

Research Article

A Novel Method for Differential Prognosis of Brain Degenerative Diseases Using Radiomics-Based Textural Analysis and Ensemble Learning Classifiers

Manju Jain ^{1,2}, C. S. Rai ¹ and Jai Jain ³

¹University College of Information, Communication and Technology, Guru Gobind Singh Indraprastha University, Dwarka Sector 16-C, New Delhi 110078, India

²Meerabai Institute of Technology Maharani Bagh, New Delhi 110065, India

³Media Agility India Ltd, New Delhi, India

Correspondence should be addressed to Manju Jain; manju.jain69@gmail.com

Received 26 April 2021; Revised 18 June 2021; Accepted 9 July 2021; Published 5 August 2021

Academic Editor: Venkatesan Rajinikanth

Copyright © 2021 Manju Jain et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a novel approach to develop a computer-aided decision support system for radiologists to help them classify brain degeneration process as physiological or pathological, aiding in early prognosis of brain degenerative diseases. Our approach applies computational and mathematical formulations to extract quantitative information from biomedical images. Our study explores the longitudinal OASIS-3 dataset, which consists of 4096 brain MRI scans collected over a period of 15 years. We perform feature extraction using Pyradiomics python package that quantizes brain MRI images using different texture analysis methods. Studies indicate that Radiomics has rarely been used for analysis of brain cognition; hence, our study is also a novel effort to determine the efficiency of Radiomics features extracted from structural MRI scans for classification of brain degenerative diseases and to create awareness about Radiomics. For classification tasks, we explore various ensemble learning classification algorithms such as random forests, bagging-based ensemble classifiers, and gradient-boosted ensemble classifiers such as XGBoost and AdaBoost. Such ensemble learning classifiers have not been used for biomedical image classification. We also propose a novel texture analysis matrix, Decreasing Gray-Level Matrix or DGLM. The features extracted from this filter helped to further improve the accuracy of our decision support system. The proposed system based on XGBoost ensemble learning classifiers achieves an accuracy of 97.38%, with sensitivity 99.82% and specificity 97.01%.

1. Introduction

Medical image processing has travelled a long journey since the last two decades. The past decade has seen the bridging of medical and information technology. It led to the development of decision support systems for early identification of various brain diseases. Age and structural changes in brain cause physiological alterations, which are reflected in routine human behaviour [1, 2]. Along the years, various studies and constant attempts have been made to study dementia.

Studies [3–5] focus on specific regions of interest in brain volumes, and these are calculated from two dimensional manually traced areas. Segmentation algorithms are used to segment out gray matter (GM), white matter (WM), and

cerebrospinal fluid (CSF). Such volumetric studies are limited to known brain structures like hippocampus and amygdala, perirhinal, entorhinal, and parahippocampal cortex.

Voxel-based studies [6–8] provide an alternative neuro-imaging method. These studies apply a general linear model (GLM) to each voxel of an MRI and statistically compare them with standard voxel values using Jacobean matrices.

Many studies [9, 10] give detailed insights on comparisons between voxel based and volumetric studies.

Several studies [11–13] use cortical thickness measurement as a biomarker for the process of identification of various brain aging diseases.

With the advancements in the machine learning techniques for image processing and image analysis and the

availability of abundance of medical imaging data, medical informatics [14, 15] has achieved great heights. The workshop MICCAI 2014 “Challenges of Computer aided diagnosis of Dementia on Structural MRI data” addresses the challenges of applying different algorithms on the same data and the same algorithm on different data. A summary of all algorithms presented in MICCAI 2014 is listed in [16]. This paper did a standardized comparison of different studies in the domain of the computer-aided decision support system for the identification of dementia-related diseases using structured MRI data. The best performing algorithm yielded an accuracy of 63% and receiver operating characteristics area under the curve with value 78.8%.

A review of various studies used for brain disorder detection using the machine learning techniques is published in [17]. Another review of the latest image processing techniques for studying brain pathology is summarized in [18].

A set of studies have been done on how oxygen supply changes the brain functioning [19, 20].

The functional modalities of medical imaging include MRI (magnetic resonance imaging), PET (positron emission tomography), and CT (computerized tomography) giving us an insight about the pathophysiology of the organ under observation. Radiologists analyze this information with their experience and knowledge. They find this time consuming and cumbersome. In this study, we explore machine learning techniques to analyze data extracted from medical images. Machine learning is the study of algorithms that solve a problem by leaning from underlying patterns in data, as opposed to statistical heuristics or rule-based programming. Radiomics [21, 22] aids in extracting imaging-based statistical biomarkers from medical images which can be used as features for machine learning methods to get accurate predictions. Ageing leads to degeneration of the brain, which may lead to dementia, further precipitating such diseases like Alzheimer’s dementia, vascular dementia, dementia with lewy body dementia, posterior cortical atrophy, and front temporal lobar degeneration. These diseases affect different regions of the brain. Clinical Dementia Rating or CDR is a five-point scale to stage dementia, ranging from 0 to 3, where 0 denotes no pathological degeneration (control patients) while any value greater than 0 indicates some pathological brain degeneration (test patients).

In this study, we propose a novel approach to develop a computational decision support system capable of differentiating control patients from test patients by analyzing features of their MRI images using Radiomics. This system can be used to assist radiologists for fast and accurate decisions.

- (1) We explore the OASIS-3 dataset [23], which is a longitudinal dataset with 4096 MRI scans. This dataset also gives specific details about how the CDR value changes for a subject with respect to changes in the subject’s MRI scan. These ratings can be used to label the MRI scans as healthy scans or scans showing signs of brain degeneration. Using these labels for a scan, a supervised machine learning binary classifier can be trained to support brain degeneration prognosis

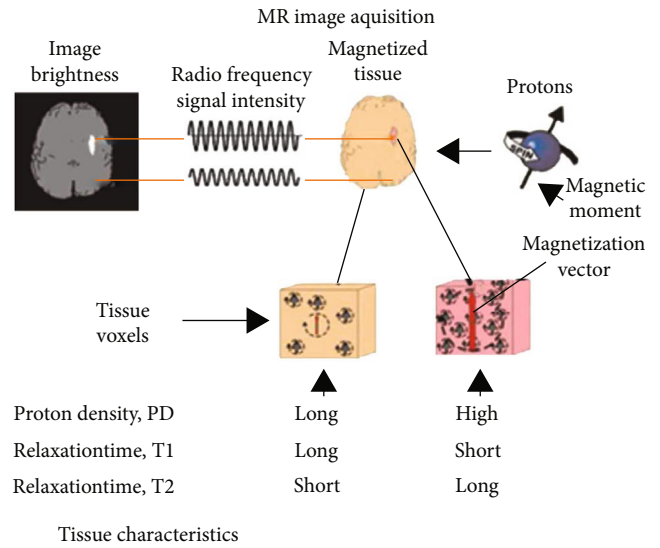


FIGURE 1: MRI acquisition process.

- (2) We employ data preprocessing best practices such as data augmentation and feature selection which help to mitigate overfitting and underfitting of the classifier and drive it to achieve optimal accuracy on our data
- (3) Feature extraction is done using Pyradiomics, which provides a python implementation of the study [24]. Pyradiomics provides a unified and standardized set of features from structured MRIs based on shape and volume as well as texture-based statistical features. Advanced Pyradiomics algorithms can handle missing data in case of low resolution MRI scans. Literature studies indicate that Radiomics has mostly been explored for oncological studies [25, 26], but not for understanding brain cognition. Our study is also an effort to determine the efficiency of Radiomics features from structural MRI scans for classification of brain degeneration diseases
- (4) We explored various ensemble learning classification algorithms such as random forests, bagging-based ensemble classifiers, and gradient-boosted ensemble classifiers such as XGBoost and AdaBoost for our classification tasks. Such ensemble learning classifiers have not been used for biomedical image classification
- (5) We propose a novel image texture analysis filter, Decreasing Gray-Level Matrix, which further improves the performance of our ensemble learning classifiers

We conclude the paper by comparing our novel solution with existing work in this field. Our results show that the proposed solution outperforms existing studies on various performance metrics such as accuracy, specificity, and sensitivity.

2. Materials and Methods

2.1. Data Acquisition. Magnetic resonance imaging is the process of acquiring images of anatomical structures using

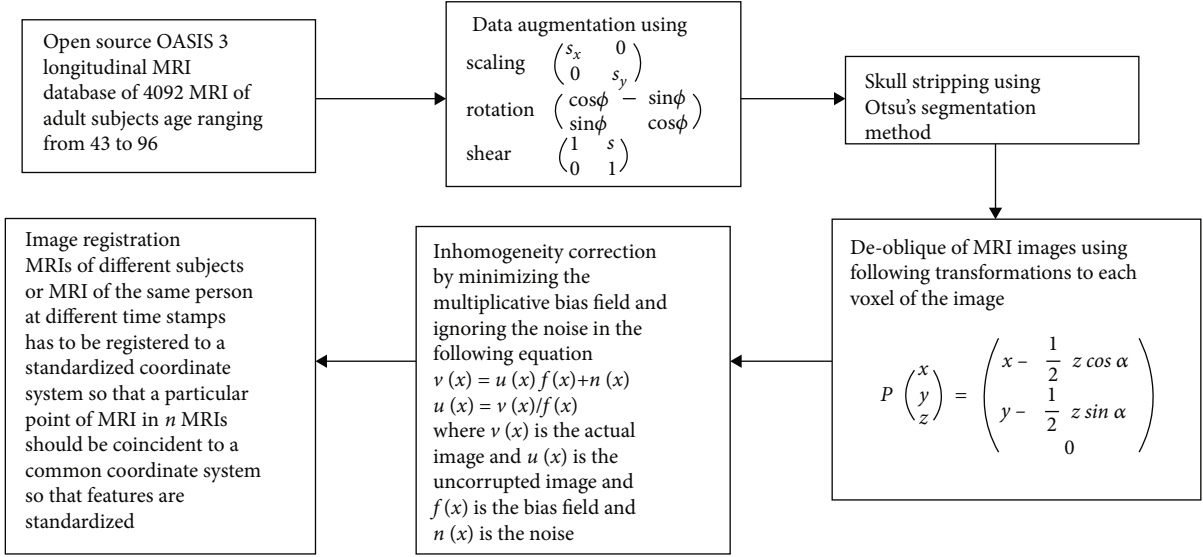


FIGURE 2: Data preprocessing flow.

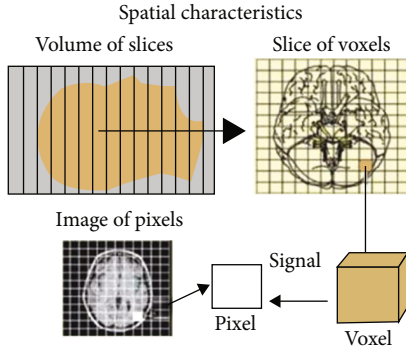


FIGURE 3: Spatial characteristics of MRI.

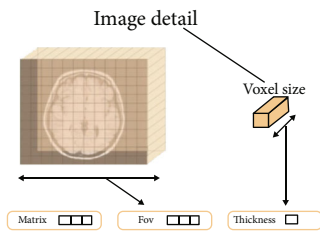


FIGURE 4: Voxel details.

magnetic field and radio frequency signals to detect diseases and functional problems. The “image is snapped” with different contrasts as different tissues and fluids react differently to magnetization signals. Tissue demagnetization time is different for different tissues. These times are identified as T1 and T2. Another characteristic of a tissue that affects an MRI is its proton density known as PD. Figure 1 depicts the complete MRI acquisition process.

In our study, we used the latest OASIS-3 dataset [23], which is an open source brain MRI database published in 2019. Most of the earlier studies have been done using ADNI

datasets, which are cross-sectional datasets and do not include more than 500 subjects. OASIS-3 is the largest longitudinal dataset of longitudinal MRI images that consists of 1068 subjects (age group of 46 to 95), collected over a period of 15 years.

“The CDR is a 5-point scale used to characterize six domains of cognitive and functional performance applicable to Alzheimer disease and related dementias: Memory, Orientation, Judgment & Problem Solving, Community Affairs, Home & Hobbies, and Personal Care. The necessary information to make each rating is obtained through a semi-structured interview of the patient and a reliable informant or collateral source (e.g. family member)” [27].

The OASIS database also provides CDR for each subject. The CDR values of a person over a particular period of time may or may not be the same. There are multiple scans of the same subject (4-5 times) in the time period of 15 years with different CDR values. These scans can be further used as samples. Hence, the database has more than 4000 MRIs.

2.2. Data Preprocessing. We performed data preprocessing using Python and FreeSurfer [28]. Main steps of data preprocessing are listed below and more visually shown in Figure 2.

2.2.1. Data Augmentation. We augmented our data to make our classifier much more tolerant towards variance in the data (prevents overfitting) and to increase dataset size (prevents underfitting). We employed 4 augmentation techniques:

- (1) *Flips*. Each image is flipped horizontally as well as vertically.
- (2) *Scaling*. Each image is scaled in either “x” or “y” direction with the help of a transform matrix $\begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix}$.

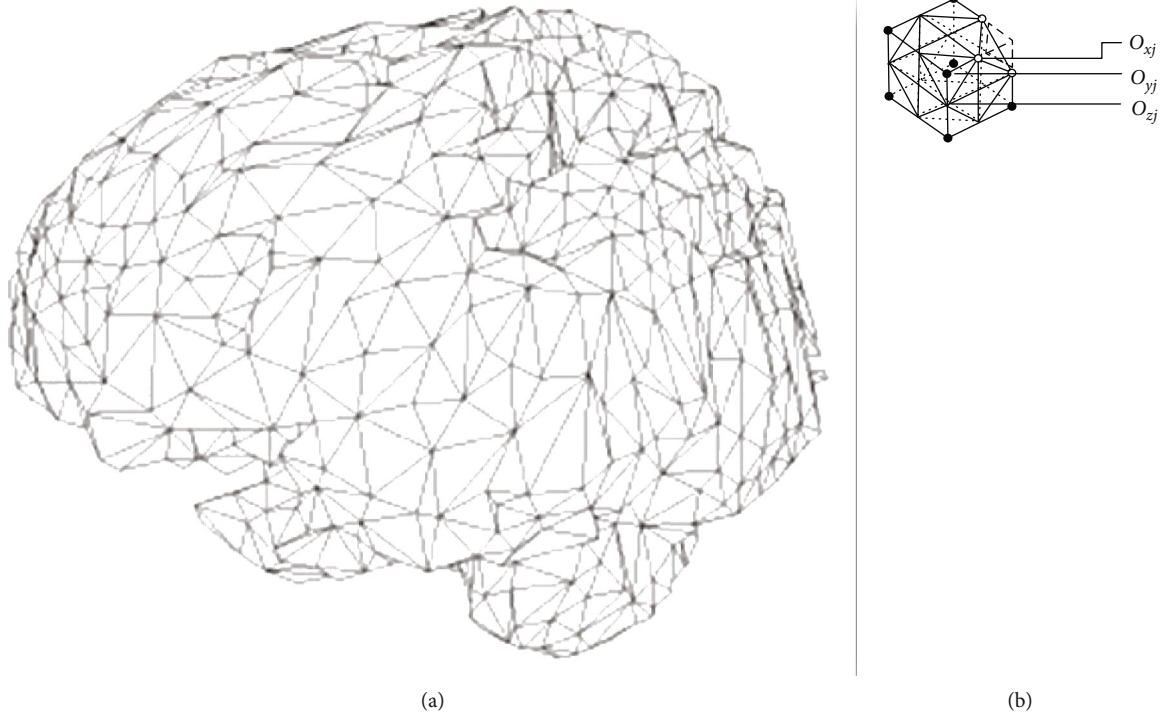


FIGURE 5: (a) Decimation of triangles as a mesh of the whole brain. (b) The trapezium points and edges. X_v is the number of voxels included in the masked region. V_m is the volume of the mesh mm^3 . A_m the surface area of the mesh in mm^2 .

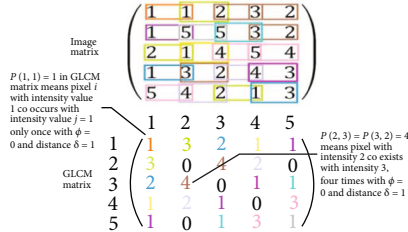


FIGURE 6: GLCM procedure. Different color schemes were used to track the steps.

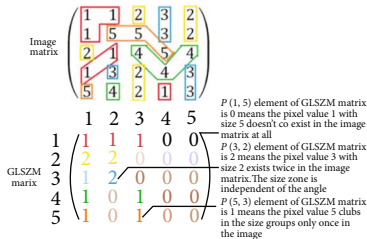


FIGURE 7: GLSZM procedure. Different color schemes were used to track the steps.

(3) *Rotations*. Affine transform matrix $\begin{pmatrix} \cos \varnothing & -\sin \varnothing \\ \sin \varnothing & \cos \varnothing \end{pmatrix}$ gives rotated MRI images in different directions. " \varnothing " was varied between 25 and 195.



FIGURE 8: GLRLM procedure. Different color schemes were used to track.

(4) *Shears*. Affine transform matrix $\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ applied to each image where Shear value changes from 0.3 to 0.7.

2.2.2. *De-Oblique*. During the MRI process, the subject's head may be tilted from to cover the whole brain or to avoid artefacts caused by water and air in the nose and eyes of the subject. This causes the MRI to be oblique and makes intersubject or intrasubject registration more difficult. The MRI images in our dataset were de-obliques using the FreeSurfer software.

2.2.3. *Inhomogeneity Correction*. Brain consists of different types of tissues like gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF), and all these tissues have different range of penetration to the magnetic field and may result into very bright or very dull artefacts in the MRI images. This may confuse a radiologist since all tissues of a

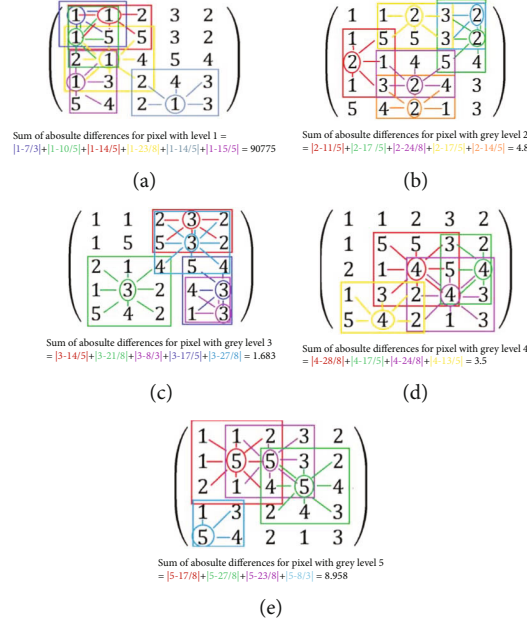


FIGURE 9: (a) NGTDM for neighbours of 1. Different color schemes were used to track NGTDM procedure to calculate the absolute sum of gray-level difference. (b) NGTDM for neighbours of 2. Different color schemes were used to track NGTDM procedure to calculate the absolute sum of gray-level difference. (c) NGTDM for neighbours of 3. Different color schemes were used to track NGTDM procedure to calculate absolute sum of gray-level difference. (d) NGTDM for neighbours of 4. Different color schemes were used to track NGTDM procedure to calculate the absolute sum of gray-level difference. (e) NGTDM for neighbours of 5. Different color schemes were used to track NGTDM procedure to calculate absolute sum of gray-level difference.

particular type should have exact intensity and brightness values. The process of correcting this is known as inhomogeneity correction.

2.2.4. Skull Stripping. The nonbrain parts (skull, neck, eyes, and nose) were removed from all MRI images to have a uniform area of study.

2.2.5. Registration. The brain consists of very fine spatial structures, due to which it is very difficult to extract and integrate the information from different images. The thickness of the cortex can be as small as 5 mm. The thickness of thalamic nuclei only extends to few millimetres. Registration is the process of aligning different MRI images in such a way that the voxels of a particular tissue from all of those images correspond to the same 3D location. We applied and adjusted the registration parameters, i.e., translations, rotations, scaling, and shear operations at voxel level to make the MRI images concurrent.

2.3. Feature Extraction

2.3.1. Features Based on Shape. To extract features based on shape, we studied spatial characteristics of MRIs as depicted in Figure 3.

Slice: an MRI is a 3D image. It consists of a set of contiguous 2D slices. These slices may either represent the axial, sagittal, or longitudinal cross section of the subject's brain.

Voxel: each slice is subdivided into rows and columns. The intersection of each row and column represents a

volume of the brain. This is known as a voxel. The field-of-view matrix of a particular size of the slice is used to determine the voxel size. The voxel details are depicted in Figure 4.

Shape features include legends of 3D size and shape. We took the whole brain area and volume as our region of interest. A triangular mesh encapsulating the whole brain area was used to extract various shape features. Figures 5(a) and 5(b) show how a brain is treated as a mesh surface [28]. The mesh has X_f number of triangles.

From this meshed surface of brain, as in Figure 5(b), we calculated the following different shape features [24].

- (i) Mesh volume $V_j = O_{xi} \cdot (O_{yj} \times O_{zj}) / 6$ (O_{xi}, O_{yj}, O_{zj} are the tetrahedral vertices)

$$V_m = \sum_{j=1}^{X_f} V_j. \quad (1)$$

- (ii) **Voxel Volume.** $\text{Voxel}_{\text{volume}} = \sum_{j=1}^{X_v} V_m$

The whole brain volume can be obtained by multiplying the voxel volume V_j with the number of voxels in the brain.

- (i) **Surface Area.** $A_j = 1/2(X_j Y_j \times X_j Z_j)$

NGTDM matrix with i grey levels and n_i pixels of the i th grey level and with p_i probability and s_i as the sum of absolute differences for that particular grey level

i	n_i	p_i	S_i
1	6	0.24	9.775
2	6	0.24	4.8
3	5	0.20	1.683
4	4	0.16	3.5
5	4	0.16	8.958

FIGURE 10: NGTDM.

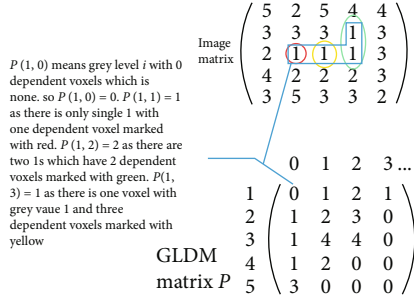


FIGURE 11: GLDM. Different color schemes were used to track GLDM procedure.

$$A_m = \sum_{j=1}^{X_f} A_j. \quad (2)$$

To calculate the surface area of the whole brain, it is divided into small mesh areas. We first calculate the surface area of each mesh and then sum all of them.

- (i) *Ratio of the Surface Area to the Volume of the Brain.* A_m/V_m
- (ii) Lower the ratio more the compactness
- (iii) *Maximum 3D Diameter.* It is the largest Euclidean distance on the various mesh surfaces on the whole brain.
- (iv) *The Maximum 2D Diameter of the Slice.* It is the defined as the largest Euclidean distance on the whole brain mesh surfaces where mesh vertices are in the axial plane.
- (v) *Major Axis length.* $4\sqrt{\gamma_{\text{major}}}$
- (vi) This feature calculates the largest axis length of the whole brain area
- (vii) *Minor Axis Length.* $4\sqrt{\gamma_{\text{minor}}}$
- (viii) This feature represents the minimum axis length of the whole brain area
- (ix) *Elongation.* This feature gives the relationship between the largest and smallest component of the whole brain.
- (x) $\text{Elongation} = \sqrt{\gamma_{\text{minor}}/\gamma_{\text{major}}}$
- (xi) $\text{Flatness} = \sqrt{\gamma_{\text{least}}/\gamma_{\text{major}}}$

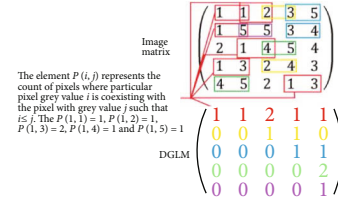


FIGURE 12: DGLM. Different color schemes were used to track DGLM procedure.

2.3.2. *First-Order Features.* These features are obtained by the statistical analysis of the whole brain based on values of voxel intensities [24].

Let S be a set of N_v voxels in the whole brain.

Let N_v be the discrete level of intensities in the whole brain then $X(i)$ is the first-order histogram.

The normalized first-order histogram $x(i) = X(i)/N_v$

- (i) $\text{Energy} = \sum_{i=1}^{N_v} (S(i) + c)^2$
- (ii) $\text{Total energy} = V_{\text{voxel}} \sum_{i=1}^{N_v} (S(i) + c)^2$
- (iii) $\text{Entropy} = -\sum_{i=1}^{N_v} x(i) \log_2(x(i) + \epsilon)$
- (iv) $\text{Minimum} = \min(S)$
- (v) 10th percentile of S
- (vi) 90th percentile of S
- (vii) $\text{Maximum} = \max(S)$
- (viii) $\text{Mean} = 1/N_v \sum_{i=1}^{N_v} S(i)$ is the average gray-level intensity of the whole brain
- (ix) $\text{Median} =$ the median gray level of the whole brain
- (x) $\text{Range} = \max(S) - \min(S)$
- (xi) $\text{Absolute mean deviation} = \sqrt{1/N_v \sum_{i=1}^{N_v} |S(i) - \bar{S}|}$
- (xii) $\text{Root mean square value of the whole brain (RMS)} = \sqrt{1/N_v \sum_{i=1}^{N_v} (S(i) + c)^2}$
- (xiii) $\text{Standard deviation} = \sqrt{1/N_v \sum_{i=1}^{N_v} (S(i) - \bar{S})^2}$
- (xiv) $\text{Skewness} = \sqrt{1/N_v \sum_{i=1}^{N_v} (S(i) - \bar{S})^3} / (\sqrt{1/N_v \sum_{i=1}^{N_v} (S(i) - \bar{S})^2})^3 = \mu_3/\sigma^3$
- (xv) $\text{Kurtosis} = \sqrt{1/N_v \sum_{i=1}^{N_v} (S(i) - \bar{S})^4} / (\sqrt{1/N_v \sum_{i=1}^{N_v} (S(i) - \bar{S})^2})^2 = \mu_4/\sigma^4$
- (xvi) $\text{Variance} = 1/N_v \sum_{i=1}^{N_v} |S(i) - \bar{S}|$

2.3.3. *Gray-Level Cooccurrence Matrix [24].* GLCM is a texture filter that gives the pixel distribution of a particular set of pixels i, j in a specific direction and distance. The

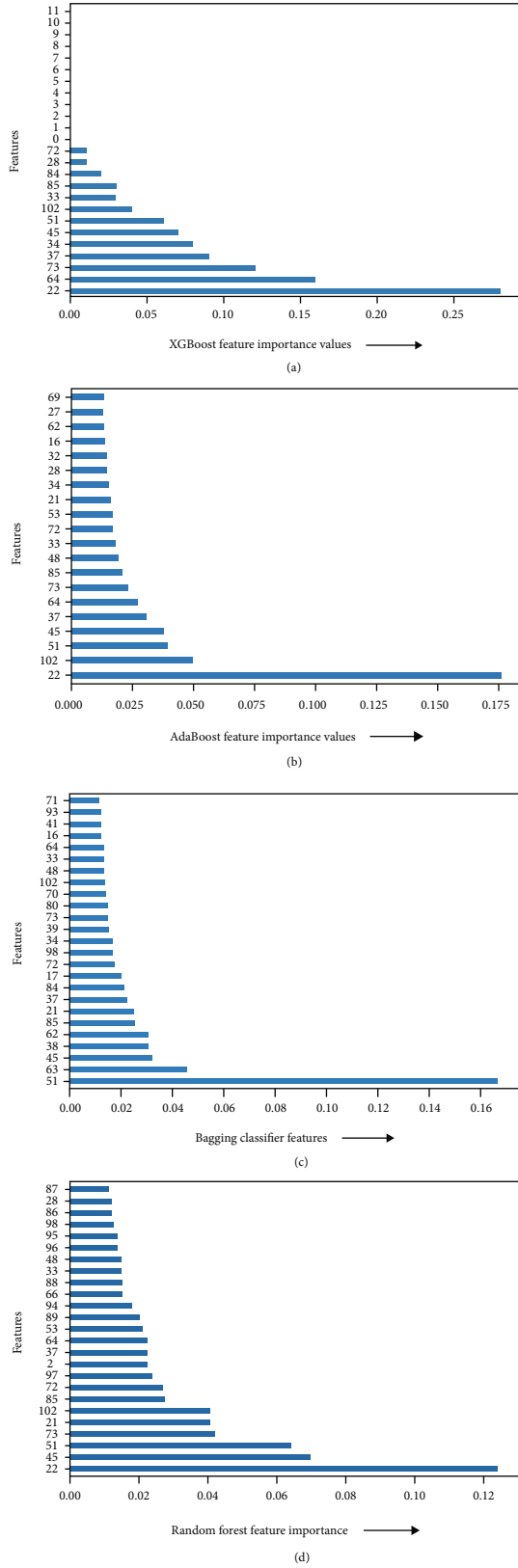


FIGURE 13: (a) Feature importance prediction by XGBoost. (b) Feature importance prediction by AdaBoost. (c) Feature importance prediction by bagging. (d) Feature importance prediction by random forest.

TABLE 1: Sensitivity, specificity, and accuracy comparison of different ensemble classifiers.

Classifiers	Sensitivity	Specificity	Accuracy
XGBoost	99.82%	97.01%	97.38%
AdaBoost	94.91%	97.76%	97.21%
Bagging classifier	74.22%	90.07%	87.56%
Random forest	94.44%	87.07%	87.72%

$p(i, j) | \emptyset, \delta |$ value of GLCM represents the number of times; the pixel with intensity i coexists with intensity j with angle \emptyset and distance δ . Figure 6 shows how GLCM can be obtained from image matrix. The different color schemes indicate a particular pixel's coexistence. Generally, the following statistical features are extracted and then averaged over GLCM for each direction (angle).

- (i) Autocorrelation
- (ii) Joint average
- (iii) Entropy
- (iv) Variance
- (v) Contrast
- (vi) Energy
- (vii) Homogeneity
- (viii) Inverse of the difference movement
- (ix) Inverse variance

2.3.4. *Gray-Level Size Zone Matrix [24]*. GLSZM quantifies different pixel intensity values in different size zones. A size zone is defined as connected pixels/voxels with the same gray-level irrespective of direction. The $P(i, j)$ element of GLSZM represents the number of times the intensity value i of the size zone j exists in the image matrix. Figure 7 depicts how GLSZM can be obtained from the image matrix. Different colors indicate different size zones of different intensity values.

GLSZM can be used to extract the following features:

- (i) Emphasis on small areas
- (ii) Emphasis on large areas
- (iii) Gray-level nonuniformity
- (iv) Normalized gray-level nonuniformity
- (v) Size zone nonuniformity
- (vi) Normalized size zone nonuniformity
- (vii) Low gray-level emphasis on small areas
- (viii) High gray-level emphasis on small areas
- (ix) Low gray-level emphasis on large areas
- (x) High gray-level emphasis on large areas

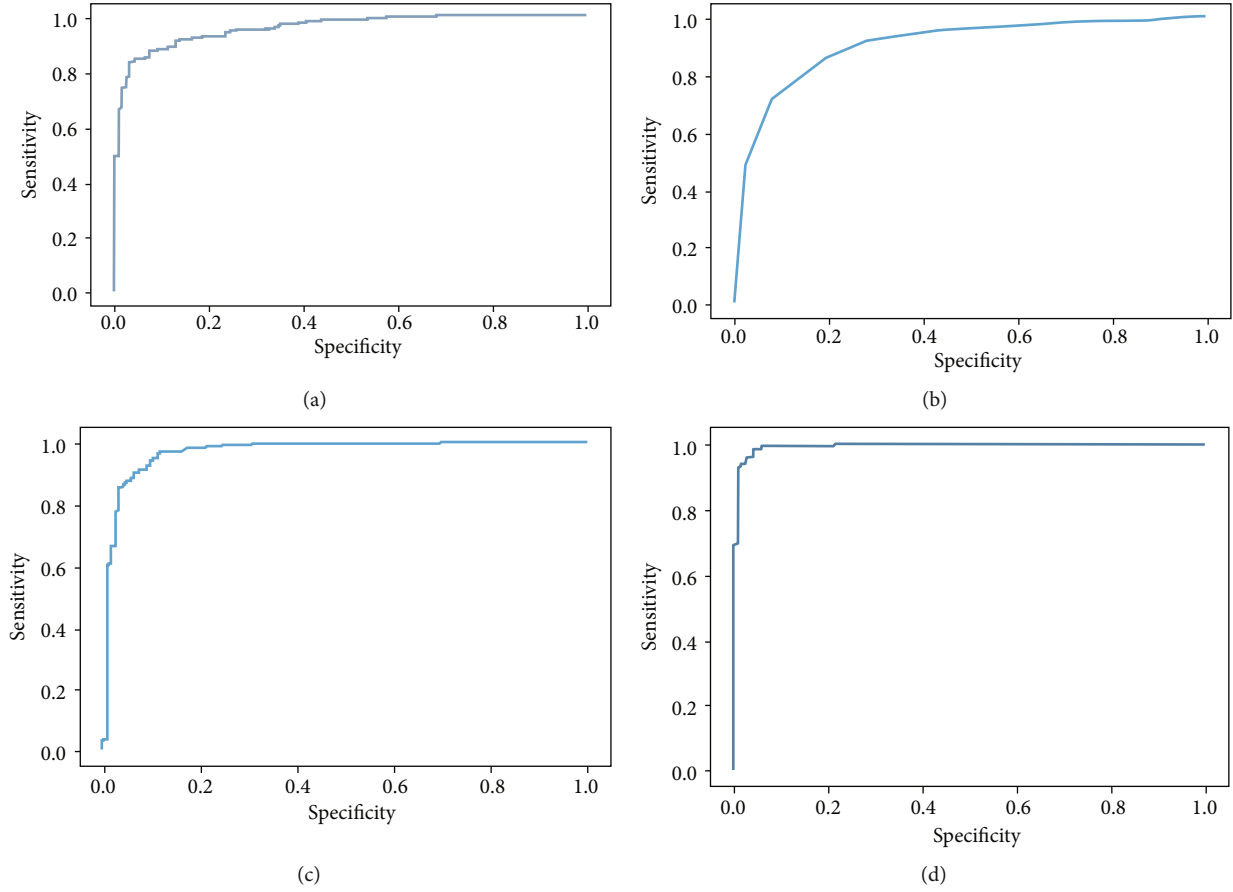


FIGURE 14: Area under the accuracy curve for different ensemble classifiers.

2.3.5. *Gray-Level Run Length Matrix (GLRLM)* [24]. The intensity runs in a GLRLM are defined as the length of connected pixels of equal intensity values and in a particular direction. A GLRLM element $P(i, j) | \emptyset$ represents the number of times a particular run length j of intensity i in direction \emptyset occurs in the image matrix. Figure 8 depicts how GLRLM can be obtained from image matrix. Different colors indicate different run lengths of the particular length in a particular direction.

The GLRLM is used to extract following features:

- (i) Emphasis on short run
- (ii) Emphasis on long run
- (iii) Nonuniform gray level
- (iv) Normalized nonuniform gray level
- (v) Nonuniform run length
- (vi) Normalized nonuniform run length
- (vii) Run percentage
- (viii) Variance of gray level
- (ix) Run variance
- (x) Run entropy

- (xi) Low gray-level run emphasis
- (xii) High gray-level emphasis
- (xiii) Short run low gray-level emphasis
- (xiv) Short run high gray-level emphasis
- (xv) Long run low gray-level emphasis
- (xvi) Long run high gray-level emphasis

2.3.6. *Neighbouring Gray Tone Difference Matrix* [24]. Here, we consider neighbouring pixels of a particular pixel at a distance ∂ of that pixel. This matrix is the set of absolute differences of the gray levels of the voxel and its neighbouring voxels. Let P_{nv} be the set of whole brain voxels; then, $p_{nv}(i_x, i_y, i_z)$ belongs to P_{nv} where p_{nv} denotes the gray level of the voxel at position (i_x, i_y, i_z) . The average gray level of the neighbourhood is given as follows: $\bar{G}_j = \bar{G}(i_x, i_y, i_z) = 1/V \sum_{k_x=-\partial}^{\partial} \sum_{k_y=-\partial}^{\partial} \sum_{k_z=-\partial}^{\partial} p_{nv}(i_x + k_x, i_y + k_y, i_z + k_z)$ where V is total number of voxels in the whole brain.

- (i) Let i denote the value of gray levels in the image
- (ii) Let n_i denote the number of voxels of gray level i
- (iii) Let p_i denote the gray-level probability

TABLE 2: Comparison with similar studies.

Research study	Year	Dataset	Brain area	Classifier	Accuracy
Ahmad Chaddad [32]	2018	OASIS-1	Hippocampus Amygdala	Random forest Random forest CNN	84.09% 92.5%
Feng Feng [33]	2018	Local hospital data	Hippocampus	SVM	86.75%
Yupeng Li and Jiehui Jiang [34]	2019	Local hospital data	Hippocampus	SVM	91.5%
Kun Zhao [35]	Jan 2020	ADNI	Hippocampus	SVM	88.21%
Tao-Ran Li [36]	Dec 2020	ADNI	Right posterior and left superior cingulate Gyrus	SVM	95.9%
Current study proposed by us	2021	OASIS-3	Whole brain	Ensemble classifiers	
				XGBoost	97.38%
				AdaBoost	97.21%
				Bagging	87.56%
				Random forest	87.72%

(iv) Let s_i be the sum of absolute difference of a gray level i

Figures 9(a)–9(d) are the required NGTDM for the pixel with intensities 1–4. Figure 9(e) describes how the absolute difference of the different gray levels is calculated. Different colors are used to track down the neighbours of a particular gray level as shown in the following example where we have 5 discrete gray levels 1 to 5. Figure 10 is the final NGTDM.

Features calculated from NGTDM are as follows:

- (i) The neighbourhood-based coarseness
- (ii) Neighbourhood-based contrast
- (iii) Rate of change of gray levels within voxels
- (iv) The complexity of neighbourhood gray levels
- (v) Strength of neighbourhood gray levels

2.3.7. Gray-Level Dependence Matrix [24]. GLDM represents the dependencies of one gray level on other gray levels. It is defined as a set of connected voxels within distance ∂ dependent on a central voxel. A voxel with a gray level i is dependent on another voxel of gray level j if

$$|i - j| \leq \gamma. \quad (3)$$

The $(i, j)^{\text{th}}$ element of GLDM $P(i, j)$ represents how often a voxel with the gray value i coexists with its dependent voxel having gray level j occurs in the whole brain image. Figure 11 describes how GLDM is obtained from brain MRI with $n = 5$, i.e., 5 discrete gray levels, $\gamma = 0$, and $\partial = 1$. The GLDM columns start from 0, and it can go to any finite number of dependent voxels.

The above GLDM is used to extract the following features:

- (i) Small dependence significance

- (ii) Large dependence significance
- (iii) Gray-level heterogeneity
- (iv) Dependence heterogeneity
- (v) Dependence heterogeneity normalized
- (vi) Gray-level deviation
- (vii) Dependence deviation
- (viii) Entropy of dependency
- (ix) The low gray-level significance
- (x) The high gray-level significance
- (xi) Small dependency and low gray-level significance
- (xii) Small dependency and high gray-level significance
- (xiii) Large dependency and low gray-level significance

2.3.8. Decreasing Gray-Level Matrix (Novel Filter). We propose a novel filter matrix to improve feature set. The $p(i, j) \mid \emptyset, \delta$ pixel of DGLM represents the occurrence of the pixel with intensity i and pixel with intensity j such that $i \leq j$. Figure 12 depicts and obtains a DGLM from the image with $\emptyset = 0$ and $\delta = 1$. Colors are used to track down the location of pixels for which the condition $i < j$ holds true.

The DGLM is used to extract the following features in four directions, i.e., 0, 45, 90, and 135. Then, the average is taken to get the summary of the following features:

- (i) Energy
- (ii) Mean
- (iii) Absolute mean deviation
- (iv) Skewness
- (v) Kurtosis

- (vi) Entropy
- (vii) Autocorrelation

2.4. Feature Selection. Feature selection is the process of evaluating and selecting the most important features from the set of all features depending on their contribution to the machine learning task at hand. This process helped us to select the features with the highest predictive relevance to our classification task. This in turn also helps to eliminate redundant features.

In our study, we focused on tree-based classification methods. These methods have intrinsic feature selection methods. Using these intrinsic methods, we found the feature relevance for that each classifier.

Figure 13(a) denotes that the novel feature “first order mean of DLGM” has the highest predictive power for XGBoost classifier; hence, it is the most important feature for this classifier. The other important features for XGBoost classifier are as follows:

- (i) Decreasing Gray-Level Matrix feature first-order mean 0.26
- (ii) Gray-Level Dependence Matrix feature high gray-level emphasis 0.16
- (iii) Gray-Level Run Length Matrix feature gray-level run emphasis 0.14
- (iv) Gray-Level Cooccurrence Matrix feature correlation 0.08
- (v) Gray-Level Cooccurrence Matrix feature cluster shade 0.05
- (vi) Decreasing Gray-Level Matrix feature information measure of correlation

Figure 13(b), denotes that the novel feature “first order mean of DLGM” has the highest predictive power for AdaBoost classifier; hence, it is the most important feature for this classifier. The other important top five features for AdaBoost classifier are as follows:

- (i) Decreasing Gray-Level Matrix feature first-order mean 0.17
- (ii) Neighbouring Gray Tone Difference Matrix feature busyness 0.05
- (iii) Decreasing Gray-Level Matrix feature maximal correlation coefficient 0.04
- (iv) Decreasing Gray-Level Matrix feature information measure of correlation 0.035
- (v) Gray-Level Cooccurrence Matrix feature correlation 0.03

Figure 13(c) denotes that the novel feature “Maximal Correlation coefficient of DLGM” has the highest predictive power for bagging classifier; hence, it is the most important

feature for this classifier. The other important top five features for bagging classifier are as follows:

- (i) Decreasing Gray-Level Matrix feature maximal correlation coefficient 0.16
- (ii) Gray-Level Dependence Matrix feature large dependence emphasis 0.05
- (iii) Decreasing Gray-Level Matrix feature information measure of correlation 0.03
- (iv) Gray-Level Cooccurrence Matrix feature difference average 0.03
- (v) Gray-Level Dependence Matrix feature high gray-level emphasis 0.03

Figure 13(d) denotes that the novel feature “first order mean of DLGM” has the highest predictive power for random forest classifier; hence, it is the most important feature for this classifier. The other important features for random forest classifier are as follows:

- (i) Decreasing Gray-Level Matrix feature first-order mean 0.12
- (ii) Decreasing Gray-Level Matrix feature information measure of correlation 0.07
- (iii) Decreasing Gray-Level Matrix feature maximal correlation coefficient 0.06
- (iv) Gray-Level Run Length Matrix feature high gray-level run emphasis 0.05
- (v) First-order mean absolute deviation 0.05

2.5. Ensemble Learning Classifiers. “Ensemble learning is a machine learning paradigm where multiple learners are trained to solve the same problem. In contrast to ordinary machine learning approaches which try to learn one hypothesis from training data, ensemble methods try to construct a set of hypotheses and combine them to use” [29]. A good number of studies [30, 31] proved that the generalization capability of a set of learners is much greater than a single learner. Ensemble classifiers have been applied in diversified fields, e.g., cyber security, intrusion detection system, face recognition system, and traffic control systems. The concept of ensemble classification proceeds in two stages:

- (a) Classifier generation
- (b) Aggregation of results of these classifiers

There are three approaches to classifier generation and aggregation.

- (i) Bagging
- (ii) Boosting
- (iii) Stacking

2.5.1. Bagging. In this method, different training datasets are generated by resampling the training dataset, i.e., replacing some of the samples randomly. Suppose we have the following dataset: (4,5,6,7,8,9,10) and we have 5 classification algorithms. A different dataset is created by randomly resampling our data and passed to each classifier for training:

Dataset for classifier 0: (4,5,5,7,8,10,10) by replacing 6 with 5 and 9 by 10.

Dataset for classifier 1: (4,5,7,7,9,9,10) by replacing 6 by 7 and 8 by 9.

Dataset for classifier 2: (5,5,7,7,9,9,6) by replacing 4 by 5 and 10 by 6.

The results of all these classifiers are aggregated when taking predictions and inference time.

2.5.2. Boosting. Boosting attempts to create chains of different classification algorithms. The chain with the best performance on training data is then used for inference, coming back to our previous example where we had our training dataset as (4,5,6,7,8,9,10) and 5 classification algorithms. If we are creating chains of 3 classifiers, we can create 10 such chains. A single chain of 3 classifiers is created in the following manner:

- (a) A batch of training dataset is passed through classification algorithm 1, i.e., classifier 0

Dataset for classifier 0: (4,5,6,7,8,9,10)

- (b) Based on the performance of classifier 0 on this training batch, the whole batch is redistributed. The incorrectly predicted samples (by classifier 0) from the training batch are chosen more often to create the training batch for classifier 1. In this manner, classifier 1 will try to improve on the mistakes done by classifier 0. This is true for each classifier in the chain

Dataset for classifier 1: (4,5,7,7,9,9,10) by replacing 6 by 7 and 8 by 9 as 7 and 8 was incorrectly predicted.

- (c) The same process will be repeated for classifier 2

Dataset for classifier 2: (10,9,7,7,9,9,10) by replacing 4 by 10 and 5 by 7 as both 4 and 5 was incorrectly predicted.

In essence, boosting will create and choose the chain which is able to collectively give better results than other chains.

In this study, we have explored two boosting ensemble classifiers XGBoost and AdaBoost. As is evident from our results, the prediction accuracy with these classifiers is much higher than bagging classifiers.

2.5.3. Stacking. Stacking is usually a 2 step approach. The classifiers in step 1 are known as base learners while the classifiers in step 2 are called stacking model learners. Each step is an ensemble of few classification algorithms. Predictions from the base learners are used as dataset for stacking model learners. Note that the predictions from base-level classifiers still maintain relationships with initial dataset which the stacking level classifiers can understand. The predictions from the stacking model learners are used at inference time.

3. Results

Along with accuracy, the most important metrics to analyze a biomedical machine learning study are sensitivity and specificity.

Sensitivity is the measure of true positives, which means accurate identification of patient with the disease. The test should have more true positives and minimum false negatives. False negatives mean we may miss out the positive identification of disease. Our study is a kind of screening test hence should have more sensitivity. Table 1 shows highest sensitivity is 99.82% hence in accordance to screening test.

Specificity is the measure true negatives, which is the ability of a test to rule out the disease accurately. Target of study is to have minimum false positives. As the study is screening test, we can have false alarms and less specific. The specificity of our study is 97.01%.

The three metrics are measured with following formulae:

- (i) Specificity = $\frac{\text{true negative outcomes}}{\text{true negative outcomes} + \text{false positive outcomes}}$
- (ii) Sensitivity = $\frac{\text{true positive outcomes}}{\text{true positive outcomes} + \text{false negative outcomes}}$
- (iii) Accuracy = $\frac{\text{true negative outcomes} + \text{true positive outcomes}}{\text{true negative outcomes} + \text{false positive outcomes} + \text{true positive outcomes}}$

3.1. Analyzing Different Ensemble Methods and Results. In our study, we observed that boosting ensemble learning classifiers such as AdaBoost and XGBoost perform better than bagging and randomized classifiers. Bagging classifiers and random forest classifiers yield almost the same accuracy of 87%. The results are listed in Table 1. The accuracy calculated from the area under the curve is depicted in Figure 14 for all four ensemble classifiers.

4. Conclusion

In this study, we have proposed to build a decision support system for radiologists in order to make fast and accurate decisions for early detection of brain degeneration by mapping CDR values to MRI images. The most important performance metrics in the field of computer-aided biomedical studies are sensitivity, specificity, and accuracy. Through this study, we have shown that better data collection and preprocessing (data augmentation and feature selection) along with gradient-boosted ensemble learning classifiers contribute to improvements in all 3 metrics.

Data is one of the most important factors for driving the accuracy of any study. In our study, we worked on the OASIS-3 dataset, which is a longitudinal dataset with 4096 MRI scans while earlier studies are performed on cross-sectional datasets with less than 500 MRI scans. This dataset also gives specific details about how the CDR value changes for a subject with respect to changes in the subject's MRI scan. Any machine learning system requires large amount of data to be optimally trained. In our study, we have also employed data augmentation techniques. Data augmentation resulted in our classifier being much more tolerant towards variance in the data; this prevents overfitting. Another major

impact of data augmentation was the increase in dataset size from 4096 to 10000 MRI scans; this prevents underfitting. Mitigating overfitting and underfitting helps to achieve optimal accuracy on any dataset, irrespective of the classifier being used.

Our domain experts (Dr. Kunal Jain and Dr. Tanu) suggested that brain degeneration is not localized and affects the brain as a whole. As such, we have utilized whole brain volumes for our study and classification.

We experimented with Radiomics features and found that, for our data, the most promising features of

- (i) GLCM are correlation, cluster shade, joint average, and cluster prominence
- (ii) GLRLM are gray-level run emphasis, short run high gray-level emphasis, short run low gray-level emphasis, and gray-level variance
- (iii) NGTD Matrix is busyness
- (iv) GLDM are high gray-level emphasis and small dependence low gray-level emphasis
- (v) GLSZM is small area low gray-level emphasis

Our study also proposes a novel texture filter DGLM. The features mean, information measure of correlation, maximal correlation coefficient, first-order entropy, and first-order skewness from novel DGLM improved the accuracy from 95.6% to 97.38%.

This study also reaffirmed the fact that ensemble learning classifiers are usually much more accurate than a single classification algorithm. The study observed that gradient-boosted classifiers do not suffer from overfitting and also help to reduce generalization error, hence improving accuracy, sensitivity, and specificity.

The study results have been compared to other different studies in this area as depicted in Table 2.

Data Availability

In this study, we used open source data. Data is available at <https://www.oasis-brains.org/>, the data was requested to Oasis-3 Brain team, and it provided the login and password to download the data; the same can be shared as and when needed.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

We sincerely acknowledge Dr. Kunal Jain (MBBS) and Dr. Tanu (BDS) for their guidance, as they clarified the facts and functionality of brain anatomy and tissue details.

References

- [1] C. L. Grady, "Functional brain imaging and age-related changes in cognition," *Biological Psychology*, vol. 54, no. 1–3, pp. 259–281, 2000.
- [2] B. J. Casey, J. N. Giedd, and K. M. Thomas, "Structural and functional brain development and its relation to cognitive development," *Biological Psychology*, vol. 54, no. 1–3, pp. 241–257, 2000.
- [3] C. R. Jack, R. C. Petersen, Y. C. Xu et al., "Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment," *Neurology*, vol. 52, no. 7, pp. 1397–1403, 1999.
- [4] J. C. Pruessner, L. M. Li, W. Serles et al., "Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: minimizing the discrepancies between laboratories," *Cerebral Cortex*, vol. 10, no. 4, pp. 433–442, 2000.
- [5] J. C. Pruessner, S. Köhler, J. Crane et al., "Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus," *Cerebral Cortex*, vol. 12, no. 12, pp. 1342–1353, 2002.
- [6] J. Ashburner and K. J. Friston, "Voxel-based morphometry—the methods," *Neuroimage*, vol. 11, no. 6, pp. 805–821, 2000.
- [7] J. C. Baron, G. Chételat, B. Desgranges et al., "In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease," *NeuroImage*, vol. 14, no. 2, pp. 298–309, 2001.
- [8] J. L. Whitwell, "Voxel-based morphometry: an automated technique for assessing structural changes in the brain," *The Journal of Neuroscience*, vol. 29, no. 31, pp. 9661–9664, 2009.
- [9] B. C. Emerton, M. Jerram, T. Deckersbach, D. D. Dougherty, C. Fulwiler, and D. A. Gansler, "A comparison of voxel-based morphometry and volumetry methods in the context of the neural basis of aggression," *Brain Imaging and Behavior*, vol. 3, no. 4, pp. 332–341, 2009.
- [10] K. M. Kennedy, K. I. Erickson, K. M. Rodrigue et al., "Age-related differences in regional brain volumes: a comparison of optimized voxel-based morphometry to manual volumetry," *Neurobiology of Aging*, vol. 30, no. 10, pp. 1657–1676, 2009.
- [11] B. Fischl and A. M. Dale, "Measuring the thickness of the human cerebral cortex from magnetic resonance images," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 20, pp. 11050–11055, 2000.
- [12] A. C. Burggren, M. M. Zeineh, A. D. Ekstrom et al., "Reduced cortical thickness in hippocampal subregions among cognitively normal apolipoprotein E e4 carriers," *NeuroImage*, vol. 41, no. 4, pp. 1177–1183, 2008.
- [13] A. Bakkour, J. C. Morris, and B. C. Dickerson, "The cortical signature of prodromal AD: regional thinning predicts mild AD dementia," *Neurology*, vol. 72, no. 12, pp. 1048–1055, 2009.
- [14] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in Medicine*, vol. 23, no. 1, pp. 89–109, 2001.
- [15] S. Tangaro, P. Inglese, R. Maglietta, and A. Tateo, "MIND-BA: fully automated method for computer-aided diagnosis of dementia based on structural MRI data," in *Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, pp. 119–128, Boston, USA, 2014.

- [16] E. E. Bron, M. Smits, W. van der Flier et al., “Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge,” *NeuroImage*, vol. 111, pp. 562–579, 2015.
- [17] L. Lazli, M. Boukadoum, and O. A. Mohamed, “A survey on computer-aided diagnosis of brain disorders through MRI based on machine learning and data mining methodologies with an emphasis on Alzheimer disease diagnosis and the contribution of the multimodal fusion,” *Applied Sciences*, vol. 10, no. 5, p. 1894, 2020.
- [18] M. Cenek, M. Hu, G. York, and S. Dahl, “Survey of image processing techniques for brain pathology diagnosis: challenges and opportunities,” *Frontiers in Robotics and AI*, vol. 5, 2018.
- [19] M. Georgescu, L. Haidar, A. F. Serb, D. Puscasiu, and D. Georgescu, “Mathematical modeling of brain activity under specific auditory stimulation,” *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 6676681, 20 pages, 2021.
- [20] A. E. Kovtanyuk, A. Y. Chebotarev, N. D. Botkin, V. L. Turova, I. N. Sidorenko, and R. Lampe, “Nonstationary model of oxygen transport in brain tissue,” *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 4861654, 9 pages, 2020.
- [21] J. J. M. van Griethuysen, A. Fedorov, C. Parmar et al., “Computational radiomics system to decode the radiographic phenotype,” *Cancer Research*, vol. 77, no. 21, pp. e104–e107, 2017.
- [22] J. E. van Timmeren, D. Cester, S. Tanadini-Lang, H. Alkadhi, and B. Baessler, “Radiomics in medical imaging—‘how-to’ guide and critical reflection,” *Insights Imaging*, vol. 11, no. 1, p. 91, 2020.
- [23] P. J. LaMontagne, T. L. Benzinger, J. C. Morris et al., “OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease,” *Journal of Chemical Information and Modeling*, vol. 53, no. 9, pp. 1689–1699, 2013.
- [24] P Organization, “Pyradiomics documentation,” 2019, <http://www.radiomics.io/>.
- [25] S. S. F. Yip and H. J. W. L. Aerts, “Applications and limitations of radiomics,” *Physics in Medicine and Biology*, vol. 61, no. 13, pp. R150–R166, 2016.
- [26] A. Traverso, L. Wee, A. Dekker, and R. Gillies, “Repeatability and reproducibility of radiomic features: a systematic review,” *International Journal of Radiation Oncology • Biology • Physics*, vol. 102, no. 4, pp. 1143–1158, 2018.
- [27] J. C. Morris, “Clinical Dementia Rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type,” *International Psychogeriatrics*, vol. 9, SUPPL. 1, pp. 173–176, 1997.
- [28] B. Fischl, “FreeSurfer,” *NeuroImage*, vol. 62, no. 2, pp. 774–781, 2012.
- [29] Z.-H. Zhou, “Ensemble learning,” in *Encyclopedia of Biometrics*, Springer, 2009.
- [30] R. Maclin, “Popular ensemble methods : an empirical study popular ensemble methods : an empirical study,” *Journal of Artificial Intelligence Research*, vol. 11, pp. 169–198, 2016.
- [31] M. P. Sesmero, A. I. Ledezma, and A. Sanchis, “Generating ensembles of heterogeneous classifiers using stacked generalization,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 5, no. 1, pp. 21–34, 2015.
- [32] A. Chaddad, C. Desrosiers, and T. Niazi, “Deep radiomic analysis of MRI related to Alzheimer’s disease,” *IEEE Access*, vol. 6, pp. 58213–58221, 2018.
- [33] F. Feng, P. Wang, K. Zhao et al., “Radiomic features of hippocampal subregions in Alzheimer’s disease and amnesic mild cognitive impairment,” *Frontiers in Aging Neuroscience*, vol. 10, pp. 1–11, 2018.
- [34] Y. Li, J. Jiang, J. Lu, J. Jiang, H. Zhang, and C. Zuo, “Radiomics: a novel feature extraction method for brain neuron degeneration disease using 18F-FDG PET imaging and its implementation for Alzheimer’s disease and mild cognitive impairment,” *Therapeutic Advances in Neurological Disorders*, vol. 12, 2019.
- [35] K. Zhao, Y. Ding, Y. Han et al., “Independent and reproducible hippocampal radiomic biomarkers for multisite Alzheimer’s disease: diagnosis, longitudinal progress and biological basis,” *Scientific Bulletin*, vol. 65, no. 13, pp. 1103–1113, 2020.
- [36] T. R. Li, Y. Wu, J. J. Jiang et al., “Radiomics analysis of magnetic resonance imaging facilitates the identification of pre-clinical Alzheimer’s disease: an exploratory study,” *Frontiers in Cell and Developmental Biology*, vol. 8, pp. 1–13, 2020.