

Swin Transformer-Based Segmentation and Multi-Scale Feature Pyramid Fusion Module for Alzheimer's Disease with Machine Learning

<https://doi.org/10.3991/ijoe.v19i04.37677>

Nasr Gharaibeh¹, Ashraf A. Abu-Ein¹, Obaida M. Al-hazaimeh¹(✉),
Khalid M.O. Nahar², Waleed A. Abu-Ain³, Malek M. Al-Nawashi¹

¹Al-Balqa Applied University, Irbid, Jordan

²Yarmouk University, Irbid, Jordan

³Applied College, Taibah University, Yanbu, Saudi Arabia

dr_obaida@bau.edu.jo

Abstract—Alzheimer Disease (AD) is the ordinary type of dementia which does not have any proper and efficient medication. Accurate classification and detection of AD helps to diagnose AD in an earlier stage, for that purpose machine learning and deep learning techniques are used in AD detection which observers both normal and abnormal brain and accurately detect AD in an early. For accurate detection of AD, we proposed a novel approach for detecting AD using MRI images. The proposed work includes three processes such as tri-level pre-processing, swin transfer based segmentation, and multi-scale feature pyramid fusion module-based AD detection. In pre-processing, noises are removed from the MRI images using Hybrid KuanFilter and Improved Frost Filter (HKIF) algorithm, the skull stripping is performed by Geodesic Active Contour (GAC) algorithm which removes the non-brain tissues that increases detection accuracy. Here, bias field correction is performed by Expectation-Maximization (EM) algorithm which removes the intensity non-uniformity. After completed pre-processing, we initiate segmentation process using Swin Transformer based Segmentation using Modified U-Net and Generative Adversarial Network (ST-MUNet) algorithm which segments the gray matter, white matter, and cerebrospinal fluid from the brain images by considering cortical thickness, color, texture, and boundary information which increases segmentation accuracy. The simulation of this research is conducted by Matlab R2020a simulation tool, and the performance of this research is evaluated by ADNI dataset in terms of accuracy, specificity, sensitivity, confusion matrix, and positive predictive value.

Keywords—Alzheimer Disease (AD), Mild Cognitive Impairment (MCI), skull stripping, segmentation, VGG-16, swin transfer

1 Introduction

Alzheimer's disease is one of the common causes of elderly death which is an irreversible and progressive neurological brain disorder that causes the brain to shrink.

It slowly destroys the memory and thinking skills then the ability to carry out the simplest task which is serious enough to hamper daily life. Alzheimer's disease is mainly caused by an abnormal build-up of protein in and around brain cells where one of the protein amyloid deposit and form plaque around the brain [1, 2]. Due to aging, the diagnoses of Alzheimer's disease are tedious task; therefore, Magnetic Resonance Imaging (MRI) is used for disease detection to help doctors [3–6].

Certain processes are involved in Alzheimer's disease detection from MRI images, the processes are pre-processing, segmentation, feature extraction and classification. Initially, the MRI images are pre-processed due to noise-prone characteristics and contain non-brain tissues (eye, fat, muscles, dura, scalp, skin, etc.) where some existing works nevermore perform skull stripping and several existing works don't consider noise reduction (i.e. Rician noise, Gaussian noise, salt and pepper noise, etc.) which moderate them with less classification accuracy [7–12]. After pre-processing, segmentation was performed to increase the classification accuracy and reduce complexity. Segmentation is one of the major processes to segment the gray matter, white matter and cerebrospinal fluid to extract the significant features for classification. Some of the existing work doesn't perform segmentation and the majority of the existing works accomplished the segmentation using automated image analyzing tool such as statistical parametric mapping (SPM), FreeSurfer and FSL-FAST4 which consume more computational time to perform full segmentation, lack of validation, automated methods in estimating the volume will not accurate and the automated tools compare intensity to atlas on to guide segmentation this augment issues of loss of fine information which leads to segmentation problem and affect the segmentation accuracy [11, 13–16].

In most of the existing work, the neural network algorithms are used for feature extraction and classification which only extract single feature or limited features that are insufficient to perform Alzheimer's disease classification [17, 18]. The existing works utilize machine learning (ML), neural network and deep learning (DL) algorithms, the ML algorithms such as support vector machine, decision tree, k-nearest neighbor, random forest etc., which has high training complexity due to large number of trees generated while feature extraction and some of the algorithms are not appropriate for large dataset [12, 17, 19]. The ML limitations are resolved by deep learning which employs neural network for feature extraction and classification such as convolutional neural network, multilayer perceptron and radial basis function [20–24]. However, this algorithm generates large number of hidden layers, high convergence weight and consumes high computational time which increases the high complexity and affects the classification accuracy. The challenges faced by automated tool while segmentation was addressed by swin transformer-based semantic segmentation then the provocation and insufficiency while feature extraction was solved by multi-scale feature extraction with effective architecture which reduces the high complexity and high false positive rate.

For Alzheimer's disease detection, most of the existing works used neural networks for feature extraction and classification. However, it increases the high complexity and reduces the accuracy of classification. Several important problems present in the existing work are described as follows in Table 1 [1].

Table 1. Most relevant problems – state-of-the-art methods

Problems	Description
Lack of Pre-processing	Most of the existing work performs pre-processing, whereas some works don't perform skull stripping and noise reduction which affects the classification accuracy due to presence of noise and non-brain tissues (eye, scalp, fat, dura, skin etc.).
Inadequate features	Many of the existing works used the texture analysis to extract texture features. And some of them extracted only the limited features (low-level features) where these features are insufficient to classify the disease and affect the true positive rate.
High Complexity	In some of the existing works the features are extracted from the entire MRI image without segmenting the region of interest which increases the high complexity. Several existing works utilized the convolutional neural network which generates numerous unwanted layers while feature extraction which leads to high complexity. Then most of the works employed support vector machine for classification which are not suitable for large dataset.

We are motivated the abovementioned problems with the aim of detecting and classifying the Alzheimer's disease (AD) using MRI images. In addition, this research identifies the problem of considering lack of pre-processing, inadequate feature, high complexity and so on. The main objective of this research is to perform mathematical modeling for classifying the Alzheimer's disease using MRI images with low false positive rate and high accuracy. The remaining objectives of this research are described as follows:

1. To enhance the quality of MRI images, pre-processing is initialized by noise reduction, removal of non-brain tissues and bias field correction that increase the true positive rate.
2. To extract the feature effectively, semantic segmentation is performed to segment the gray matter, white matter and cerebrospinal fluid which reduce the high complexity.
3. To increase the classification accuracy, significant features are extracted in multi-scale from the segmented image which reduces the high false positive rate.

To make it clear, this research mainly focuses on detecting AD accurately using MRI images with high accuracy. The major contributions of this research are listed as follows:

1. Initially, we perform tri-level pre-processing, in first level we perform noise removal by HKIF algorithm which removes the noise and preserves the edge information of the images, in second level we perform skull stripping by extracting brain images using GAC which increases accuracy and reduces complexity. In third level, we perform bias filed correction to increase the intensity of the images which is done by EM algorithm that helps to increase the quality of the images.
2. Swin transformer-based segmentation is performed to segment gray matter, white matter, and cerebrospinal fluid of brain which is done by ST-MUNet algorithm, this process increases the detection accuracy of AD, and MCI.
3. Multi-scale feature extraction based AD detection is performed by MSFP-VGG16 which extracts multiple features (i.e, structural, statistical, edge, blobs, color,

and contour) in multi-scale format that provides additional information of the features which increases classification accuracy. The proposed MSFP-VGG16 classifies the images into three classes namely normal, AD, and MCI.

The performance of this research is evaluated by various performance metrics such as accuracy, sensitivity, specificity, confusion matrix, and positive predictive value. This paper is structured as follows; Section 3 explains the detailed summary of the literature review and its research gap. Section 2 explains the major problem statement which are addressed from the existing works and solved by the proposed work. Section 4 illustrates the research methodology which explains the research methodology, pseudocode, algorithm, and mathematical formulations. Section 5 illustrates the detailed description of the experimentation study of the proposed work. Section 6 concludes the conclusion.

2 Problem statement

In [23], an automated detection model for Alzheimer's detection using T2 weighted brain MRI images is proposed. Here, Alzheimer disease is classified by Support Vector Machine (SVM) Poly-1 and Random Forest algorithm. The major problems of this research are expressed as follows:

1. In this work, the features are extracted from the entire image without focusing on region of interest (GM, WM, and CSF) that increase the high complexity.
2. The local binary pattern (LBP) was employed for feature fusion which produces rather long histogram that leads to slow down the model performance and affect the true positive rate.
3. Here, the model does not allow for training larger dataset due to its low computation capacity therefore the model does not suitable for large dataset that affects the classification accuracy.

The author proposed a deep learning method for Alzheimer classification using T1 weighted MRI images [25]. The multilayer perceptron (MLP) was used for classifying cognitive normal (CN), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI) and AD. The research issues are listed as follows,

1. In this work, the pre-processing was performed for bias field correction, B1 non-uniformity and grad-warp where skull stripping is the main process, lack of consideration leads to less true positive rate.
2. Here, only texture features were considered for Alzheimer's disease classification. However, lack of considering other significant features (i.e., structural, statistics, edge etc.) will affect the classification performance and its accuracy.
3. The multilayer perceptron (MLP) was employed for classification, where it generates many hidden nodes and convergence of the weight will be very slow which degrades the model performance and it leads to high complexity.

The author proposed a deep fusion method for Alzheimer's detection using T1 MRI image from ADNI dataset [26]. Searchlight was utilized for feature extraction from specific position of brain and the SVM classifies the images into AD and normal. The main research problems of this research are listed as follows,

1. Here, the segmentation was performed on raw image, where the presence of non-brain tissues, noise and intensity inhomogeneity affect the image quality and leads to high false positive rate.
2. In this work, statistical parametric mapping (SPM) was utilized for segmentation, where it was an automatic image analysing tool, which is not suitable for any other dataset's segmentation process this affects the classification accuracy.
3. The support vector machine (SVM) was employed for Alzheimer's disease classification which doesn't perform well on large dataset due to its high training complication that leads to high complexity.

The author proposed a method for early diagnosis of Alzheimer's detection using T2 weighted MRI from ADNI database [27]. Convolutional Neural Network (CNN) was utilized to perform feature extraction and classification into three classes such as normal, MCI, and AD.

1. In this work, the pre-processing was performed to enhance the image quality and skull stripping where bias field correction and noise reduction were not considered which leads to affect the image quality and classification accuracy.
2. Here, the clustering-based segmentation was performed where random initialization of centroid of cluster lack of consistency in determination of abnormalities that leads to high false positive rate.
3. The feature extraction and classification were performed by Convolutional Neural Network (CNN) which generates large number of unwanted layers while feature extraction that leads to high complexity.

3 Related work

In paper [28], the authors proposed a machine-learning model for Alzheimer's disease diagnosis using T1 MRI images. Initially, the pre-processing was performed using FreeSurfer software for skull stripping, image registration, cortical segmentation, and thickness estimation then the brain regions are segmented using the same software and 63 volumetric features are obtained. The Welch's test was used for feature selection and 22 significant features are selected for classification. Finally, the classification was performed by Random Forest which classifies into three classes as Normal, MCI and Alzheimer's disease. Here, the classification was performed using Random Forest which generates a large number of trees for a large dataset during classification increases high complexity and high latency. Authors in [29], proposed a model for Alzheimer's disease detection using T1 weighted MRI images. Initially, the pre-processing was performed for skull removal using FreeSurfer, then sampling, clipping, and intensity normalization are executed. The feature extraction and classification were accomplished using Modified U-net, which has skip-connection to fuse low-level information and high-level semantic features. Finally, the structural features are used for classification which classifies normal, early cognitive impairment, late mild cognitive impairment, and Alzheimer's disease. In this work, the feature extraction was performed on the overall image instead of focusing on the region of interest which leads to high complexity in extracting features from the overall image.

The authors proposed a deep learning model for Alzheimer's and Dementia disease diagnosis using MRI images [5]. Initially, the pre-processing was performed to balance the dataset using Synthetic minority over-sampling technique (SMOTE) and the normalization was executed. The DEMNET network was implemented for feature extraction and disease detection where the convolutional layers extract the discriminate features from the pre-processed image. Finally, the softmax layer classifies into four classes such as non-AD, very mild AD, mild AD, and moderate AD. Here, the pre-processing was only performed to balance the dataset where the presence of non-brain tissues, noise, and intensity in homogeneity affects the classification accuracy. In paper [30], the author's deep and conventional machine learning model for Alzheimer's disease detection using T1 MRI images. Initially, the N4 algorithm was used to correct the non-uniformity and skull stripping, and then a non-rigid B-spline transformation model was implemented for pair wise registration. The gray matter was segmented and the feature map was extracted using a statistical parametric mapping from the pre-processed image. Finally, the deep convolutional neural network and conventional SVM classify into normal, mild cognitive impairment, and Alzheimer's disease. The gray matter segmentation and feature map were extracted by statistical parametric mapping which is an analyzing tool and it will not suitable for all the datasets that leads to high false positive rate. Authors in [31], proposed an automatic classification method for Alzheimer's detection using T1 MRI images. Initially, the FMRIB's software library (FSL) and brain extraction tool were used for skull stripping then the gray matter, white matter, and CSF are segmented using FSL-FAST4 automated segmentation tool. Then FreeSurfer software was implemented for feature extraction based on cortical thickness and 63 features are extracted. The Six classifiers are executed separately for classification into normal, early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease. Finally, the SVM with radial basis function obtained high accuracy. In this work, the feature extraction was performed using FreeSurfer, which resulted in errors in terms of precise borders of a different brain region, which will affect the true positive rate.

In paper [32], authors proposed a deep learning method for early diagnosis of Alzheimer's disease using MRI images. Initially, the pre-processing was performed using non-local mean algorithm for normalization, standardization, and resize then the data augmentation was performed to avoid data scarcity. The feature extraction and classification were performed using two methods, the first method was employed by CNN and the second method was using the transfer learning method. Finally, the features are extracted and classified into normal, EMCI, LMCI, and AD. In this work, the features are extracted from the overall images without segmenting the GM, WM, and CSF which leads to high complexity. In paper [27], the authors proposed a deep learning method for the early detection of Alzheimer's disease using T2 MRI images from two datasets. Initially, the pre-processing was performed to remove the non-brain tissues using the skull stripping algorithm. The Enhanced independent component analysis was used for segmenting the gray matter which is used to estimate the atrophy changes. The CNN based on inception type of block was proposed for feature extraction and classification which extract the deep features. Finally, the softmax layer classifies into normal, AD, and MCI. The convolutional neural network was used for feature

extraction and classification, where it takes long time to train, especially with large dataset which leads to high false positive rate. The authors proposed a deep transfer learning model for Alzheimer's disease detection using MRI images [33]. Initially, the images are channelized, resize and the images were cropped to remove the white space and enhance the data. Then, the images are augmented to balance the dataset and the CNN was utilized for feature extraction, the convolutional layers at starting of CNN's filters extract the low-level features such as edge, blobs, and color and the network end learns the global, high level and complicated features. Finally, by using CNN based transfer learning technique the softmax classifies into AD, MCI, and normal. The pre-processing was performed to channelize, resized, and crop to remove the white space, where lack of skull stripping, noise reduction, and bias field correction affects the classification accuracy.

Feature selection-based AD classification was performed by using optimization and machine learning algorithms [34]. Initially, the input images were pre-processed and normalized for improving the quality of the images. The pre-processed images were sent to feature selection phase which were done by genetic with logistic regression algorithm. The selected features were sent to classification phase which was done by support vector machine, extreme gradient boosting, and random forest algorithm which classified the images into AD/NC, AD/MCI. The performance of this research was evaluated by ADNI dataset that prove the proposed work achieved better performance in AD detection and classification. Here, machine learning algorithm was proposed for AD detection which takes high training time for large dataset that reduces the training performance of AD detection and classification.

Deep learning-based AD detection using MRI images [35]. This research includes three processes such as pre-processing, feature extraction, and classification. Here, 3D MRI images were taken as an input, and then pre-processing was performed for image slicing. The sliced images were sent to the CNN algorithm which performs feature extraction. Here, transfer learning was applied to the CNN for efficient feature extraction. The extracted features were fed into RNN which classified the images into AD or NC. The performance of this research was evaluated by ADNI dataset. AD detection was performed by incorporating CNN and VGG-16 using MRI images [36]. Initially, MRI images were collected from kaggle for AD detection. After that, feature extraction was performed by VGG-16, based on the extracted features neural network was classified the images into four classes such as mild, moderate, very mild, and normal. The experimental result demonstrates that the proposed work achieved superior performance in accuracy, recall, precision, f-measure, AUC and ROC curve. Here, features were extracted from the noisy images which lead to less classification and detection accuracy. Automatic detection of AD and MCI was performed using MRI images [37]. Initially, MRI images were collected from ADNI dataset for AD and MCI detection. Then perform image resizing and slicing for accurate detection of AD. Here, deep convolutional neural network extracted the features automatically from the sliced MRI images. The proposed convolutional neural network includes depth wise and point wise convolutions for increasing the performance of feature extraction which extracts the features in multiple views. The classification includes two classes such as AD and MCI. Finally, the experimental results show that the proposed work achieved good performance in terms of accuracy, AUC, and ROC curve.

4 System model

In this research, we mainly focus on the Mathematical modeling of Alzheimer’s disease detection using Magnetic Resonance Imaging (MRI). This proposed work improves the classification accuracy by performing pre-processing, segmentation and feature extraction. We have considered T1 weighted MRI image from Alzheimer’s disease Neuro imaging Initiative (ADNI) database. Figure 1 represents the architecture of the proposed ST-Seg model. This proposed work consists of three sequential phases such as: Tri-level Pre-processing, Swin Transformer based Segmentation using MU-net and GAN, and Multi-scale feature pyramid fusion module and Classification.

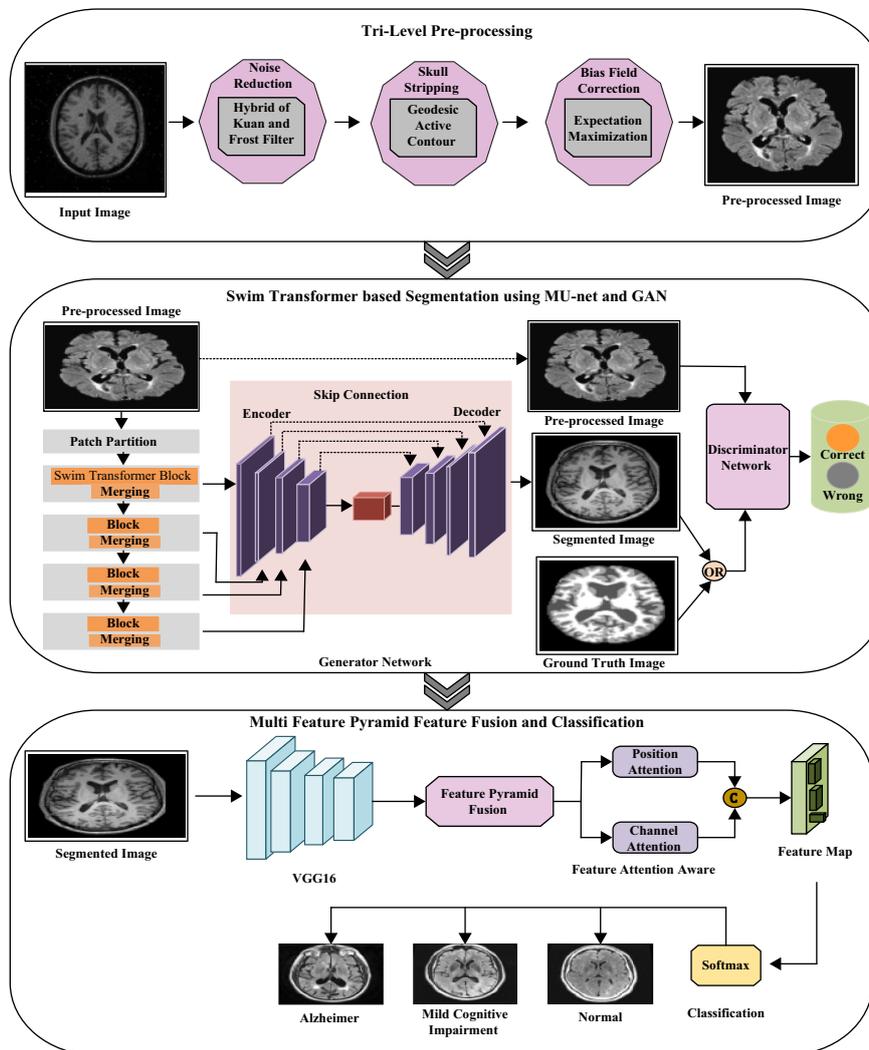


Fig. 1. Architecture of proposed ST-Seg model

4.1 Tri-level Pre-processing

In this work, we have taken MRI images from ADNI dataset where we perform tri-level pre-processing to enhance the image quality by accomplishing skull stripping, noise reduction and bias field correction.

Noise reduction. The MRI images have certain noises such as Speckle noise, Gaussian noise, and so on. Removal of these noises is required to enhance the MRI image quality. For this purpose, we proposed Hybrid of Kuan and Improved Frost filters (HKIF) where these filters remove noise and smoothen the image without removing edge features which is shown in Figure 2. Here, the noisy images are fed into HKIF whereas the Kuan filter and frost filter provides the noise removed images which are incorporated and provided the final noise removal image. The Kuan filter depends on the effective count of looks which is the main parameter since that needs to adjust for every image to tune the parameters of the filter for obtaining efficient results in noise removal. Here, the Kuan filter automatically evaluated the optimal parameter of the filter based upon edge pixels that provides effective results. Let consider W is the window centre of the filter with pixel coordinates (p, q) . C_w is the centre pixels in W and δ is weighted value. The Kuan filter gray level pixel positions are given as follows,

$$F(p, q) = \varepsilon_w + \delta(C_w - \varepsilon_w) \quad (1)$$

where ε_w represent the mean value inside the filter window W . The weight value of the filter is evaluated as follows,

$$\delta = \frac{1 - C_\varepsilon^2 / C_i^2}{1 + C_\varepsilon^2} \quad (2)$$

$$C_i = \rho_w / \varepsilon_w \quad (3)$$

where ρ_w is represent the standard deviation in W , and $C_\varepsilon = 1 / \sqrt{look_n}$. The $look_n$ metric efficiently manages the smoothing volume which is applied in the image by changing the coefficient of noise variation C_ε . The value of $look_n$ is represent the efficient number of looks (γ) of the given image which is closure to the actual $look_n$ of the noisy image. If the $look_n$ has smaller values, then it represents more smoothening, otherwise it preserves the edge information of the image. The value of $look_n$ is adjusted to control the performance of the filter. The calculation of γ is defined as follows,

$$\gamma = (\mu_r / \rho_r)^2 \quad (4)$$

where μ_r represent the mean value of the even image region (r), and ρ_r represent the standard deviation of the even image region. For identifying the uniform areas of the image, the image is divided into 25×25 sub images, and calculated the γ for each sub images, finally average two sub images value of γ for obtaining final γ value. Let consider $look_n = \gamma$ at initial stage, and then the values are adjusted by manually which is calculated as follows,

$$look_n = (1 + a)\gamma \tag{5}$$

where a represent the coefficient of adjustment parameter, and the value of a changes with variance of noise (a_n) and detail of image (a_d) which is formulated as follows,

$$a = a_d + a_n \tag{6}$$

In this way, image noises are removed by using kuan filter. Parallely, the noisy images are sent to the improved frost filter for noise removal. Here, the window size and tuning factors are adaptively changed based on regional characteristics. The mathematical representation of frost filter is defined below,

$$\mathbb{I}(i, j) = \sum_x \sum_y p_{xy} \mu_{xy} / \sum_x \sum_y \mu_{xy} \mu_{xy} = e^{-kh_i^2 D_{xy}} \tag{7}$$

where (i, j) represent the current pixel location, $\mathbb{I}(i, j)$ is filter output, and p_{xy} is the pixels values between window centre at (i, j) , $k(k > 0)$ represent the factor of tuning parameter, and h_i represent the variation coefficient which is defined as ration between sample mean and standard deviation, and D_{xy} is distance between the pixels to centre pixels. The value of k is small, then improved frost filter perform noise removal efficiently, but it does not preserve edge information accurately. The value of k is increased, then it effectively preserves the edge information effectively. Hence, the size of the window is adjusted dynamically based on adaptive windowing method. Let assume (i, j) represent the current location of the pixel, and the threshold calculation of the current window is defined as follows,

$$h_{ij} = \frac{\rho_{ij}}{\mu_{ij}} \tag{8}$$

$$Tr_{ij} = \beta \left(1 + \sqrt{\frac{1 + 2\rho_f^2}{(S_{w_{ij}} - 1)}} \right) \times \rho_f \tag{9}$$

where μ_{ij} represent the mean value, and ρ_{ij} is the standard deviation of the current window. Tr_{ij} is represent the threshold value which effectively smoothed the image. ρ_f^2 denotes the noise variance and $S_{w_{ij}}$ represent the size of the current window, β is denotes the system parameter. Here, window centre size is changed dynamically that is expressed as follows,

$$S_{w_{ij}} = \begin{cases} \text{Min}[S_{w_{ij}} + 2, S_{w_{Max}}], & \text{if } S_{w_{ij}} \leq Tr_{ij} \\ \text{Max}[S_{w_{ij}} - 2, S_{w_{Min}}], & \text{if } S_{w_{ij}} > Tr_{ij} \end{cases} \tag{10}$$

where S_{wMin} represent the minimum window size, and S_{wMax} and maximum window size. Here, variation coefficient h_{ij} is not bigger than Tr_{ij} , then the window size is gradually increases for obtaining better results. In this way, improved frost filter performs noise removal and edge preservation. Finally, two outputs of kuan filter and improved frost filter are combined and produce final noise removed images with effective edge preservation.

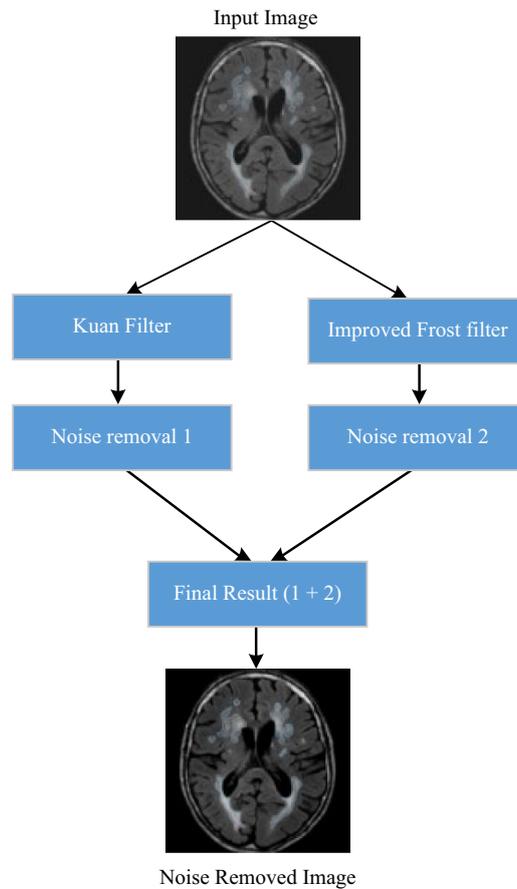


Fig. 2. Process of noise reduction

Skull stripping. The main process of pre-processing is skull stripping in Alzheimer’s disease detection, which is the process of segmenting the brain tissues by removing non-brain tissues (fat, scalp, muscles, dura etc.) from surrounding region which improves the robustness of the disease detection. For that purpose, we proposed GAC is a contour model that adjusts the smooth curve and constructs closed contour for region. Figure 3 represents the skull stripping input and output. Initially, we defined the edge indicator of the image (E_t) which is expressed as follows,

$$E_l = \frac{1}{1 + |\nabla g_\rho \times I|^2} \quad (11)$$

where $g_\rho \times I$ is used to remove the noise by performing Gaussian smoothing, and ∇ represent the function of gradient. The function of edge indicator takes small values on image boundaries compared to other regions. So, we can introduce level set function $(lsf)\mathcal{G} : \Omega \rightarrow r$ which is formulated as follows,

$$\mathcal{G}(s) = \begin{cases} -\text{Inf}_{t \in \partial B} \|s - t\|_2, & s \in B_{in} \\ 0, & s \in \partial B \\ \text{Inf}_{t \in \partial B} \|s - t\|_2, & s \in B_{out} \end{cases} \quad (12)$$

where $\|s - t\|_2$ is represent the distance between s and t , and B_{in} , B_{out} and ∂B represent the outside, boundary, and inside of the object respectively. Here, segmentation contour \mathcal{G} has the value of zero. The energy calculation of proposed GAC is defined as follows,

$$\mathfrak{Z}(\mathcal{G}) = \int_{\Omega} E_1 \dot{\eta}_\epsilon(\mathcal{G}) |\nabla \mathcal{G}| dx \quad (13)$$

$$\dot{\eta}_\epsilon = \frac{\epsilon}{\pi(\mathcal{G}^2 + \epsilon^2)} \quad (14)$$

where ∇ represent the gradient operator, here energy is calculated the edge indicator line integral with zero level set contour, which results slight length curve. The GAC energy function is integrated with penalty and area which are expressed as follows,

$$\mathfrak{L}(\mathcal{G}) = \int_{\Omega} E_l \mathfrak{h}(-\mathcal{G}) dx \quad (15)$$

$$\mathfrak{R}(\mathcal{G}) = \frac{1}{2} \int_{\Omega} (|\nabla \mathcal{G}| - 1)^2 dx \quad (16)$$

where $\mathfrak{h}(\mathcal{G})$ represent the function of Heaviside, the weight value of area and initial contour values are essential. If the initial value of contour is present inside the objects, the weight values of area must be positive so that shrink the zero lsf of contour through evolution. On the divergence, if the initial contour is pinched inside the object, the area weight values are negative so that it enlarges to the zero-level set contour. These types of shrinking and expanding procedure are followed in GAC for extracting brain regions, and removing skull regions in the MRI images.

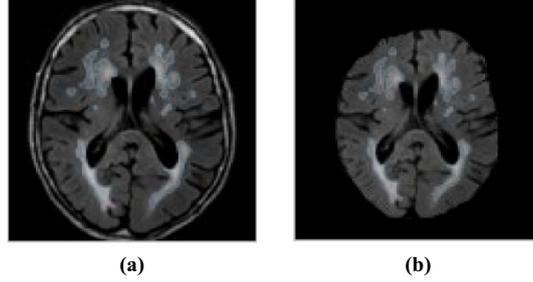


Fig. 3. (a) Noise removed image (b) Skull stripping image

Pseudocode for Pre-processing
<p>Input: Input MRI image I Output: Pre-processed Image (\mathcal{P}_I) Begin</p> <p style="padding-left: 20px;">For each/do</p> <p style="padding-left: 40px;">Initialize Kuan filter</p> <p style="padding-left: 40px;">Perform noise removal by Kuan filter from Eq. (1)–Eq. (6)</p> <p style="padding-left: 40px;">Adjust γ by manually using Eq. (4)</p> <p style="padding-left: 40px;">Obtain noise removed image $NR_I(1)$ by Kuan filter</p> <p style="padding-left: 40px;">Initialize improved frost filter (IF)</p> <p style="padding-left: 40px;">Perform noise removal by IF using Eq. from (7)–Eq. (10)</p> <p style="padding-left: 40px;">Adjust the S_{wij} dynamically using Eq. (10)</p> <p style="padding-left: 40px;">Obtain noise removed image image $NR_I(2)$ by IF</p> <p style="padding-left: 40px;">$NR_I \leftarrow$ image $NR_I(1)$ + image $NR_I(2)$</p> <p style="padding-left: 20px;">End For</p> <p style="padding-left: 40px;">Return NR_I</p> <p style="padding-left: 20px;">For each NR_I do</p> <p style="padding-left: 40px;">Perform skull stripping using Eq. (11)–Eq. (16)</p> <p style="padding-left: 40px;">Perform bias field correction by EM using Eq. (12) and Eq. (13)</p> <p style="padding-left: 40px;">Obtain pre-processed image \mathcal{P}_I</p> <p style="padding-left: 20px;">End for</p> <p style="padding-left: 40px;">Return \mathcal{P}_I</p> <p>End</p>

Bias field correction. The bias field correction is the step of removing intensity in homogeneity or non-uniformity that occurs from magnetic field. The bias field is removed by cleaning the spatial frequencies. In MRI images, the useful information is presented in the higher spatial frequency compared to intensity non-uniformity, hence we need to eliminate low spatial frequency non-uniformity. In this research, the bias field correction is performed using by Expectation-Maximization (EM) algorithm which makes the difference of brain tissues clearly and it is mainly performed to reduce classification errors. The main objective of EM is to detect maximum probability solution for bias field correction. The observed intensity non-uniformity from N images which are expressed as $N_I = \{n_i(1), n_i(2), \dots, n_i(m)\}$. The latent elements of the observed non-uniformity are $V_m = \{V_m(1), V_m(2), \dots, V_m(N)\}$. Let $\{N_I, V_m\}$ is a complete data and its likelihood function is expressed as $\ln l(N_I, V_m | \alpha)$, whereas α represent the total metrics of the model. The proposed EM algorithm is used to maximize the

probability and acquire entire metrics in the model by performing two steps which are *istep* and *jstep*. In *istep*, the distribution of posterior for V_m is obtained by using the present model metric α^{old} . The complete posterior distribution-based likelihood function is defined as follows,

$$Z(\alpha, \alpha^{old}) = i[\log I(N_i, V_m | \alpha) | N_i, \alpha^{old}] \quad (17)$$

In *jstep*, the likelihood function is maximized to inform the model metrics which are expressed as follows,

$$\alpha^{old} = \text{Arg max}_{\alpha} Z(\alpha, \alpha^{old}) \quad (18)$$

where α^{old} represent the current metric of the model, step *i* and *j* are performed alternatively until the meeting constraint is met. In this way, bias filed correction is performed in this research which increases the intensity of the brain tissues that helps to detect AD accurately.

Swin transformer based segmentation. After successful pre-processing, we perform Swin transformer-based segmentation to improve the classification accuracy. In this research, Swin Transformer based Segmentation using Modified U-net and Generative Adversarial Network was implemented for segmentation by considering cortical thickness, color, texture and boundary information, the Gray matter (GM), White matter (WM) and Cerebrospinal fluid (CSF) are segmented from the pre-processed image. The detailed architecture of proposed GAN is shown in Figure 4. In this research, we proposed a hierarchical transformer that is computed with shifted windows which consist of swin based Multi-head Self Attention (MSA) module. The Swin transformer is a type of vision transformer, it built the hierarchical feature map by merging image patches in deep layers and it has low computation complexity and high accuracy. The input of the swin is divided into multiple patches of (h', w', d') and dimension of $h' \times w' \times d' \times u$. Initially, we used the patch partition to made an order of 3D tokens with dimension of $\left[\frac{h}{h'} \right] \times \left[\frac{w}{w'} \right] \times \left[\frac{d}{d'} \right]$ and develop them into implanting space with dimension c .

The swin transfer windowing method used a window size $\mathfrak{H} \times \mathfrak{H} \times \mathfrak{H}$ to regularly divide the 3D tokens into $\left[\frac{h'}{\mathfrak{H}} \right] \times \left[\frac{w'}{\mathfrak{H}} \right] \times \left[\frac{d'}{\mathfrak{H}} \right]$ sections at a given layer in the transformer encoder.

Later in layer $L+1$, then the window is partitioned and shifted by $\left(\left[\frac{\mathfrak{H}}{2} \right], \left[\frac{\mathfrak{H}}{2} \right], \left[\frac{\mathfrak{H}}{2} \right] \right)$ voxels. The output of layers in encoder is expressed as follows,

$$\hat{O}(L) = w - \text{MSA}(N_L(O(L-1))) + O(L-1) \quad (19)$$

$$O(L) = \text{MLP}(N_L(\hat{O}(L))) + \hat{O}(L) \quad (20)$$

$$\hat{O}(L+1) = sw - \text{MSA}(N_L(O(L))) + O(L) \quad (21)$$

$$O(L+1) = \text{MLP}(N_L(\hat{O}(L+1))) + \hat{O}(L+1) \quad (22)$$

where w -MSA, and sw -MSA represent the even and window partitioning MSA module respectively, and $\hat{O}(L)$ represent the output of w -MSA, and $\hat{O}(L+1)$ is the output of sw -MSA, N_L represent the normalization of layer. The output of the swin transfer is fed to the encode which is placed in the Generator.

The Modified U-net (MU-net) is used for semantic segmentation where the skip-connection characteristics will append every upsampled feature map at decoder with corresponding encoder and the Generative adversarial network (GAN) is deep learning-based architecture that can train and classify the medical data more effectively. The GAN's architecture consists of generator and discriminator network, the generator network includes MU-net where the swin transformer technique is applied in encoder and the decoder executes the result of segmentation. The proposed MUNet performed down sampling in encoder and up sampling in decoder with skip connections. The generator pursues to learn a map $G : e \rightarrow f$ which produces a binary map of segmentation f from input image e based on training data. The map of discriminator $\{e, f\}$ is covered of segmentation map and input image with the range in between 0 to 1, in which the discriminators evaluate the f is a ground truth mask or predicted map from generator. For segmentation, the objective function of GAN is expressed as follows,

$$Z(\alpha, \alpha^{\text{old}}) = i[\log l(N_j, V_m | \alpha) | N_j, \alpha^{\text{old}}] \quad (23)$$

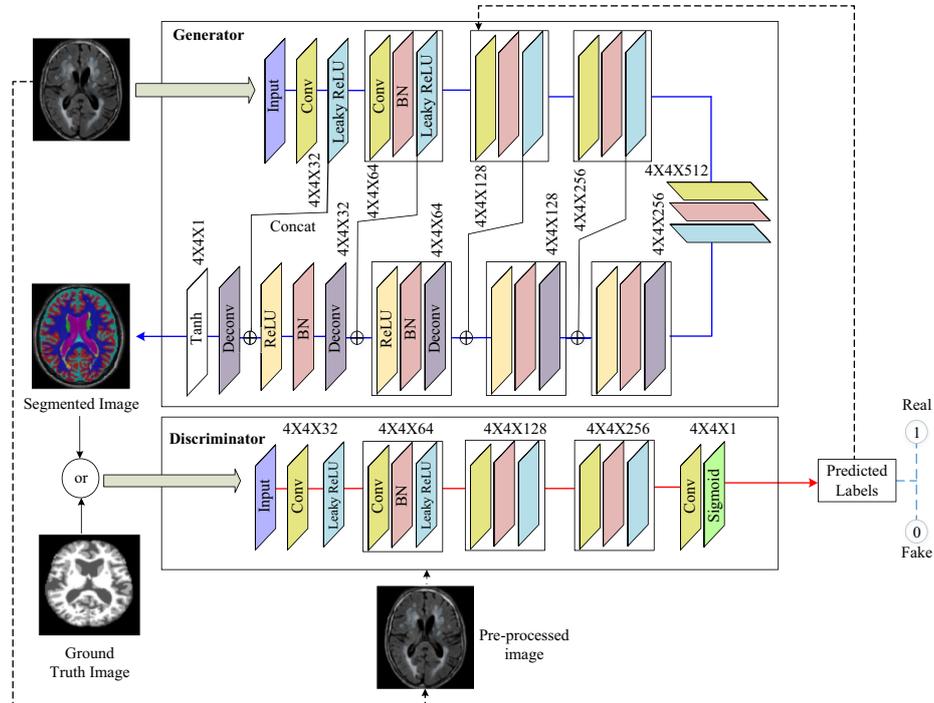


Fig. 4. Architecture of GAN for segmentation

The objective function aligns with maximum of $\check{D}(e, f)$, and minimum of $\check{D}(e, \hat{G}(e))$ that pursue to train the discriminator \check{D} for making correct decision. The generator generates the output from real data to hinder the discriminator \check{D} that makes correct decision and also minimizes the objective function. Here, loss function is calculated based on binary cross entropy which is common in segmentation which is defined as follows,

$$L_{seg}(\hat{G}) = \mathbb{E}_{e, f \sim \text{pdata}(e, f)} [-f \cdot \log \hat{G}(e) - (1 - f) \cdot \log(1 - \hat{G}(e))] \quad (24)$$

By integrating both segmentation loss and Objective function of GAN, we get the segmentation results from optimal generator is defined as follows,

$$\hat{G}^* = \text{Arg min}_{\hat{G}} [\max_{\check{D}} L_{GAN}(\hat{G}, \check{D})] + \gamma L_{seg}(\hat{G}) \quad (25)$$

where γ represent the weighting parameters which can balanced the objective functions. Here, the skip connections of the U-Net are connected between decoder and encoder allowing inserting of information in the encoder for every resolution. The proposed MU-Net includes batch normalization, Leaky ReLU function in encoder and ReLU function for decoder, and pooling layers. Here, convolutional layers with size of 4×4 tracked by leaky ReLU function and batch normalization, in the expanding arm. For mapping every feature vector to chosen number of classes such as affected region and non-affected region. We used the final convolutional layer with the size of 4×4 and tanh activation function to map the estimated values for classification. The mathematical formulation of ReLU and Leaky ReLU are expressed as follows,

$$\text{ReLU}(j) = \begin{cases} 0 & \text{if } j \leq 0 \\ j & \text{if } j > 0 \end{cases} \quad (26)$$

$$\text{Leaky ReLU}(j) = \begin{cases} j & \text{if } j > 0 \\ \infty j & \text{if } j \leq 0 \end{cases} \quad (27)$$

where ∞ denotes the small constant range between 0.1 to 0.3. The ground truth image and estimated segmentation image is given to the discriminator for detecting output is real (one) of fake (zero). The discriminator also used seventeen convolutional layers with the size of 4×4 , and batch normalization and leaky ReLU activation function with the size of 2×2 . The final output of discriminator is between zero to one that determines the probability of segmentation results. Then, the discriminator network will evaluate the segmentation result with ground-truth image whether the segmentation performs accurately. And if the segmentation doesn't perform precisely then the process repeated again to implement accurate segmentation, by this way the segmentation accuracy increase.

Pseudocode for Segmentation
<p>Input: Pre-processed Image (\mathcal{P}_I)</p> <p>Output: Segmented Image</p> <p>Begin</p> <p style="padding-left: 20px;">Initialize segmentation features</p> <p>For each \mathcal{P}_I do</p> <p style="padding-left: 20px;">Train G and D based on features</p> <p style="padding-left: 20px;">Initialize Encoder \mathbb{E}</p> <p style="padding-left: 20px;">Perform swin transfer mechanism ()</p> <p style="padding-left: 20px;">Divide \mathcal{P}_I into multiple patches with h', w', d'</p> <p style="padding-left: 20px;">Output of encoder</p> <p style="padding-left: 20px;">Learn binary map of segmentation by G</p> <p style="padding-left: 20px;">Compute objective function of GAN</p> <p style="padding-left: 20px;">Compute loss function of GAN</p> <p style="padding-left: 20px;">Obtain segmentation results by G</p> <p style="padding-left: 20px;">Compute activation function for encoder</p> <p style="padding-left: 20px;">Compute activation function for decoder</p> <p style="padding-left: 20px;">Evaluate the segmentation results by D</p> <p>End For</p> <p style="padding-left: 20px;">Return S_I</p> <p>End</p>

4.2 Multi-scale feature pyramid fusion module based AD detection

The feature extraction is performed from the segmented region, to improve the classification accuracy multi-scale feature extraction was performed. For that purpose, we proposed MSFP-VGG16 which is shown in Figure 5. Here, the VGG16 is the backbone of the architecture, which utilizes for feature extraction and classification. The VGG16 is type of CNN architecture that performs efficiently with deep network and small convolution filter which has low training cost and high performance. Initially, the VGG16 was implemented to extract features such as textural, statistical, structural, edge, blobs, color and contour then the feature map was obtained from feature pyramid fusion module which was implemented to fuse the local and global information based on multi-scale feature extraction. This will improve the feature representation and classification accuracy. Initially, we select first ten layers of VGG16 for feature extraction layer in MSFP network. After that, feature maps are attained through the feature extraction layer and forward to fusion of feature pyramid network which perform multi scale feature extraction. The multi scale feature maps are sent to the attention module for fusing both global and local information of the features. The generated features maps are divided into multiple blocks and perform group convolutions (GC) 3×3 with various dilations. The Figure 5 represents the architecture of feature pyramid fusion model, which divide the feature maps into four blocks with the size of 3×3 and dilation rate of $R(R = 1, 2, 3, 4)$ whereas the group number is increased with 2^3 in pyramid structure. The output feature map (f_m) is defined as follows,

$$f_{m_i}(b) = \begin{cases} GC(b, nb_i, G_i, R_i), & i = 1 \\ GC(f_{m_1}(b), nb_i, G_i, R_i), & i = 2 \\ \vdots \\ GC(f_{m_{k-1}}(b), nb_i, G_i, R_i), & i = k \end{cases} \quad (28)$$

where $GC(b, nb_i, G_i, R_i)$ represent the pyramid group convolution, k is the count of layer, nb_i is denotes the count of blocks, R_i is represent the dilation rate, G_i count of groups in every convolution layer. In this research, we reduce the computational cost by using group convolution operation as a replacement of vanilla convolution. The computational cost of GC is defined as follows,

$$\text{Cost}(G, t, f_{m(in)}, f_{m(out)}) = \frac{t^2 \times f_{m(in)} \times f_{m(out)} \times h \times w}{G} \quad (29)$$

where t represent the kernel size of convolution, and $f_{m(in)}$ represent the count of input in feature map, and $f_{m(out)}$ represent the count of output in feature map, $h \times w$ represent the height and weight of the feature map. The feature maps are fed into attention layers which helps to enhance the robustness of feature extraction. In this research, we used both position and channel attention mechanism.

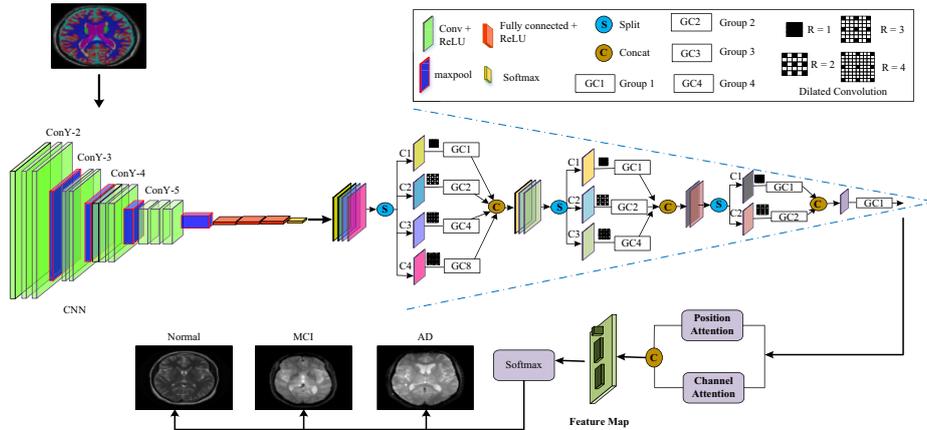


Fig. 5. AD detection using MSFP-VGG16

The position attention is used to shows the spatial relationship between every position in the image. The channel attention provides the relationship between various channels in the feature map by selecting different weight values. The output feature

map from feature pyramid fusion model is send to convolutional layer for obtaining two feature maps (G_1, G_2) which are reshaped to $R^{nc \times n}$ whereas nc represent the number of feature map channels, and $n = h \times w$ represent the image pixel counts. The final feature map output is sent to the softmax layer for getting spatial relationship between matrix $\mathfrak{z}(s) \in R^{n \times n}$ which is expressed as follows,

$$\mathfrak{z}_s^{ij} = \frac{e(G_2^j \cdot G_1^i)}{\sum_i^n e(G_2^j \cdot G_1^i)} \quad (30)$$

where \mathfrak{z}_s^{ij} represent the relationship between i th and j th position of the feature map. We multiplied the $G_1 \times \mathfrak{z}(s)$ and reshaped the result to $R^{h \times w \times nc}$ for obtaining position attention-based map. Next the spatial parameter ($\#$) is multiplied with $G(s)$ which is defined as follows,

$$G'(s) = \#(G_1^{nc \times n} \times (G_1^{nc \times n})^T \times G_1^{nc \times n}) + G^{nc \times h \times w} \quad (31)$$

Finally, we got two feature maps $(F_1, F_2 \in R^{nc \times h \times w})$ from feature pyramid fusion model. The final global connection matrix $u(c) \in R^{c \times c}$ is calculated by softmax layer that is defined as follows,

$$\mathfrak{z}_c^{ij} = \frac{e(F_2^j \cdot F_1^i)}{\sum_i^n e(F_2^j \cdot F_1^i)} \quad (32)$$

where \mathfrak{z}_c^{ij} represent the weight value of channel, after we perform matrix multiplication $u(c) \times F_2$. We multiply the scale parameter of the channel \mathfrak{J} and perform element-based addition which is defined as follows,

$$F'(C) = \mathfrak{J}((F_2^{c \times n} \times (F_1^{c \times n})^T) \times F_2^{c \times n}) + F^{c \times h \times w} \quad (33)$$

The final feature map of feature attention-based module is defined as follows,

$$A^{2c \times h \times w} = F'(C) \oplus G'(s) \quad (34)$$

where $F'(C)$, and $G'(s)$ represent the channel connection feature map and spatial connection feature map respectively, and \oplus is the concatenation operation. Finally, based on the feature extraction the softmax layer classify into three classes such as normal, AD, and MCI.

Pseudocode for AD detection	
Input:	Segmented image (S_i)
Output:	Disease Detection
Begin	
	Initialize features f
	$f \leftarrow \{t_p, st_p, s_p, e_p, b_p, c_p, Cr_p\}$
	Initialize MSFP-VGG16
For each S_i do	
	Extract f from S_i using VGG16
	Divide feature map into 4 blocks
	Obtain feature map output using eqn (23)
	Calculate computation cost of GC
	Forward the output feature map into convolution layer
	Get two feature map (G_1, G_2)
	Get spatial relationship of feature map
	Get channel relationship of feature map
	Get final feature map
	Obtain classification results (N, AD, MCI)
End For	
End	

5 Experimental results

This section illustrates the experimental results of proposed ST-Seg model which also includes four sub divisions such as dataset description, simulation setup, comparative analysis, and research summary. This section also proved that the proposed work achieved superior performance in AD detection compared to existing approaches.

5.1 Dataset description

In this research, we have used Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset for AD detection. This dataset includes T1 weighted brain MRI images. This dataset provides the data of AD which also includes 3D volumetric data. That becomes lightly difficult to work on 3D images, hence we have considered only 2D images which are extracted from the ADNI dataset that includes three classes of data such as AD, MCL, and normal subjects. The details of dataset are shown in table. Figure 6 represents the sample brain images of ADNI dataset. Table 2 depicts the training and testing images presented in the ADNI dataset.

Table 2. Training and test images in ADNI

Classes	Test	Train	Total
AD	26	145	171
Normal	87	493	580
MCI	35	198	233
Total	148	836	984

5.2 Simulation setup

This section illustrates the simulation setup of the proposed ST-Seg model which is implemented by Matlab R2020a simulation tool that helps to enhance the performance of the simulation results. The simulation of this research is conducted for pre-processing, segmentation, and AD detection by extracting multi scale features from the segmented image. Table 3 illustrates the system specification which includes both hardware and software specifications.

Table 3. System specifications

Hardware Specifications	CPU	3.00GHz
	RAM	6GB
	Storage of Hard Disk	1 TB
Software Specifications	Operating system	Windows 10 pro
	Simulation tool	Matlab R2020a
	Processor	Intel (R) core™ i5-4590S

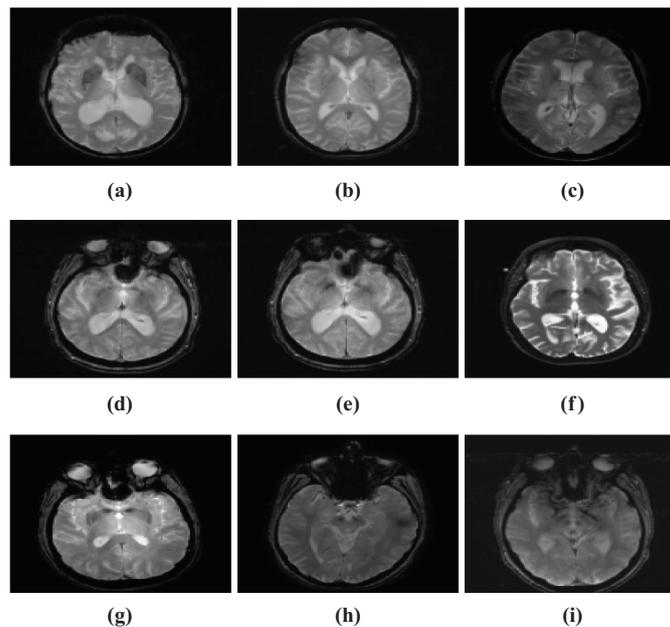


Fig. 6. (a)–(c) sampled normal brain images, (d)–(f) sample AD brain images, (g)–(i) sample MCI images

5.3 Comparative analysis

This section explains the comparative analysis of the proposed and existing models. In this research, we compare two existing works such as BEMD [1], and Ex-PFE [38] with our proposed ST-Seg model. The performance of the proposed work is compared

with existing works in terms of accuracy, sensitivity, specificity, positive predictive value, and confusion matrix.

Impact of accuracy. This metric is used to evaluate the accuracy of AD detection and classification. It is calculated by dividing the sum of true positive rate and true negative by sum of all samples. The mathematical expression of accuracy is defined as follows,

$$\bar{A} = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (35)$$

where t_p is true positive, t_n is true negative, f_p represent the false positive rate, and f_n represent the false negative. Figure 7 represents the comparison of accuracy with number of images. The comparison proved that the proposed ST-Seg model achieves high accuracy compared to existing models. The existing BEMD model does not perform noise removal or skull stripping, hence it performed AD detection with the presence of noise and non-brain regions that leads to poor detection accuracy. In addition, features are extracted by local binary pattern algorithm to extract the features from the whole brain image which leads to high complexity and misclassification that reduces detection accuracy. In Ex-PFE model perform AD detection by considering both brain and non-brain images which leads to misclassification and poor accuracy. Additionally, it performs AD detection by considering single feature (i.e., textural) which is not enough for accurate AD detection which leads to less accuracy compared to our proposed ST-Seg model. In our proposed ST-Seg model consider many features (i.e., textural, structural, edge, blobs, color, and contour) for detecting AD which leads to high accuracy, additionally, we remove non-brain (i.e., skull stripping) tissues from the MRI image for increasing accuracy. The proposed ST-Seg model achieves (0.98) higher than existing models.

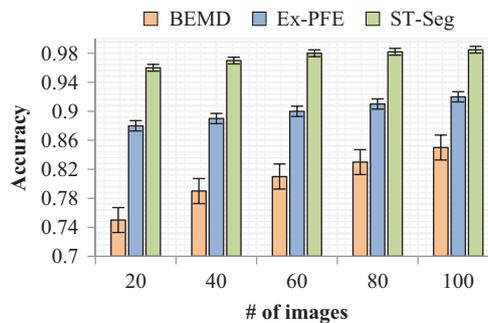


Fig. 7. Comparison of accuracy

Impact of specificity. This metric is used to measure the correctness of detecting such as AD, normal, and MCI using MRI images. The mathematical representation of specificity is defined as follows,

$$\mathcal{S} = \frac{t_n}{t_n + f_p} \quad (36)$$

where β represent specificity, the definition of specificity is denoted as the true positive rate is divided by the sum of true negative and false positive. Figure 8 illustrates the comparison of specificity with respect to number of images which proved that the proposed work achieved superior performance in terms of specificity compared to existing works. The proposed ST-Seg model perform noise removal, skull stripping, and bias field correction before initiating segmentation which increases true positive rate because of reducing unwanted tissues and noises form the image that leads to high specificity. The existing BEMD model perform only contrast enhancement and Ex-PFE model perform only bias field correction which reduces correctness of detection due to poor quality of images. The Ex-PFE model consider only texture feature for disease detection which increases misclassification and reduces specificity. In addition, the BEMD model does not focusing on segmentation for detection which is one of the significant processes to accurately classified the disease, hence it achieves less specificity. The proposed ST-Seg model achieves high (0.93) specificity compared to existing models.

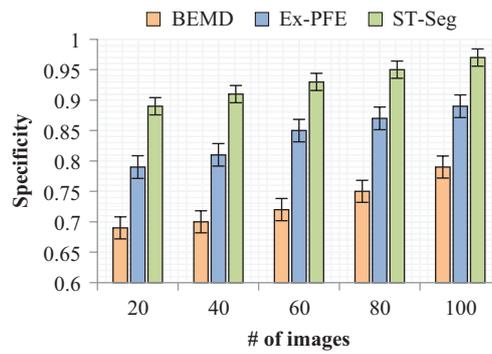


Fig. 8. Comparison of specificity

Impact of sensitivity. This metric is used to calculate the detection ability of positive (i.e., affected by disease) samples, which is also known as recall. In addition, it helps to calculate the applicability from the detected samples. The mathematical expression of sensitivity is defined as below,

$$\mathcal{S} = \frac{t_p}{t_p + f_n} \quad (37)$$

where \mathcal{S} represent the sensitivity, Figure 9 represents the comparison of sensitivity with respect to number of images which shows that the proposed ST-Seg model achieves high sensitivity compared to existing models. The proposed ST-Seg model extract multi scale features for improving detection accuracy, but the existing works used single scale features which provide limited information about the features that leads to less detection accuracy and sensitivity. In addition, we have used attention layers (i.e., spatial and channel) in VGG-16 for given an importance for every feature that increases sensitivity value during detection and classification of Alzheimer disease. The proposed ST-Seg model achieves high (0.9) sensitivity compared to existing models.

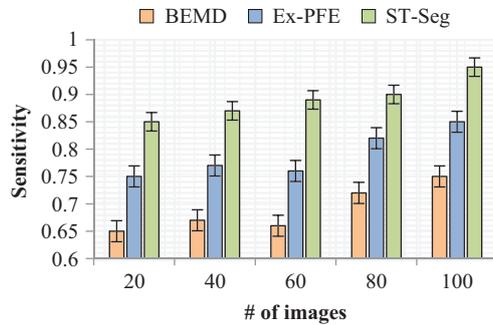


Fig. 9. Comparison of sensitivity

Impact of positive predictive value. This metric is used to successfully identify the positive results. In other words, the proportion between truly identified as positive to all the samples that had positive results (i.e., normal persons). The mathematical representation of positive predictive value (\mathbb{P}) is defined as follows,

$$\mathbb{P} = \frac{t_p}{t_p + f_p} \quad (38)$$

Figure 10 represents the comparison of proposed and existing positive predictive value with respect to number of images. The comparison result shows that the proposed work achieved better performance in positive predictive value compared to existing works. In this research, we have taken multiple features for segmenting GM, WM, and CSF using ST-MUNet algorithm which performs semantic segmentation that improves high positive predictive value, because discriminator evaluate the segmentation result to ground truth to improve segmentation accuracy which also increases positive predictive values. In addition, the proposed MSFP-VGG-16 takes multiple features (i.e., statistical, textural, structural, edge, color, blobs, and contour) in multi scale for improving classification accuracy. But the existing works only consider limited features for segmentation and classification with single scale for detecting AD which reduces the positive predictive value compared to out proposed work. The ST-Seg model achieves high positive prediction value (0.7) in AD detection.

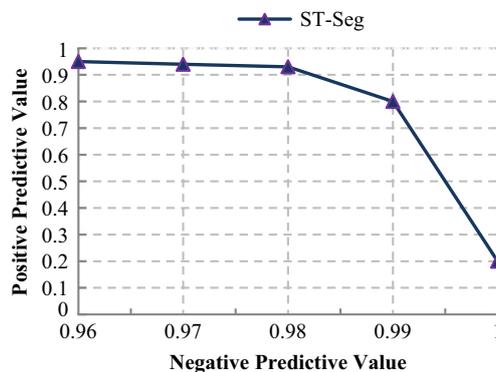


Fig. 10. Comparison of positive predictive value

Impact of confusion matrix. It is the matrix to represent the performance of AD detection and classification achieved by the proposed ST-Seg model. It helps to predict the tiny errors in the detection or classification. The confusion matrix represent the matrix of square format which includes $C_i(i, j)$ represent the number of images which are detected in the dataset whereas i represent the true label and j represent the predicted label. Table 4 represents the confusion matrix of the proposed ST-Seg model.

Table 4. Confusion matrix

True Label	AD	0.98	0	3
	Normal	0	0.97	0
	MCI	3	0	0.95
		AD	Normal	MCI
Predicted Label				

To make it clear, Table 5 illustrates the average values of the performance metrics of both proposed and existing models.

Table 5. Numerical analysis of performance metrics

Performance Metrics	Scenario	Proposed vs. Existing Systems		
		BEMD	Ex-PFE	ST-Seg
Accuracy	# of images	0.80	0.9	0.98
Specificity	# of images	0.73	0.84	0.93
Sensitivity	# of images	0.69	0.79	0.9
Positive predictive value	Negative predictive value	–	–	0.7

6 Conclusion

In this research, we proposed a deep learning approach for detecting AD using MRI images. Initially, preprocessing is performed to increase the quality of the images, for that noise removal is performed by HKIF algorithm which removes the noise from the brain MRI images, and then perform skull stripping to eliminate non-brain tissues and extracting brain tissues which increases the detection accuracy which is done by GAC algorithm. After completed skull stripping, bias field correction is performed to correct the intensity of the non-uniformity which increases the quality of the brain images which is done by EM algorithm. The preprocessed images are fed into segmentation for segmenting gray matter, white matter, and cerebrospinal fluid by considering cortical thickness, color, texture, and boundary information which is done by ST-MUNet. Here, the preprocessed images are divided into multiple patches for performing segmentation which increases the accuracy of segmentation. After completed segmentation, AD detection and classification is done by MSFP-VGG16 algorithm which extracts the features from multi scales and given an attention to features in terms of spatial and channel which increases the accuracy of detection. Finally, the MRI brain images are classified into three classes such as normal, AD, and MCI.

7 References

- [1] A. Deshmukh, M. V Karki, B. SR, and H. JP, “Deep neural network model for automated detection of Alzheimer’s disease using EEG signals,” *International Journal of Online & Biomedical Engineering*, vol. 18, 2022. <https://doi.org/10.3991/ijoe.v18i08.29867>
- [2] S. PC, V. Sherimon, P. SP, R. V. Nair, and R. Mathew, “A systematic review of clinical decision support systems in Alzheimer’s disease domain,” *International Journal of Online & Biomedical Engineering*, vol. 17, 2021. <https://doi.org/10.3991/ijoe.v17i08.23643>
- [3] R. Hedayati, M. Khedmati, and M. Taghipour-Gorjikolaie, “Deep feature extraction method based on ensemble of convolutional auto encoders: Application to Alzheimer’s disease diagnosis,” *Biomedical Signal Processing and Control*, vol. 66, p. 102397, 2021. <https://doi.org/10.1016/j.bspc.2020.102397>
- [4] B. A. Mohammed, E. M. Senan, T. H. Rassem, N. M. Makbol, A. A. Alanazi, Z. G. Al-Mekhlafi, *et al.*, “Multi-method analysis of medical records and MRI images for early diagnosis of dementia and Alzheimer’s disease based on deep learning and hybrid methods,” *Electronics*, vol. 10, p. 2860, 2021. <https://doi.org/10.3390/electronics10222860>
- [5] S. Murugan, C. Venkatesan, M. Sumithra, X.-Z. Gao, B. Elakkiya, M. Akila, *et al.*, “DEMNET: A deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images,” *IEEE Access*, vol. 9, pp. 90319–90329, 2021. <https://doi.org/10.1109/ACCESS.2021.3090474>
- [6] M. Raju, T. Sudila, V. P. Gopi, and V. Anitha, “Classification of mild cognitive impairment and Alzheimer’s disease from magnetic resonance images using deep learning,” in *2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, 2020, pp. 52–57. <https://doi.org/10.1109/RTEICT49044.2020.9315695>
- [7] S. Aruchamy, V. Mounya, and A. Verma, “Alzheimer’s disease classification in brain MRI using modified kNN algorithm,” in *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*, 2020, pp. 1–6. <https://doi.org/10.1109/iSSSC50941.2020.9358867>
- [8] R. De Feo, A. Shatillo, A. Sierra, J. M. Valverde, O. Gröhn, F. Giove, *et al.*, “Automated joint skull-stripping and segmentation with Multi-Task U-Net in large mouse brain MRI databases,” *NeuroImage*, vol. 229, p. 117734, 2021. <https://doi.org/10.1016/j.neuroimage.2021.117734>
- [9] O. M. Al-hazaimh, A. A. Abu-Ein, N. M. Tahat, M. m. A. Al-Smadi, and M. M. Al-Nawashi, “Combining Artificial Intelligence and Image Processing for Diagnosing Diabetic Retinopathy in Retinal Fundus Images,” *International Journal of Online & Biomedical Engineering*, vol. 18, 2022. <https://doi.org/10.3991/ijoe.v18i13.33985>
- [10] A. Hoopes, J. S. Mora, A. V. Dalca, B. Fischl, and M. Hoffmann, “SynthStrip: Skull-Stripping for Any Brain Image,” *arXiv preprint arXiv:2203.09974*, 2022. <https://doi.org/10.1016/j.neuroimage.2022.119474>
- [11] O. Al-hazaimh, S. A. Alomari, J. Alsakran, and N. Alhindawi, “Cross correlation–new based technique for speaker recognition,” *Int J Acad Res*, vol. 6, pp. 232–239, 2014. <https://doi.org/10.7813/2075-4124.2014/6-3/A.33>
- [12] N. Gharaibeh, O. M. Al-Hazaimh, B. Al-Naami, and K. M. Nahar, “An effective image processing method for detection of diabetic retinopathy diseases from retinal fundus images,” *International Journal of Signal and Imaging Systems Engineering*, vol. 11, pp. 206–216, 2018. <https://doi.org/10.1504/IJSISE.2018.093825>
- [13] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, L. Shen, *et al.*, “A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer’s disease,” *NeuroImage*, vol. 208, p. 116459, 2020. <https://doi.org/10.1016/j.neuroimage.2019.116459>

- [14] E. Dicks, W. M. van der Flier, P. Scheltens, F. Barkhof, B. M. Tijms, and A. s. D. N. Initiative, "Single-subject gray matter networks predict future cortical atrophy in preclinical Alzheimer's disease," *Neurobiology of Aging*, vol. 94, pp. 71–80, 2020. <https://doi.org/10.1016/j.neurobiolaging.2020.05.008>
- [15] S. Basheera and M. S. S. Ram, "A novel CNN based Alzheimer's disease classification using hybrid enhanced ICA segmented gray matter of MRI," *Computerized Medical Imaging and Graphics*, vol. 81, p. 101713, 2020. <https://doi.org/10.1016/j.compmedimag.2020.101713>
- [16] A. Ma'moun, O. M. Al-hazaimeh, N. Alhindawi, and S. M. Hayajneh, "A dual curvature shell phased array simulation for delivery of high intensity focused ultrasound," *Computer and Information Science*, vol. 7, p. 49, 2014. <https://doi.org/10.5539/cis.v7n3p49>
- [17] Z. Xia, G. Yue, Y. Xu, C. Feng, M. Yang, T. Wang, *et al.*, "A novel end-to-end hybrid network for Alzheimer's disease detection using 3D CNN and 3D CLSTM," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 1–4. <https://doi.org/10.1109/ISBI45749.2020.9098621>
- [18] O. M. Al-Hazaimeh, M. Al-Nawashi, and M. Saraee, "Geometrical-based approach for robust human image detection," *Multimedia Tools and Applications*, vol. 78, pp. 7029–7053, 2019. <https://doi.org/10.1007/s11042-018-6401-y>
- [19] N. Gharaibeh, O. M. Al-hazaimeh, A. Abu-Ein, and K. M. Nahar, "A hybrid svm naïve-bayes classifier for bright lesions recognition in eye fundus images," *Int J Electr Eng Inform*, vol. 13, pp. 530–545, 2021. <https://doi.org/10.15676/ijeei.2021.13.3.2>
- [20] P. Khan, M. F. Kader, S. R. Islam, A. B. Rahman, M. S. Kamal, M. U. Toha, *et al.*, "Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances," *IEEE Access*, vol. 9, pp. 37622–37655, 2021. <https://doi.org/10.1109/ACCESS.2021.3062484>
- [21] C. S. Eke, E. Jammeh, X. Li, C. Carroll, S. Pearson, and E. Ifeakor, "Early detection of Alzheimer's disease with blood plasma proteins using support vector machines," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, pp. 218–226, 2020. <https://doi.org/10.1109/JBHI.2020.2984355>
- [22] E. Hussain, M. Hasan, S. Z. Hassan, T. H. Azmi, M. A. Rahman, and M. Z. Parvez, "Deep learning based binary classification for Alzheimer's disease detection using brain MRI images," in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2020, pp. 1115–1120. <https://doi.org/10.1109/ICIEA48937.2020.9248213>
- [23] T. A. Tuan, T. B. Pham, J. Y. Kim, and J. M. R. Tavares, "Alzheimer's diagnosis using deep learning in segmenting and classifying 3D brain MR images," *International Journal of Neuroscience*, pp. 1–10, 2020. <https://doi.org/10.1080/00207454.2020.1835900>
- [24] O. M. Al-Hazaimeh and M. Al-Smadi, "Automated pedestrian recognition based on deep convolutional neural networks," *International Journal of Machine Learning and Computing*, vol. 9, pp. 662–667, 2019. <https://doi.org/10.18178/ijmlc.2019.9.5.855>
- [25] H. S. Zaina, S. B. Belhaouari, T. Stanko, and V. Gorovoy, "An exemplar pyramid feature extraction based Alzheimer disease classification method," *IEEE Access*, vol. 10, pp. 66511–66521, 2022. <https://doi.org/10.1109/ACCESS.2022.3183185>
- [26] J. E. Arco, J. Ramírez, J. M. Górriz, M. Ruz, and A. s. D. N. Initiative, "Data fusion based on searchlight analysis for the prediction of Alzheimer's disease," *Expert Systems with Applications*, vol. 185, p. 115549, 2021. <https://doi.org/10.1016/j.eswa.2021.115549>
- [27] S. Basheera and M. S. S. Ram, "Deep learning based Alzheimer's disease early diagnosis using T2w segmented gray matter MRI," *International Journal of Imaging Systems and Technology*, vol. 31, pp. 1692–1710, 2021. <https://doi.org/10.1002/ima.22553>
- [28] M. Song, H. Jung, S. Lee, D. Kim, and M. Ahn, "Diagnostic classification and biomarker identification of Alzheimer's disease with random forest algorithm," *Brain Sciences*, vol. 11, p. 453, 2021. <https://doi.org/10.3390/brainsci11040453>

- [29] Z. Fan, J. Li, L. Zhang, G. Zhu, P. Li, X. Lu, *et al.*, “U-net based analysis of MRI for Alzheimer’s disease diagnosis,” *Neural Computing and Applications*, vol. 33, pp. 13587–13599, 2021. <https://doi.org/10.1007/s00521-021-05983-y>
- [30] E. E. Bron, S. Klein, J. M. Papma, L. C. Jiskoot, V. Venkatraghavan, J. Linders, *et al.*, “Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer’s disease,” *NeuroImage: Clinical*, vol. 31, p. 102712, 2021. <https://doi.org/10.1016/j.nicl.2021.102712>
- [31] V. S. Rallabandi, K. Tulpule, M. Gattu, and A. s. D. N. Initiative, “Automatic classification of cognitively normal, mild cognitive impairment and Alzheimer’s disease using structural MRI analysis,” *Informatics in Medicine Unlocked*, vol. 18, p. 100305, 2020. <https://doi.org/10.1016/j.imu.2020.100305>
- [32] H. A. Helaly, M. Badawy, and A. Y. Haikal, “Deep learning approach for early detection of Alzheimer’s disease,” *Cognitive computation*, vol. 14, pp. 1711–1727, 2022. <https://doi.org/10.1007/s12559-021-09946-2>
- [33] A. Ashraf, S. Naz, S. H. Shirazi, I. Razzak, and M. Parsad, “Deep transfer learning for Alzheimer neurological disorder detection,” *Multimedia Tools and Applications*, vol. 80, pp. 30117–30142, 2021. <https://doi.org/10.1007/s11042-020-10331-8>
- [34] R. Divya and R. Shantha Selva Kumari, “Genetic algorithm with logistic regression feature selection for Alzheimer’s disease classification,” *Neural Computing and Applications*, vol. 33, pp. 8435–8444, 2021. <https://doi.org/10.1007/s00521-020-05596-x>
- [35] A. Ebrahimi, S. Luo, R. Chiong, and A. s. D. N. Initiative, “Deep sequence modelling for Alzheimer’s disease detection using MRI,” *Computers in Biology and Medicine*, vol. 134, p. 104537, 2021. <https://doi.org/10.1016/j.combiomed.2021.104537>
- [36] S. Sharma, K. Guleria, S. Tiwari, and S. Kumar, “A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans,” *Measurement: Sensors*, vol. 24, p. 100506, 2022. <https://doi.org/10.1016/j.measen.2022.100506>
- [37] F. U. R. Faisal and G.-R. Kwon, “Automated detection of Alzheimer’s disease and mild cognitive impairment using whole brain MRI,” *IEEE Access*, vol. 10, pp. 65055–65066, 2022. <https://doi.org/10.1109/ACCESS.2022.3180073>
- [38] J. E. W. Koh, V. Jahmunah, T.-H. Pham, S. L. Oh, E. J. Ciaccio, U. R. Acharya, *et al.*, “Automated detection of Alzheimer’s disease using bi-directional empirical model decomposition,” *Pattern Recognition Letters*, vol. 135, pp. 106–113, 2020. <https://doi.org/10.1016/j.patrec.2020.03.014>

8 Authors

Nasr Gharaibeh is an Assistance Professor in the Department of electrical Engineering. He received his PhD in Electrical engineering in 1990. Now, he is a lecturer at Al-Balqa Applied University – Al-huson University College, Jordan. He can be contacted at email: nas@bau.edu.jo.

Ashraf A. Abu-Ein is an Associate Professor in the Department of Electrical Engineering. He has completed his PhD at National Technical University of Ukraine, Computer Engineering. “Computers, Computing Systems and Networks”, 2007. Now, he is a lecturer at Al-Balqa Applied University – Al-huson University College, Jordan. He can be contacted at email: ashraf.abuain@bau.edu.jo.

Obaida M. Al-hazaimeh earned a BSc in Computer Science from Jordan's Applied Science University in 2004 and an MSc in Computer Science from Malaysia's University Science Malaysia in 2006. In 2010, he earned a PhD in Network Security (Cryptography) from Malaysia. He is an Full professor at Al-Balqa Applied University's Department of Computer Science and Information Technology. Cryptology, image processing, machine learning, and chaos theory are among his primary research interests. He has published around 55 papers in international refereed publications as an author or co-author. He can be contacted at email: dr_obaida@bau.edu.jo.

Khalid M.O. Nahar is an Associate Professor in the Department of Computer Sciences-Faculty of IT, Yarmouk University, Irbid-Jordan. He received his BS and MS in Computer Sciences from Yarmouk University in Jordan, in 1992 and 2005, respectively. He was awarded a full scholarship to continue his PhD in Computer Sciences and Engineering from King Fahd University of Petroleum and Minerals (KFUPM), KSA. In 2013, he completed his PhD and started his job as an Assistant Professor at Tabuk University, KSA for two years. In 2015, he backs to Yarmouk University and from now he is the Assistant Dean for Quality Control, Jordan. He can be contacted at email: Khalids@yu.edu.jo.

Waleed A. Abu-Ain his PhD in Artificial Intelligence in 2016 from the Artificial Intelligence Technology Center of the National University of Malaysia, Department of Computer Science (UKM). Since 2016, he is currently working as a full-time assistant professor in the Department of Computer Science at Taibah University, Applied College, Saudi Arabia. His research interests include artificial intelligence, machine learning and optimization, data science, computer vision, deep learning, and the Internet of Things. He can be contacted at email: wabuain@taibahu.edu.sa.

Malek M. Al-Nawashi is an a Lecturer in the Department of Computer Science and Information Technology at Al-Balqa Applied University–Al-huson University College, Jordan. He has completed his PhD at University of Salford Manchester in Computer Science in 2019. His main research interests are image processing and machine learning. He can be contacted at email: nawashi@bau.edu.jo.

Article submitted 2022-12-30. Resubmitted 2023-02-01. Final acceptance 2023-02-02. Final version published as submitted by the authors.