{Precision+Recall}$

links with the neurons. The solid black line indicates the Out of the bag error, and the colored lines are error rate for the binary class (0 and 1).

After implementing the stacking generalization model, the algorithm is tested for its efficiency through the confusion matrix, which has various metrics such as Accuracy, Sensitivity, Specificity, Precision, and FMeasure. Figure 6, 7, 8, 9 shows the confusion matrix of MLP, SVM, RF, and stacking ensemble model, respectively.

Also, we have implemented the AUC (Area Under the Receiver Operating Characteristic Curve) curve to evaluate the model further. AUC-ROC is th4e most widely used method for assessing the classification model. In AUC-ROC, AUC represents the measure of separability, and ROC represents the probability curve of the model. When the AUC is high, then it implies that the model has high classification accuracy. The AUC-ROC curve is plotted with False Positive Rate (FPR) on the x-axis and True Positive Rate (TPR) on the y-axis.

The AUC of the Stacking ensemble is found to be 98%. The result is plotted in figure 10. It is found that the prediction accuracy of ensemble stacking generalization is considerably better than the prediction accuracy of individual base learners. We have done 50 iterations to normalize the accuracy. Also, we have implemented the proposed approach with various ratios of training and testing data, such as 90-10, 80-20, 60-40, and 50-50. The maximum accuracy is attained at the 80-20 ratio. We have used a k-fold cross-validation method for validating the performance of the implemented model on unseen data. K-fold cross-validation is the process of constructing and evaluating k number of models to calculate the prediction accuracy and model error. We have

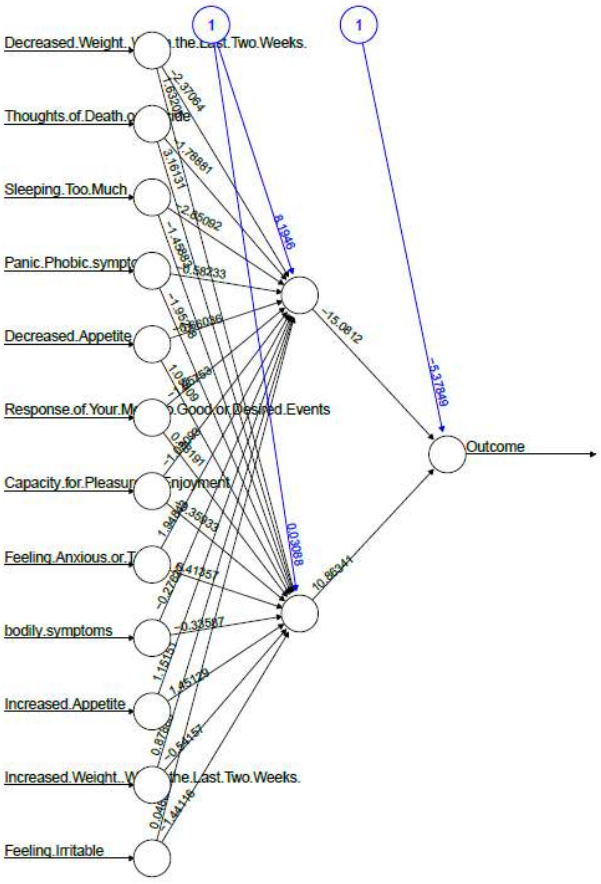

FIGURE 4. MLP Network.

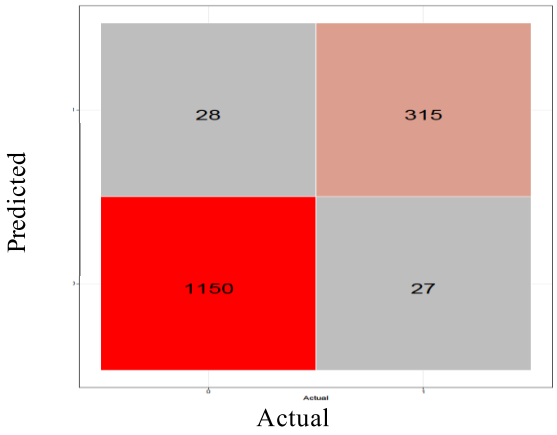

FIGURE 5. MLP Confusion Matrix.

used 25 folds in the k-fold cross-validation to estimate the accuracy. Table 3 shows the performance evaluations of the individual base learners and the ensemble stacking algorithm.

From Table 5, we can infer that the prediction accuracy of the stacking ensemble is considerably higher than the individual base learners. The stacking ensemble compromises the issues with implementing individual learners and improves the prediction accuracy. The accuracy of the

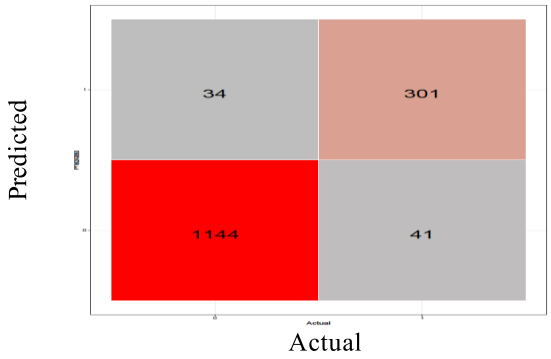

FIGURE 6. SVM Confusion Matrix.

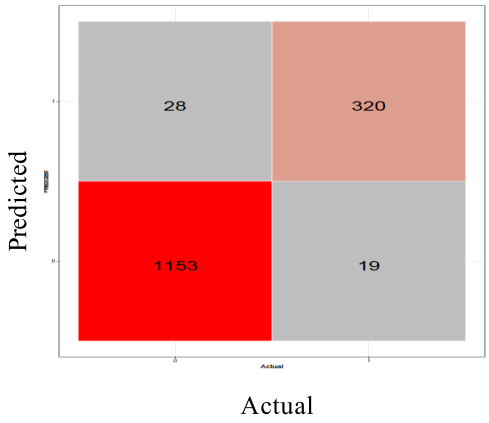

FIGURE 7. RF Confusion Matrix.

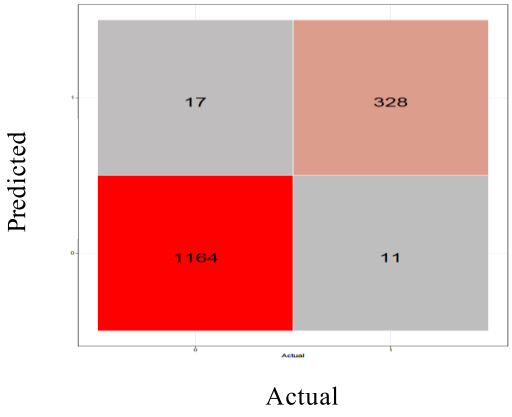

FIGURE 8. Stacking Confusion Matrix.

ensemble stacking model is 98.16%, which shows that by combining the errors of individual base learners using Stacking, the accuracy is also improved. The plot in figure 11 shows the comparisons of performance metrics of individual and stacking models.

In comparison with the existing work on Stacking [46] for predicting depression among the elderly, the proposed stacking algorithm performs better. In the current work, they have implemented Stacking with Decision Tree (DT), Neural Network (NN), Logistic Regression (LR), and Support Vector

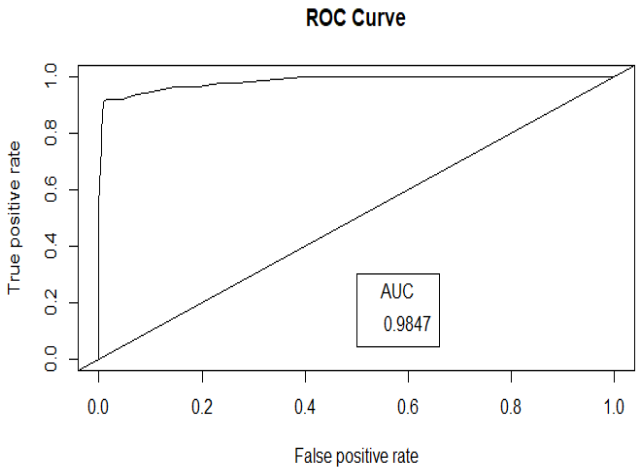

FIGURE 9. AUC-ROC Curve for Stacking Generalization Model.

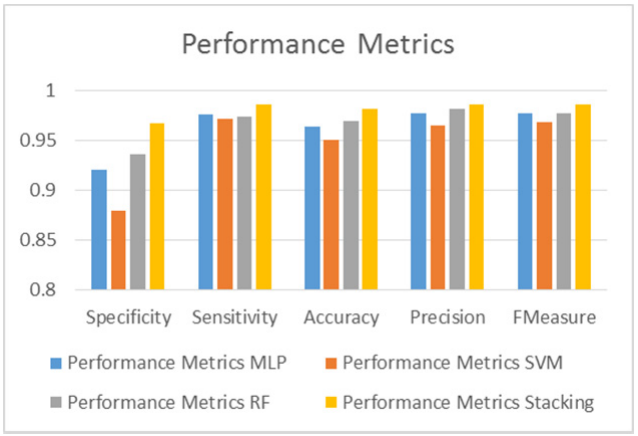

FIGURE 10. Performance Evaluation of MLP, SVM, RF, and Stacking.

Machine as Meta-Level Learners and Logistic Regression as Base-Level Learner. It offers classification with an accuracy of 86%, and the AUC is 81%, whereas the proposed method provides classification with an accuracy of 93%, and AUC is 0.9125

In comparison with existing work on Stacking [47] for automatic prediction of depression in older age, the proposed stacking algorithm shows better results. In the existing work, they have implemented Stacking with Random Forest, which created an ensemble with the decision trees. The classification accuracy of 88% and AUC is 0.894, whereas the proposed method offer classification accuracy of 91% and AUC is 0.938. Table 5 shows the comparative results of the proposed work with the present state-of-the-art works.

VI. DISCUSSION

Many of the existing works related to MDD are mostly based on bagging and boosting. The boosting method decreases the bias, thereby increasing the predictive power, and bagging reduces the variance in the prediction model. Bias is the tendency to learn the wrong relations by not considering all

TABLE 5. Comparison of various state-of-the-art techniques of MDD Prediction models with the proposed model.

MDD Prediction Models	Dataset Source/ Dataset used	Accuracy (%)	AUC
Proposed Model	Major Depressive Disorder (MDD) Among Adults.	98%	0.984
Stacking Ensemble (LR, DT, NN, SVM)	Korea National Health and Nutrition Examination Survey (KNHNE)	86%	0.81
Proposed Model implemented using KNHNE Dataset.	Korea National Health and Nutrition Examination Survey (KNHNE)	93%	0.9125
Stacking Ensemble (RF)	English Longitudinal Study of Ageing (ELSA) – Wave 7	88%	0.894
Proposed Model implemented using ELSA – Wave 7	English Longitudinal Study of Ageing (ELSA) – Wave 7	91%	0.938

TABLE 6. Performance metrics.

Performance Metrics	MLP	SVM	RF	Stacking
Specificity	0.9210	0.8801	0.9356	**0.9675**
Sensitivity	0.9762	0.9711	0.9738	**0.9856**
Accuracy	0.9638	0.9506	0.9690	**0.9816**
Precision	0.9770	0.9654	0.9812	**0.9856**
FMeasure	0.9765	0.9681	0.9774	**0.9856**

the features. The model with high bias will not learn the correlation between the attributes efficiently. Variance is the measure of indicating, how dependent the developed model is on the training data. If the dependency is high, then that would make the model unstable. High Variance and High Bias will significantly affect the accuracy of the model.

As focus on improving the predictive power and also decrease the variance of the model, we have implemented the stacking generalization ensemble model. The stacking ensemble model reduces the variance as well as the bias and increases the predictive power of the model. In this study, we have implemented a Stacking Ensemble model to improve prediction accuracy as well as decrease the variance. We have used MLP, RF, and SVM as base-learners and MLP as the meta-learner.

We have combined these three to compensate each other model's limitations and offer better accuracy than implementing them as individual models. We have tabulated the results in Table 6. We have compared the implemented Stacking model with the existing State-of-the-art techniques and found that our proposed approach performs significantly better. In comparison, we have applied our model on the existing work datasets and proved that the proposed approach achieved enhanced prediction accuracy than the existing works on those datasets. The results of the comparison are shown in Table 5.

VII. CONCLUSION

In this paper, we have developed a stacking generalization model for improving the accuracy in predicting MDD. In the first step, we have implemented a KNN Imputation preprocessing technique for handling the missing values in the data. Then in the next step, we have used Random Forest-Based Backward Elimination, which is a wrapper based feature selection method for reducing the feature dimension, which would reduce the feature interactions and helps in increasing the prediction accuracy. The initial number of features was 22, and then RF-BE has reduced to 12 features with which further process. We have split the data into training and validation data. The training of the stacking generalization is done with the training data, and the prediction is made using the validation data. The stacking generalization is made by combining three low learners MLP, SVM, and RF and then averaging them into a Meta level learner (MLP). The classifiers are also implemented individually to compare the results. The accuracy of individual classifiers MLP, SVM, RF is 96.38%, 95.06%, and 96.90%, respectively. The accuracy of the stacking generalization model is 98.16%. The results show that the stacking generalization model has improved the prediction accuracy when compared with individual base learners.

REFERENCES

[1] E. McElroy,P. Fearon, J. Belsky, P. Fonagy, andP. Patalay, "Networks of depression and anxiety symptoms across development," *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 57, no. 12, pp. 964–973, 2018, doi: 10.1016/j.jaac.2018.05.027.

[2] Mayo Clinic Staff. *Depression*. Accessed: Jul. 22, 2015. [Online]. Available: https://www.org/diseasesconditions/depression/basics/definition/con-20032977

[3] D. Collerton, "Psychotherapy and brain plasticity," *Frontiers Psychol.*, vol. 4, p. 548, 2013. Accessed: Sep. 6, 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3764373/

[4] I. E. Fried and M. R. Nesse, "Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study," *J. Affect. Disorders*, vol. 172, pp. 96–102, Feb. 2015, doi: 10.1016/j.jad.2014.10.010.

[5] WHO. *Health Statistics and Information Systems—Estimates for 2000-2015 [website]*. Geneva, Switzerland, World Health Organization, 2017. Accessed: Jan. 2, 2019. [Online]. Available: http://www.who.int/healthinfo/global_burden_disease/estimates/en/

[6] *Depression and Other Common Mental Disorders: Global Health Estimates*, Geneva, Switzerland, World Health Organization, 2017. Accessed: Jan. 2, 2019. [Online]. Available: http://apps.who.int/iris/handle/10665/254610

[7] E. I. Fried, R. M. Nesse, K. Zivin, C. Guille, andS. Sen, "Depression is more than the sum score of its parts: Individual DSM symptoms have different risk factors," *Psychol. Med.*, vol. 44, no. 10, pp. 2067–2076, Dec. 2013, doi: 10.1017/S0033291713002900.

[8] B. S. McEwen. *Stress, Depression and Brain Structure Dbsalliance*. Accessed: May 27, 2019. [Online]. Available: https://org/site/PageServer?pagename=education_anxiety_stress_brain_structure

[9] H. Klakk, P. L. Kristensen, L. B. Andersen, K. Froberg, N. C. Møller, andA. Grøntved, "Symptoms of depression in young adulthood is associated with unfavorable clinical- and behavioral cardiovascular disease risk factors," *Preventive Med. Rep.*, vol. 11, pp. 209–215, Sep. 2018, doi: 10.1016/j.pmedr.2018.05.017.

[10] A. Jan, H. Meng, Y. F. B. A. Gaus, and F. Zhang, "Artificial intelligent system for automatic depression level analysis through visual and vocal expressions," *IEEE Trans. Cognit. Develop. Syst.*, vol. 10, no. 3, pp. 668–680, Sep. 2018, doi: 10.1109/TCDS.2017.2721552.

[11] C. M. Cummings, N. E. Caporino, andP. C. Kendall, "Comorbidity of anxiety and depression in children and adolescents: 20 years after.," *Psychol. Bull.*, vol. 140, no. 3, pp. 816–845, 2014, doi: 10.1037/a0034733.

[12] A. M. Chekroud, R. J. Zotti, Z. Shehzad, R. Gueorguieva, M. K. Johnson, M. H. Trivedi, T. D. Cannon, J. H. Krystal, andP. R. Corlett, "Cross-trial prediction of treatment outcome in depression: A machine learning approach," *Lancet Psychiatry*, vol. 3, no. 3, pp. 243–250, 2016, doi: 10.1016/S2215-0366(15)00471-X.

[13] C. Tack, "Artificial intelligence and machine learning | applications in musculoskeletal physiotherapy," *Musculoskeletal Sci. Pract.*, vol. 39, pp. 164–169, Feb. 2019, doi: 10.1016/j.msksp.2018.11.012.

[14] A. Stetco, F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, M. Barnes, J. Keane, and G. Nenadic, "Machine learning methods for wind turbine condition monitoring: A review," *Renew. Energy*, vol. 133, pp. 620–635, Apr. 2019, doi: 10.1016/j.renene.2018.10.047.

[15] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review," *Expert Syst. Appl.*, vol. 97, pp. 205–227, May 2018, doi: 10.1016/j.eswa.2017.12.020.

[16] S. B. Kotsiantis, D. Kanellopoulos, andP. E. Pintelas, "Data preprocessing for supervised leaning," *Int. J. Comput. Sci.*, vol. 1, no. 2, pp. 111–117, 2006.

[17] A. Khan, A. Sohail, U. Zahoora, andA. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," 2019, *arXiv:1901.06032*. [Online]. Available: http://arxiv.org/abs/1901.06032

[18] A. S. Qureshi, A. Khan, A. Zameer, andA. Usman, "Wind power prediction using deep neural network based meta regression and transfer learning," *Appl. Soft Comput.*, vol. 58, pp. 742–755, Sep. 2017.

[19] J. Huang, Y.-F. Li, andM. Xie, "An empirical analysis of data preprocessing for machine learning-based software cost estimation," *Inf. Softw. Technol.*, vol. 67, pp. 108–127, Nov. 2015, doi: 10.1016/j.infsof.2015.07.004.

[20] N. Mahendran andD. R. Vincent, "Effective classification of major depressive disorder patients using machine learning techniques," *Recent Patents Comput. Sci.*, vol. 12, no. 1, pp. 41–48, Jan. 2019.

[21] Z. M. Hira andD. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Adv. Bioinf.*, vol. 2015, Jun. 2015, Art. no. 198363, doi: 10.1155/2015/198363.

[22] R. Polikar, *Ensemble Learning*. Boston, MA, USA: Springer, 2012, pp. 1–34.

[23] G. Wang, J. Hao, J. Ma, andH. Jiang, "A comparative assessment of ensemble learning for credit scoring," *Expert Syst. Appl.*, vol. 38, no. 1, pp. 223–230, 2011, doi: 10.1016/j.eswa.2010.06.048.

[24] N.-F. Jie, M.-H. Zhu, X.-Y. Ma, E. A. Osuch, M. Wammes, J. Theberge, H.-D. Li, Y. Zhang, T.-Z. Jiang, J. Sui, and V. D. Calhoun, "Discriminating bipolar disorder from major depression based on SVM-FoBa: Efficient feature selection with multimodal brain imaging data," *IEEE Trans. Auto. Mental Develop.*, vol. 7, no. 4, pp. 320–331, Dec. 2015, doi: 10.1109/TAMD.2015.2440298.

[25] M. P. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemometrics Intell. Lab. Syst.*, vol. 83, no. 2, pp. 83–90, 2006, doi: 10.1016/j.chemolab.2006.01.007.

[26] Y. Xia, C. Liu, B. Da, andF. Xie, "A novel heterogeneous ensemble credit scoring model based on bstacking approach," *Expert Syst. Appl.*, vol. 93, pp. 182–199, Mar. 2018, doi: 10.1016/j.eswa.2017.10.022.

[27] M. H. Abdelaal, N. A. Elmahdy, A. A. Halawa, andA. H. Youness, "Improve the automatic classification accuracy for Arabic tweets using ensemble methods," *J. Elect. Syst. Inf. Technol.*, vol. 5, no. 3, pp. 363–370, 2018, doi: 10.1016/j.jesit.2018.03.001.

[28] A. Kaya, "Cascaded classifiers and stacking methods for classification of pulmonary nodule characteristics," *Comput. Methods Program. Biomed.*, vol. 166, pp. 77–89, Nov. 2018, doi: 10.1016/j.cmpb.2018.10.009.

[29] F. Anifowose, J. Labadin, andA. Abdulraheem, "Improving the prediction of petroleum reservoir characterization with a stacked generalization ensemble model of support vector machines," *Appl. Soft Comput.*, vol. 26, pp. 483–496, Jan. 2015, doi: 10.1016/j.asoc.2014.10.017.

[30] K. D. Kedarisetti, L. Kurgan, and S. Dick, "Classifier ensembles for protein structural class prediction with varying homology," *Biochem. Biophys. Res. Commun.*, vol. 348, no. 3, pp. 981–988, 2006, doi: 10.1016/j.bbrc.2006.07.141.

[31] S. M. Hamze-Ziabari andT. Bakhshpoori, "Improving the prediction of ground motion parameters based on an efficient bagging ensemble model of M5' and CART algorithms," *Appl. Soft Comput.*, vol. 68, pp. 147–161, Jul. 2018, doi: 10.1016/j.asoc.2018.03.052.

[32] J. Rubin, C. Potes, M. Xu-Wilson, J. Dong, A. Rahman, H. Nguyen, and D. Moromisato, "An ensemble boosting model for predicting transfer to the pediatric intensive care unit," *Int. J. Med. Inform.*, vol. 112, pp. 15–20, Apr. 2018, doi: 10.1016/j.ijmedinf.2018.01.001.

[33] Doreswamy, I. Gad, andB. R. Manjunatha, "Performance evaluation of predictive models for missing data imputation in weather data," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2017, pp. 1327–1334, doi: 10.1109/ICACCI.2017.8126025.

[34] J. S. Shah,S. N. Rai, A. P. DeFilippis, B. G. Hill, A. Bhatnagar, and G. N. Brock, "Distribution based nearest neighbor imputation for truncated high dimensional data with applications to pre-clinical and clinical metabolomics studies," *BMC Bioinf.*, vol. 18, no. 1, p. 114, Feb. 2017.

[35] S. Salcedo-Sanz, L. Cornejo-Bueno, L. Prieto, D. Paredes, and R. García-Herrera, "Feature selection in machine learning prediction systems for renewable energy applications," *Renew. Sustain. Energy Rev.*, vol. 90,pp. 728–741, Jul. 2018, doi: 10.1016/j.rser.2018.04.008.

[36] K. Z. Mao, "Orthogonal forward selection and backward elimination algorithms for feature subset selection," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 34, no. 1, pp. 629–634, Feb. 2004, doi: 10.1109/TSMCB.2002.804363.

[37] S. Mabu, M. Obayashi, andT. Kuremoto, "Ensemble learning of rule-based evolutionary algorithm using multi-layer perceptron for supporting decisions in stock trading problems," *Appl. Soft Comput.*, vol. 36, pp. 357–367, Nov. 2015, doi: 10.1016/j.asoc.2015.07.020.

[38] A. Delgoshaei andC. Gomes, "A multi-layer perceptron for scheduling cellular manufacturing systems in the presence of unreliable machines and uncertain cost," *Appl. Soft Comput.*, vol. 49, pp. 27–55, Dec. 2016, doi: 10.1016/j.asoc.2016.06.025.

[39] X. Huang, W. Wu, H. Qiao, andY. Ji, "Brain-inspired motion learning in recurrent neural network with emotion modulation," *IEEE Trans. Cognit. Develop. Syst.*, vol. 10, no. 4, pp. 1153–1164, Dec. 2018, doi: 10.1109/TCDS.2018.2843563.

[40] I. Guyon, J. Weston, andS. Barnhill, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, p. 389, Jan. 2002, doi: 10.1023/A:1012487302797.

[41] Y. Hou, J. Xu, Y. Huang, andX. Ma, "A big data application to predict depression in the university based on the reading habits," in *Proc. 3rd Int. Conf. Syst. Informat. (ICSAI)*, Nov. 2016, pp. 1085–1089, doi: 10.1109/ICSAI.2016.7811112.

[42] Y. Zhou andG. Qiu, "Random forest for label ranking," *Expert Syst. Appl.*, vol. 112, pp. 99–109, Dec. 2018.

[43] B. P. O. Lovatti, M. H. C. Nascimento, Á. C. Neto, E. V. R. Castro, andP. R. Filgueiras, "Use of Random forest in the identification of important variables," *Microchemical J.*, vol. 145, pp. 1129–1134, Mar. 2019, doi: 10.1016/j.microc.2018.12.028.

[44] J. Xiao, "SVM and KNN ensemble learning for traffic incident detection," *Phys. A, Stat. Mech. Appl.*, vol. 517, pp. 29–35, Mar. 2019, doi: 10.1016/j.physa.2018.10.060.

[45] M. Sokolova andG. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.

[46] E. S. Lee, "Exploring the performance of stacking classifier to predict depression among the elderly," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Aug. 2017, pp. 13–20.

[47] H. Yang andP. A. Bath, "Automatic prediction of depression in older age," in *Proc. 3rd Int. Conf. Med. Health Informat. (ICMHI)*, 2019, pp. 36–44.

NIVEDHITHA MAHENDRAN received the M.S. degree in software engineering from the Vellore Institute of Technology (VIT), Vellore, India, in 2016, and the M.Tech. (by research) degree from VIT, in 2019, where she is currently pursuing the Ph.D. degree with the School of Information Technology and Engineering. Her current research interests include machine learning predictive analytics and feature engineering.

P. M. DURAI RAJ VINCENT (Member, IEEE) received the B.E. and M.E. degrees in electronics and computer science from Anna University, Chennai, India, and the Ph.D. degree from the Vellore Institute of Technology (VIT), Vellore, India, in 2015. He is currently working as an Associate Professor with the School of Information Technology and Engineering, VIT. He has more than 12 years of teaching and research experience. He is a Motivated Researcher with more than 45 publications in Scopus Database. He has delivered a keynote in few reputed conferences and also delivered invited talks in leading institutes including one in Anna University, Chennai, India. He has organized a few conferences, including one virtual conference. He has excellent industry connectivity and also handled sessions for Wipro employees. His current research interests include security, machine learning, the Internet of Things, and data analytics. He is a Technical Committee Member of several conferences, including the upcoming Indo-Taiwan International Conference on Computing, Analytics, and Networks. He also acts as a Doctoral Committee Member in a few universities. He is a Lead Guest Editor in one of the special issue in the *Journal of Mobile Multimedia*. He is a Reviewer in a few reputed journals including IEEE Access.

KATHIRAVAN SRINIVASAN (Senior Member, IEEE) received the B.E. degree in electronics and communication engineering, the M.E. degree (Hons.) in communication systems engineering, and the Ph.D. degree in information and communication engineering from Anna University, Chennai, India. He was previously working as a Faculty/Lecturer with the Department of Computer Science and Information Engineering and also as the Deputy Director, Office of International Affairs, with National Ilan University, Taiwan. He is currently an Associate Professor with the School of Information Technology and Engineering, Vellore Institute of Technology (VIT), Vellore, India. He has around 15 years of research experience in the area of machine learning, artificial intelligence, and its applications. His research interests include machine learning, artificial intelligence, deep learning, communication systems and networks, computer vision and multimedia, data analytics, and feature engineering.

Moreover, he received the Best Service Award as the Deputy Director at the Office of International Affairs, National Ilan University, in 2016, the Best Service Award from the Department of Computer Science & Information Engineering, National Ilan University, the Best Paper Award at the IEEE International Conference on Applied System Innovation, Sapporo, Japan, on May 13–17, 2017, the Best Paper Award at International Conference on Communication, Management and Information Technology (ICCMIT 2017), Warsaw, Poland, and the Best Conference Paper Award at the IEEE International Conference on Applied System Innovation, Chiba, Tokyo, April 13–17, 2018. He is presently serving as the Editor for the IEEE Future Directions and *KSII Transactions on Internet and Information Systems* (TIIS), an Associate Editor for IEEE Access, *IET Networks*, the *Journal of Internet Technology*, and Editorial Board Member and a Reviewer for various IEEE Transactions, SCI, SCIE, and Scopus Indexed Journals. He is the Guest Editor *Algorithms* (MDPI), *Future Internet*, the *Journal of Mobile Multimedia*, the *International Journal of Distributed Sensor Networks*, and *Recent Patents on Computer Science*. He has played an active role in organizing several International Conferences, Seminars, and Lectures. He has been a Keynote Speaker in many International Conferences and IEEE events.

VISHAL SHARMA (Member, IEEE) received the B.Tech. degree in computer science and engineering from Punjab Technical University, in 2012, and the Ph.D. degree in computer science and engineering from Thapar University, in 2016. He has been with Thapar University as a Lecturer since 2016. From 2016 to 2017, he was a Joint Postdoctoral Researcher with the MobiSec Laboratory, Department of Information Security Engineering, Soonchunhyang University, and also with Soongsil University, South Korea. He is currently a Research Assistant Professor with the Department of Information Security Engineering, Soonchunhyang University, South Korea. He has authored or coauthored more than 90 journal/conference papers and book chapters and co-edited two books with Springer. His current research interests include 5G networks, UAVs, estimation theory, and artificial intelligence. He has been the TPC member of ITNAC-IEEE TCBD'17, ICCMIT'18, CoCoNet'18, ITNAC-IEEE TCBD'18, ETIC-2019, and WiMO-2019. He serves as the Program Committee Member for the *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* (JoWUA). He is a Professional Member of ACM and a Past Chair for ACM Student Chapter-TIET Patiala. He received the Three Best Paper Awards from the IEEE International Conference on Communication, Management and Information Technology (ICCMIT), Warsaw, Poland, in 2017, from CISC-S'17 South Korea, in 2017, and from IoTaas Taiwan, in 2017. He was the Track Chair of MobiSec'16 and AIMS-FSS'16, and a PC member and a Reviewer of MIST'16 and MIST'17, respectively. He also serves as a Reviewer for various IEEE Transactions and other journals.

DUSHANTHA NALIN K. JAYAKODY (Senior Member, IEEE) received the M.Sc. degree in electronics and communications engineering from the Department of Electrical and Electronics Engineering, Eastern Mediterranean University, Turkey, in 2010 (under the university full graduate scholarship), and the Ph.D. degree in electronics, electrical, and communications engineering from the University College Dublin, Ireland. From 2014 to 2016, he was a Postdoctoral Research Fellow with the Institute of Computer science, University of Tartu, Estonia, and also with the Department of Informatics, University of Bergen, Norway. Since 2016, he has been a Professor with the School of Computer Science & Robotics, National Research Tomsk Polytechnic University (TPU), Russia. He also serves as the Head of the Research and Educational Center on Automation and Information Technologies and also a Founder of Tomsk Infocom Lab, TPU, Russia. Since 2019, he has also been serving as the Head of the School of Postgraduate & Research, Sri Lanka Technological Campus (SLTC), Padukka Sri Lanka, and also the Founding Director of the Centre of Telecommunication Research, SLTC. He is currently supervising eight Ph.D. students and many bachelor's and master's students. He has published more than 120 international peer-reviewed journal and conference papers and books. His research interests include PHY and NET layer perspective of 5G communications technologies such as NOMA for 5G, cooperative wireless communications, the device to device communications, LDPC codes, and unmanned aerial vehicles.

He spent a short research stay at the Centre for Telecommunications Research, University of Sydney, Australia, in 2015, and Texas A&M University, in 2018. He received the Best Paper Award from the IEEE International Conference on Communication, Management and Information Technology (ICCMIT), in 2017, and the International Conference on Emerging Technologies of Information and Communications, Bhutan, March 2019. In July 2019, he received the Education Leadership Award from the World Academic Congress, in 2019. In 2017 and 2018, he received the Outstanding Faculty Award by the National Research Tomsk Polytechnic University, Russia. He also received a Distinguished Researcher in wireless communications from the International Journal for Research Under Literal Access, India, in 2019. He has organized or co-organized more than 20 workshops and special sessions of various IEEE conferences. He also served as a Chair, a Session Chair, or a Technical Program Committee Member for various international conferences, such as the IEEE PIMRC, from 2013 to 2019, the IEEE WCNC, from 2014 to 2018, and the IEEE VTC, from 2015 to 2018. He currently serves as an Area Editor for the *Physical Communications* Journal (ELSEVIER), *Information* Journal (MDPI), and *Internet of Technology Letters* (Wiley). In his career, so far, he has attracted research funding of nearly U.S. $6 million. Also, he serves as a Reviewer for various IEEE Transactions and other journals.