

Kernelized Normal Discriminant Feature Selection and Borda Count Bootstrap Aggregating Classification for Risk Factor Identification and Disease Diagnosis



P. S. Renjeni, B. Mukunthan, G. Rakesh

Abstract: Automatic detection of disease is crucial in health care management to evaluate large patient data. The early diagnosis and treatment of disease are an important task to prevent the patient from death. The various researchers have contributed to the development of disease diagnosis. But still, it causes the more risk for identifying the patient health conditions. In order to improve the disease diagnosing accuracy, A Kernelized Normal discriminant Feature Selection based Borda count bootstrap aggregating Classification (KNDFS-BCBAC) technique is introduced for identifying the patient health condition and critical factor analysis with higher accuracy and lesser time. At first, radial basis kernelized normal discriminant analysis is used to identify the relevant feature for minimizing the complexity of disease diagnosis. After selecting the relevant features, Borda count bootstrap aggregating Classifier is applied to classify the patient data as abnormal or normal by constructing the weak learner as bivariate correlated regression tree. Then, the diseased data is considered as a training sample for analyzing the critical factor and classifies the patient data level as the initial stage, critical stage based on the threshold range of features value. By applying the Borda count voting scheme, the weak learner results are combined into strong. In this way, disease diagnosis and critical factor analysis of patient data are performed with greater accuracy and minimal time complexity (TC). Experimental is performed with tumor dataset on metrics namely disease diagnosis accuracy (DDA), false alarm rate (FAR), and TC. The observed results evident that KNDFS-BCBAC technique achieves higher DDA with lesser complexity and FAR than the conventional methods.

Keywords: Disease diagnosis, feature selection, radial basis kernelized normal discriminant analysis, bootstrap aggregating Classification, bivariate correlated regression tree, Borda count voting scheme.

I. INTRODUCTION

Healthcare industry comprises the big and complex patient data that used to determine the pattern of diseases and also makes efficient decisions in the diagnostic process.

Revised Manuscript Received on April 27, 2020.

* Correspondence Author

P. S. Renjeni*, Research Scholar, Jairams Arts & Science College, (Affiliated to Bharathidhasan University) Karur – 639003, Tamilnadu.
<https://orcid.org/0000-0002-5494-0650> Email: s.renjeni@yahoo.com

B. Mukunthan, Research Supervisor, Jairams Arts & Science College, (Affiliated to Bharathidhasan University) Karur - 639003. Tamilnadu.
<https://cid.org/0000-0001-8452-3164> Emorail: dr.mukunthan.bmk@gmail.com

G. Rakesh, Dean of Science, Department of Computer Science, Jairams Arts and Science College, (Affiliated to Bharathidhasan University) Karur - 639003, Tamilnadu, India. <https://orcid.org/0000-0001-6978-418X> Email: rakesh.gnana@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Early detection of diseases and continuous care of clinicians reduce the mortality rate. However, accurate detection of diseases in all cases and discussion of a patient disease is difficult since it requires more time and expertise. There are several mining algorithms employed for heart disease prediction in the past. In the existing model, researchers used dataset as input which may or may not be an appropriate format. The various data mining approaches are utilized to achieve higher efficiency of the disease diagnosis using the data mining techniques like classification, clustering, and association rule mining. While performing the disease diagnosing, the feature selection is carried out for minimizing the time complexity. In [1], long-term recurrent convolutional network (LRCN) was introduced to identify epileptic seizures from EEG signals. However, the TC was higher. To attain accurate and robust detection with higher classification accuracy, an Epileptic Seizure Detection using Long-Short-Term Memory (ESD-LSTM) was designed in [2]. The designed method failed to provide improved performance with the more patient's data.

To categorize the different disease, Hybrid Cascade Forward Neural Network with Elman Neural Network (HECFNN) was designed in [3]. The designed method failed to perform the feature selection for improving performance accuracy. An ensemble empirical mode decomposition (EEMD) model was developed in [4] to identify the influences of factors on epileptic seizures. However, it does not accurately detect disease with minimum complexity. In [5], a multiscale radial basis functions (MRBF) and modified particle swarm optimization (MPSO) and Support Vector Machine (SVM) was designed to categorize epileptic seizures. But, it does not offer more accurate and reliable classification model with lesser complexity.

An Eigenspace Time-Frequency Based Feature selection was designed in [6] to identify Seizure with EEG Data. But, the early detection of onsets of seizures was not carried out. For disease pattern classification, a multi-features and multilayer perceptron neural network (MLPNN) classifier was designed in [7]. However, the accuracy was not improved with more patients' data.

To discover the brain tumor, an Artificial Neural Network Input Gain Measurement Approximation based hybrid feature selection and classification were designed in [8]. The error rate of disease classification was more.

In [9], K-Nearest Neighbor Classifier was designed to predict and diagnose the epilepsy level at diverse ages of patients. However the classification accuracy was not enhanced.

For identifying epileptic seizures, a novel matrix determinant feature selection was introduced in [10] with EEG signals. TC performance of epileptic seizures detection remained unsolved.

1.1 Paper contribution

The existing techniques have a few limitations in the disease diagnosis such as lesser accuracy, more complexity, higher error rate and so on. To overcome such kind of issues, the KNDFS-BCBAC technique is introduced. The major contribution of KNDFS-BCBAC technique is described as follows,

- ◆ KNDFS-BCBAC technique is introduced to enhance the DDA by applying discriminant analysis based feature selection and bootstrap aggregating classifier.
- ◆ RBKNDAs is performed to identify the more relevant features based on the similarity measure. The relevant features subset is chosen for disease diagnosis hence it minimizes the TC.
- ◆ The bootstrap aggregating classifier is used for classifying the patient data by constructing the bivariate regression tree. Based on predefined threshold value, the regression tree analyzes patient data and classifying the disease level. Ensemble classifier combines regression tree results and calculates generalization error. After that, weak learner’s results are ranked and identified majority vote of results using Borda count approach. This assists to enhance DDA and lessen the FAR.

1.2 Outline of the paper

The rest of article is ordered as follows: A brief review is presented in Section 2. Section 3 briefly describes the KNDFS-BCBAC technique employed in our research. Section 4 details the experiments undertaken in the framework. The evaluation procedures and results are provided in section 5. The conclusion is summarized in Section 6.

II. LITERATURE REVIEWS

A hybrid intelligent system was developed in [11] using Fuzzy Min-Max (FMM) neural network for classifying the medical data. The robustness of the system was not improved. An automatic generation of medical report scheme was introduced in [12] for identifying epilepsy. But, the feature selection is not considered to lessen the complexity.

For brain tumor disease prediction, The Fuzzy entropy (FuzzyEn) and distribution entropy (DistEn) was designed in [13]. However, the disease prediction accuracy was not improved. In [14], a machine learning system was designed to identify the whole-brain seizure. The designed system does not lessen the complexity of seizure detection since it failed to perform feature selection.

In [15], a new feature selection technique based on support vector machine (SVM) was designed to enhance medical classification performance. For identifying the epilepsy seizure detection, the nonlinear sparse extreme learning machine (SELM) was designed in [16]. The computational complexity was higher.

In [17], an adaptive multi-parent crossover Genetic Algorithm (GA) was introduced to identify the epileptic seizures with the selected features. In [18], the Random Forest classifier was designed for seizure diagnosing with higher accuracy. But the risk factor analysis was not

performed. An automated seizure detection systems were presented in [19] for identifying the type of seizure. The complexity of seizure detection was not minimized. A logistic model tree (LMT) was developed in [20]. But, the error rate was not reduced for identifying epileptic seizure.

The major issues of conventional methods are overcome by introducing a HDFS-BCBS technique which is presented in the next section.

III. PROPOSED METHODOLOGY

A KNDFS-BCBAC technique is introduced for disease diagnosis with higher accuracy and lesser time. KNDFS-BCBAC technique comprises two major processes namely feature selection and classification. With the assist of RBKNDAs, the feature selection is presented to discover a linear combination of features. The resultant selection of the features is more commonly used for dimensionality reduction before the classification. After the feature selection, the classification is done using Borda count bootstrap aggregating classifier for diagnosis the disease and identifies the risk factors by classifying the normal patient or abnormal patient. These two major processes of the proposed KNDFS-BCBAC technique are described in the following subsections.

3.1 Radial basis kernelized Normal discriminant analysis

The feature selection assists to select less number of features from the number of features since the processing of more features consumes more computation time, large space, and so on. Therefore, the KNDFS-BCBAC technique performs the feature selection using RBKNDAs, because it provides more accurate classification results. Consider the dataset D_p with number of attributes i.e. features $a_1, a_2, a_3, \dots, a_n$. Among them, the feature which is more relevant to the disease is identified for minimizing the complexity. The RBKNDAs initially defines the separation function to split the features into two subsets namely relevant feature subset (f_{s_r}) and irrelevant feature subset ($f_{s_{ir}}$). The separation function defines the ratio of the which is mathematically expressed as follows,

$$s_p = \frac{D_b}{D_w} = \frac{v m_w d_t}{v m_b d_t} \quad (1)$$

Where, s_p represents the separation function, D_b denotes the variance between the subsets, D_w denotes a variance within the subsets, v is the linear discriminant vector used for projecting the relevant and irrelevant features into the subsets, m_w, m_b denotes a scatter matrix within and between the

subsets. The proposed technique uses the radial basis kernel function to measure the similarity between two samples (i.e. features, objective function). Here the objective function is the tumor disease symptoms. The radial basis kernel function is given below,

$$K_{RBF} = \exp(-\rho \|a_i - o_b\|^2) \quad (2)$$

$$\rho = -\frac{1}{2 D^2} \quad (3)$$

In (2), (3) K_{RBF} is the radial basis kernel function, $\|a_i - o_b\|^2$ is the squared Euclidean distance between the two samples, D denotes a deviation parameter $D > 0$, ρ is the parameter,



a_i denotes a attributes, o_b is the disease symptoms. The radial basis kernel function provides the similarity value between 0 and 1. Based on the similarity, the discriminant vector projects the features into the subsets.

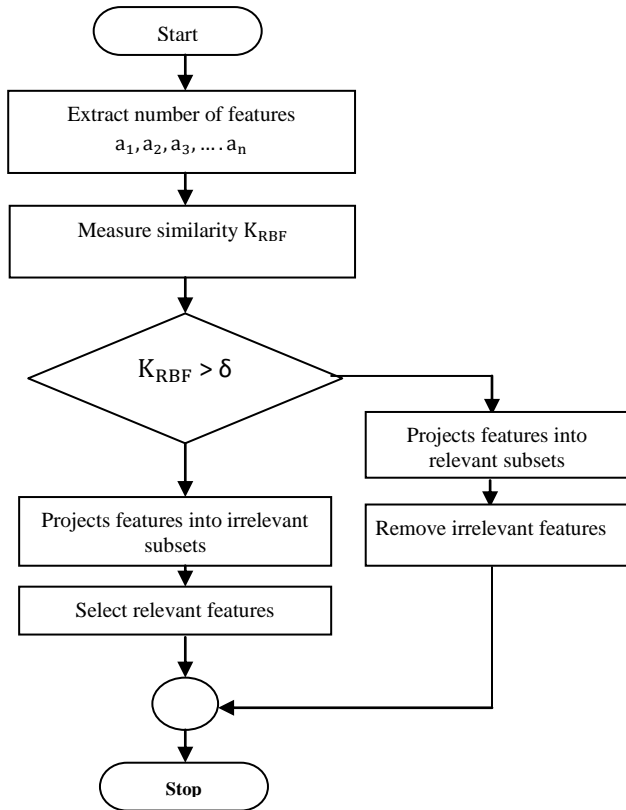


Figure 1 Flow chart of Radial basis kernelized Normal discriminant analysis

The relevant feature selection process is portrayed in figure 1. With the relevant feature selection, KNDFS-BCBAC technique takes minimal time for disease classification as compared to conventional works.

Input: Dataset D_p , number of attributes i.e. features $a_1, a_2, a_3, \dots, a_n$.

Output: Select relevant feature subsets

Begin

1. **Extract** the number of features $a_1, a_2, a_3, \dots, a_n$ from D_p
 2. **Define** the separation function s_p
 3. **for** each feature a_i and objective function o_b
 4. Measure the radial basis kernel function K_{RBF}
 5. **if** ($K_{RBF} > \delta$) then
 6. Discriminant vector projects the features into relevant subsets ' f_{s_r} '
 7. **else**
 8. Discriminant vector projects the features into irrelevant subsets ' $f_{s_{ir}}$ '
 9. **end if**
 10. Select relevant features subset ' f_{s_r} '
 11. Remove the irrelevant features subset ' $f_{s_{ir}}$ '
 12. **end for**
- end**

Algorithm1 Radial basis kernelized Normal discriminant analysis

As given in the above algorithm, RBKNDA based feature selection process is performed for disease diagnosis. The

linear discriminant vector in the given dimensional is used for projecting the features as relevant and irrelevant subsets based on the radial basis kernel function. The RBKNDA selects the relevant feature subsets and removes the irrelevant feature subsets. The relevant feature subsets are used for minimizing the complexity involved in disease diagnosing and hence the accuracy is improved.

3.2 Borda count bootstrap aggregating classifier

After selecting the relevant feature subset, the patient data is classified using Borda count bootstrap aggregating classifier. The bootstrapped aggregating classifier is called as bagging technique is a machine learning ensemble classifier to improve the accuracy by constructing the number of weak learners. Weak learner is a base classifier which does not offer accurate classification results. In contrast, bootstrapped aggregating is a strong ensemble classifier provides the accurate results through the voting scheme. The proposed KNDFS-BCBAC technique uses the Borda count voting scheme to provide the accurate classification results through the ranking procedure. The BCBAC structure is given below,

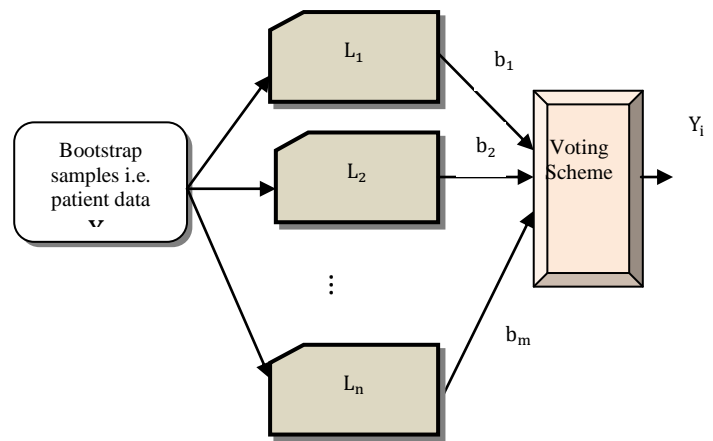


Figure 2 Borda count bootstrap aggregating classifier

Figure 2 shows the BCBAC. The classifier considers the training sets $\{X_i, Y_i\}$ where X_i denotes a bootstrap samples (i.e. input patient data) and Y_i represents an ensemble classification results. The BCBAC builds empty set of 'm' weak learners $\{L_1, L_2, L_3, \dots, L_n\}$. The bivariate correlated regression tree is used as a weak learner to classify the patient data. The regression tree is used to analyze the patient data with the disease feature value for identifying the patient as normal or abnormal. The regression tree comprises root node, branch node, and leaf node. The root node in the tree measures correlation among training feature value (i.e. patient data) and testing feature value (i.e. disease feature value) to perform the classification.

$$\beta = \frac{(\sum d_i a_t) - (\sum d_i)(\sum a_t)}{\sqrt{[\sum d_i^2 - (\sum d_i)^2][\sum a_t^2 - (\sum a_t)^2]}} \quad (4)$$

From (4), β indicates a correlation coefficient, d_i represents patient data, a_t denotes a disease feature value, $\sum d_i a_t$ denotes a sum of product of paired score, $\sum p_i$ is sum of p_i score, $\sum a_t$ is sum of a_t score, $\sum d_i^2$ is the sum of squared score of $\sum d_i$ and $\sum a_t^2$ is the sum of squared score of $\sum a_t$. Bivariate correlation coefficient offers value among +1 and -1 where '+1' indicates positive correlation and '-1'



indicates negative correlation. The positive correlation indicates the patient data is classified as abnormal (i.e. disease) and the negative correlation is the normal. After the classification, the critical factor analysis of patient data is carried out in the branch node by classifying the different stages based on the threshold range of feature value. Finally, classification results are obtained in the leaf nodes. In order to analyze the critical risk factor of the patient data, the threshold is set to the abnormal disease feature value (d_p). If the disease feature value is greater than the threshold, then the data is classified as a critical stage. Otherwise, the disease patient data is classified as an initial stage. The classification of the bivariate correlated regression tree is shown in figure 3.

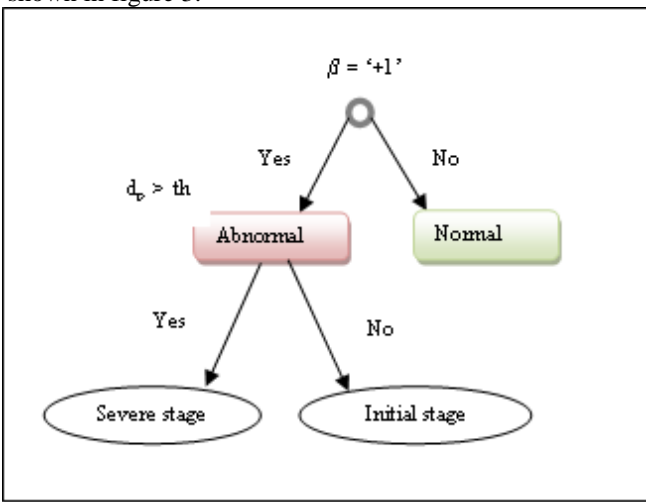


Figure 3 bivariate correlated regression tree

Based on the correlation coefficient values, the root node classifies the patient data as normal or abnormal. Then the branch node set the predefined threshold value for the disease feature value in order to identify the risk factor of the patient. In this way, the patient data are classified and identified the disease as well as risk. The weak learner includes training error in classification results. To attain strong classification results, the classification results of each weak learners are combined. The strong classification results are obtained as follows,

$$Y_i = \sum_{i=1}^m b_i \quad (5)$$

From (5), Y_i is the strong classifier output and b_i indicates the weak learners output. The generalization error is measured for each weak learner to predict accurate classification results. The generalization error is calculated as the differentiation between the expected and predicted error.

$$e_g = \{e_x - e_p\} \quad (6)$$

In (6), e_g denotes a generalization error, e_x is the expected error, e_p represents the predicted error. By applying the borda count voting scheme, the weak classifiers are arranged and ranked based on the generalization error.

$$e_g(b_1) \leq e_g(b_2) \leq e_g(b_3) \leq \dots \leq e_g(b_m) \quad (7)$$

Compared to the conventional bootstrap aggregating classifier, the proposed bootstrap aggregating classifier gives the first rank to their most preferred weak classifier results i.e. lesser generalization error and second rank to

their second most preferred, and so on. The higher-ranked classifiers are taken to the final classification results. From the classification results, all votes of the samples in the higher-ranked classifiers are counted and identify the majority to be elected.

$$Y = \arg \max_m v(d_i) \quad (8)$$

In (8), Y symbolizes strong classification results, $\arg \max$ indicates argument of maximum function to discover majority vote (v) of samples (i.e. patient data 'd') whose decision known to m^{th} classifier. At last, bootstrap aggregating classifier provides majority of samples as strong classification results therefore it enhances the DDA and lessen the false positive rate.

Algorithm:

Input: Patient data $d_1, d_2, d_3, \dots, d_n$

Output: Improve disease diagnosing accuracy

Begin

1. Construct 'm' regression trees with training Patient data $d_1, d_2, d_3, \dots, d_n$
2. for each data d_i
3. Measure the correlation
4. **if** ($\beta = +1$) **then**
5. Positive correlation
6. data classified as abnormal
7. **else**
8. Negative correlation
9. data classified as normal
10. **end if**
11. **if** ($d_p > th$) **then**
12. Patient data classified as critical stage
13. **else**
14. Patient data classified as initial stage
15. **end if**
16. Combine a set of weak learners $Y_i = \sum_{i=1}^m b_i$
17. **For** each b_i
18. Calculate the generalization error ' e_g '
19. **end for**
20. Assign the rank based on error rate
21. Select high ranked classifier
22. Identify the majority votes of the samples $Y = \arg \max_m v(p_i)$
23. Obtain strong classification results

end

Algorithm 2 Borda count bootstrap aggregating classifier

The process of BCBAC is described in algorithm 2 for diagnosing the disease. The bootstrap aggregating classifier builds empty set of regression tree with patient data. After that, the root node calculates the correlation among training feature value (i.e. patient data) and disease feature value. Based on correlation results, the normal and abnormal patient condition is identified. The risk factor analyzes is done by setting the threshold of the abnormal disease feature value. The various stages of the patient health condition are identified. The weak classifications results are combined to creating strong by minimizing the generalization error. The ensemble classifiers ranked the classifier which having lesser generalization error and obtains the strong results by a majority vote of all the samples. As a result, bootstraps aggregating classifiers improves DDA and lessen the FAR.



$$Disease\ diagnosing\ accuracy = \left(\frac{no.of\ data\ correctly\ classified}{n} \right) * 100 \tag{9}$$

IV. EXPERIMENTAL SETUP AND PARAMETER SETTINGS

The Experimental evaluation is carried out using Java language with the help of the OASIS dataset taken from the <https://www.oasis-brains.org/>. The three methods are KNDFS-BCBAC and existing [1] and [2]. The OASIS dataset includes 13 features and 416 instances. The OASIS dataset contains the Longitudinal MRI Data of the patient for diagnosing the tumor disease with different stages such as normal, abnormal, mild, and very mild. The thirteen features description is shown in table 1.

Table 1 feature description

Serial No.	Features	Description
1.	MR sessions	MRI imaging sessions
2.	Subject	Patients
3.	M/F	Gender Male/ female
4.	Hand	Right-handed (R)
5.	Age	Patient age
6.	Education	years of education
7.	SES	Socioeconomic status since evaluated by the Hollingshead Index of Social Position and categorized into categories from 1 (highest status) to 5 (lowest status)
8.	CDR	Clinical Dementia Rating 0 = Normal 0.5 = very mild 1 = mild
9.	MMSE	Mini-Mental State Examination score (range is from 0 [worst] to 30 [best])
10.	eTIV	estimated Total Intracranial Volume (cm ³)
11.	nWBV	normalized Whole Brain Volume
12.	ASF	Atlas Scaling Factor (unitless)
13.	Scans	MRI scans were obtained

The performance of KNDFS-BCBAC technique and existing methods are evaluated with different metrics are listed below,

- Disease diagnosing accuracy
- False alarm rate
- Time complexity

V. PERFORMANCE ANALYSIS AND RESULTS DISCUSSION

The analysis of KNDFS-BCBAC, existing [1] and [2] are discussed in this section with different metrics namely DDA, FAR and TC. The comparative analyses are performed with the graphical illustration.

5.1 Disease diagnosing accuracy

DDA is measured by patient data classification. Thus, the accuracy is the ratio of number of (i.e. no. of) patient data are correctly classified to number of patient data. DDA is calculated as below,

From (9), *n* indicates a total number of data. DDA is calculated in percentage (%).

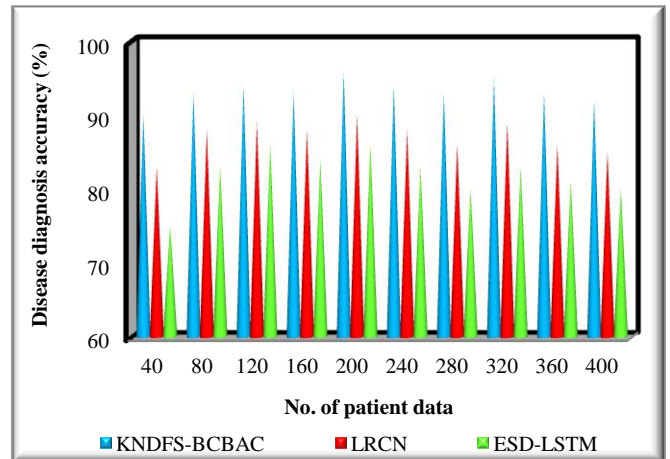


Figure 4 Disease diagnosing accuracy versus no. of patient data

DDA is illustrated in figure 4 using three techniques with number of patient data. The graphical results prove that the KNDFS-BCBAC improves the brain tumor diagnosing accuracy. By applying bootstrap aggregating classifier, the accuracy is improved. Ensemble classifier builds regression tree with patient data and correlated with disease features value. From the regression analyses, the normal patient and tumor affected patient are identified. Followed by, the different stages of the brain tumor such as mild and very mild are also identified. Through lessen the generalization error, the weak classifiers are combined to offer strong classification results. The ensemble classifier improves the disease diagnosing accuracy. This is evident from the sample mathematical calculation. With ‘40’ patient data, ‘36’ patient data are correctly diagnosed by KNDFS-BCBAC technique. In a similar manner, with ‘40’ patient data, ‘33’ patient data correctly diagnosed using LRCN [1] and ‘30’ patient data correctly diagnosed using ESD-LSTM [2], therefore the overall disease diagnosing accuracy was found to be ‘83%’ and ‘75%’ respectively. The comparison results show that the brain tumor disease diagnosing accuracy of KNDFS-BCBAC is improved by 7% compared to LRCN [1] and 14% compared to ESD-LSTM [2].

5.2 False alarm rate

FAR is defined as ratio of number of (i.e. no. of) patient data are incorrectly classified to the number of patient data. FAR is calculated as below,

$$FAR = \left(\frac{no.of\ data\ incorrectly\ classified}{n} \right) * 100 \tag{10}$$

From (10), *FAR* indicates a False alarm rate, *n* indicates a total number of data. *FAR* is calculated in percentage (%).

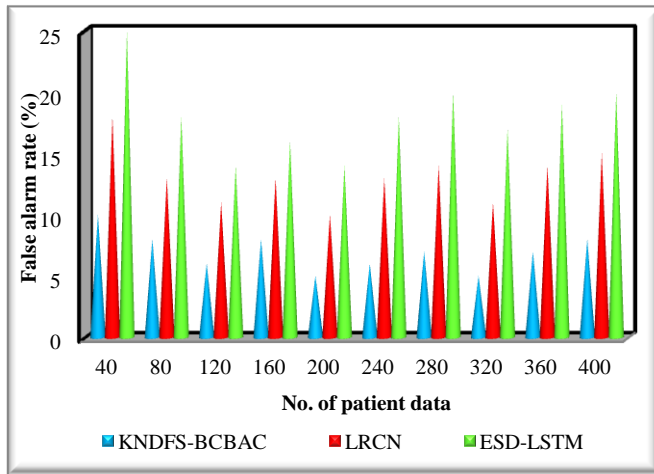


Figure 5 False alarm rate versus no. of patient data

The result of FAR is depicted in figure 5 with 400 patient data. From figure 5, the FAR of KNDFS-BCBAC technique is minimal. This improvement of KNDFS-BCBAC technique is achieved by applying the Borda count bootstrap aggregating classifier. The bootstrap aggregating classifier uses the Borda count voting scheme to identify the weak classifier with minimum generalization error based on the ranking procedure. As a result, the high ranked classifiers are selected to obtain strong classification results. Majority votes of results are taken from higher-ranked classifiers. This process of KNDFS-BCBAC technique minimizes the incorrect identification of the diseased patient. FAR of KNDFS-BCBAC technique is minimized by 47% and 61% as compared to existing [1] and [2].

5.3 Time complexity

TC is defined as the amount of time consumed to diagnosis the tumor with number of patient data. TC is formulated as below,

$$TC = n * T(\text{diagnosis single patient data}) \quad (11)$$

In (11), T indicates a time for diagnosing one patient data. TC is calculated in milliseconds (ms).

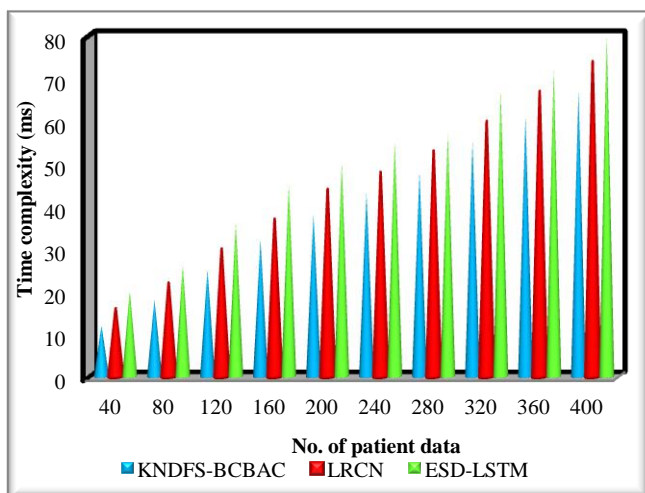


Figure 6 Time complexity versus no. of patient data

The result of TC is shown in figure 6 with number of patient data. From figure 6, the TC of KNDFS-BCBAC technique is found to be minimized. This is because of the application of RBKND A based feature selection. The discriminant vector employed to project more similar features for brain tumor diagnosis in feature subset. The classification of patient as normal, mild and very mild is performed with the

Clinical Dementia Rating (CDR) value and selected features. This, lessens the TC in tumor disease diagnosis. This is evident from the sample calculation. With ‘40’ number of patient data, the time involved in diagnosing is ‘12ms’ using KNDFS-BCBAC technique, the TC using LRCN [1] and ESD-LSTM [2] are ‘16ms’ and ‘20ms’ respectively. The various runs are carried out with the patient data. TC of KNDFS-BCBAC technique is reduced by 19% and 24% as compared to existing [1] and [2].

VI. CONCLUSION

In this paper, a KNDFS-BCBAC technique is developed to enhance DDA with lesser time. KNDFS-BCBAC technique applied in this work uses two different processes. The relevant features are selected by the radial basis kernelized discriminant analysis, thus removing irrelevant features subset and obtaining a more effective relevant subset of features. This is helpful for reducing the TC and therefore improving the disease diagnosis accuracy. Furthermore, bootstrap aggregating ensemble technique is applied to change the weak classification results into strong by minimizing the generalization error. As a result, the accurate classification is performed resulting in improves the diagnosing results. From the results, it is evident that the KNDFS-BCBAC technique using OASIS dataset provides better performance in terms of DDA, FAR and TC.

REFERENCES

- Xiaoyan Wei, LinZhouBZhenZhangCZiyiChendYiZhoua, “Early prediction of epileptic seizures using a long-term recurrent convolutional network”, Journal of Neuroscience Methods, Elsevier, Volume 327, 2019, Pages 1-39
- Ramy Hussein , Hamid Palangi , Rabab K. Ward, Z. Jane Wang, “Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals”, Clinical Neurophysiology, Volume 130, Issue 1, 2019, Pages 25-37
- Mutasem Sh. Alkhasawneh, “Hybrid Cascade Forward Neural Network with Elman Neural Network for Disease Prediction”, Arabian Journal for Science and Engineering, Springer, 2019, Pages 1-12
- Jhe-Nan LinaFu-YiYangbWei-ZenSunCdMeng-HanYang “Using methods of time series data mining to recognize the influences of environmental factors on epileptic seizures”, Computers & Electrical Engineering, Elsevier, Volume 68, May 2018, Pages 437-446
- Yang Li, Xu-Dong Wang, Mei-Lin Luo, Ke Li, Xiao-Feng Yang, and Qi Guo, “Epileptic Seizure Classification of EEGs using Time-Frequency Analysis based Multiscale Radial Basis Functions”, IEEE Journal of Biomedical and Health Informatics, Volume 22, Issue 2, March 2018, Pages 386 – 397
- M.DericheS.ArafatS.Al-InsafM.Siddiqui , “Eigenspace Time Frequency Based Features for Accurate Seizure Detection from EEG Data”, IRBM , Elsevier, Volume 40, Issue 2, 2019, Pages 122-132
- N. Sriraam, S. RaghuKadeeja Tamanna, Leena Narayan, Mehraj Khanum, A. S. Hegde, Anjani Bhushan Kumar, “Automated epileptic seizures detection using multi-features and multilayer perceptron neural network”, Brain Informatics, Springer, Volume 5, Issue 8, 2018, Pages 1-10
- Shamsul Huda; John Yearwood; Herbert F. Jelinek; Mohammad Mehedi Hassan; Giancarlo Fortino; Michael Buckland, “A Hybrid Feature Selection With Ensemble Classification for Imbalanced Healthcare Data: A Case Study for Brain Tumor Diagnosis”, IEEE Access, Volume 4 , 2016, Pages: 9145 – 9154
- Md. Kamrul Hasan, Md. Asif Ahamed, Mohiuddin Ahmad, and M. A. Rashid, “Prediction of Epileptic Seizure by Analysing Time Series EEG Signal Using k-NN Classifier”, Applied Bionics and Biomechanics, Hindawi, Volume 2017, August 2017, Pages 1-12.

10. S Raghu, Natarajan Sriraam, Alangar Sathyaranjan Hegde, Pieter L Kubben, "A novel approach for classification of epileptic seizures using matrix determinant", Expert Systems with Applications, Elsevier, Volume 127, 2019, Pages 323-341
11. Manjeevan Seera, Chee Peng Limb, "A hybrid intelligent system for medical data classification", Expert Systems with Applications, Elsevier, Volume 41, Issue 5, 2014, Pages 2239-2249
12. Jefferson Tales Oliva and João Luís Garcia Rosa, "Classification for EEG report generation and epilepsy detection", Neurocomputing, Elsevier, Volume 335, 2019, Pages 81-95
13. Peng Li, Chandan Karmakar, John Yearwood, Svetha Venkatesh, Marimuthu Palaniswami, Changchun Liu, "Detection of an epileptic seizure based on entropy analysis of short-term EEG", PLoS ONE, Volume 13, Issue 3, 2018, Pages 1-17
14. P.Fergus, A.Hussain, David Hignett, D.Al-Jumeily, Khaled Abdel-Aziz, Hani Hamdan, "A machine learning system for automated whole-brain seizure detection", Applied Computing and Informatics, Elsevier, Volume 12, Issue 1, 2016, Pages 70-89
15. Qiong Liu, Qiong Gu, Zhao Wu, "Feature selection method based on support vector machine and shape analysis for high-throughput medical data", Computers in Biology and Medicine, Elsevier, Volume 91, 2017, Pages 103–111
16. Yuanfa Wang, Zunchao Li, Lichen Feng, Chuang Zheng, and Wenhao Zhang, "Automatic Detection of Epilepsy and Seizure Using Multiclass Sparse Extreme Learning Machine Classification", Computational and Mathematical Methods in Medicine, Hindawi, Volume 2017, June 2017, Pages 1-10
17. Salah Al-Sharhan, Andrew Bimba, "Adaptive multi-parent crossover GA for feature optimization in epileptic seizure identification", Applied Soft Computing, Elsevier, Volume 75, February 2019, Pages 575-587
18. Nhan Duy Truong, Levin Kuhlmann, Mohammad Reza Bonyadi, Jiawei Yang, Andrew Faulks, Omid Kavehei, "Supervised learning in automatic channel selection for epileptic seizure detection", Expert Systems With Applications, Elsevier, Volume 86, 2017, Pages 199–207
19. A. Ulate-Campos, F. Coughlin, M. Gai'nza-Lein, I. Sa'ñchez Ferna'ndez, P.L. Pearl b, T. Loddenkemper, "Automated seizure detection systems and their effectiveness for each type of seizure", Seizure, Elsevier, Volume 40, 2016, Pages 88–101
20. Enamul Kabir, Siuly, Yanchun Zhang, "Epileptic seizure detection from EEG signals using logistic model trees", Brain Informatics, Springer, Volume 3, Issue 2, Pages 93–100, June 2016

College, Karur. He is a member of IEEE & IEEE computer society since 2009, a life member of the MISTE since 2010. He has published more than 10 research papers in reputed international journals including Thomson Reuters (SCI & Web of Science). His main research work focuses on Algorithms, Bioinformatics, Big Data Analytics, Data Mining, IOT and Neural Networks. He has 12 years of teaching experience and 10 years of Research Experience.

AUTHOR PROFILE



P.S. Renjeni received B.Sc Chemistry from Sree Devi Kumari College for women, Manonmaniam Sundaranar University, India and obtained MCA from Noorul Islam College of Engineering, Manonmaniam Sundaranar University, India. Presently, she is doing Ph.D in Bharathidasan University, Trichy, India. She has 15 years of teaching experience and working as Assistant

Professor at the Department of Computer Science in V.T.M. College of Arts and Science, Arumanai, Tamil Nadu, India. Her research interest is in Data Mining.



B. Mukunthan pursued Bachelor of Science in Computer Science from Bharathiar University, India in 2004 and Master of Computer Applications from Bharathiar University in year 2007 and Ph.D from Anna University - Chennai in 2013. He is currently working as Research Advisor in Department of Computer Science, Jairams Arts & Science College, Affiliated to Bharathidasan

University, Tiruchirapalli since 2016. He is a member of IEEE & IEEE computer society since 2009, a life member of the MISTE since 2010. He has published more than 10 research papers in reputed international journals. He is also Microsoft Certified Solution Developer. His main research work focuses on Algorithms, Bioinformatics, Big Data Analytics, Data Mining, IOT and Neural Networks. He also invented a Novel and Efficient online Bioinformatics Tool and filed for patent. He has 12 years of teaching experience and 10 years of Research Experience.



G. Rakesh pursued Bachelor of Computer Application from Bharathidhasan University, India in 2004 and Master of Computer Applications from Bharathiar University in year 2007 and Ph.D from Bharathiar University - Coimbatore on 2019. He is currently working as Dean of science in Jairams Arts & Science