A hybrid feature extraction based optimized random forest learning model for brain stroke prediction

G Vijayadeep[1], Dr N Naga Malleswara Rao[2]

Research Scholar, Department of Computer Science and Engineering
[1]University College of Engineering, Acharya Nagarjuna University, Guntur, India
[2]Professor, Department of Information Technology, RVR and JC College of Engineering, Guntur, India.

Abstract:
Brain stroke detection (SD) is one of the major chronic diseases which can be detected using the MRI images. Most of the severe patterns in the SD database are detected using the machine learning models. Most of the traditional machine learning models such as random tree, random forest, SVM, Naïve bayesetc are difficult to find the feature extraction and difficult to detect essential features for disease classification. Also, the classification of the SD patterns based on severity is difficult in the traditional models due to complexity in feature selection and ranking. Hence, it is essential to develop an appropriate and an effective automatic test in order to diagnose stroke for better disease management. However, as the number of stroke features increases, traditional models require high computational memory and time for feature selection and pattern evaluation. Also, these models generate high false positive rate and error rate due to high feature space and data uncertainty. In order to overcome these issues, a hybrid feature selection-based classification learning model is designed and implemented on high dimensional dataset. Experimental results proved that the present model improves the true positivity and error rate compared to the traditional models.

Index Terms— **Brain stroke, feature extraction, stroke classification, machine learning models.**

I INTRODUCTION

Due to many difficulties, computer-assisted brain tumor identification, stroke lesions, hemorrhage lesions, and multiple sclerosis lesions are the most difficult problems in the area of irregular tissue segmentation. Brain injuries are of various types and also affect other components in normal and healthy tissues. The distribution of strength of normal tissues is very complex, and some overlaps occur between various tissue types. Both forms of segmentation of brain disorder use the dogma of the discrepancy between the abnormal brain MRI and its normal counterpart. Different approaches have been suggested for the same over the past decade. Others interpreted the segmentation task as a question of tissue recognition, which meant using a well-trained model that could decide whether a pixel belongs to a normal or abnormal tissue based on an approach to machine learning. Brain tumors are one of the most common brain diseases, so it is important for medical diagnosis to detect and segment brain tumors in MRI. Existing methods leave ample space for greater precision, applicability and accuracy. This disease has different stages from non-harmful initial stage to worse terminal stage. If the disease progresses to the terminal stage, then it may lead to death of the patient. Extraction of features is the method of extracting certain characteristic attributes from an image and producing a collection of meaningful descriptors. Mass effect and hyperintense areas were observed in clusters of spinal fluids and anomaly clusters, both of which were the characteristics derived in this process. A wavelet-based approach to remove the texture characteristics is proposed in [1]. The mean and range of measurements of each 22 features were used as features over the four offset angles. Therefore, 44 characteristics were removed. To reduce the number of functions, the Genetic Algorithm (GA) was used. The optimal characteristics were average contrast, average homogeneity, average sum average, mean sum variance and auto correlation range. Eight texture features were

identified to each stroke image. They were energy, entropy, inverse differential measure(IDM), inertia, shadow, prominence, correlation, and variance. Attributes from the left and right side of the brain picture were removed.
A texture characteristic using the cooccurrence matrix of the gray point is defined in [2]. The features extracted were Angular Second Moment (ASM), comparison, entropy, IDM, and distinction. It was used in the classification of 4 types of tumor astrocytoma. Kumaret al.[3] derived 14 features from three groups, including four wavelet-based features, four first-order histogram-based features, and six gray-level cooccurrence matrix-based features. This approach used the method of optimizing particle swarm for the selection of features. The optimal feature set consisted of wavelet-based features that include cluster hue, cluster prominence; first-order histogram-based features that include skewedness, kurtosis, and entropy; Gray Level Co-Occurrence Matrix (GLCM) features that include comparison, IDM, correlation, and entropy of variance. In this paper, the following GLCM features were extracted. The characteristics of the intensity histogram were mean, variance, skewedness, kurtosis, entropy, and energy. In this paper, to pick the best features, a hybrid approach of fuzzy roughest method with GA was suggested. For classification, the selected optimal features were considered. The overall characteristics extracted were 40, the selected characteristics were eight from the extracted features using GA. Area, centroid, minor axis length, autocorrelation, sum entropy, difference variance, mean and energy were the characteristics selected by the GA system. Kurtosis, standard deviation, sum average and sum variance were the characteristics identified by the rough set process.Various research works have given emphasis on training systems along with respiratory data to retrieve relevant and important features for the process of diagnosis and prediction. Traditional machine learning approaches which are used in the diagnosis process of asthma disease are thoroughly studied and analyzed. Data sets for total 100 numbers of patients are considered here.
Fuzzy rules are integrated with neural networks in order to develop an efficient and effective prediction tool that will detect asthma and stroke more accurately. In this experimental analysis, total 455 numbers of patients are considered. Such pre-processing techniques have significant roles in order to achieve exact classification accuracy. There exist some other techniques those integrate machine learning approaches with feature significance ranking techniques. All the traditional machine learning approaches can be categorized under two basic categories, those are:- linear machine learning approaches and non-linear machine learning approaches. Linear and non-linear machine learning approaches are analyzed and compared with each other. As logistic regression can be included under the category of linear and random forests can be included under the category of non-linear machine learning technique. Both logistic regression and random forests techniques are implemented together in order to predict the stroke. Random forests approach shows best and optimized performance as compared to every other pre-existing classification technique. This technique is very much efficient to analyze the severity of abnormal modifications in case of visual examination of stroke images. Most of the traditional stroke detection models are categorized in to three subcategories, those are:-

1. Density-based approach: The density-based approach has the responsibility to detect the relationship among the previously defined threshold and the ratio of stroke tissue area.
2. Model-based approach: Model-based approaches usually depend upon the brain stroke anatomy. All the anatomical characteristics are retrieved initially to predict the stroke patterns.
3. Texture based approach: These approaches are generally used to calculate different measures to analyze the stroke characteristics.

Different efficient and effective classification schemes are implemented to train the system. In the subsequent time, another generalize system is developed which has the responsibility of regional stroke classification. Here, each and every small region is classified into one of the six categories. This approach considers all five patterns with equal training set which is quite impossible in the practical scenario[4]. In order to resolve the above mentioned classification issue, a new method is introduced that will include two sets of wavelet filters. DWF are a special kind of filter which can be easily isolatable and these are 2-D wavelet filters. RWF will allow freedom during the whole process of orientation tuning for image content analysis. Discrete wavelet transform and random wavelet function are combined together in order to develop a new tool that will help to retrieve directional texture features. In order to resolve the above mentioned issues, texture classification scheme is introduced. These techniques are implemented to classify different patches of brain stroke tissue, or regions of interest (ROIs) in the image[5]. The outcomes of this classification phase are integrated to evaluate the probability of stroke. All supervised techniques need manual annotated ROIs. This process is very much complex and expensive. There is another method in order to use weakly labeled medical images. A global image label is used in order to train an image classifier. When labeled ROIs are not present, this image label is propagated to its ROIs. The training of ROI classifier is performed in the traditional way. The above technique is known as Simple multi-instance level(MIL) approach.

It enhances the noise label in case of ROI classifier. The approach is implemented in order to handle learning. This proposed technique involves weakly labeled data and it is known as multiple instance learning (MIL) technique. The prime objective of this classification scheme is to construct an efficient classifier for a group (bag) of feature vectors. These feature vectors are also known as instances. A bag is considered as positive, only when it contains at least a single positive instance. It can be assumed that, all positive instances are detected in the area of feature space which is also known as the concept.

The GLCM is a calculation of the second order texture. Texture features were extracted and used to improve the result of classification as a parameter. This paper provides a study of three types of texture used in the identification of stroke abnormalities.

L. I. Kuncheva, et.al, developed a new ensemble classification approach in order to diagnose stroke patterns[1]. The diagnosis process of stroke depends upon its symptoms, test results along with some other factors too. Large numbers of people are ignorant about the above disease so that, they never pay attention to its effects.

A. Corlateanu, et.al, emphasized on multi-lateral assessment of stroke [2]. stroke is considered as multi-level disease from the perspective of clinical, cellular, and molecular view. In this paper, a thorough survey is performed from the basic model to high-level advanced model.

*E. M. van Rikxoorta, et.al,*implemented a novel machine classification algorithm which depends on various feature selection metrics and stroke patterns [3].

S. A. Khan, et.al, developed an advanced approach for stroke detection and classification [4]. In this paper, an advanced region-based ACM with Otsu pre-processing approach is introduced. Support Vector Machine classification algorithm is implemented in the above scenario.

V. Cheplygina, et.al, introduced an advanced classification scheme for stroke with the help of multiple instance learning [5]. stroke disease is considered as most dangerous kind of cancer. The chance of survival of patient increases, if the disease can be identified at an early stage.

A.Sh. Abdalla, et.al, presented a new computer-aided diagnosis system for classification of brain tumors[6]. In this technique, artificial neural network concepts are used in order to classify among malignant or benign tumor. The Gaussian texture features algorithm shows better performance as compared to intensity features. An efficient weighting technique which depends upon classifiers can distinguish among scans from various domains. This technique can definitely enhance the results of the traditional classification approaches.

$$f(g) = \frac{Pr_1}{\sqrt{2\pi}\sigma_1} e^{\frac{-(I-\mu_1)^2}{2\sigma_1^2}} + \frac{Pr_2}{\sqrt{2\pi}\sigma_2} e^{\frac{-(I-\mu_2)^2}{2\sigma_2^2}}$$

$$aT^2 + bT + c = 0$$

with

$$a = \sigma_1^2 + \sigma_2^2$$

$$b = k(\mu_1\sigma_2^2 - \mu_2\sigma_1^2); k \in \Pr ob(i \mid x_j, \theta) = \frac{p_i f_i(x_j \mid \alpha_i)}{\sum_{k=1}^{K} p_k f_k(x_j \mid \alpha_k)}$$

$$c = \mu_2^2\sigma_1^2 - \mu_1^2\sigma_2^2 + \sigma_1^2\sigma_2^2 \ln(\max\{a, b\}.\frac{Pr_1}{Pr_2})$$

where P1 and P2 are class probabilities. The optimal threshold (T) can be found as solving the a,b,c parameters.

P. Ghosh, developed a new and advanced unsupervised segmentation of stroke regions from brain images [7]. Here, a new cellular automation-oriented unsupervised segmentation approach is presented which is also known as unsupervised grow-cut (UGC).At first, the algorithm automatically identifies each and every abnormal region in case of stroke field.

R. J. Huijsmans,et.al, focused clinical utility of the GOLD classification of stroke disease severity in pulmonary rehabilitation [8]. GOLD classification scheme can be further divided into four phase classification. Retrospective analysis is carried out on 253 patients. Analysis of variance technique is usually implemented to identify the differences among all GOLD phases. Hence, the classification scheme is not at all beneficial to choose candidates for pulmonary rehabilitation.

P. Kohlmann, et.al, implemented a new segmented based automatic stroke detection in limited datasets [9]. The above presented approach is completely automatic lung segmentation technique which considers images of magnetic resonance (MR). In most of the traditional techniques, Segmentation process is the initial phase of automatic technique that involves morphological MR images along with high spatial resolution. In the last phase, the total mass for both left as well as right lungs are distributed into numbers of different partitions. 14 numbers of patients having two time points results 28 perfusion data sets. These data sets are required during the initial evaluation phase of the above suggested approach.

Most of these machine learning models are executed in two sub-systems, those are:-

1. Feature extraction: In the feature selection process, they adapted Gabor filter. This feature extraction method has the responsibility to retrieve patterns like local edges and segments from input textures. In the recognition phase, they implemented a new boosting approach. Boosting approach can be defined as a new technique which performs voting with the help of some classifiers in order to enhance decision precision. In this case, AdaBoost algorithm is implemented as an efficient boosting technique. Initially, they calculated every individual boosting component of classifier and verified that there is no

significant performance. After implementing the said boosting technique, again the performance is evaluated [10].

2. Classification: An advanced classification scheme is implemented in order to classify the stroke tissues.

In this research work, we have emphasized on the following features:-

1) The True positive prediction rate of the proposed approach is optimized compared to conventional learning approaches.
2) Since traditional approaches are sensitive to high dimensional datasets with limited data size; proposed model is applicable to high dimensional datasets with large size dataset.
3) Proposed technique optimizes the feature selection mechanism from high dimensional dataset to optimize training accuracy.

The rest of the paper is organized as follows. Section 2 describes the hybrid stroke detection using feature extraction and classification models. Section 3 describes the simulation outputs of stroke disorder prediction model and its performance measures. Finally, in section 4, we conclude the paper with performance improvement.

2.Proposed Model

Texture is one of the important features used to identify objects or areas of interest in an image. The gray level co-occurrence matrix (GLCM), also known as the gray spatial dependency matrix, is a statistical method for testing texture that considers spatial connection of pixels. The GLCM functions define an image's texture by measuring how often pairs of pixels with certain values occur, generating a GLCM, and then extracting statistical measurements from this range. Let F be a rectangular, discrete image containing a finite number of gray levels. F may be defined over the domain. The feature extraction should also be related to how the disease progresses and how the disease affects the life of the patient.The strength of consistency of the independent relationship between the stroke and feature extraction is also analysed. How the feature extraction relates to the clinical outcomes such as the patient being hospitalized and death is also analysed and the strength determined. Determining if the feature extraction can be modified using interventions should be evidently found out by carrying out controlled trials. Lastly, the facts from randomized and controlled tests should be able to indicate whether the changes in feature extraction status can result in hard clinical outcomes such as death. Whether a feature extraction is short-term or long-term should also be known. It should also be scientifically provable if changes in the feature extraction have a significant effect. Lesions in medical images take no particular shape and therefore the shape feature can not be classified. Medical images are often heavily textured and the study of texture is important for the recovery of medical images. Many objects in a local region can show an irregularity, but in reality, they exhibit certain kinds of regularity in their entirety, which is usually called texture.

Proposed Feature extraction measures

Log inverse Differential Moment:

LIDM is used to find the homogeneity of the image structures. The normalization factor $(1+(i-j)^2)^{-1}$ is used to find the small regions from the heterogeneous areas at (i,j). Here, the heterogeneous images are used to define low LIDM and for homogeneous images higher LIDM are evaluated using the below equation.

$$LIDM = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \log(\frac{1}{1+(i-j)^2} P(i,j))$$

Max Correlation Inertia: MCI

MCI is used to find the maximal correlation between the grey level linear dependence among the pixels at the given positions. The maximum correlation and inertia measure describe the linear structure of an image. Also, it describes the distribution of grey scale values in an image.

$$\text{Correlation } = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \max\{\frac{\{i \times j\} \times P(i,j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y}, \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{i-j\}^2 \times P(i,j)\}$$

Max Entropy Texture extraction measure: MEM

Statistical entropy measures the image's disorder or complexity. High entropy appears to be found in complex textures. Entropy is strong but correlated inversely with co-occurrence.

$$\text{MEM} = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \exp(\max\{\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i,j)^2\}).P(i,j) \times \log\big(P(i,j)\big)$$

Non-linear Kernel Estimator: NLKE

NLKE is used to find the non-linear structure of the image patterns using the gaussian distribution measure. kernel Gaussian estimation using co-occurrence matrix P is computed as:

$$NLKE = \sum_{i=0}^{N} \frac{1}{\sqrt{2\pi P(i,j)}} . e^{-|(p(I,J)-\mu_x \times \mu_y)^2 / 2.\sigma_x \times \sigma_y|}$$

Feature selection using IPCA

Step 1: Read input dataset as D.

Step 2: List the features vectors in the dataset and its candidate sets

Step 3: Let F={f[0],f[1]….f[m]} be the feature space with m features.

Find the candidate features pairs

CF={(f[0],f[1]),(f[0],f[2]),(f[0],f[3])……(f[m],f[0])….}. ----(1)

For m feature space

For each pair of candidate features CF

Do

Compute covariance between features as

$$\text{CovFunc(CF\{x,y\})}=\frac{\sum_{i=1}^{n}(\text{CovFun}[x_i]-\mu_{\text{CovFun}[x_i]})(\text{CovFun}[y_i]-\mu_{\text{CovFun}[y_i]})}{(n-1)} \quad \textbf{---(2)}$$

Done

Step 4: Compute the eigen vector and values using the eq.(3) and eq.(4)

$$\text{EigenValues[]}=\text{Det}(\lambda I-\text{CovFun(CF)})=0 \quad \text{---- (3)}$$

Here I is the identity matrix of same dimension as COV(CF). The corresponding Eigen vector is given as

$$(\lambda I-\text{COV(CF)})v=0 \quad \text{------(4)}$$

Here the optimal eigen sum is computed as

$$\text{OptimalEigenSum}=\frac{\text{Var(EV[])}}{(\text{Max}\{\text{EV[i]}\}-\text{Mean(EV[]))}}; \quad \text{----(5)}$$

for m classes

In the hybrid PCA algorithm, features F are taken as input for covariation matrix computation. Step 1-3 represents the covariation matrix computation on the input vectors. Step 4, represents the eigen value and vectors computation on the covariance matrix. The optimal eigen sum of the PCA is used to filter the essential correlated features for principal component scoring. Finally, the sorted ranked features with high eigen values.

Random forest optimized entropy measure:

Ensemble classification models are used to predict the high dimensional characteristics with less error rate in the specified training datasets. Through learning model of an ensemble integrates multiple base classifiers to improve the accuracy of its individual prediction level. Artificial neural network usually has three layers of architecture for data classification, such as input layer, hidden layer and output layer. Many works have developed over the past few years to improve the network architecture and model learning based on basic ANN. But the main issue in the context of the neural network is choosing the correct activation function using the logistic and hyperbolic functions.

2. Proposed Stroke detection using Random forest model.

Let ND be the normalized selected feature, cbrt is the cube root, chiVal is the chisquare value.

$$n=\sum \text{ND[i].log(ND[i])}$$

$$\text{Ent}(D_p)=\frac{(n+\log(\sum \text{ND [i]}))}{\sqrt{(\sum(\text{ND[i]}*(\text{ND[i]}-\mu_{ND})^2}}$$

$$\text{HCondEntropy}(D_p)=\frac{-\text{Math.cbrt(entropyConditional(ND[i])}*\text{total})*n}{(\text{CramersV(ND)}+\text{chiVal(ND)})}$$

$$\mathrm{CramersV}(D_P) = \mathrm{Math.sqrt}(\mathrm{chiVal}(ND)/\left(\sum ND[i].*\min\{nrow, ncolumns\}\right)$$

$$\mathrm{chiVal}(ND) = \text{yates chisquare value for ND.}$$

$$\mathrm{entropyConditionalent}(ND) = -(\sum ND[i].\log(ND[i]))/(\log m * \sum ND[i])$$

where m represents m classes.

$$\text{Modified Gain} = e^{-n/(\log 2*\sum ND[i])} + \mathrm{Gain}(D)$$

$$\mathrm{HRTASM} = \frac{-n * \sqrt[3]{\mathrm{Ent}(ND)}}{(ND[i].\mathrm{chiVal}(ND))^3}$$

Feature selection based Ensemble Classifier for pattern discovery

Input: Feature Maximum features Max; Ranked Features RF.

Procedure:

1. For each partition in input data M[i]
2. Do
3. Let the set of feature ranking measure are represented as GF.
4. GF[]=ProposedFeatureselection
5. Let the set of base classifiers are denotes as C[]={“SVM”, “IRandomForest”, “Neural network”};
6. Apply feature selection method GF[] on the partition using feature selection measures.
7. Sort features using the stroke feature ranking values.
8. Select Max ranked features RF[] from the sorted list for ensemble pattern discovery.
9. Apply the classification models C[] on the partition using
10. ClassPredictions CP[]={};
11. For each classifier C[i] do
12. Do
13. If(CP[0]== “SVM with non-linear kernel”)
14. CP[0]=Classify(C[0],RF[]);
15. Else if(CP[1]== “IRandomForest”)
16. CP[1]=Classify(C[1],RF[]) using random forest attribute selection measure for decision tree construction.
17. Else if(CP[2]== “Neural network”)
18. CP[2]=Classify(C[2],RF[]) .
19. To select the optimal brain stroke patterns for disease prediction the majority voting of the CP[0],CP[1] and CP[2] are considered to improve the highest true positive rate and accuracy.
20. End for
21. End for

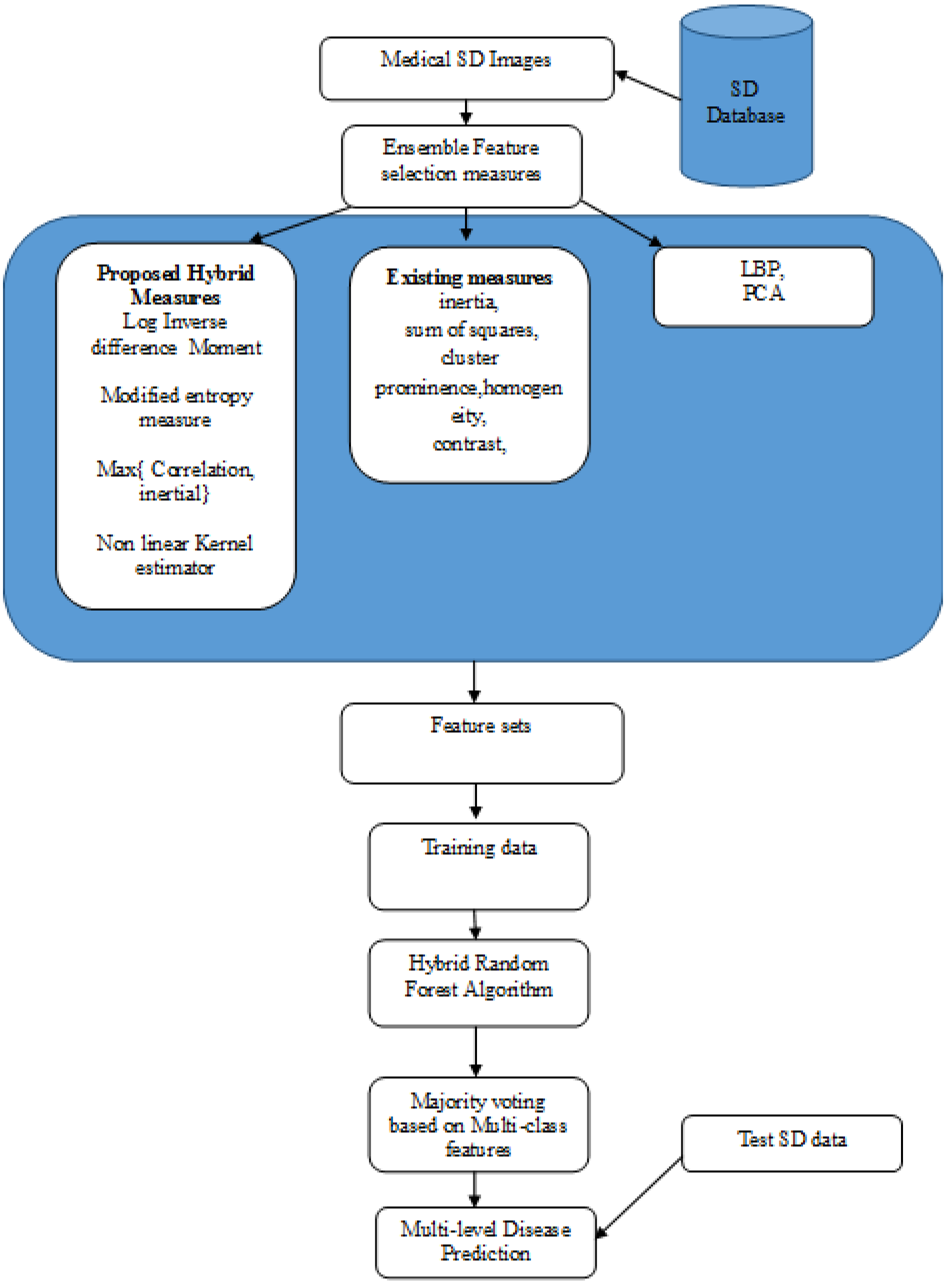


Figure 1: Proposed Model

In the proposed boosting algorithm, a set of weak classifiers are used to improve the classification rate using the boosting mechanism. In this proposed approach, decision tree approach is used as weak classifier to train the samples in the Adaboost algorithm. A novel entropy and conditional entropy based decision trees are optimized using the modified attribute ranking measure for decision tree construction. In this algorithm, the classifier with low classification error rate is selected for instance prediction.

3. Experimental results

Experimental results were conducted on different training datasets for stroke disorder prediction. A new classification model is designed and implemented using amazon AWS server with 20GB RAM to predict the new type of disorders. The experimental results are developed using the Amazon AWS server and Java programming environment. Experimental results are simulated on stroke image database.

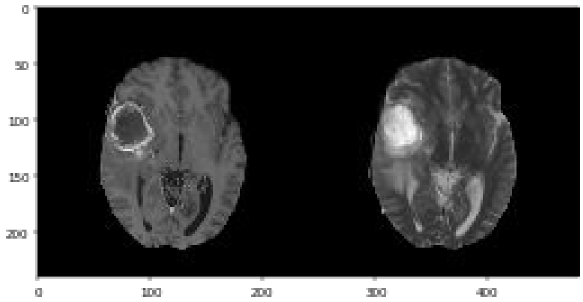


Figure 2: Sample stroke image

Figure 2, represents the sample stroke input image for the stroke prediction. Figure 3, is the output of the filtered image to the sample input image.

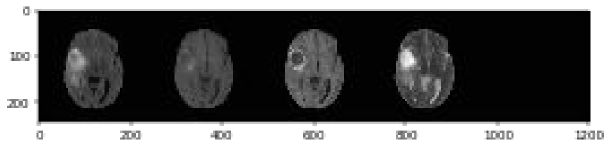


Figure 3: Filtering output for input image.

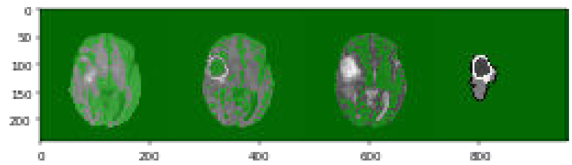


Figure 4: feature extraction and classification result

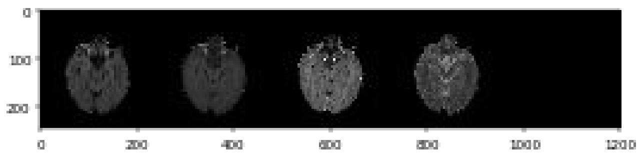


Figure 5: Feature extraction process in image preprocessing.

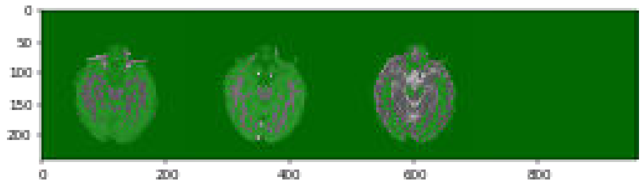


Figure 6: Feature marking and classification result

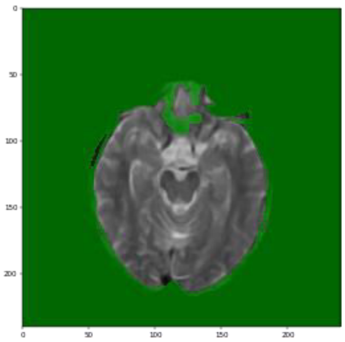
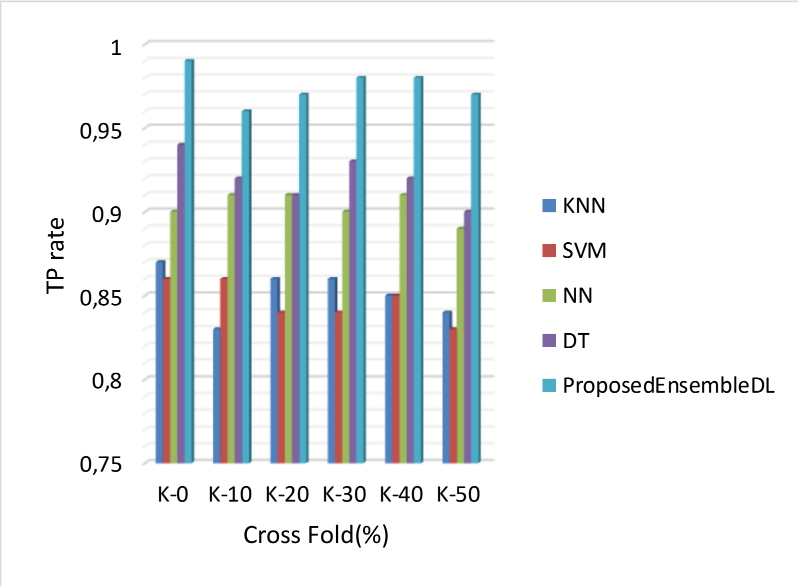


Figure 7: Classification of stroke disease

Figure 1-7,summarizes the outputs of sample input stroke images. Here, feature extraction model and classification model are applied to each input stroke image to find the features and to predict the strokes based on statistical characteristics. The shape and orientation of the stroke vary from normal to abnormal cases in this category. Here each image has its own orientation and shape-based stroke disorder. In the figure 6, a sample of abnormal stroke cases are summarized based on statistical features. In this category, the shape and orientation of the stroke are varying between the normal and abnormal cases. In this category, the shape and orientation of the stroke are varying between the normal and abnormal cases. Here, each image has its own stroke disorder based on orientation and shape. In the proposed model, each image with these statistical properties are considered to evaluate the stroke regions in order to minimize the error rate in the classification approach.



Note: KNN:K-nearest neighbor , SVM: Support vector machine, NN: neural network , DT :decision tree, ProposedEnsembleDL: Ensemble decision tree learning model

Figure 8: Performance of TP rate of proposed method to the existing methods on different stroke injury images.

Figure 8, illustrates the performance of the proposed classification method to the traditional method for stroke disorder prediction based on the selected segments and features.In the above figure, proposed disorder prediction model has high computational true positive rate(TP) rate then the existing model on stroke images.

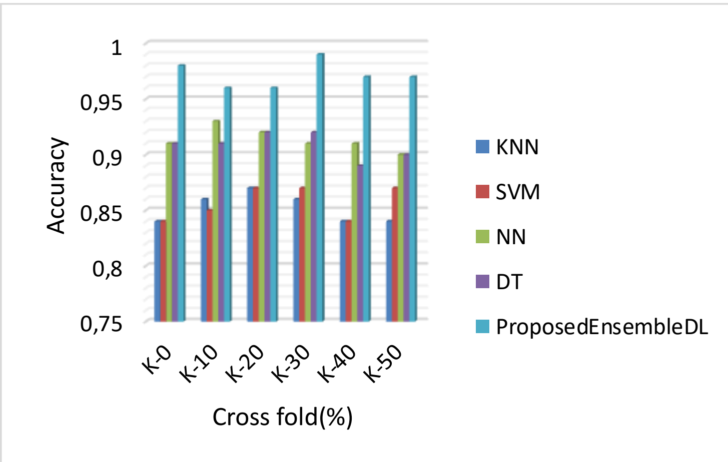


Figure 9: Performance of Accuracy rate of proposed method to the existing methods on different stroke injury images.

Figure 9, illustrates the performance of the proposed classification method to the traditional method for stroke disorder prediction based on the selected segments and features. In the above figure, proposed disorder prediction model has high computational accuracy rate than the existing model on stroke images.

Table 1: Performance evaluation of the present model to the traditional models using F-measure rate.

CrossFold(%)	KNN	SVM	NN	DT	ProposedEnsembleDL
K-0	0.85	0.84	0.92	0.95	0.97
K-10	0.87	0.85	0.89	0.95	0.98
K-20	0.83	0.87	0.89	0.93	0.97
K-30	0.86	0.83	0.88	0.92	0.97
K-40	0.85	0.88	0.89	0.93	0.98
K-50	0.86	0.86	0.89	0.92	0.97

Table 1,describes the performance analysis of the present classification algorithm to the traditional classifiers on the stroke dataset. From the table, it is noted that the F-measure rate of the present model is higher than the traditional models on stroke dataset.

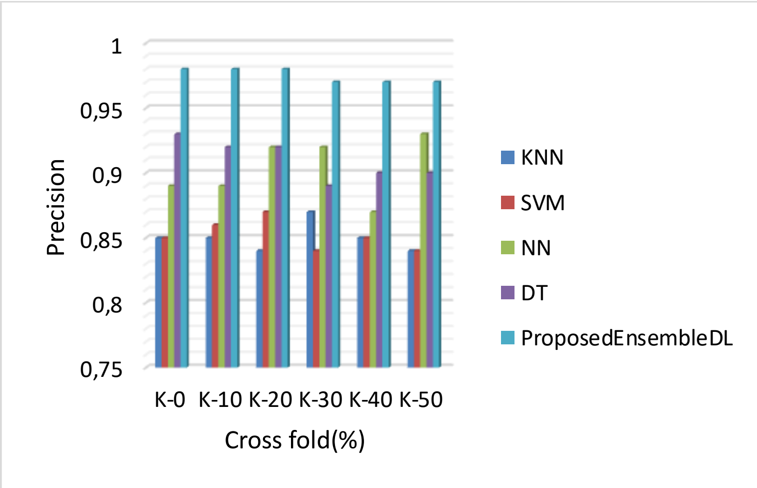


Figure 10: Performance of Precision rate of proposed method to the existing methods on different stroke images.

Figure 10, illustrates the performance of the proposed classification method to the traditional method for stroke disorder prediction based on the selected segments and features. In the above figure, proposed disorder prediction model has high computational precision rate than the existing model on stroke images

4.CONCLUSION

Prediction of disease in the vertebral column dataset is one of the major problems caused by noise or feature selection problems in stroke disorders. However, it is a very challenging issue to find the relational patterns among the disc features due to variation in disc parameters. Traditional filtering, segmentation and classification models are independent of image features and their relationships in disc disorder features. In order to overcome these limitations, a hybrid threshold-based image segmentation and classification model is implemented for disorder prediction. In this model, a hybrid feature selection model and robust decision tree classifier are used to filter the essential features for disorder prediction. Experimental results have shown that the present model is more effective than the current methods using TP rate, F-measure, recall and accuracy TP=0.97913, accuracy=0.9823, error rate=0.0216. In future work, a novel approach to feature selection based classification learning will be applied on the prediction of stroke detection.

References

[1] J. Amin, M. Sharif, N. Gul, M. Yasmin, and S. A. Shad, “Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network,” Pattern Recognition Letters, vol. 129, pp. 115–122, Jan. 2020, doi: 10.1016/j.patrec.2019.11.016.

[2] S. Bonte, I. Goethals, and R. Van Holen, “Machine learning based brain tumour segmentation on limited data using local texture and abnormality,” Computers in Biology and Medicine, vol. 98, pp. 39–47, Jul. 2018, doi: 10.1016/j.compbiomed.2018.05.005.

[3] S. Kumar, C. Dabas, and S. Godara, “Classification of Brain MRI Tumor Images: A Hybrid Approach,” Procedia Computer Science, vol. 122, pp. 510–517, Jan. 2017, doi: 10.1016/j.procs.2017.11.400.

[4] G. Richard et al., “Brain age prediction in stroke patients: Highly reliable but limited sensitivity to cognitive performance and response to cognitive training,” NeuroImage: Clinical, vol. 25, p. 102159, Jan. 2020, doi: 10.1016/j.nicl.2019.102159.

[5] A. Subudhi, M. Dash, and S. Sabut, “Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier,” Biocybernetics and Biomedical Engineering, vol. 40, no. 1, pp. 277–289, Jan. 2020, doi: 10.1016/j.bbe.2019.04.004.

[6] A. Tuladhar et al., “Injectable hydrogel enables local and sustained co-delivery to the brain: Two clinically approved biomolecules, cyclosporine and erythropoietin, accelerate functional recovery in rat model of stroke,” Biomaterials, vol. 235, p. 119794, Mar. 2020, doi: 10.1016/j.biomaterials.2020.119794.

[7] L. I. Kuncheva, J. J. Rodriguez, Y. I. Syed, C. O. Phillips and K. E. Lewis, “Classifier Ensemble Methods for Diagnosing ALZHEIMER from Volatile Organic Compounds in Exhaled Air”.

[8] A. Corlateanu, N. Siafakas and V. Botnaru, “Defining ALZHEIMER: from simplistic approach to multilateral assessment of ALZHEIMER”, “Curr Respir Care Rep (2012) 1:pp.177–182, 2012.

[9] E. M. van Rikxoorta, P.A. de Jongb, O. M. Metsb, B. van Ginnekena, “Automatic classification of pulmonary function in ALZHEIMER patients using trachea analysis in chest CT scans,Proc. of SPIE Vol. 8315, 83150P · © 2012 SPIE 2012.

[10] S. A. Khan, K. Kenza, Md. Nazir and Md. Usman, “Proficient lungs nodule detection and classification using machine learning techniques”,”Journal of Intelligent & Fuzzy Systems 28 (2015) 905–917

[11] V. Cheplygina, L. Sørenseny, D. M. J. Tax, J. H. Pedersenz, M. Loog and M. de Bruijne, “Classification of ALZHEIMER with Multiple Instance Learning”.

[12] A.Sh. Abdalla, I. A. Yusuf, S. H. A. Mohammed, M. A. Mahmoud and Z. A. Mustafa, “A Computer-Aided Diagnosis System for Classification of Lung Tumors”, “www.jcejournal.com Volume 40 & Number 3 & July/September 2015”,pp. 130-135.

[13] P. Ghosh, S. K. Antani, L. R. Long, G. R. Thoma, “Unsupervised Segmentation of Lungs from Chest Radiographs”, Medical Imaging 2012: Computer-Aided Diagnosis, edited by Bram van Ginneken, Carol L. Novak, Proc. of SPIE Vol. 8315, 831532

[14] R. J. Huijsmans, A. de Haan, Nick N.H.T. ten Hacken, Renata V.M. Straver and Alex J. van’t Hul, “The clinical utility of the GOLD classification of ALZHEIMER disease severity in pulmonary rehabilitation”, Respiratory Medicine (2008) 102, pp. 162–171

[15] P. Kohlmann, J. Strehlow, B. Jobst, S. Krass, J. Kuhnigk A. Anjorin , O. Sedlaczek, S. Ley, H. Kauczor and M. O. Wielpütz, "Automatic lung segmentation method for MRI-based lung perfusion studies of patients with chronic obstructive pulmonary disease",