ISSN 2278-3091
Volume 8, No.6, November – December 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse151862019.pdf
https://doi.org/10.30534/ijatcse/2019/151862019

Diagnosis of Brain Hemorrhage Using KNN based Radial Basis Classifier

***Jathin Desan[1]**
[1]11th grade Independence High School, Frisco, Texas- 75035, United States, jathindesan95@gmail.com

ABSTRACT

There occurs a peak interest in medical diagnosis with the aid of machine learning. One of the common old age disease is the Alzheimer disease and it is significant for accurate diagnosis of Brain hemorrhage in AD patients on the basis of image processing and machine learning approaches. Brain hemorrhage is the leading death cause between the age group of fifteen and twenty four. Lifesaving of such kind of patients is crucial and totally depends on the identification of the exact location and hemorrhage type in very early stage. The proposed method designed a fully automated machine learning approach for the classification by the novel KNNRBF method of disease with respect to its severity levels (Brain hemorrhage, PICS and Alzheimer disease) even in the absence of an expert. The performance analysis of the proposed system with respect to the performance measures like accuracy, specificity, sensitivity, Jaccard coefficient, kappa and dice coefficient respectively. The future work deals with the implementation of the model in deep learning and in realtime as a radiology optimization tool.

Key words: Brain hemorrhage, PICS, Alzheimer disease, machine learning, KNN.

1. INTRODUCTION

Disease diagnosis and its early detection is now becoming a main research aspect with the aim of reducing the disease progression. The structural image gives detailed information regarding the structural features like shape, volume, size and anatomy of the disease. Among the brain related disease Alzheimer disease is very common in old age people [1]. It is necessary to accurately diagnose the Brain hemorrhage in AD patients on the basis of image processing and machine learning approaches [2]. Most powerful and recent classification method were need to be investigated with the aim of enhancing the accuracy of the diagnosis. Computed tomography, an alternative rapid to MRI imaging utilizes X-Rays with the help of computer processing for the generation of 3-D images and cross sectional observations of the brain. The paper mainly deals with the image processing of Brain hemorrhage that occurs due to bleeding around the tissues of the brain. The bleeding in case is caused as a result of artery rupture. In this case when trauma blood disturbs the tissues of the brain, it forms swelling leading to cerebral edema. On the basis of pressure and bleeding region. BH have been classified as subdural hemorrhage, subarachnoid hemorrhage, intra cerebral hemorrhage and Extradural hemorrhage. In CT imaging, the rotation of X-ray tube around the patients head and captures several images. The CT images need expertized system for precise investigation and analysis.[3, 4] The machine learning algorithms which is suitable to train large number of multi-layer neural networks is engaged for classification of various automated tasks. In this study preprocessing by means of Gaussian Filtering, Segmentation with K-means clustering, Feature extraction by GLCM and DWT followed by Classification with the novel KNNRB classifier. This process of classification outperforms the existing models and enable effective diagnosis by the radiologists.

1.1 Objectives

- To classify the extracted features of Brain image with the robust KNNRBF classifier.
- To develop a prospective machine learning system capable of detecting Blood Hemorrhage on a large database of CT studies.
- To analyze the severity levels on the basis of the characteristics of the affected region.

1.2 Organization

The continuing subdivision of this paper is prearranged as surveys; Section II clarifies the related works on the segmentation and classification technique. Section III illuminates the implementation of the proposed NMSVM classifier. Section IV presents the experimental results of the proposed data collection. Section V shows the performance analysis of the proposed approach. Finally, section VI concluded the proposed NMSVM classifier technique remarks, respectively.

2. RELATED WORKS

The following section comprises of the related works.
[5] The abnormality identification and efficient BH recognition in MRI with the help of NBPKC was suggested in this paper. The work conducted a preprocessing followed by binary thresholding that applied an image an image mask over the affected region. A modified segmentation algorithm have

been implied with the use of GLCM and minimal LBP(local binary pattern).Here classification have been performed by NB-PKC that precisely determines the position thereby decreased the human errors.[6] The suggested technique implemented statistical features, SVM and Gabor texture. The MRI images was classified into degenerative, inflammatory, cerebrovascular, normal and and neoplastic. It was trained on a full dataset of Brain Atlas Harvard Medical School. Apart from that a locally developed dataset for achieving high accuracy have been utilized. It leads to the further classification of tumorous slices into primary and secondary tumor followed by other such types like sarcoma, glioma, bronchogenic carcinoma, meningioma and adenocarcinoma. It should be noticed that the mentioned subtypes were not determined without biopsy. [7] Detecting the type of haemorrhage is very crucial step in the medical treatment to save life of the patient.In this study the segmenting and quantifying the region were performed on the basis of the watershed algorithm (WSA).The computation of background and foreground markers have been performed before the application of WSA. The ANN would reduce the output error that enabled the BH detection efficiently. This framework helps the doctors and radiologist for accurate verification. The system used in this suggested allows even the non-technical user to process the setup after initialization. It would be accurate and simple for BH diagnosis.[8] The paper detects the existence of BH in CT brain scan with machine learning approach. With the suggested approach, the 100% existence or absence of BH was obtained followed by the achievement of ninety two classification accuracy. [9] The study attempted for the brain CT image segmentation with the population and distribution of pixel intensity. The work designed a mathematic set up for the identification of unpredicted variation found in the intensity population in hemorrhage CT image. Here the accuracy of suggested method was evaluated by segmenting various kinds of hemorrhage in carious patients. The system utilized a completely automatic and speedy CAD for automatic segmentation in the expert absence thereby further checking of volume, shape, severity, shape, size and the hemorrhage types were performed. The CAD competence was tested with frequently [10] The study indicated that ML algorithm enhanced the prediction of DCI in a population of aSAH patients.The work enforced three main approached for the prediction of DCI development in aSAH patients. The first one is the use of predictors known from logistic regression and the next, with the use of ML algorithm with all the possible variables followed by integration of clinical and imaging data.[11] The prescribed study tested and trained a ML alogorithm of KDSQ for the differentiation of cognitive normal and impaired with CREDOS data that was suggested as screening tool of CI.The integrated combination of KDSQ and MMSE yielded high accuracy and proved that the suggested system was more efficient screening tool [12]. The author employs an automatic BH detection and classification model on the CT images. After the process of preprocessing, MDRLSE (a modified version) have been utilized for the separation of hemorrhage areas. A flawless texture and shape feature set from all the identified hemorrhage areas have been extracted. Further, a synthetic feature named as weighted grayscale histogram was defined. The work classified the BH into four types like ICH, EDH, IVH and SDH by means of ordered classification structure. The suggested algorithm was assessed on a perfect CT scan image set and obtained high range of slightly varying accuracy for the two classifiers respectively. [13] The author estimated the asymmetry on the basis of intensity variation found between the right and left of mid sagittal plane. Here a novel technique on the basis of low intensity IF region for solving the problem when the object is tilted or rotated have been employed. It employed K-means clustering algorithm with concerned component label for the determination of size and location of tumor. The result depicted that effectiveness of the suggested method outperforms with respect to consumption time and computational cost. [14] The plan focus on finding and analyzing the brain stroke present in MRI images of various patients. Few preprocessing methods such as removal of noise, filtering and segmentation have been utilized for extraction of brain stoke. The brain segmentation was applied with the use of Fuzzy C-means clustering in which the edge detection was utilized for the accurate identification of segmented region in the brain stroke edges. The results prove that the suggested technique effectively helps the doctors in decision making in a short time. [15] presented a new prototype by combining various text mining tools with different machine learning techniques for classifying stroke. Various machine learning approaches were trained in a suitable manner and were utilized in important regions like the security surveillance, management of data and mainly in the medical field. Along with the machine learning approach, data mining technique was also utilized in the proposed work for tracking the informations depending upon the syntactic and semantic point of view. Various symptoms of stroke condition in the patients were mined accurately from the information obtained from the patients' case sheets and the obtained data were trained accordingly. The data were gathered from Sugam Multispecialty Hospital in the Kumbakonam district and case sheets of 507 patients were taken into consideration. The common and very unique set of features required for classifying the symptoms of strokes were extracted. By using different machine learning techniques like ANN, SVM, random forest classification and boosting and bagging, the processed data were classified. By comparing the performances of all these classifiers, ANN trained by using the stochastic gradient descent algorithm exhibits enhanced performance of increasing the accuracy to about 95% and decreasing the standard deviation to value of 14.69. [16] The study implemented ML and computer vision techniques for developing a system that detects CT-identifiable lesions automatically. It employed SIFT and CNN for feature identification that differentiates TBI lesions from the available background. The paper engaged TDA to reveal the correlation found between pathological and healthy data. It manually segmented the large lesions found in every positive scans. High accuracy indicating the prospective clinical use have been achieved with eighty percent training data and twenty percent testing data. [17] Various medical imaging techniques like MRI, CT scan etc. can be utilized for detecting the brain images. However brain hemorrhage could not be properly diagnosed by these imaging techniques, due to the influence of some factors like noise removal, accuracy and high segmentation. Therefore neural networks were utilized in the proposed work as it is comparatively advanced by means of robustness, accuracy and speed. Intelligent computing approaches and its applications were listed in the survey section. Numerous computing techniques were discussed both with their merits and demerits. Finally enhanced delivery of healthcare with cost efficiency and improved structure was obtained for the ever growing medical systems.

3. PROPOSED WORK

The overall flow of the proposed system have been depicted in the figure 1. The input in the form of brain computed tomography image is subjected to preprocessing with the aid of Gaussian Filter. The filter will effectively blur the edges and reduces the contrast thereby decreasing the noise. After that the preprocessed image is segmented by K-means clustering that easily processes a huge number of variables in a short time. It easily adapts to the various cluster types with respect to size and shape. It processes by means of first selecting the number of clusters followed by randomly setting the initial cluster center. The image object were brought closer to the cluster center as far as possible for effective segmentation. After the affected region was segmented the texture features were extracted through GLCM feature extraction and Discrete Wavelet Transform. The extracted text features were classified with the help of robust proposed classifier named KNNRBF that use the probability (with the implementation of KNN classifier) as bias value for Radial basis classifier. The obtained results were processed for the performance analysis in terms of accuracy.

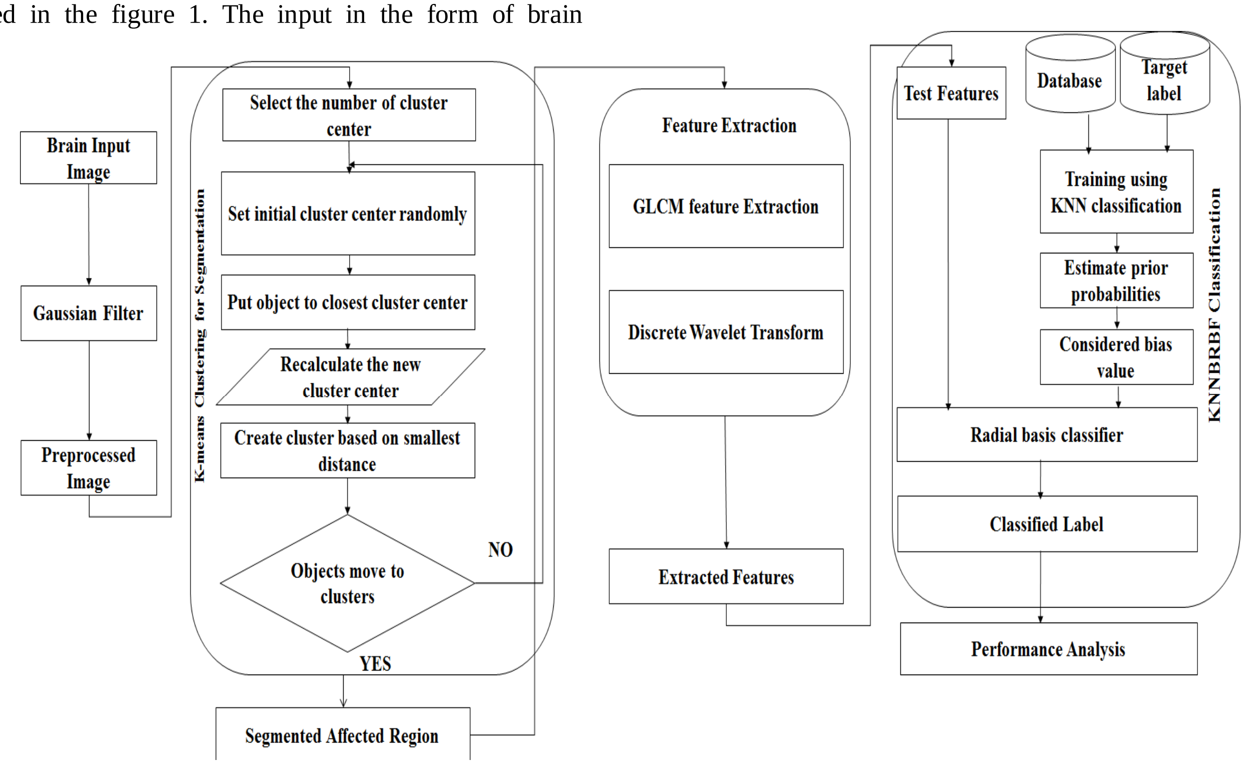

Figure 1: Overall flow of the proposed (kNNRBF) classifier system

3.1 Gaussian Filtering

The Gaussian filter is a type of image filtering that uses a Gaussian function (which also expresses the normal distribution in statistics) for calculating the transformation to apply to each pixel in the image.

$$G(i,j) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{i^2+j^2}{2\sigma^2}} \qquad (1)$$

Where i is the distance from the origin in the horizontal axis, j is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution.

3.2 Segmentation by k-means clustering

K means clustering algorithm is an unsupervised algorithm that detects more number of segments in the data. At first the number of the clusters K were selected. The k points are selected randomly from the concerned dataset and all the data points are assigned to the nearby centroid which forms the centroid. Finally computation followed by placing the new centroid of all the cluster have been done. The parameter k is used for deciding the number of segmentations. Specifying a larger k-value leads to a cluster of pixels in the space to be segmented into smaller pieces of points groups. The k means iteratively explored the centroid of all the clusters followed by pickuping and reorganizing the nearby points into a new group.

3.3 Feature Extraction

a) GLCM

The Gray level co occurrence Matrix extracts the texture features with the use of connected regions. The features thus obtained will be given to a classifier for training and testing.The features such as Auto correlation, Correlation, Entropy, Sum of entropy, Sum of variance , Cluster prominence, Contrast, Homogeneity, Maximum probability were extracted.Few of the formulas are depicted below.

a) Entropy - The irregularities found in the gray level image size with some variation in each co-occurrence matrix elements is characterized by the entropy. The equation for this measure is shown below:

$$\text{Entropy} = \sum_{s=1}^{M}\sum_{d=1}^{N}\left(I_{seg}(s,d) * \log\left(I_{seg}(s,d)+\varepsilon\right)\right) \quad (2)$$

where ε is a constant and its value is $2.2204*10^{-16}$

Energy - It is represented as the total image color uniformity. It is measured by,

$$\text{Energy} = \sum_{a=1}^{M}\sum_{b=1}^{N}\left(I_{seg}(a,b)\right)^2 \quad (3)$$

b) Discrete Wavelet Transform.

Here the segmented input image is separated into two sub-bands such as high frequency (H) and low frequency (L) and Further it is segregated into High-Low (HL), low-low (LL), High-high (HH) and Low-High (LH) sub-band coefficients with the use of Discrete Haar Wavelet Transform.

c) KNNRBF Classifier

KNNRBF Algorithm

Input: Train features f, test features t_f and target label Ta

- Calculate distance between each samples in the trained features with the nearest feature data points, $d_i = \sqrt{(f_i - f_j)^2}$ i =1, 2, ….., **n**; where **n** represents the number of samples and d_i denotes the Euclidean distance between the points,f_i is the traine feature of i[th] data points.
- Transform the distance d_i to a probability using logistic regression with optimal coefficients β_0 and β_1.

$\Pr(d) = \frac{1}{1+\exp(\beta_0+\beta_1 d_i)} \; \epsilon \; [0.1]$ (4)

- Use this probability as bias value for Radial basis classifier,

Radial basis classifier probability is expressed by Gaussian distribution as follows,

$p(f - c_i) = \exp(-\beta\|f - c_i\|^2)$

c_i - Center vector for neuron i

β – is the bias value which is considered from the probability of KNN

- Estimate weights with the test features and target label.

$$W = (ff^T)^{-1} t_f T a^T$$

- Finally predict the label.

$Cl = Ta(\max(W))$

4. DATASET DESCRIPTION

The dataset of 82 CT scans was collected, including 36 scans for patients diagnosed with Brain hemorrhage. Each CT scan for each patient includes about 30 slices with 5 mm slice-thickness. The mean and standard deviation of patients' age were 27.8 and 19.5, respectively. 46 of the patients were males and 36 of them were females. Each slice of the non-contrast CT scans was by two radiologists who recorded AD, PICS and Brain hemorrhage.

5. PERFORMANCE MEASURES

The performance of the proposed system was investigated in this section.

The performance measures described in this section was provided in detail. The performance analysis section investigates the results of the proposed technique with respect to accuracy, recall, precision, Dice, Jaccard, and Kappa coefficients. The performance analysis of the proposed was compared with the existing techniques and proved that the proposed system outperforms in terms of efficiency.

Sensitivity is termed as the as the fraction of positives which are appropriately recognized or also named as true positive rate.

$Sen = \frac{TP}{TP+FN}$ (5)

Specificity is well-defined as which estimates the fraction of negatives that are appropriately recognized or also called as true negative rate.

$Speci = \frac{TN}{TN+FP}$

Accuracy is defined as the nearness of a measured value to the known or standard value.

$Acc = \frac{TP+TN}{P+N} \; or \; \frac{TP+TN}{TP+TN+FP+FN}$

Jaccard Coefficient

Jaccard Coefficient (JC) is the measurement of the uneven information in a binary variable. Jaccard coefficient is calculated while considering matching items the customer buy in a store, there are numerous products that the customer does buy. In which the negative value is not essential and counting the non-existing value is not needed. The coefficient is calculated as:

$$JC_{ij} = \frac{a}{a + b + c}$$

Dice Coefficient

The Sorensen–Dice is otherwise termed as the F1 score or Dice Similarity Coefficient (DSC). It is commonly used in retrieving information using the statistics such as precision (p) and recall (r). Dice Coefficient is given by:

$$DC = 2.\frac{precision\ * recall}{Precision + recall}$$

6. PERFORMANCE ANALYSIS

The performance of the proposed system was investigated in this section.

The performance measures described in this section was provided in detail. The performance analysis section investigates the results of the proposed technique with respect to accuracy, recall, precision, Dice, Jaccard, and the “Kappa” coefficients. The performance analysis of the proposed NMSVM with the existing KNN and PNN classifier regarding FAR, FRR, and GAR is done for CT images respectively.

Table 1: Comparison of performance value of the proposed system in Alzheimer diseases, PICS and healthy individuals

	Alzheimer	**PICS**	**Healthy**
Accuracy	97.894	90.2345	89.2364
Sensitivity	95.311	87.345	83.4651
Specificity	93.3215	89.6521	87.1354
PPV	78.654	75.1324	76.4132
NPV	65.4123	62.16541	67.1463
Jaccard	0.9124	0.8532	0.8481
Dice	0.8987	0.8589	0.7798

Table 1 describes the comparison of proposed system performance value in Alzheimer diseases, PICS and healthy individuals. The performance measures are accuracy, sensitivity, specificity, PPV, NPV, Jaccard and Dice co-efficient. In Alzheimer disease the proposed method contains the accuracy of 97.89%, 90.23% for PICS condition and 89.24% for healthy individuals.

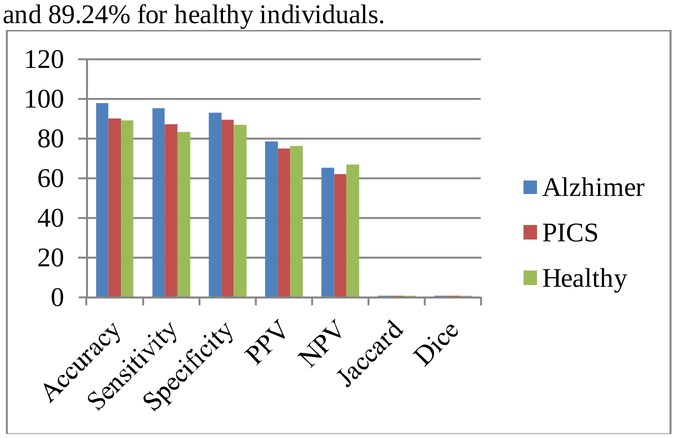

Figure 2: Comparison of performance value of the proposed system in Alzheimer diseases, PICS and healthy individuals

Figure 2 defines the statistical analysis of proposed system measures in Alzheimer diseases, PICS and healthy individuals with accuracy, sensitivity, specificity, PPV, NPV, Jaccard and Dice co-efficient.

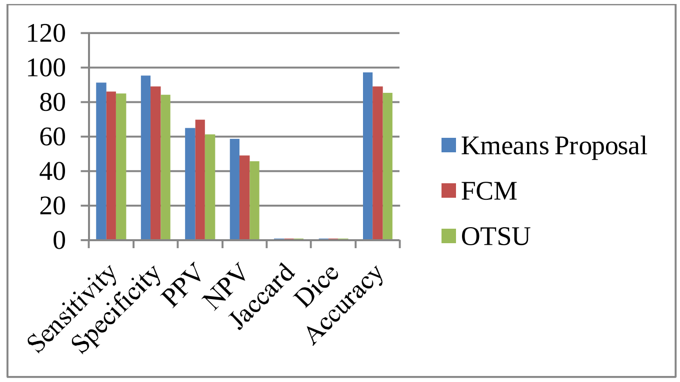

Figure 3: Segmentation comparison of proposed system with existing methods

Figure 3 defines the comparison of proposed system with existing methods such as FCM [18], OTSU [19]. Proposed system had the accuracy of 89.22%, sensitivity contains 86.24%, and specificity contains 89.21%. These analyses proved the proposed method had highest segmentation accuracy compared than existing methods.

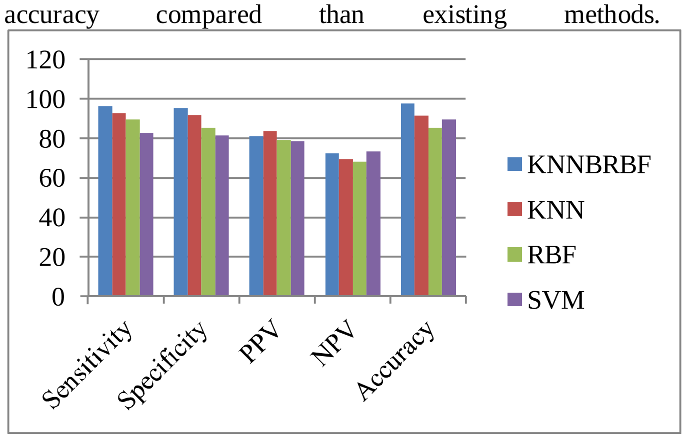

Figure 4: Classification comparison of proposed system with existing methods

Figure 4 defines the comparison of proposed system with existing classification methods such as KNN, RBF [20] and SVM [21]. Proposed system had the accuracy of 97.56%, sensitivity contains 96.35%, and specificity contains 95.46%. These analyses proved the proposed method had highest classification accuracy compared than existing methods.

Table 2: Comparison of jaccard and Dice co-efficient measures for proposed method with existing methods

	KNNBRBF	KNN	RBF	SVM
Jaccard	0.9575	0.8548	0.8789	0.7981
Dice	0.9789	0.8123	0.7164	0.8754

Table 2 describes the comparison of jaccard and Dice co-efficient measures for proposed KNNBRBF method with existing methods such as KNN, RBF, and SVM.

Table 3: Accuracy of Testing and Training in proposed classification

	Testing Accuracy	Training Accuracy
1-fold	72.00%	95.00%
2-fold	73.20%	94.56%
3-fold	75.25%	93.65%
4-fold	78.45%	92.58%
5-fold	79.78%	91.56%
6-fold	81.12%	89.87%
7-fold	85.43%	88.56%
8-fold	87.75%	87.62%
9-fold	91.89%	86.45%
10-fold	93.13%	85.21%

Table 3 defines Accuracy of proposed method classification with 10-fold crosses Validation. The classification accuracy involves the testing and training accuracy.

7. CONCLUSION

The paper implements machine learning approach for effective diagnosis of Brain hemorrhage in AD patients with respect to its severity levels. The order of severity levels of Brain Hemorrhage (PICS, AD patients) and healthy individuals were diagnosed and compared. The preprocessing is performed through Gaussian Filtering, segmentation by k-means clustering, feature extraction by GLCM and DWT followed by the novel proposed KNNRBF classified the extracted features with K-nearest neighbor based Radial basis function (KNNRBF) which is used to classify the input image as healthy, pics and Alzheimer. Finally the performance of the proposed segmentation and classification with state-of-the-art methods (KNN, RBF and SVM) by some metrics such as sensitivity, specificity, accuracy and similarity metrics, etc.

REFERENCES

1. J. E. Villanueva-Meyer, *et al.*, "Machine learning in neurooncology imaging: from study request to diagnosis and treatment," *American Journal of Roentgenology,* vol. 212, pp. 52-56, 2019. https://doi.org/10.2214/AJR.18.20328
2. D. Shivananda and G. A. Attimarad, "Detection of Exudates and Microaneurysms in Diabetic Retinopathy."
3. M. R. Arbabshirani, *et al.*, "Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration," *npj Digital Medicine,* vol. 1, p. 9, 2018/04/04 2018.
4. A. S. Kumar, "ENSEMBLE ONLINE SEQUENTIAL EXTREME LEARNING MACHINE AND SWARM INTELLIGENT BASED FEATURE SELECTION FOR CLEVELAND HEART DISEASE PREDICTION SYSTEM," *International Journal,* vol. 6, 2017.
5. N. Kakhandaki, *et al.*, "Classification of Brain Hemorrhages in MRI Using Naïve Bayes-Probabilistic Kernel Approach," *Journal of Cases on Information Technology (JCIT),* vol. 21, pp. 51-65, 2019. https://doi.org/10.4018/JCIT.2019070104
6. G. Gilanie, *et al.*, "Computer aided diagnosis of brain abnormalities using texture analysis of MRI images," *International Journal of Imaging Systems and Technology*, vol. 29, pp. 260-271, 2019. https://doi.org/10.1002/ima.22312
7. V. Davis and S. Devane, "Diagnosis of Brain Hemorrhage Using Artificial Neural Network," 2017.
8. M. Sridevi, *et al.*, "Diagnosis and classification of brain hemorrhage using CAD system," *Indian Journal of Scientific Research,* pp. 121-126, 2015.
9. S. Ray, *et al.*, "Intensity population based unsupervised hemorrhage segmentation from brain CT images," *Expert Systems with Applications,* vol. 97, pp. 325-335, 2018.
10. L. A. Ramos, *et al.*, "Machine learning improves prediction of delayed cerebral ischemia in patients with subarachnoid hemorrhage," *Journal of neurointerventional surgery,* vol. 11, pp. 497-502, 2019.
11. Y. C. Youn, *et al.*, "Detection of cognitive impairment using a machine-learning algorithm," *Neuropsychiatric disease and treatment,* vol. 14, p. 2939, 2018. https://doi.org/10.2147/NDT.S171950
12. B. Shahangian and H. Pourghassem, "Automatic brain hemorrhage segmentation and classification algorithm based on weighted grayscale histogram feature in a hierarchical classification structure," *Biocybernetics and Biomedical Engineering,* vol. 36, pp. 217-232, 2016.
13. F. Baji, *et al.*, "Brain Tumor Detection Based on Asymmetry and K-Means Clustering MRI Image Segmentation," *Journal of Engineering Science and Technology*, vol. 13, pp. 4145-4159, 2018.
14. S. Keerthana and K. Sathiyakumari, "Brain stroke segmentation using fuzzy C-means clustering," *International Journal of Computer Applications,* vol. 154, pp. 26-30, 2016.
15. P. Govindarajan, *et al.*, "Classification of stroke disease using machine learning algorithms," *Neural Computing and Applications,* pp. 1-12, 2019.
16. K. N. Keshavamurthy, *et al.*, "Machine learning algorithm for automatic detection of CT-identifiable hyperdense lesions associated with traumatic brain injury," in *Medical Imaging 2017: Computer-Aided Diagnosis,* 2017, p. 101342G.
17. R. Mahajan and P. Mahajan, "Survey on diagnosis of brain haemorrhage by using artificial neural network," *International Journal of Scientific Research Engineering & Technology,* vol. 5, pp. 378-381, 2016.

18. H. Huang, *et al.*, "Brain image segmentation based on FCM clustering algorithm and rough set," *IEEE Access*, vol. 7, pp. 12386-12396, 2019.
19. A. Sharma, *et al.*, "Brain tumor segmentation using DE embedded OTSU method and neural network," *Multidimensional Systems and Signal Processing*, vol. 30, pp. 1263-1291, 2019. https://doi.org/10.1007/s11045-018-0603-3
20. [20] M. A. Khan, *et al.*, "Brain tumor detection and classification: A framework of marker-based watershed algorithm and multilevel priority features selection," *Microscopy research and technique*, vol. 82, pp. 909-922, 2019. https://doi.org/10.1002/jemt.23238
21. M. Padlia and J. Sharma, "Fractional Sobel filter based brain tumor detection and segmentation using statistical features and SVM," in *Nanoelectronics, circuits and communication systems*, ed: Springer, 2019, pp. 161-175. https://doi.org/10.1007/978-981-13-0776-8_15