

Article

Realizing an Integrated Multistage Support Vector Machine Model for Augmented Recognition of Unipolar Depression

Kathiravan Srinivasan ¹, Nivedhitha Mahendran ¹, Durai Raj Vincent ¹,
Chuan-Yu Chang ^{2,*} and Shabbir Syed-Abdul ^{3,4,*}

¹ School of Information Technology and Engineering, Vellore Institute of Technology (VIT), Vellore 632 014, India; kathiravan.srinivasan@vit.ac.in (K.S.); nivedhitha.m2019@vitstudent.ac.in (N.M.); pmvincent@vit.ac.in (D.R.V.)

² Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan

³ International Center for Health Information Technology (ICHIT), Taipei Medical University, Taipei 110, Taiwan

⁴ Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei 110, Taiwan

* Correspondence: chuanyu@yuntech.edu.tw (C.-Y.C.); drshabbir@tmu.edu.tw (S.S.-A.)

Received: 11 March 2020; Accepted: 13 April 2020; Published: 15 April 2020



Abstract: Unipolar depression (UD), also referred to as clinical depression, appears to be a widespread mental disorder around the world. Further, this is a vital state related to a person's health that influences his/her daily routine. Besides, this state also influences the person's frame of mind, behavior, and several body functionalities like sleep, appetite, and also it can cause a scenario where a person could harm himself/herself or others. In several cases, it becomes an arduous task to detect UD, since, it is a state of comorbidity. For that reason, this research proposes a more convenient approach for the physicians to detect the state of clinical depression at an initial phase using an integrated multistage support vector machine model. Initially, the dataset is preprocessed using multiple imputation by chained equations (MICE) technique. Then, for selecting the appropriate features, the support vector machine-based recursive feature elimination (SVM RFE) is deployed. Subsequently, the integrated multistage support vector machine classifier is built by employing the bagging random sampling technique. Finally, the experimental outcomes indicate that the proposed integrated multistage support vector machine model surpasses methods such as logistic regression, multilayer perceptron, random forest, and bagging SVM (majority voting), in terms of overall performance.

Keywords: multistage support vector machine model; multiple imputation by chained equations; SVM-based recursive feature elimination; unipolar depression

1. Introduction

In recent years, depression seems to be a very prevalent disorder around the globe, having a presence among approximately 264 million individuals. Psychiatrists usually claim that this disorder is unique, and it is unlike mood swings or ephemeral emotions and their reactions. Usually, when such a depressive condition is prevalent for a long duration among individuals, it might be a somber state of health. Additionally, the causes and effects of such cases are severe, and it critically rescinds the day to day functioning of individuals. In the worst scenario, it might stimulate suicidal tendencies in an individual.

In the millennial (born 1981–1996) era, depression is found to be on the rise; the reason is not apparent. Research shows that depression is greater among the younger Millennials, which results in

many risk factors such as substance abuse and behavioral failures [1]. It is found that the depression symptoms have gone to 15 percent from 9 percent between 2005 and 2015, which is very shocking [1]. The three main parts of the brain that are affected by depression are the hypothalamus, the prefrontal cortex, and the amygdala. Some of the common reasons for being depressed are hormonal imbalance, stress, or genetic [2]. The symptoms of depression involve prolonged feelings of regret, sadness and hopelessness, irregular appetite, weight gain or weight loss, and many others. These days, more than the physical health issues, mental health issues are increasing exponentially [3]. It seems like almost everyone is affected by stress, anxiety, and depression [4].

More than physical health, mental health is essential, as it would directly affect physical health too. It will be easier if we have proper techniques to identify mental health as well [5]. There are few significant issues in diagnosing and treating the individual affected with unipolar depression such as, it is not easy for a depressed individual to seek expert help due to motivation and cost, and in some cases, the individual fails to take the mental health seriously [6].

In order to treat the depressed individuals better, we have proposed a machine learning classification algorithm, integrated multistage support vector machine model. It is an ensemble-based classification algorithm, where the support vector machine (SVM) classifiers are integrated with the help of the SVR-based weighted voting method to produce the outcome. Machine learning techniques are the best in identifying the patterns in the dataset and predict the outcome. We gathered data with the help of a questionnaire and preprocessed it to handle the missing values. The preprocessed data is then processed with a feature selection technique to select the relevant features.

The key contributions of this work include the following:

- The multiple imputation by chained equations (MICE) method is deployed for preprocessing and cleaning the gathered dataset
- The feature selection process is accomplished by employing the support vector machine-based recursive feature elimination (SVM RFE).
- The UD classification is performed using the proposed integrated multistage support vector machine classifier, which is built by employing the bagging random sampling approach.

The significant motivation of this research is to devise a random sampling-based integrated multistage SVM model for classifying the unipolar depression dataset, and we also attempt to enhance the overall performance of the proposed model. The rest of this research work is organized as follows. Section 2 elucidates the methodology formulation process and provides a detailed outlook into the individual modules of the proposed integrated multistage SVM model. Section 3 focuses on the experimental results. Section 4 provides information regarding the conclusion and future work.

2. Materials and Methods

2.1. Utilized Dataset

The dataset we used in this study was collected from various individuals with an average age of 30. We framed a questionnaire based on the “Hamilton Depression Rating Scale” [7] and prepared a self-rating report. The dataset collected had 3040 samples with 22 features, including the target variable. The features are the demographic attributes and symptom scores. For processing, we split the dataset into training and testing sets (75-25 rule). The model was trained with the training set then tested with the test set and evaluated with specific performance metrics. Essential features are portrayed in Table 1.

Table 1. The portrayal of essential features.

Features	Description
Age	Average Age 30.
Gender	Male/Female.
Sleep Quotient	Time taken to fall asleep.
Early Wake-Up	The irregular waking uptime.
Sleeping excessively	Irregular sleep hours.
Gloomy	Prolonged feelings of sadness, sometimes in the day or all the time.
Exasperation	Prolonged feelings of irritation, sometimes in the day or all the time.
Apprehensive or Nervous	Prolonged feelings of anxiousness or tension, sometimes in the day or all the time.
Response of the Individual to Preferred Happenings	Reactions, mood-wise to the events happening in life.
Relation between an Individual's Mood and Time	Moods at different time of the day.
Mood Quality	If the individual is sad, is it because of something happened or sad for no reason.
Reduced Desire for food	Not eating enough food.
Augmented Desire for food	Eating more than enough food.
Weight Reduction	Losing more weight in two weeks without any reason.
Weight Increase	Gaining weight at a specific time.
Ability to make Decisions/Attentiveness	Failure in making decisions and losing focus.
Future Perspective	Positive and Negative thoughts about the future.
Suicidal Contemplations	Attempting to harm oneself.
Happiness Quotient	Feeling good or extremely annoyed with pleasure and enjoyment in life.
Fidgety	Constant pacing and difficulty in concentrating.
Physical Indications	Sweating, increased heartbeat, blurred vision, shivering, chest pain or none at all.
Paranoid Signs	Constant panic attacks or none at all.
Result	Depressed or Not Depressed.

2.2. Data Cleaning and Preprocessing

The data cleaning and preprocessing were performed by utilizing the multivariate imputation by chained equations (MICE) [8]. MICE is a flexible, advanced method in handling the missing values [9]. This technique handles the missing values by imputing multiple values [10]. The primary assumption in MICE is that the imputation variables were from the observed values, not from the unobserved values [11]. The process of chained equations involves various steps as follows,

Step 1: For every missing value in the dataset, the mean imputation technique was performed. These mean imputations were considered as placeholders.

Step 2: The mean imputation placeholder for any one of the variables say “var” was set back to null.

Step 3: In Step 2, the values that were observed for the “var” variable, which was made null, were regressed with other variables present in the imputation model and might or might not have all the variables from the dataset. In simpler terms, in this regression model, “var” was considered to be the dependent variable, and the other variables were considered to be independent.

Step 4: The variable “var”, which was made as null, was now replaced with the actual imputations or predictions from the regression model. In the later stages, when “var” was used as an independent variable for other variables in the regression model, the observed values, as well as the imputed values, were used.

Step 5: For every missing value in the dataset, steps 2–4 were repeated to impute values. This process was continued to one iteration or one cycle. At the end of the first cycle, all the missing values would have been handled and imputed with the predictions from the regression model that can be seen in the observed data.

Step 6: Steps 2–4 were repeated for several cycles; the iterations depend on one's requirement. The imputation values would be updated at the end of each cycle. The final imputation values were retained at the end, which formed an imputed dataset. The most common number of cycles used was ten.

2.3. Selection of Features

In the dataset we collected the Hamilton Depression Rating Scale based self-rating report, there were about 22 features, including the target variable with 3040 samples. In the dataset, we found that there were features that interacted with each other. The features that were dependent on each other directly affected the accuracy of the model. In order to reduce the interaction between the features and remove the irrelevant or redundant variables, we implemented a wrapper based feature selection algorithm, support vector machine-based recursive feature elimination (SVM RFE).

Using this approach, nine features were selected. Alternatively, it has to be witnessed that choosing extra features will not assure higher accuracy levels in classification scenarios. Table 2 demonstrates the selected features and their indices for UD classification.

Table 2. Selected features and their indices for unipolar depression (UD) classification.

Selected Features	Index
Gloomy	1
Exasperation	2
Apprehensive or Nervous	3
Response of the Individual to Preferred Happenings	4
Relation between an Individual's Mood and Time	5
Suicidal Contemplations	6
Happiness Quotient	7
Physical Indications	8
Paranoid Signs	9

2.4. Machine Learning Approaches Considered

2.4.1. Logistic Regression Approach

Logistic regression (LR) is a statistical approach, borrowed by machine learning in predictive analysis. This approach is mainly used when the dependent or the target variable is categorical. In logistic regression, the dependent variable must be dichotomous (i.e., Binary, Yes or No) [12]. The main assumptions made in logistic regression are that there are no outliers in the data, and that there is no multicollinearity between the predictor variables. Logistic regression is an extension of linear regression when the target variable seems to be categorical [13]. In this work, the penalized logistic regression uses a Glmnet in RStudio for predicting the unipolar depression. Table 3 presents the parameter settings for the logistic regression approach. Logistic regression is calculated through the probability of event occurrence with the help of the following the logistic function.

$$\text{logit}(p) = \log\left(\frac{p(z=1)}{1-(p=1)}\right) = \alpha_0 + \alpha_1 y_{j2} + \cdots \alpha_x y_{jn} \quad (1)$$

where, p is the probability of event occurrence, for $j = 1, \dots, n$.

Table 3. Logistic regression approach—parameter settings.

Parameters	Settings
fdev	0.00001
devmax	0.999
eps	0.000001
big	9.9×10^{35}
mnlam	5
pmin	0.00001
exmx	250
prec	0.0000000001
mxit	100
factory	FALSE

2.4.2. Multilayer Perceptron Approach

Usually, when there is an increase in the complexity of the problem, the complexity of the theoretical understanding of the problem also upsurges. In that case, traditional statistical approaches are sought after. Currently, the studies show that neural networks, multilayered perceptron (MLP) in particular, seem to be replacing traditional statistical approaches. Multilayered perceptron does not make any prior assumptions about the data distribution, unlike the statistical models, and it can model even a highly non-linear function to accuracy by training it with new unseen data [14]. Multilayered perceptron is a model with interconnected nodes or neurons, which are connected by connection links with weights and the output signals [14]. We implemented the MLP in RStudio by deploying the RSNNS package. Table 4 shows the parameter settings for the multilayer perceptron approach. The input and the output signals are connected with the help of these neurons and connection links. The net input is calculated by,

$$PA = \sum_{k=1}^n Wt_k I_k + b \quad (2)$$

where,

PA —preactivation function or Net input;

Wt_k —the weight associated with the connection link;

I_k —inputs (I_1, I_2, \dots, I_n);

B —bias.

Based on the error rate at every iteration, the weights of the neurons can be adjusted. The perceptron weight adjustment is calculated by,

$$\Delta Wt = L \times P \times I \quad (3)$$

where

ΔWt —change in weights of the neurons;

L —learning rate;

P —predicted or desired output.

Table 4. Multilayer perceptron approach—parameter settings.

Parameters	Settings
Max. output unit error	0.2
Learning function	Rprop Backprop
Modification	None
Print covariance and error	No
Cache the unit activations	No
Prune new hidden unit	No
Min. covariance change	0.040
Candidate patience	25
Max. no. of covariance updates	200
Activation function	LogSym
Error change	0.010
Output patience	50
Max. no. of epochs	200

2.4.3. Random Forest Approach

In this work, we utilized the tuneRanger package in RStudio for the quick deployment of the random forests. Table 5 presents the hyperparameter settings of the random forest (RF) approach in this work. Random forest is an ensemble approach; it uses a recursive partitioning method to produce

numerous trees, which are then aggregated to get the results [15–17]. Every tree in the random forest was constructed independently with the help of bootstraps of the training data. In random forest, each tree was constructed using two-thirds of the training data and the remaining one-third was used for testing the tree. The error rate of the forest depends on the strength of the individual trees and the correlation between each tree. The main advantage of using random forest is that there is no need to use any cross-validation methods, as the random forest approach itself has a built-in method called the out-of-bag errors to determine the test set errors in an unbiased manner. When compared with decision tree, random forest seems to have better accuracy and was less dependent on the training set and more tolerant to noise.

Table 5. Random forest approach—hyperparameter settings.

Hyperparameter	Settings
mtry	3
sample size	3040
replacement	TRUE
node size	1
number of trees	1000
splitting rule	random

2.4.4. SVM Classifier

Support vector machine (SVM) is a machine learning algorithm that can be modeled for both regression and classification problems but it is majorly used for classification of a binary class problem [18,19]. In this work, we utilized the e1071 package in RStudio for the deployment of the SVM classifier. Table 6 illustrates the hyperparameter settings for the SVM classifier in this work.

Table 6. Support vector machine (SVM) classifier—hyperparameter settings.

Hyperparameter	Settings
Kernel	RBF
Problem type	Classification
$\log_2 C$	−5, 15, 2
$\log_2 \gamma$	3, −15, −2

When a labeled training data is given as an input, the model gives an optimal hyperplane as an output, which categorizes the samples. It is easy to maintain a linear hyperplane between two classes. However, when there is no precise classification between the vector points, manual separation is not possible [20]. For such situations, SVM has a strategy called the kernel. Kernel techniques convert a non-separable space to a separable space, which is called kernels used in non-linear separation models. Some of the commonly used kernel techniques are Gaussian kernel, Polynomial kernel, and many more [21,22].

2.5. Integrated MultiStage Support Vector Machine Classification Model

The proposed integrated multistage support vector machine classification model comprises of two segments: the first one being the design of the SVM classifier, and the second is the UD feature selecting and ranking.

2.5.1. Design of Integrated Multistage Support Vector Machine Classifier

In the proposed model, we were combining the individual SVM classifiers into a stronger and accurate model to improve the robustness and the generality of the SVM classifier. The deployment of this integration model depends on two factors: (i) the efficient way to build the member classifiers, aligning with the integration technique, and (ii) how to make all the member classifiers fuse to end up

with a robust classifier. Therefore, to form a group of member classifiers, a random sampling method based on bagging is applied repeatedly [23]. For every individual SVM classification member classifier, around 75% from the original data sample is selected randomly for the training set, and the rest of the samples are used as a test or validation set to evaluate the performance of the model. A grid search utilizing the factor ranges $C = \{1, 2, 3, 4, 5, \dots, 30\}$ and $\gamma = \{0.1, 0.2, 0.3, 0.4, 0.5, \dots, 5\}$ is accomplished, for determining the optimum values of C and γ . Later, without considering the optimal number of members in an integrated classifier, in this study, we implemented 10 different SVM classifiers with data from 10 random samplings and validated using the 10-fold cross-validation. This technique uses SVM RFE as the base learner. Thus, we constructed an SVM classifier with ten members in this study. In the SVM RFE, the features will be selected by the member classes based on their rankings in the support vector ratio-based ranking criteria. As the member classifiers are built with different random samples, they tend to have behaviors different from each other, and also, they will have different classification outcomes for the same data. As the final step, to integrate all the decisions by the individual classifiers to form ensemble SVM classifiers, the SVR-based weighted voting technique is implemented. The overall design of the proposed method is shown in Figure 1. Once the integrated classifier is built, it can be used for any classification tasks, as shown in Figure 1. In Figure 1, we can see that, once the member classifier was trained, the rest of the samples, which was 25% from the training set, was used as a temporary validation set for evaluating the performance of the model. In order to maintain the diversity of the classifiers and the simplicity of the integrated model, we used $m = 10$ member classifiers in this study.

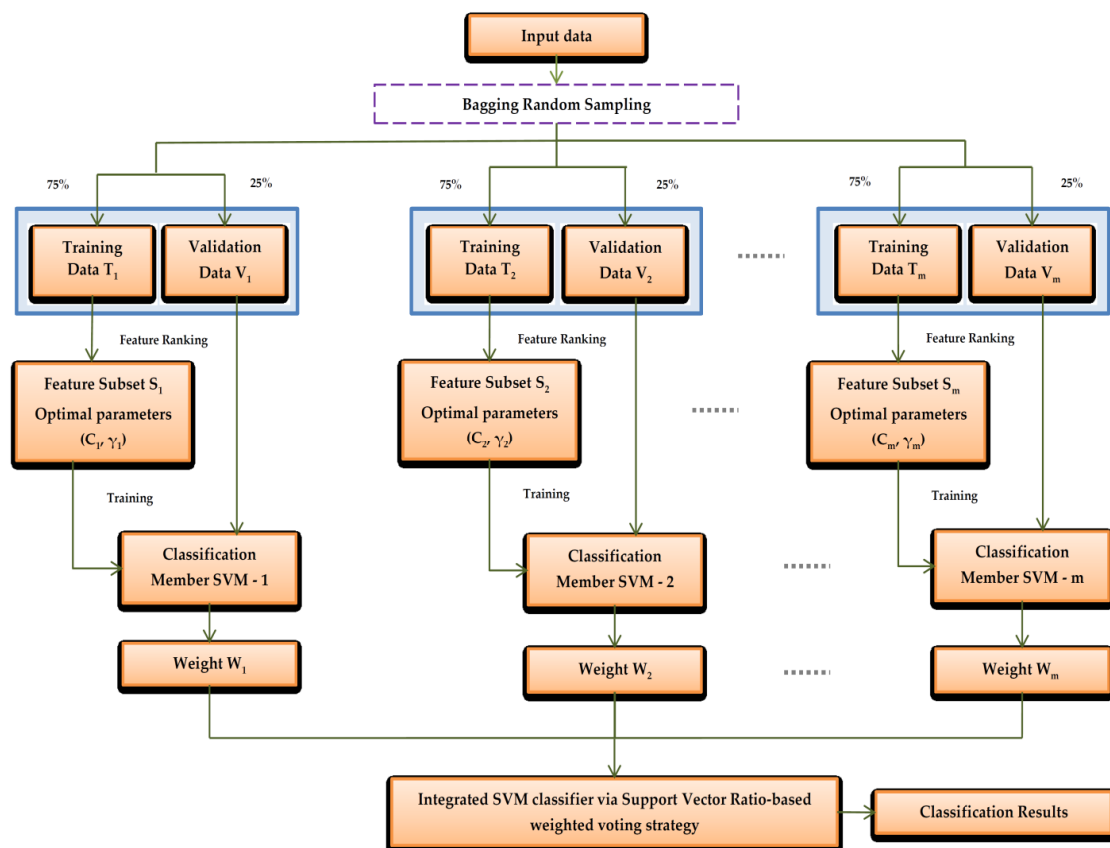


Figure 1. Proposed support vector ratio-based integrated multistage support vector machine classification model—architectural framework.

2.5.2. Ranking and Selection of Features

The essential step in implementing the integrated multistage SVM classification model is selecting the feature subset, which eventually enhances the performance of the member classifiers. Figure 2

represents the flow diagram of the support vector ratio-based support vector machine recursive feature ranking—the irrelevant variables and the variables that interact with each other usually slow down the overall performance of the model concerning computation and storage, during training or prediction. Sometimes, the irrelevant features can make drastic effects on the learning phase of the model. To improve the performance of the SVM classifier, we implemented an effective feature ranking and feature selection method to remove the irrelevant features from the 22 available features in the dataset, which can be seen in Table 1. The commonly used feature selection algorithms come under two categories, the filter methods and wrapper methods [24]. As simple as the filter methods look, they are not considered most of the time because they do not take into account the interaction between the features, which reduces the optimality of the feature subset, though they are computationally effective. On the other hand, wrapper methods evaluate the features jointly and iteratively, which results in effective capturing of interaction between the features [25].

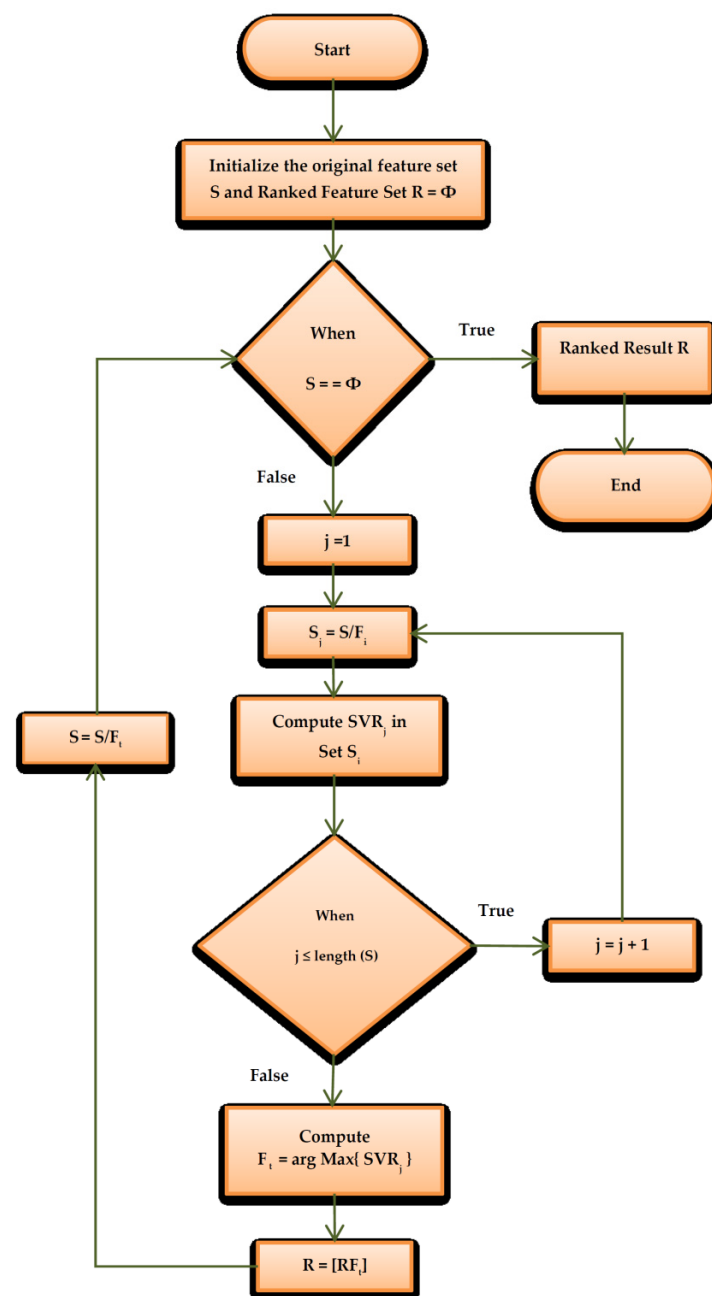


Figure 2. Support vector ratio-based support vector machine recursive feature ranking—flow diagram.

Due to the above-mentioned advantage, we used a wrapper method for feature selection in constructing the ensemble-based SVM classifier. Among all the existing wrapper-based feature selection methods, SVM RFE is considered as the most effective [26]. In this study, we implemented RFE as a part of the RBF SVM classification with the help of the support vector rate (SVR) metric for ranking all the 22 features shown in Table 1. The SVR is given by,

$$\text{Support Vector Ratio} = \frac{\text{The number of support vectors}}{\text{the number of total training samples}} \times 100\% \quad (4)$$

The features are the support vectors in SVM; it is known that some of the support vectors help in minimizing the computational load of SVM and also improve its efficiency during the training. The ranking process is illustrated in Figure 2. The algorithm for the ranking process is as follows,

Step 1: Initialize the feature set, define S with all the 22 features from the dataset.

Step 2: Assume the ranked feature set as R.

Step 3: Eliminate one feature from the set and train the SVM model with 21 features. The classifier is initialized with empirical parameters, in order to calculate the SVR, which allows us to find out the contribution of the removed feature.

Step 4: Repeat step 3 for all the 22 features in the dataset. The feature with higher SVR after removal is placed in the ranked set R. It implies that the feature is not a support vector and is far away from the hyperplane.

3. Results and Discussion

The collected dataset had 3040 samples with 22 features, including the outcome variable. We preprocessed the dataset for removing the missing values using the MICE technique. Once the missing values were handled, we applied a wrapper-based feature selection technique, SVM RFE, to eliminate the less relevant and low performing features from the set. The algorithm removed the features in iteration and ranked them based on the SVR score. From the total 22 features, the algorithm selected nine features as the most important ones. These nine features did not depend on each other and also there was no interaction among them. The dataset was then divided into training and testing sets, where the model was trained with a training set and evaluated with the testing set. The composition was 75-25 for the training and testing dataset, respectively, with 10-fold cross-validation. In the numerical implementation, we implemented the proposed method with 10-member SVM classifiers and then integrated them with the help of the SVR-based weighted voting technique, as explained in the previous section.

To evaluate the proposed model, we have used the confusion matrix [27]. The confusion matrix was used to validate the performance of the model, which was tested with test data and whose true values were known. The technical terms involved in the confusion matrix are the true positive TP (model prediction—positive, actual outcome—positive), true negative TN (model prediction—negative, actual outcome—negative), false positive FP (model prediction—positive, actual outcome—negative), and false negative FN (model predicted—negative, actual outcome—positive). From the confusion matrix, different performance metrics can be calculated, such as accuracy, specificity, precision, sensitivity, and FMeasure [27]. The respective formulas for the metrics can be seen in Table 7. The results are tabulated in Table 8; the proposed model was compared with other methods such as logistic regression (LR), multilayer perceptron (MLP), random forest (RF), and bagging SVM (majority voting). Figure 3 represents the confusion matrix for LR, MLP, RF, and bagging SVM (majority voting), the proposed model, respectively. A comparison of evaluation metrics of the proposed model with other approaches is illustrated in Figure 4. It can be witnessed that the proposed model surpasses all other compared approaches in terms of performance and superior accuracy. The receiver operating characteristic (ROC) curve for the LR, MLP, RF, bagging SVM (majority voting), and the proposed model is depicted in Figure 5, Figure 6, Figure 7, Figure 8, and Figure 9, respectively. Stability comparison between the integrated SVM classifier and the member classifiers is shown in Figure 10.

Actual	No	625	39
	Yes	36	60
		No	Yes
		Predicted	

(a)

Actual	No	636	32
	Yes	29	63
		No	Yes
		Predicted	

(b)

Actual	No	639	25
	Yes	22	74
		No	Yes
		Predicted	

(c)

Actual	No	645	19
	Yes	17	79
		No	Yes
		Predicted	

(d)

Actual	No	655	9
	Yes	6	90
		No	Yes
		Predicted	

(e)

Figure 3. Confusion matrix: (a) Logistic regression (LR), (b) multilayer perceptron (MLP), (c) Random Forest (RF), (d) bagging SVM (majority voting), and the (e) proposed model.

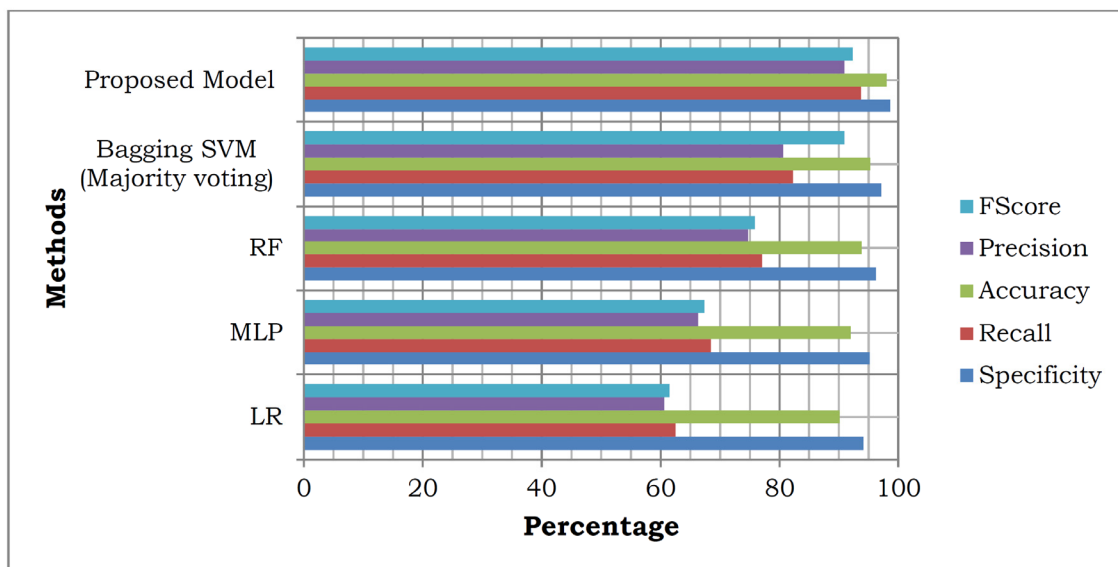


Figure 4. Comparison of evaluation metrics of the proposed model with other approaches.

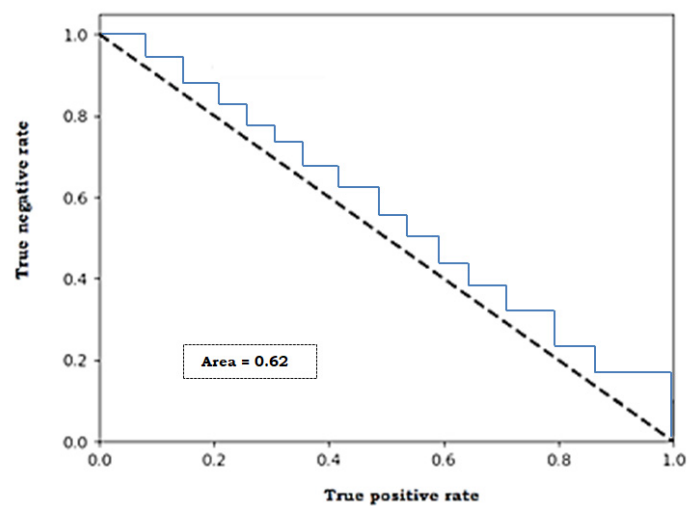


Figure 5. Receiver operating characteristic (ROC) curve for logistic regression.

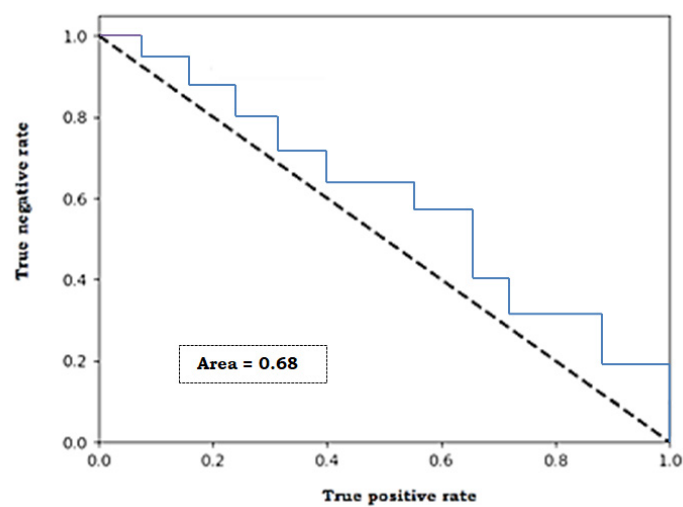


Figure 6. Receiver operating characteristic curve for multilayer perceptron.

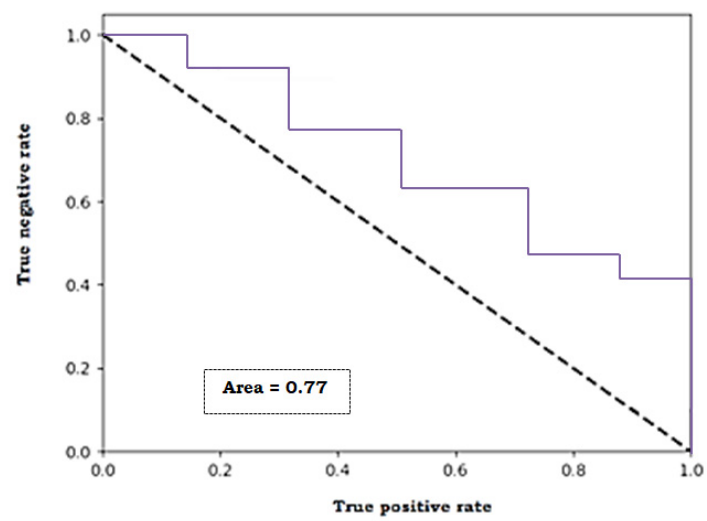


Figure 7. Receiver operating characteristic curve for random forest.

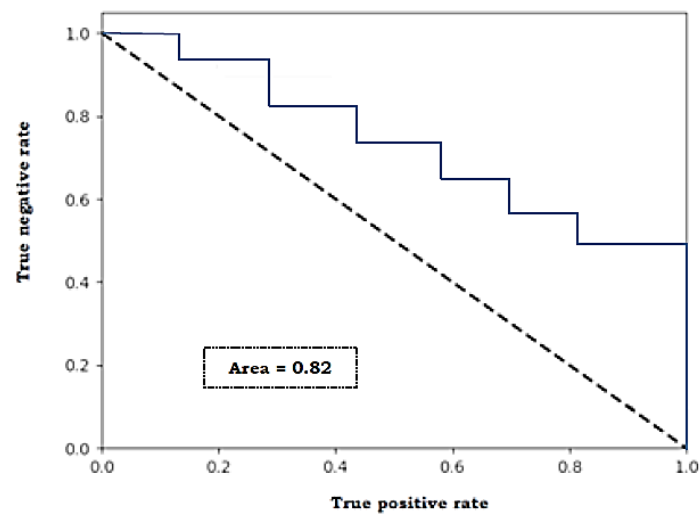


Figure 8. Receiver operating characteristic curve for bagging SVM (majority voting).

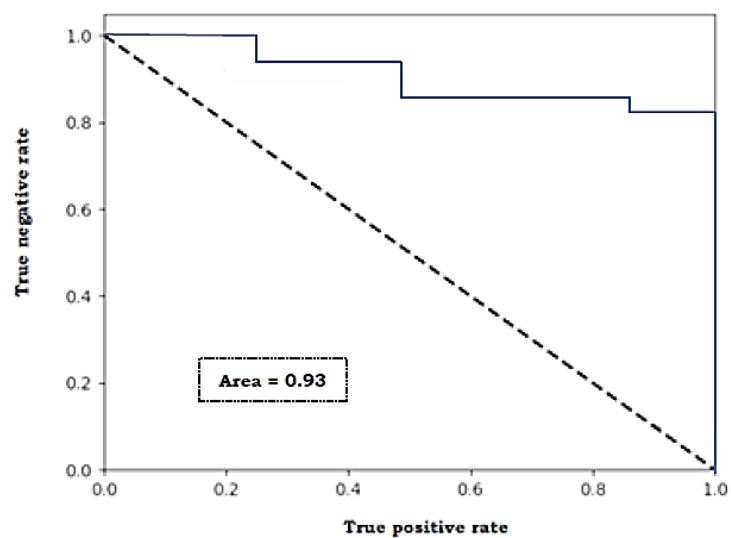


Figure 9. Receiver operating characteristic curve for the proposed model.

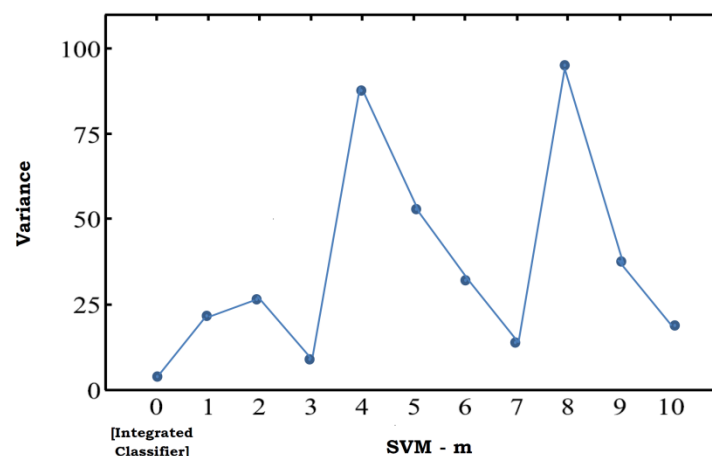


Figure 10. Stability comparison between the integrated SVM classifier and the member classifiers.

Table 7. Confusion matrix related metrics.

Confusion Matrix	Formula
Specificity	$TN / (TN + FP)$
Recall	$TP / (TP + FN)$
Accuracy	$(TN + TP) / (TN + FP + FN + TP)$
Precision	$TP / (TP + FP)$
FScore	$2 \times (Precision \times Recall) / (Recall + Precision)$

Table 8. Comparison of evaluation metrics of the proposed model with other approaches.

Evaluation Metric	LR (%)	MLP (%)	RF (%)	Bagging SVM (Majority Voting)	Proposed Model (%)
Specificity	94.12	95.20	96.23	97.13	98.64
Recall	62.5	68.47	77.08	82.29	93.75
Accuracy	90.13	91.97	93.81	95.26	98.02
Precision	60.6	66.31	74.74	80.61	90.91
FScore	61.53	67.37	75.89	81.44	92.31

4. Conclusions

In this study, we proposed an effective ensemble-based classification model, integrated multistage support vector machine classification model for enhancing the predicting accuracy of UD. As the first step, we cleaned the data with MICE for handling the missing values. Then we implemented SVM RFE, a wrapper-based feature selection technique in order to reduce the feature dimension and select the necessary features, which are not dependent on each other, which eventually improves the accuracy of the model. The initial number of features in the original dataset was 22 on which the feature selection technique was applied. We used a 75-25 composition for training and testing datasets. The results proved that the proposed methodology had improved the prediction accuracy of UD when compared with other classification models. It could be observed that the proposed model was better than all other compared approaches in terms of performance and also offered greater accuracy.

Author Contributions: Conceptualization, K.S., N.M. and D.R.V.; Methodology, K.S. and N.M.; Software, K.S.; Validation, C.-Y.C. and S.S.-A.; Formal Analysis, K.S.; Investigation, K.S.; Resources, C.-Y.C. and S.S.-A.; Data Curation, N.M. and D.R.V.; Writing—Original Draft Preparation, K.S.; Writing—Review & Editing, K.S., N.M., C.-Y.C., and S.S.-A.; Visualization, K.S.; Supervision, C.-Y.C.; Project Administration, C.-Y.C. and S.S.-A.; Funding Acquisition, C.-Y.C. and S.S.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the “Intelligent Recognition Industry Service Center” from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan and The APC was funded by Ministry of Education (MOE) in Taiwan.

This work was partly supported by Ministry of Science and Technology (MOST), Taiwan (106-2923-E-038-001-MY2, 107-2923-E-038-001-MY2, 106-2221-E-038-005, 108-2221-E-038-013); Taipei Medical University (106-3805-004-111, 106-3805-018-110, 108-3805-009-110); Wanfang hospital (106TMU-WFH-01-4).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Patalay, P.; Gage, S.H. Changes in millennial adolescent mental health and health-related behaviours over 10 years: A population cohort comparison study. *Int. J. Epidemiol.* **2019**, *48*, 1650–1664. [\[CrossRef\]](#) [\[PubMed\]](#)
2. McElroy, E.; Fearon, P.; Belsky, J.; Fonagy, P.; Patalay, P. Networks of Depression and Anxiety Symptoms Across Development. *J. Am. Acad. Child Adolesc. Psychiatry* **2018**, *57*, 964–973. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Fried, E.I.; Nesse, R.M.; Zivin, K.; Guille, C.; Sen, S. Depression is more than the sum score of its parts: Individual DSM symptoms have different risk factors. *Psychol. Med.* **2013**, *44*, 2067–2076. [\[CrossRef\]](#)
4. Fried, E.I.; Nesse, R.M. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *J. Affect. Disord.* **2014**, *172*, 96–102. [\[CrossRef\]](#)
5. Klakk, H.; Kristensen, P.L.; Andersen, L.B.; Froberg, K.; Møller, N.C.; Grøntved, A. Symptoms of depression in young adulthood is associated with unfavorable clinical- and behavioral cardiovascular disease risk factors. *Prev. Med. Rep.* **2018**, *11*, 209–215. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Papakostas, G.I.; Petersen, T.; Mahal, Y.; Mischoulon, D.; Nierenberg, A.A.; Fava, M. Quality of life assessments in major depressive disorder: A review of the literature. *Gen. Hosp. Psychiatry* **2004**, *26*, 13–17. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Hamilton, M. A Rating Scale For Depression. *J. Neurol. Neurosurg. Psychiatry* **1960**, *23*, 56–62. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Azur, M.; Stuart, E.; Frangakis, C.; Leaf, P.J. Multiple imputation by chained equations: What is it and how does it work? *Int. J. Methods Psychiatr. Res.* **2011**, *20*, 40–49. [\[CrossRef\]](#)
9. Raghunathan, T.W.; Lepkowski, J.M.; Van Hoewyk, J.; Solenbeger, P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv. Methodol.* **2001**, *27*, 85–95.
10. Van Buuren, S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat. Methods Med. Res.* **2007**, *16*, 219–242. [\[CrossRef\]](#)
11. Schafer, J.L.; Graham, J.W. Missing data: Our view of the state of the art. *Psychol. Methods* **2002**, *7*, 147–177. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Peduzzi, P.; Concato, J.; Kemper, E.; Holford, T.R.; Feinstein, A.R. A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* **1996**, *49*, 1373–1379. [\[CrossRef\]](#)
13. Press, S.J.; Wilson, S. Choosing between logistic regression and discriminant analysis. *J. Am. Stat. Assoc.* **1978**, *73*, 699–705. [\[CrossRef\]](#)
14. Gardner, M.; Dorling, S. Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [\[CrossRef\]](#)
15. Breiman, L. Random Forests. *Mach. Learn.* **2011**, *45*, 5–32. [\[CrossRef\]](#)
16. Kandaswamy, K.K.; Chou, K.-C.; Martinetz, T.; Möller, S.; Suganthan, P.N.; Sridharan, S.; Pugalenth, G. AFP-Pred: A random forest approach for predicting antifreeze proteins from sequence-derived properties. *J. Theor. Biol.* **2011**, *270*, 56–62. [\[CrossRef\]](#)
17. Zhou, Q.; Zhou, H.; Zhou, Q.; Yang, F.; Luo, L. Structure damage detection based on random forest recursive feature elimination. *Mech. Syst. Signal Process.* **2014**, *46*, 82–90. [\[CrossRef\]](#)
18. Hamed, T.; Dara, R.; Kremer, S.C. An Accurate, Fast Embedded Feature Selection for SVMs. In Proceedings of the 2014 13th International Conference on Machine Learning and Applications, Detroit, MI, USA, 3–5 December 2014; pp. 135–140.
19. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [\[CrossRef\]](#)
20. Jarrett, K.; Kavukcuoglu, K.; Ranzato, M.A.; LeCun, Y. What is the best multi-stage architecture for object recognition? In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, September 27–October 4 2009; pp. 2146–2153.
21. Chang, C.-Y.; Srinivasan, K.; Chen, M.-C.; Chen, S.-J. SVM-Enabled Intelligent Genetic Algorithmic Model for Realizing Efficient Universal Feature Selection in Breast Cyst Image Acquired via Ultrasound Sensing Systems. *Sensors* **2020**, *20*, 432. [\[CrossRef\]](#)

22. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **2002**, *46*, 389–422. [[CrossRef](#)]
23. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
24. Tuia, D.; Pacifici, F.; Kanevski, M.; Emery, W.J. Classification of Very High Spatial Resolution Imagery Using Mathematical Morphology and Support Vector Machines. *IEEE Trans. Geosci. Remote. Sens.* **2009**, *47*, 3866–3879. [[CrossRef](#)]
25. Atas, M.; Yardimci, Y.; Temizel, A. A new approach to aflatoxin detection in chili pepper by machine vision. *Comput. Electron. Agric.* **2012**, *87*, 129–141. [[CrossRef](#)]
26. Wang, R.; Li, R.; Lei, Y.; Zhu, Q. Tuning to optimize SVM approach for assisting ovarian cancer diagnosis with photoacoustic imaging. *Bio-Med. Mater. Eng.* **2015**, *26*, S975–S981. [[CrossRef](#)] [[PubMed](#)]
27. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).