Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Classification For Risk Factor Identification And Disease Diagnosis Using Kernelized Normal Discriminant Feature Selection And Borda Count Bootstrap Aggregating Classification

P.S RENJENI[1] , B. MUKUNTHAN[2] , B. SENTHILKUMARAN[3] , L. JAYA SINGH DHAS[4]

[1]Department of Computer Science, Jairams Arts and Science College (Affiliated to Bharathidasan University, Tiruchirappalli) Karur – 639003, Tamil Nadu, India.

[2]Assistant Professor, Department of Computer Science, Jairams Arts and Science College (Affiliated to Bharathidasan University, Tiruchirappalli) Karur - 639003, Tamil Nadu, India.

[3]Assistant Professor, Department of Computer Science, Jairams Arts and Science College (Affiliated to Bharathidasan University, Tiruchirappalli) Karur - 639003, Tamil Nadu, India.

[4]Assistant Professor, Department of Computer Science, Scott Christian College (Affiliated to Manonmaniam Sundaranar University, Tirunelveli) Nagercoil - 629003, Tamil Nadu, India.

Abstract

In order to examine vast amounts of patient data, automatic illness detection is critical in health care administration. The goal of early disease identification and treatment is critical in preventing the patient's death. The development of disease diagnosis has been aided by a number of researchers. However, it increases the danger of misdiagnosing a patient's health condition. A Kernelized Normal Discriminant Feature Selection based Borda count bootstrap aggregating Classification (KNDFS-BCBAC) technique is introduced to increase illness diagnosing accuracy by detecting the patient's health state and crucial factor analysis with higher accuracy and less time. To reduce the complexity of disease diagnosis, radial basis kernelized normal discriminant analysis is initially utilised to locate the relevant feature. By creating the weak learner as a bivariate correlated regression tree, Borda count bootstrap aggregating Classifier is used to categorise the patient data as abnormal or normal after picking the appropriate features. The diseased data is then used as a training sample for analysing the crucial factor, and the patient data level is classified as either initial or critical depending on the feature value threshold range. The Borda count voting technique is used to aggregate the weak learner results into strong results. Disease diagnosis and crucial factor analysis of patient data are performed with greater accuracy and less time complexity in this manner. With respect to a variety of patient data, an experimental evaluation is conducted with a tumour dataset on parameters such as illness diagnosis accuracy, false alarm rate, and time complexity. The findings reveal that the KNDFS-BCBAC strategy achieves higher illness diagnostic accuracy with less complexity and a lower false-alarm rate than current methods.

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

1. Introduction

The healthcare sector includes large and complex patient data that is used to determine disease patterns and make effective diagnostic judgments. The mortality rate is reduced by early diagnosis of disorders and clinicians' ongoing care. However, proper disease identification in all circumstances and discussion of a patient's ailment is challenging due to the time and skill required. In the past, numerous mining algorithms were used to forecast heart disease. Researchers used dataset as input in the existing model, which may or may not be an appropriate format. Using data mining techniques such as classification, clustering, and association rule mining, various data mining approaches are used to improve the efficiency of disease diagnosis. The feature selection is carried out while doing the disease diagnosis in order to reduce the time complexity.

1.1 Paper Contribution:

Existing disease diagnostic techniques have a number of drawbacks, including lower accuracy, increased complexity, and a larger error rate, among others. A unique technique known as the KNDFS-BCBAC technique is introduced to tackle such difficulties. The following is a list of the KNDFS-BCBAC technique's key contributions:

- The KNDFS-BCBAC technique uses discriminant analysis-based feature selection and a bootstrap aggregating classifier to increase disease diagnosis accuracy.
- Based on the similarity measure, radial basis kernelized normal discriminant analysis is used to select the more significant characteristics. For disease diagnosis, the appropriate characteristics subset is picked, which reduces the time complexity.
- By generating a bivariate regression tree, the bootstrap aggregating classifier is utilised to classify the patient data. The regression tree examines patient data and categorises disease levels depending on a predetermined threshold value. The ensemble classifier calculates the generalisation error by combining the regression tree outputs.
- The outcomes of the weak learner are then ranked, and the majority vote of the results is determined using the Borda count method. This helps to increase disease diagnosis accuracy while lowering the rate of false alarms.

1.2 Paper Organization:

The paper is organised as follows: A brief literature overview of some illness diagnosis methodologies is presented in Section 2. The proposed KNDFS-BCBAC approach used in our study is briefly described in Section 3. Section 4 describes the experiments that were carried out as part of the framework. Section 5 details the evaluation processes followed and the outcomes obtained. The conclusion is summarised in Section 6.

2. Literature survey

In the recent research, the performance evaluation of brain tumor diagnosis is done by [1] for the patients with symptoms. Automatic Seizure Detection in Children with Epilepsy was introduced in [2] for determining the epilepsy of the children. An automated seizure detection systems were presented in [3] for identifying the type of seizure. The complexity of seizure detection was not minimized. The survey of Brain Tumor Detection is done in [4],[8] and the classification and segmentation studies are carried out in [5],[6] and also developed an algorithm. A logistic model tree (LMT) was developed in [7] for identifying the epileptic seizure from EEG signals. However, the error rate was not minimized for identifying the epileptic seizure.

An automatic generation of medical report scheme was introduced in [9] for identifying epilepsy with EEG signals. The designed scheme failed to perform the feature selection for minimizing complexity. An ensemble empirical mode decomposition (EEMD) model was developed in [10] to identify the influences of factors on epileptic seizures. But the model failed to accurately detect the disease with minimum complexity. Classification and Segmentation of Brain Tumor MRI images using Deep Learning is carried out in [11] and the subtype classification is supported in [12]. A hybrid intelligent system was developed in [13] using Fuzzy Min-Max (FMM) neural network for classifying the medical data. The robustness of the system was not improved. K-Nearest Neighbor Classifier was developed in [14] for predicting and diagnosing the epilepsy level at various ages of the patients. But the classification accuracy was not improved. An Eigenspace Time-Frequency Based Feature selection was developed in [15] for identifying the Seizure with the EEG Data. However, the early detection of onsets of seizures was not performed.

Brain Tumor Analysis and future challenges was developed in [16]. The Hybrid Cascade Forward Neural Network with Elman Neural Network (HECFNN) was introduced in [17] to categorize the different disease. The designed methods and analysis failed to perform the feature selection for improving performance accuracy. The Random Forest classifier was developed in [18] for diagnosing the seizure with higher accuracy. But the risk factor analysis was not performed. A multi-features and multilayer perceptron neural network (MLPNN) classifier was developed in [19] for disease pattern classification. Though the designed classifier minimizes the false detection rate, the accuracy was not improved with more patients' data. The Fuzzy entropy (FuzzyEn) and distribution entropy (DistEn) was developed in [20] for brain tumor disease prediction. However, the disease prediction accuracy was not improved. A machine learning system was introduced in [21] for automatically identifying the whole-brain seizure. The designed system failed to minimize the complexity of seizure detection since it failed to perform the feature selection.

A new feature selection technique based on support vector machine (SVM) was developed in [22] for improving the medical classification performance. The designed method failed to solve the multi-class problem with the various criterions. An Epileptic Seizure Detection using Long-Short-Term Memory (ESD-LSTM) was introduced in [23] for accurate and robust detection with higher classification accuracy. The designed method failed to provide

improved performance with the more patient's data. An adaptive multi-parent crossover Genetic Algorithm (GA) was designed in [24] for selecting the features to identify the epileptic seizures. The designed algorithm failed to accurately identify epileptic seizures. An Artificial Neural Network Input Gain Measurement Approximation based hybrid feature selection and ensemble classification were introduced in [25] for identifying the brain tumor. The approach failed to accurately minimize the error rate in the disease classification.

A novel matrix determinant feature selection was introduced in [26] for identifying the epileptic seizures using EEG signals. The performance of time complexity of epileptic seizures detection remained unsolved. Novel Approach for Brain Tumor Classification Using Convolutional Neural Network was performed in [27] and this approach was failed in classification. The long-term recurrent convolutional network (LRCN) was developed in [28] for identifying the epileptic seizures from EEG signals. But the time complexity of the disease prediction remained unsolved.

A multiscale radial basis functions (MRBF) and a modified particle swarm optimization (MPSO) and Support Vector Machine (SVM) was developed in [29] to classify the epileptic seizures. The designed method failed to obtain more accurate and reliable classification model with minimum computational complexity. The nonlinear sparse extreme learning machine (SELM) was introduced in [30] for identifying the epilepsy seizure detection. The computational complexity was not minimized.

The major problems are recognized from the literature review are overcome by introducing a innovative technique called KNDFS-BCBAC technique. The process of KNDFS-BCBAC technique is presented in the following section.

3. Methodology

A KNDFS-BCBAC approach has been developed to diagnose the disease more accurately and quickly at an earlier stage. The two key steps in the KNDFS-BCBAC approach are feature selection and classification. A machine learning technique used to identify a linear combination of features is feature selection using the radial basis kernelized normal discriminant analysis. The resulting feature selection is frequently used for dimensionality reduction before to classification. Following feature selection, a Borda count bootstrap aggregating classifier is used to diagnose the disease and identify risk variables by classifying the normal patient or abnormal patient. The suggested KNDFS-BCBAC technique's two key processes are outlined in the following subsections.

3.1 kernelized Normal discriminant analysis

Because processing more features uses more computation time, big space, and so on, the feature selection aids in selecting a smaller number of features from a larger number of features. As a result, the KNDFS-BCBAC technique uses Radial Basis Kernelized Normal Discriminant Analysis (RBKNDA) to pick features since it produces more accurate classification results.

Consider the dataset Ds, which contains a number of attributes, namely features a_1,a_2,a_3, ….. a_n. For decreasing the complexity, the feature that is more relevant to the condition is identified among them. The RBKNDA begins with defining the separation function, which divides the features into two subsets: relevant feature subset (fs_r)and irrelevant feature subset (fs_{ir}). The separation function defines the ratio, which can be stated mathematically as follows:

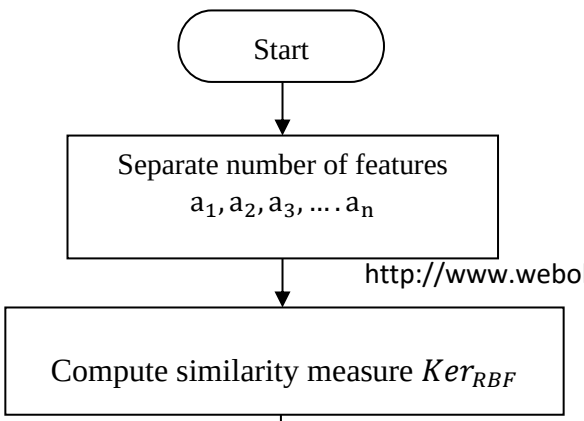
$$sep_f = \frac{V_b}{V_w} = \frac{L\, m_w d_t}{L\, m_b d_t} \qquad (1)$$

Where sep_f stands for separation function, V_b for variance between subsets, V_w for variance within subsets, L for linear discriminant vector used for projecting relevant and irrelevant features into subsets, and m_w, m_b for scatter matrix within and between subsets. To quantify the similarity between two samples, the suggested technique employs the radial basis kernel function (i.e. features, objective function). The tumour illness symptoms are the objective function in this case. Below is the radial basis kernel function.

$$Ker_{RBF} = exp(-\rho \, \|at_i - ds_b\|^2) \qquad (2)$$

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

$Ker_{RBF}$is the radial basis kernel function, $\|at_i - ds_b\|^2$ is the squared Euclidean distance between the two samples, and D is a deviation parameter in (2), (3) where D>0, ρ is the parameter, at_i represents an attribute, and ds_b represents disease symptoms. The radial basis kernel function returns a number between 0 and 1 for similarity. The discriminant vector divides the features into subsets based on their similarity. The highest similarity features are projected into the applicable feature subsets. Otherwise, the features are projected onto feature subsets that are irrelevant. The following figure 1 depicts the flowchart of the proposed technique.



Yes

No

Figure 1 Flow chart of kernelized Normal discriminant analysis

Figure 1 shows the steps involved in selecting relevant features using a radial basis kernelized model. Discriminant analysis in the traditional sense. In comparison to previous efforts, the KNDFS-BCBAC technique examines a smaller number of features for disease classification and thereby reduces the time complexity. The following is an explanation of the feature selection algorithm.

Algorithm 1: kernelized Normal discriminant analysis
Input: Dataset D_s, number of attributes or features$a_1, a_2, a_3, \ldots . a_n$.

Output: Select relevant feature subsets

Begin

1. **Select** the number of features $a_1, a_2, a_3, \ldots . a_n$ from D_s
2. Define the separation function sep_f
3. **for** each feature calculate the attribute at_i and objective function ds_b
4. Compute the radial basis kernel functionKer_{RBF}
5. **if** $(Ker_{RBF} > \delta)$ then
6. Discriminant vector projects the features into relevant subsets 'fs_r'

7. **else**
8. Discriminant vector projects the features into irrelevant subsets 'fs_{ir}'
9. **end if**
10. Select relevant features subset 'fs_r'
11. Remove the irrelevant features subset ' fs_{ir}'
12. **end for**

end

For disease diagnosis, a radial basis kernelized normal discriminant analysis-based feature selection procedure is used, as shown in the above algorithm. The radial basis kernel function is used to project the features into relevant and irrelevant subsets using the linear discriminant vector in the provided dimensions. The RBKNDA selects the relevant feature subsets while discarding the unnecessary ones. The relevant feature subsets are employed to reduce the complexity of disease diagnosis, resulting in increased accuracy.

3.2 Borda count bootstrap aggregating classifier

The patient data is categorised using the Borda count bootstrap aggregating classifier after the relevant feature subset has been selected. The bagging technique, also known as the bootstrapped aggregating classifier, is a machine learning ensemble classifier that improves accuracy by creating a large number of weak learners. The weak learner is a type of base classifier that has a hard time delivering appropriate classification results. Bootstrapped aggregating, on the other hand, is a powerful ensemble classifier that produces correct results using a voting mechanism. The Borda count voting mechanism is used in the proposed KNDFS-BCBAC technique to produce reliable classification results through the ranking phase. Below is the Borda count bootstrapped aggregating classifier structure.

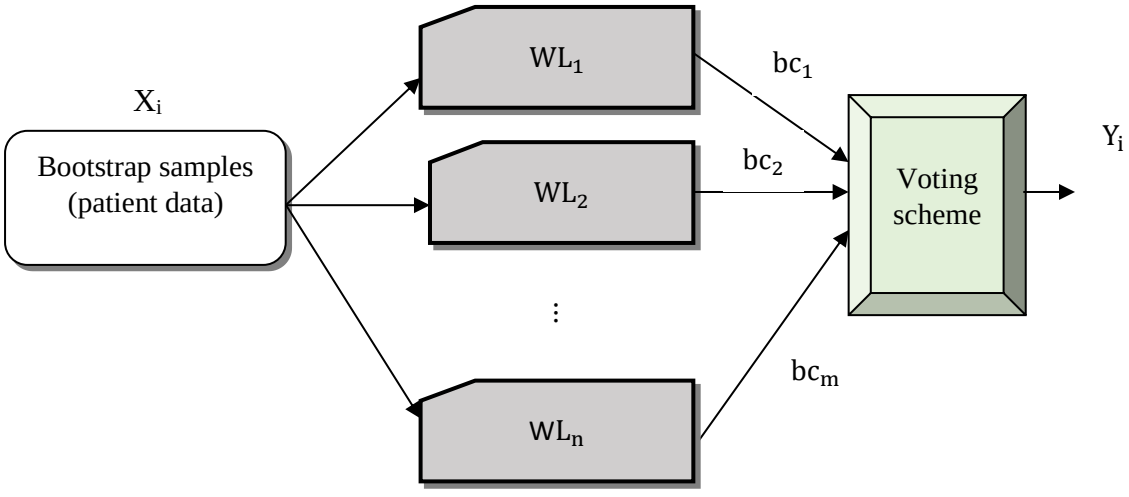


Figure 2 Borda count bootstrap aggregating classifier

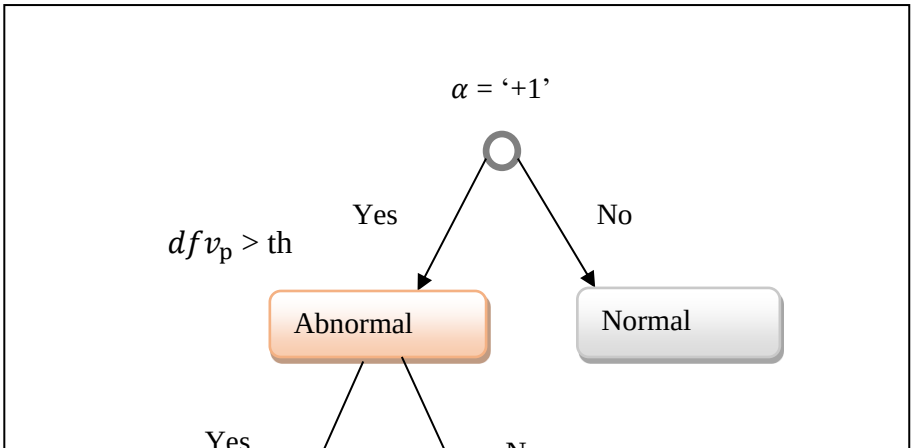
The Borda count bootstrap aggregating classifier is shown in Figure 2. The classifier takes into account the training sets (X_i, Y_i), where X_i stands for bootstrap samples (i.e. input patient data) and Y_i stands for ensemble classification results. An empty set of' 'm' weak learners left (WL_1, WL_2, WL_3, ….WL_n) right is created via the Borda count bootstrap aggregating classifier (bc_1,

bc_2, …bc_m). To classify the patient data, a bivariate correlated regression tree is utilised as a weak learner. The regression tree is used to assess patient data in conjunction with the illness feature value in order to determine if the patient is normal or abnormal. The root node, branch node, and leaf node make up the regression tree. The tree's root node assesses the relationship between the training feature value (patient data) and the testing feature value (disease feature) for analysis.

$$\alpha = \frac{(\sum pd_i df_t) - (\sum pd_i)(\sum df_t)}{\sqrt{[\sum {pd_i}^2 - (\sum pd_i)^2][\sum {df_t}^2 - (df_t)^2]}} \qquad (4)$$

Where, α denotes a correlation coefficient, pd_i represents patient data, df_tdenotes a disease feature value, $\sum pd_i df_t$ denotes a sum of the product of paired score, $\sum p_i$ is the sum of p_v score, $\sum df_t$ is the sum of df_t score, $\sum {pd_i}^2$is the sum of the squared score of $\sum pd_i$ and $\sum {df_t}^2$ is the sum of the squared score of $\sum df_t$. The bivariate correlation coefficient provides the value between +1 and -1 where '+1' indicates the positive correlation and '-1' indicates the negative correlation. The positive correlation implies that the patient data is aberrant (i.e. disease), whereas the negative correlation suggests that the data is normal. After classification, the branch node performs a critical factor analysis of patient data by classifying the distinct phases based on the feature value threshold range. At the leaf nodes, the classification results are obtained. The threshold is set to the aberrant illness feature value (dfv_p) in order to examine the critical risk factor of the patient data. The data is categorised as critical stage if the disease feature value exceeds the threshold. Aside from that, the disease patient data is categorised as early stage. Figure 3 depicts the classification of the bivariate correlated regression tree.

Figure 3. Classification of the bivariate correlated regression tree



A bivariate linked regression tree is shown in the above figure. Based on the correlation coefficient values, the root node in the tree decides whether the patient data is normal or abnormal. The branch node then established a specified threshold value for the illness feature value to identify the patient's risk factor. The patient data is categorised in this fashion, and the disease and risk are detected. In the classification results of the weak learner, there is some training error. The classification results of all the weak learners are pooled in order to get strong classification results. The following are the strong classification results:

$$Y_i = \sum_{i=1}^{m} bc_i \qquad (5)$$

The output of the strong classifier is Y_i, while the output of the weak learners is bc_i. The generalisation error is calculated for each weak learner in order to predict appropriate classification outcomes. The difference between the expected and predicted error is used to compute the generalisation error.

$$eg_{err} = \{ex_{err} - pre_{err}\} \qquad (6)$$

In (6),eg_{err} denotes a generalization error, $ex_{err}$is the expected error, pre_{err} represents the predicted error. By applying the Borda count voting scheme, the weak classifiers are arranged and ranked based on the generalization error.

$$\mathrm{eg_{err}}\ (\mathrm{bc_1}) \leq \mathrm{eg_{err}}\ (\mathrm{bc_2}) \leq \mathrm{eg_{err}}\ (\mathrm{bc_3}) \leq \cdots . \mathrm{eg_{err}}\ (\mathrm{bc_m})) \qquad (7)$$

In contrast to the conventional bootstrap aggregating classifier, the proposed bootstrap aggregating classifier assigns the first rank to their most favoured weak classifier results, i.e. those with the lowest generalisation error, and so on. The results of the higher-ranked classifiers are used in the final classification. The votes of the samples in the higher-ranked classifiers are counted and the majority to be chosen is determined based on the classification results.

$$\mathrm{Y} = \arg\max_{\mathrm{m}} \mathrm{vot}\ (\mathrm{pd_i}) \qquad (8)$$

arg max denotes an argument of the maximum function to obtain the majority vote (vot) of the samples (i.e. patient data' pd') whose choice is known to the m^{th} classifier in (8). Finally, the bootstrap aggregating classifier delivers good classification results for the majority of the samples, improving disease diagnosis accuracy and lowering false positive rates. The proposed ensemble classification results' algorithmic approach is shown below.

Algorithm 2: Borda count bootstrap aggregating classifier
Input: Patient data $pd_1, pd_2, pd_3, \ldots pd_n$
Output: To determine disease diagnosing accuracy
Begin

1. Design 'm' regression trees with training Patient data $pd_1, pd_2, pd_3, \ldots pd_n$
2. for each data pd_i
3. Compute the correlation
4. **if** $(\alpha = +1)$ **then**
5. Calculate the optimal positive correlation
6. Classify the abnormal data's
7. **else**
8. Determine the negative correlation
9. Classify the normal data's
10. **end if**
11. **if** $(dfv_p > th)$ **then**
12. Separate the patient data as critical stage
13. **else**
14. Separate the patient data as preliminary stage
15. **end if**
16. Combine a set of weak learners $Y_i = \sum_{i=1}^{m} bc_i$
17. **For** each bc_i
18. Compute the generalization error 'eg_{err}'
19. **end for**
20. Allocate the rank based on error rate
21. Select the highest ranked classifier
22. Categorize the majority votes of the samples $Y = \arg\max_{m} vot\,(pd_i)$
23. Obtain the optimal classification output

End

Algorithm 2 depicts the Borda count bootstrap aggregating classifier's step-by-step procedure for diagnosing the disease at an early stage. With the patient data, the bootstrap aggregating classifier creates an empty set of regression trees. The regression tree's root node then calculates the correlation between the training feature value (i.e. patient data) and the illness feature value. The normal and abnormal patient conditions are determined based on correlation data. The risk factor analysis is carried out by determining the abnormal disease feature value's threshold. The several stages of the patient's health are noted. By minimising the generalisation error, the weak classification results are combined to create strong classifications. By a majority vote of

all the samples, the ensemble classifiers ranked the classifier with the lowest generalisation error and the best results. As a result, the bootstrap aggregating classifier enhances disease diagnosis accuracy while lowering the percentage of false positives.

4.Experimental 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 namely KNDFS-BCBAC, LRCN [1] and MRBF-MPSO-SVM [2] are implemented in the Java language. The OASIS dataset includes 13 features and 416 instances.

Table 1 Feature Description

Serial no.	Features	Description
1.	MR Sessions	MRI Imaging sessions
2.	Subject	Patients
3.	M/F	Gender Male or female
4.	Hand	Right-handed (R)
5.	Age	Age of the patient
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 analysis of KNDFS-BCBAC technique and existing methods are evaluated with different parameters are listed below,

- Disease diagnosing accuracy
- False positive rate
- Time complexity

5. Result Analysis and Discussion

In this section, the comparative result analysis of proposed KNDFS-BCBAC, LRCN [1] and ESD-LSTM [2] are discussed with different parameters such as disease diagnosing accuracy,

false positive rate, and time complexity. The comparative performance analysis are done with the help of tables and graphical representation.

5.1 Disease diagnosing accuracy

Disease diagnosis accuracy is measured through the classification of patient data. Therefore, the accuracy is the ratio of a number of (i.e. no. of) patient data are correctly classified to the number of patient data. The Disease diagnosis accuracy is calculated as follows,

$$\mathrm{DDA} = \left(\frac{\text{No.of data correctly classified}}{\mathrm{n}}\right) * 100 \qquad (9)$$

Wheren denotes a total number of data. The Disease Diagnosis Accuracy(DDA) is measured in the unit of percentage (%). The mathematical formula for calculating the tumor diagnosis accuracy is given below,

Sample calculation:

- **Existing LRCN:** No. of data correctly classified is 32 and the total no. of patients is 40. The disease diagnosis accuracy is mathematically calculated as,

$$\mathrm{DDA} = \left(\frac{32}{40}\right) * 100 = 80\%$$

- **Existing ESD-LSTM:** No. of data correctly classified is 25 and the total no. of patients is 40. The disease diagnosis accuracy is mathematically calculated as,

$$\mathrm{DDA} = \left(\frac{25}{40}\right) * 100 = 62.5\%$$

- **Proposed KNDFS-BCBAC**: No. of data correctly classified is 35 and the total no. of patients is 40. The disease diagnosis accuracy is mathematically calculated as,

$$\mathrm{DDA} = \left(\frac{35}{40}\right) * 100 = 87.5\%$$

Table 2. Disease diagnosing accuracy

No. of Patient Data	KNDFS-BCBAC	LRCN	ESD-LSTM
40	89	82	74
80	92	87	82
120	93	88	85
160	92	87	83
200	95	89	85
240	93	87	82
280	92	85	79

320	94	88	82
360	92	85	80
400	91	84	79

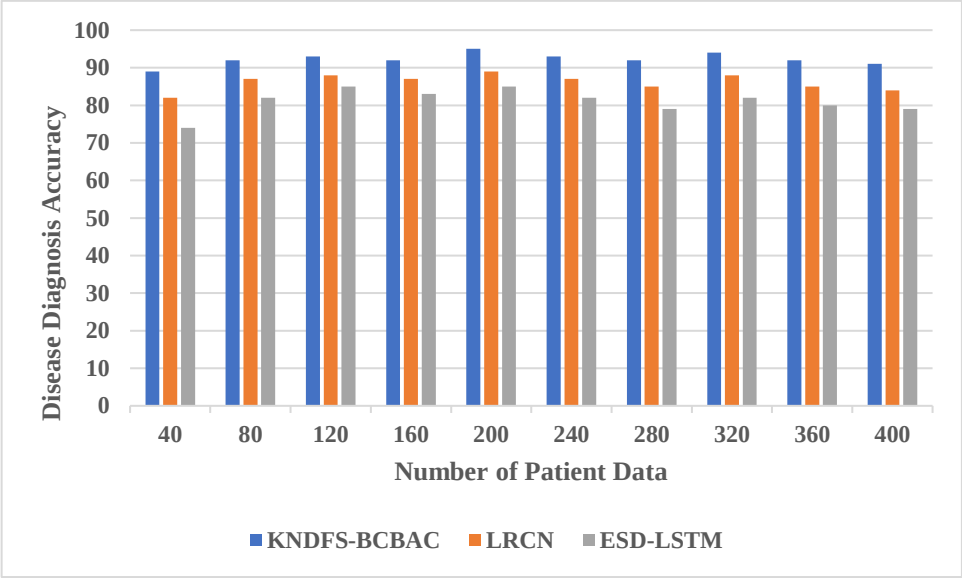


Figure 4. Disease diagnosing accuracy

Figure 4 shows a comparison of disease diagnostic accuracy using three methodologies based on the number of patient data. The graphs show that the KNDFS-BCBAC enhances the accuracy of brain tumour diagnosis. By using the bootstrap aggregating classifier, the accuracy rate can be improved. With the patient data, the ensemble classifier creates a regression tree that is correlated with the illness characteristics value. The normal patient and tumor-affected patient are determined using regression analyses. Following that, the various phases of a brain tumour are identified, such as mild and extremely mild. By minimising the generalisation error, the weak classifiers are combined to produce strong classification results. The ensemble classifier improves the accuracy of disease diagnosis. The sample mathematical computation demonstrates with '40' patient data being considered for testing and '36' patient data being successfully diagnosed utilising the KNDFS-BCBAC approach. With '40' patient data, '33' patient data successfully diagnosed using LRCN [1] and '30' patient data correctly diagnosed using ESD-LSTM [2], the total disease diagnosing accuracy was determined to be 82% and 74 % respectively. The results suggest that KNDFS-BCBAC improves the accuracy of identifying brain tumour disease by 6% when compared to LRCN [1] and 13% when compared to ESD-LSTM [2].

5.2 False Positive Rate

The false positive rate is calculated as the ratio of the number of erroneously classified patient data to the total quantity of patient data. The following is the mathematical formula for estimating the false positive rate is,

$$FPR = \left(\frac{\text{No.of data incorrectly classified}}{n}\right) * 100 \qquad (10)$$

FPR is stands for False Positive Rate and n stands for the total number of data points. The FPR is expressed in percentage terms (%). This performance is stated to be improved if the false positive rate is lower.

Sample calculation:

- **Existing LRCN**: No. of data incorrectly classified is 6 and the total no. of patients is 40. The false positive rate is mathematically calculated as,

$$\text{FAR} = \left(\frac{6}{40}\right) * 100 = 15\%$$

- **Existing ESD-LSTM**: No. of data incorrectly classified is 12 and the total no. of patients is 40. The false positive rate is mathematically calculated as,

$$\text{FAR} = \left(\frac{12}{40}\right) * 100 = 30\%$$

- **Proposed KNDFS-BCBAC**: No. of data incorrectly classified is 5 and the total no. of patients is 40. The false positive rate is mathematically calculated as,

$$\text{FAR} = \left(\frac{5}{40}\right) * 100 = 12.5\%$$

Table 3. False Positive Rate

No. of Patient Data	KNDFS-BCBAC	LRCN	ESD-LSTM
40	11	19	25
80	9	14	19
120	7	12	15
160	9	14	17
200	6	11	15
240	7	14	19
280	8	15	20
320	6	12	17
360	8	15	20
400	9	16	21

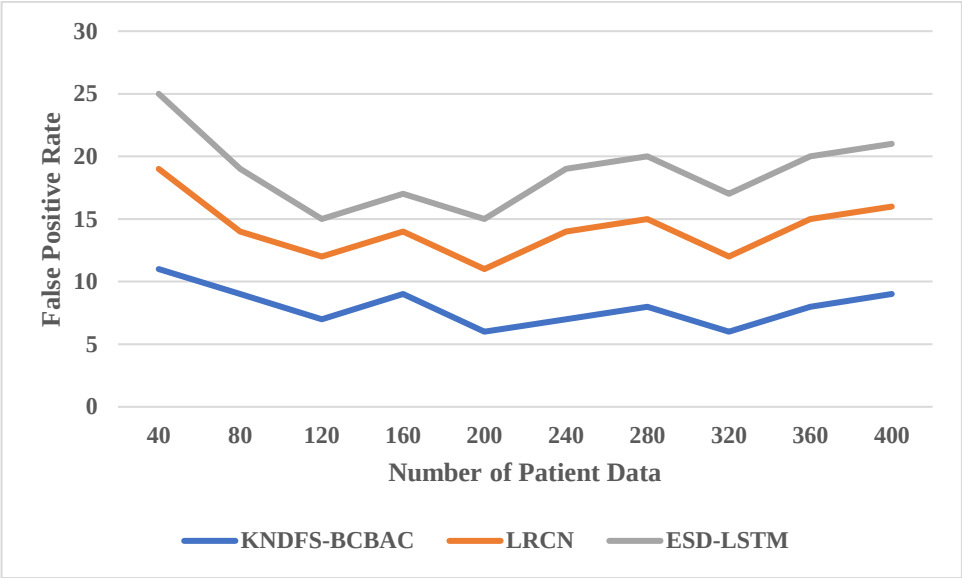


Figure 5. Graphical representation of false positive rate

Figure 5 shows the false positive rate performance measure in relation to 400 patient data. The suggested KNDFS-BCBAC approach with the OASIS dataset has a low false positive rate, as demonstrated in the above figure. The Borda count bootstrap aggregating classifier has been used to improve the KNDFS-BCBAC approach. The Borda count voting system is used by the bootstrap aggregating classifier to determine the weak classifier with the lowest generalisation error based on the ranking algorithm. As a result, high-ranking classifiers are chosen to produce reliable classification results. The higher-ranked classifiers receive the majority of the votes in the results. The KNDFS-BCBAC approach reduces the chances of the ill patient being misidentified. When compared to the two state-of-the-art approaches, using LRCN [1] and ESD-LSTM [2], the KNDFS-BCBAC technique reduces the false positive rate by 47 % and 61 % respectively.

5.3 Time complexity

Time complexity is measured as the amount of time taken by the algorithm to diagnosis the tumor with the number of patient data. The time complexity is calculated as follows,

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

In (11) TC denotes the time complexity and T represents a time for diagnosing one patient data. Time complexity is measured in milliseconds (ms).

Sample Calculation:

- **Existing LRCN:** No. of patient data is 40, and the time taken to diagnosis the single data is 0.4 ms. Then the overall time complexity is computed as follows,

$$TC = 40 * 0.4 \text{ ms} = 16 \text{ ms}$$

- **Existing ESD-LSTM**: No. of patient data is 40, and the time taken to diagnosis the single data is 0.5 ms. Then the overall time complexity is computed as follows,

$$TC = 40 * 0.5 \text{ ms} = 20 \text{ ms}$$

- **Proposed KNDFS-BCBAC**: No. of patient data is 40, and the time taken to diagnosis the single data is 0.3 ms. Then the overall time complexity is computed as follows,

$$TC = 40 * 0.3 \text{ ms} = 12 \text{ ms}$$

Table 4. Time Complexity

No. of Patient Data	KNDFS-BCBAC	LRCN	ESD-LSTM
40	13	17	21
80	19	23	27
120	25	31	37
160	31	38	46
200	39	45	51
240	44	49	56
280	49	54	58
320	56	61	68
360	62	68	73
400	68	75	81

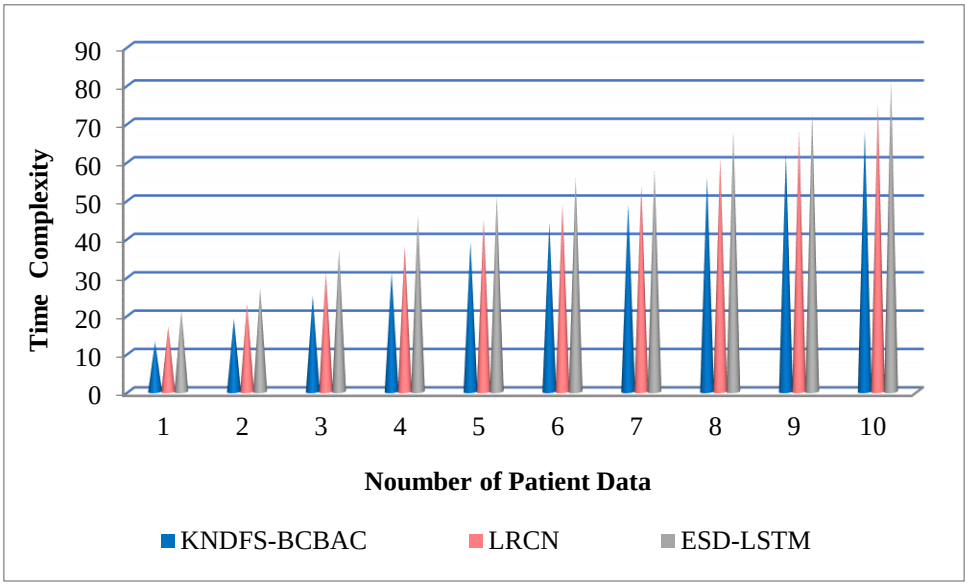


Figure 6. Graphical representation of time complexity

The performance results of time complexity of disease diagnosis using a number of patient data are shown in Figure 6. The x-axis depicts the amount of patient data points, while the y-axis reflects the time complexity in milliseconds(ms). As the volume of patients' data grows, so

does the amount of time it takes to diagnose a condition. The temporal complexity associated in brain tumour disease diagnosis was shown to be lowered employing the KNDFS-BCBAC technique, as shown in the image. This is due to the usage of radial basis kernelized data. Feature selection based on normal discriminant analysis. The discriminant vector was utilised to project the more similar features in the feature subset for brain tumour diagnosis. The Clinical Dementia Rating (CDR) score is used to classify the patient as normal, mild, or extremely mild based on the selected criteria. As a result, the time complexity of tumour disease diagnosis is reduced. The sample computation demonstrates this. The time complexity using LRCN [1] and ESD-LSTM [2] is '20ms' and '16ms' correspondingly, with '40' patient data considered for testing and the time required in diagnosing is '12ms' using KNDFS-BCBAC approach. The patient data is used to execute the various runs. The KNDFS-BCBAC technique's findings are compared to those of other approaches. In comparison to previous procedures, the time complexity is lowered by 20% and 25%, respectively, according to the comparison results.

The above comparison findings and discussion clearly demonstrate that the suggested KNDFS-BCBAC technique performs better in terms of disease diagnosis accuracy, false positive rate, and time complexity.

6. Conclusion

In this study, a KNDFS-BCBAC technique based on machine learning is created with the goal of enhancing disease diagnosis accuracy in a short amount of time. Two methods are used in the KNDFS-BCBAC methodology used in this study. The radial basis kernelized discriminant analysis selects the relevant features, reducing the irrelevant features subset and obtaining a more effective relevant subset of features. This aids in lowering time complexity and, as a result, boosting disease diagnosis accuracy. Furthermore, by minimising the generalisation error, the bootstrap aggregating ensemble technique is used to transform the weak classification results into strong classification results. As a result, accurate classification is carried out, which enhances diagnosis outcomes. The OASIS dataset was used to test the accuracy of KNDFS-BCBAC and existing approaches in terms of disease diagnosis, false alarm rate, and time complexity. In terms of disease diagnosis accuracy, false positive rate, and time complexity, the quantitative results reveal that the proposed KNDFS-BCBAC strategy outperforms previous state-of-the-art methods.

References:

1. Asma Naseer, Tahreem Yasir, Arifah Azhar, Tanzeela Shakeel, and Kashif Zafar, “Computer-
Aided Brain Tumor Diagnosis: Performance Evaluation of Deep Learner CNN Using Augmented Brain MRI”, International Journal of Biomedical Imaging Volume 2021, Pages 1-11
https://doi.org/10.1155/2021/5513500

2. Ahmed Abdelhameed and Magdy Bayoumi "A Deep Learning Approach for Automatic Seizure Detection in Children with Epilepsy", Frontiers in Neuroinformatics, Volume 15, Issue 650050, April 2021, Pages 1-15 https://doi.org/10.3389/fncom.2021.650050

3. A. Ulate-Campos, F. Coughlin , M. Gaı´nza-Lein, I. Sa´nchez 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 DOI: 10.1016/j.seizure.2016.06.008

4. Dhanashri Joshi, Hemlata Channe, A Survey on Brain Tumor Detection Based on Structural MRI Using Machine Learning And Deep Learning Techniques, International Journal of Scientific & Technology Research, Volume 9, Issue 04, April 2020 ISSN 2277-8616

5. Erena Siyoum Biratu, Friedhelm Schwenker, Yehualashet Megersa Ayano, Taye Girma Debelee, A Survey of Brain Tumor Segmentation and Classification Algorithms, Journal of Imaging, 2021, Pages 1-30 https://doi.org/ 10.3390/jimaging7090179

6. Francisco Javier Díaz-Pernas, Mario Martínez-Zarzuela, Míriam Antón-Rodríguez and David González-Ortega, A Deep Learning Approach for Brain Tumor Classification and Segmentation Using a Multiscale Convolutional Neural Network, Healthcare 2021, 9, 153. https://doi.org/10.3390/ healthcare9020153

7. 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 DOI:10.1007/s40708-015-0030-2

8. Javaria Amin, Muhammad Sharif, Anandakumar Haldorai, Mussarat Yasmin, Ramesh Sundar Nayak, Brain tumor detection and classification using machine learning: a comprehensive survey, Complex & Intelligent Systems, Nov 2021, pages 1-23 https://doi.org/10.1007/s40747-021-00563-y

9. Jefferson Tales Oliva and João Luís Garcia Rosa, "Classification for EEG report generation and epilepsy detection", Neurocomputing, Elsevier, Volume 335, 2019, Pages 81-95 https://doi.org/10.1016/j.neucom.2019.01.053

10. Jhe-Nan Lina Fu-Yi Yangb Wei-Zen Suncd Meng-Han Yang "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 DOI:10.1016/j.compeleceng.2018.04.020

11. Likith Sr,Nalini N, Classification and Segmentation of Brain MRI images using Deep Learning, IEEE Xplore 2021 DOI: 10.1109/GCAT52182.2021.9587460

12. Linmin Pei, Lasitha Vidyaratne, Md Monibor Rahman & Khan M. Iftekharuddin, "Context aware deep learning for brain tumor segmentation subtype classification and survival prediction using radiology images", Scientific Reports | (2020) 10:19726 | Pages: 1-11 https://doi.org/10.1038/s41598-020-74419-9

13. 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 DOI:10.1016/j.eswa.2013.09.022

14. 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 https://doi.org/10.1155/2017/6848014

15. M. Deriche S. Arafat S. Al-Insaif M. Siddiqui, "Eigenspace Time Frequency Based Features for Accurate Seizure Detection from EEG Data", IRBM , Elsevier, Volume 40, Issue 2, 2019, Pages 122-132 https://doi.org/10.1016/j.irbm.2019.02.002

16. Muhammad Waqas Nadeem, Mohammed A. Al Ghamdi, Muzammil Hussain, Muhammad Adnan Khan, Khalid Masood Khan, Sultan H. Almotiri and Suhail Ashfaq Butt "Brain Tumor Analysis Empowered with Deep Learning: A Review, Taxonomy, and Future Challenges", Brain Sciences, 2020,10(2),118, Pages 1-33 doi: 10.3390/brainsci10020118

17. 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 DOI:10.1007/s13369-019-03829-3

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 https://doi.org/10.1016/j.eswa.2017.05.055

19. N. Sriraam, S. Raghu Kadeeja 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 DOI:10.1186/s40708-018-0088-8

20. 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 https://doi.org/10.1371/journal.pone.0193691

21. 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 https://doi.org/10.1016/j.aci.2015.01.001

22. 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 DOI: 10.1016/j.compbiomed.2017.10.008

23. RamyHussein, 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 25-37https://doi.org/10.1016/j.clinph.2018.10.010

24. 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 DOI:10.1016/j.asoc.2018.11.012

25. 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 DOI: 10.1109/ACCESS.2016.2647238

26. 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 DOI:10.1016/j.eswa.2019.03.021

27. Tahia Tazin, Sraboni Sarker, Punit Gupta, Fozayel Ibn Ayaz, Sumaia Islam, Mohammad Monirujjaman Khan, Sami Bourouis, Sahar Ahmed Idris, and Hammam Alshazly, "A Robust and Novel Approach for Brain Tumor Classification Using Convolutional Neural Network", Computational Intelligence and Neuroscience Volume 2021, Pages 1-11 https://doi.org/10.1155/2021/2392395

28. Xiaoyan Wei, Lin Zhoub Zhen Zhangc Ziyi Chend Yi Zhoua, "Early prediction of epileptic seizures using a long-term recurrent convolutional network", Journal of Neuroscience Methods, Elsevier, Volume 327, 2019, Pages 1-39 DOI: 10.1016/j.jneumeth.2019.108395

29. 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 DOI: 10.1109/JBHI.2017.2654479

30. 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 https://doi.org/10.1155/2017/6849360