Diagnostic Framework for Automatic Classification and Visualization of Alzheimer's Disease with Feature Extraction Using Wavelet Transform

M. Anitha[1*], V. Karpagam[2], P. Tamije Selvy[3]

Abstract

Alzheimer's Disease (AD) is a serious disease that destroys brain and is classified as the most widespread type of dementia. Manual evaluation of image scans relies on visual reading and semi-quantitative investigation of various human brain sections, leading to wrong diagnoses. Neuroimaging plays a significant part in AD detection, using image processing approaches that succeed the drawback of traditional diagnosis methods. Feature extraction is done through Wavelet Transform (WT). Feature selection is an important step in machine learning, where best features set from all possible features is determined. Mutual Information based feature selection (MI) and Correlation-based Feature Selection (CFS) captures the 'correlation' between random variables. Machine Learning techniques are broadly used in a classification problem, as it is simple, effective mechanisms and capability to train to contribute intelligence to the arrangement. Classifiers used in this proposed work are Artificial Neural Network (ANN), Random Forest, Convolutional Neural Network (CNN), and Wavelet-based CNN. The superior ability of ANN is high-speed processing achieved through extensive parallel implementation, and this has emphasized necessity of research in this field. CNN has encouraged tackling this issue. This work proves that wavelet-based CNN performs better with a classification accuracy of 91.87%, the sensitivity of 0.94 for normal brain and 0.88 for AD affected brain, the positive predictive value of 0.91 for normal brain and 0.92 for AD affected brain, and F measure of 0.92 for normal brain and 0.90 for AD affected brain on ADNI MRI dataset of the human brain in detecting AD.

Key Words: Alzheimer's Disease (AD), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Correlation-based Feature Selection (CFS), Feature Selection, Machine-Learning Methods, Mutual Information (MI), Neuroimaging, Random Forest (RF), Wavelet-based CNN.

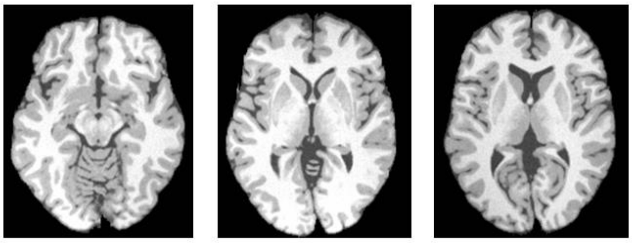
DOI Number: 10.14704/nq.2021.19.7.NQ21088 **NeuroQuantology 2021; 19(7):84-95**

Introduction

Currently, Alzheimer's Disease (AD) is a widespread problem of dementia. As per the World Alzheimer Report (Patterson 2018), in 2018, around 50 million people were afflicted with this Disease and by 2050, this number is anticipated to triple. Frequently, after 60 years of age, Alzheimer's symptoms will become visible. In some cases, individuals with gene mutations can experience the early (30–50 years of AGE) onset of certain AD types. The brain's functions and structures are changed due to AD. Among AD patients, period between healthy states to Alzheimer's Disease can range over several years. The patients will initially have Mild Cognitive Impairment (MCI), slowly developing into AD. Though, every MCI patient does not become an AD patient.

Corresponding author: M. Anitha
Address: [1*]Research Scholar, Department of Information Technology, Sri Ramakrishna Engineering College, Coimbatore, India; [2]Professor, Information Technology Department, Sri Ramakrishna Engineering College, Coimbatore, India; [3]Professor, Computer Science and Engineering Department, Sri Krishna College of Technology, Coimbatore, India.
E-mail: [1*]anitha_murugaiyan@yahoo.com; [2]karpagam.vilvanathan@srec.ac.in; [3]p.tamijeselvy@skct.edu.in
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Received: 20 March 2021 **Accepted:** 30 June 2021

Hence, the major aim of existing research is the prediction of the MCI's conversion to AD. Techniques such as medical imaging and blood plasma spectroscopy are applied to assess these variations.



(a) (b) (c)

Figure 1. MRI Scan of Normal Brain (a), MCI (b) and Alzheimer Disease Brain (c) with Variations in Hippocampal Region, Ventricles and Cerebral Cortex

The changes in the Cerebrospinal Fluid (CSF) are an essential indicator for AD, as seen in Figure 1. AD development consists of three different stages: Preclinical AD, MCI, and AD. The individuals will have measurable variations in the blood (biomarkers), CSF, or/and brain in the preclinical AD stage (Soliman et al., 2019). The transitional stage between normal and AD is known as MCI. MCI patients can carry out their day-to-day activities, even though they may experience mild cognitive and memory issues. Amnestic MCI (aMCI) and non-amnestic MCI (naMCI) are the two distinct kinds of MCI. Research shows that MCI can affect about 10% to 20% of individuals of ages sixty-five or older. Alzheimer's disease stage is severe as the individual will experience behavioral changes, reduced thinking capability, and memory loss. Individuals with Alzheimer's are not even able to perform their daily activities.

Magnetic Resonance Imaging (MRI) is an effective medical imaging method for analyzing the human brain, which provides high-quality soft tissue anatomy images that enhanced brain pathology diagnosis and treatment. Physicians implement brain MRI for AD diagnosis as the total brain atrophy, and primarily the hippocampal atrophy is acknowledged as clear indicators of AD. T1 weighted MRIs measure the amount of atrophy. Neuroimaging has helped machine learning (Beheshti et al., 2016); (Matsuda et al., 2017) for the early diagnosis of AD. Researchers use neuroimaging methods to assess the brain's uncontrollable changes during Alzheimer's disease development.

Features in the frequency domain are more efficient in machine learning, and Wavelet filters analysis can uncover data features overlooked by other image analysis methods, like, discontinuities, breakdown points, and trends in self-similarity and higher derivations. The construction of classification systems is a significant problem in feature selection. Misclassification occurs due to the interdependency and redundancy of features. At the same time, high costs are correlated with extra computations. It is very beneficial to restrict the classifier's number of input features.

Artificial Neural Networks (ANNs) are commonly applied for universal function approximation in numerical algorithms due to their outstanding input attributes to output mapping, non-linearity, fault tolerance, adaptivity, and self-learning. An excellent benefit of ANNs application is its ability to produce easily utilizable models and better accuracy from complex natural systems that have huge inputs (Jahnavi 2017); (Tamije Selvy et al., 2019); (Kim et al., 2019). ANN is the latest most suitable model employed for machine learning and problem resolutions. Artificial Neural Networks in which the individual neurons are organized such that they respond to overlapping areas in the visual field. This feed-type type of ANN is known as Convolutional Neural Network (CNN).

Contribution of Proposed Work

This work proposes a structure for the wavelet-based CNN structure for distinguishing Alzheimer's Disease in MRI brain images. The second section describes the related works in literature. The third section details the techniques which are used. The techniques used include Wavelet Transform for Feature Extraction, Correlation-based Feature Selection (CFS), and Mutual Information (MI) based Feature Selection for selecting the features. Feature Selection is accompanied by a classification mechanism, including Random Forest (RF), ANN, CNN and Wavelet-based CNN. The fourth section analyzes the results of the experiments. Consequently, the fifth section concludes the proposed work with points to eventual improvement.

Literature Survey

For AD classification using Computer-Aided Diagnosis (CAD) systems, various machine learning methods, including Decision tree, Naïve Bayes (NB), K-Nearest Neighbor (KNN), Fuzzy, Deep Learning (DL), ANN, and Support Vector Machines (SVM), have been investigated in the literature.

Masoumi et al., (2019) introduced Wavelet Brain for representing the white and gray matter of the brain. The proposed method extracted the local shape data from a global harmonic representation. Experimental results showed that the proposed technique shows substantial improvement in classifying AD.

Acharya et al., (2019) presented an automatic AD identification system that classified T2-weighted brain MRI images. The work mainly focused on evaluating the efficacy of various wavelet transforms, like Discrete Wavelet Transform (DWT), Contourlet, Curvelet, Complex Wavelet Transform (CWT), Dual-Tree Complex Wavelet, Empirical, and Shearlet Transform for feature extraction. The median filter uses for pre-processing of the images, and KNN for classifying the features extracted. The experimental results showed that the wavelet transforms were effective as feature extractors for AD classification, and Shearlet transform performed the best.

Jha et al., (2018) presented a CAD based on dual-tree complex wavelet transforms for identifying the AD images. The proposed technique applied linear discriminant analysis on the wavelet coefficients and extreme learning machine for detecting AD. The proposed methods were evaluated using ADNI and OASIS datasets, and experiments demonstrated the efficiency of the methods.

Long et al., (2017) presented a technique to distinguish patients with AD or MCI from normal. The brain's regional morphological differences are analyzed within groups. The asymmetric diffeomorphic registration for pair of subjects is computed, accompanied by embedding algorithm and learning method for classification.

Tanveer et al., (2020) reviewed 165 papers from 2005 to 2019, which utilized several machine learning and feature extraction methods. The three major classification algorithms used in literature were ANN, SVM, and DL and ensemble techniques. A presentation did on the comprehensive survey of these three techniques for Alzheimer's and potential directions for the future.

Khagi et al., (2019) applied the deep layers feature extraction assisted by DNN architecture, with no comprehensive hardware resource training and image classification based on a simplistic algorithm that chooses the best features to decrease hardware utilization time, classification error, and workload. The CNN layer applies to utilize the same architecture as Alexnet with certain variations in the parameters, for automatically extracting features of images gained through the entire brain MRI's slice extraction whilst the extraction of 13 manual features which are based on the Gray level co-occurrence matrix was also done for evaluating features' ranking impact.

The mobility data of Alzheimer's patients can be collected by existing smartphones utilizing accelerometer integration. A technique was proposed by Bringas et al., (2019) for accelerometer data processing and developing a CNN, which classified the disease stage as per the patient's mobility patterns. In a case study with 35 AD patients, this technique demonstrated a 91% accuracy rate for classification.

Wang et al., (2018) investigated the use of inter-class variance measures for choosing single slice from 3D volumetric data to classify AD images. The proposed method was based on wavelet entropy, MLPNN, and biogeography-based optimization. It was compared with six different techniques and was found to perform better in detecting AD.

A new CNN architecture with Multi-Resolution Analysis was presented by Fujieda et al., (2018), which used a combination of CNNs with a multi-resolution analysis. The proposed observation was that CNN could be considered a multi-resolution analysis limited form. As per this observation, the wavelet transform is used to augment the multi-resolution analysis's absent portions and merge them into the overall architecture as additive components. In traditional CNNs, spectral information, which is beneficial in various image processing applications, gets lost. This spectral information gets used in Wavelet CNNs. The image annotation and texture classification performance of wavelet CNNs are assessed. The experimental results demonstrate that, compared to the current models, wavelet CNNs are more accurate in image annotation and texture classification. Also, compared to the traditional CNNs, the wavelet CNNs has a lower number of parameters.

The literature studies show that multi-resolution analysis is effective in feature extraction from MRI images, and Neural Networks and their derivatives give a higher accuracy compared to other classifiers. Based on the literature survey, this work proposed a framework to identify AD from MRI images.

Methodology

Alzheimer's Disease Neuroimaging Initiative (ADNI) Dataset

The longitudinal multisite observational research of aged individuals with normal cognition, MCI, or AD is associated with Alzheimer's Disease Neuroimaging Initiative (ADNI) study (Jack et al., 2008). The ADNI study's main aims are: to enhance procedures for clinical trials in AD and MCI, to validate biomarker and imaging data through correlation with concurrent clinical and psychometric evaluations, to determine the optimum techniques for image acquisition and analysis, to establish technical benchmarks for imaging in longitudinal studies, and to link all the data at every point of time. The ADNI study as a whole is partitioned into cores, where each core handles ADNI-related activities inside its expertise sphere: imaging, biomarkers, biostatistics, informatics and clinical (Hampel et al., 2017). The proposed framework incorporates a collection of ADNI datasets of MRI brain images from feature extraction by Wavelet
Transform (WT) and feature selection by CFS and MI-based Feature Selection (MI). The selected features are subjected to classification techniques that include RF, ANN, CNN, and Wavelet-CNN, represented in figure 2.

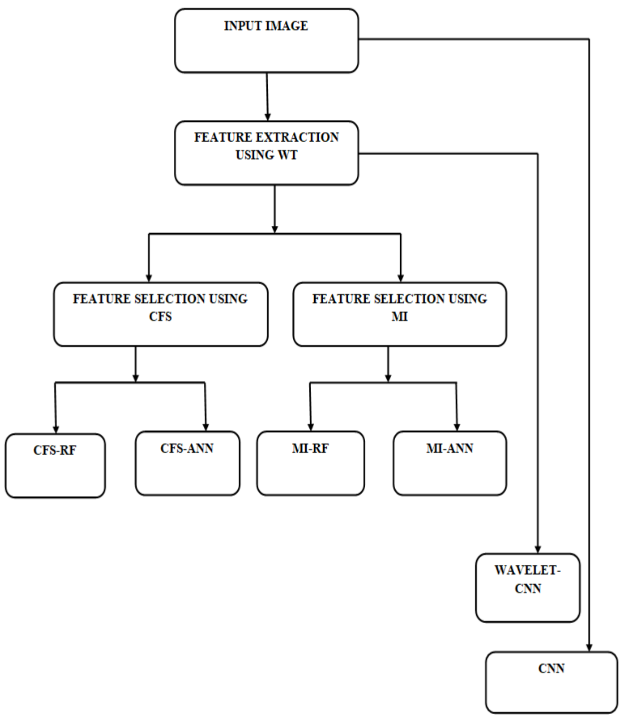


Figure 2. Framework of the Proposed Method

Feature Extraction

Various image and signal processing works have been prepared in the preceding ten years, utilizing wavelets (Ravi et al., 2018). Due to the time-frequency domain analysis scope, these methods are very suitable in the fields of fractals, resolution enhancement, image enhancement, denoising, and image or signal compression, and so on. Signal analysis, as per the scale, is the basic concept of this method (Karpagam & Rangarajan 2012). Compared to the conventional Fourier Transforms, Wavelet Transforms (WTs) are more beneficial for representing functions with sharp peaks and discontinuities and for the accurate deconstruction and reconstruction of signals that are finite non-periodic, and non-stationary. As a result of attributes like multi-resolution structure and sparsity, wavelet-based image analysis, denoising, enhancement, and so on works excellently.

The multi-resolution analysis is an attribute of a wavelet transform (Udhaya Suriya & Rangarajan 2017). Expression of the signal's local feature in the time and frequency domain is another attribute of the wavelet transform. Due to this, the wavelet transform is quite fit for recognizing the signal's flash state and abnormality and leaving on its composition. A wavelet ψ with compact support and n vanishing moments is represented as per Equation (1):

$$\int_{-\infty}^{+\infty} t^k \psi(t)dt = 0, \quad for \ 0 \le k \le n \tag{1}$$

Also, θ is a function with fast decay and provided as per Equation (2):

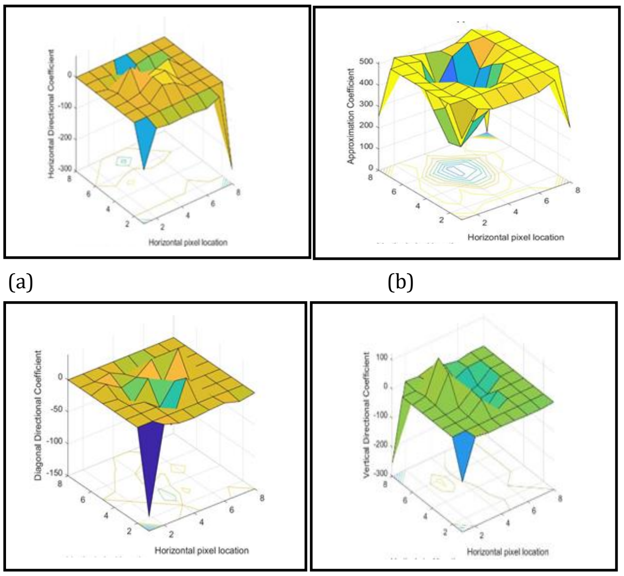
$$\psi(t) = (-1)^n \frac{d^n\theta(t)}{dt^n} \tag{2}$$

Later, for a signal f, the Wavelet transform for signal f is provided as per Equation (3):

$$Wf(u,s) = s^n \frac{d^n}{du^n}(f * \overline{\theta}_s)(u) \tag{3}$$

Where, $\overline{\theta}_s(t) = s^{-1/2}\theta(-t/s)$ is a time or space coordinate. Scale is denoted by s. The outcomes of this are, within these conditions, wavelet transform Wf (u, s) will become f's nth-order derivative that has on a domain which is proportional to s.

In this work, the statistical features of Mean, Median, Standard Deviation, Kurtosis, min value, and max value of the LL, HH, HL and LH bands were obtained. Figure 3 shows the feature vectors obtained for an AD image. The extracted features are reduced using CFS and MI is described in subsequent sections.



(a) (b) (c) (d)

Figure 3. Surface Plot of the Wavelet Features (a) Approximation, (b)Horizontal, (c) Vertical, (d) Diagonal Direction Coefficients Respectively

Feature Selection

1) Correlation- based Feature Selection (CFS)

CFS assesses an attribute subset's value by taking into consideration the degree of redundancy between each feature together with their predictive capability. It is employed to evaluate the correlation between class and subset of attributes and inter-correlations between features. The features group's weights decreases with inter-correlation growth and increases with the correlation between classes and features (Karegowda et al., 2010); (Vergara & Estévez 2014). The best feature subset is determined utilizing the CFS.

A feature filtering method like the CFS has been utilized to identify the feature subset possibly applied for a certain purpose. Equation (4) states the CFS is a heuristic regarding assessing the feature subsets' value or utility. The individual features and their utility have been depicted to forecast a class label concurrently with inter-correlation level between attributes as per Equation (4):

$$Merits = \frac{k\overline{r}_{cf}}{\sqrt{k + k(k-1)\overline{r}_{ff}}} \quad (4)$$

Where, Merits indicate a heuristic "merit" of s, which is the feature subset that consists of k features. Average feature to feature inter-correlation is denoted by $\overline{r}_{ff}$. Class correlation between an average feature is denoted by $\overline{r}_{cf}$ The heuristic discards every irrelevant and redundant feature as they are the poor predictors of the class. To evaluate the correlation between features, it is very important to discretize the numeric features by the CFS with the usage of the Symmetric Uncertainty (SU) measures the amount of its corelation among distinct features. The SU is obtained by normalizing the mutual information to its feature values entropies and target classes. It is used for assessing features goodness for classification by several researchers. Entropy of the Y both before and after the observation of X, where, the X and the Y indicates the discrete random variables according to Equations (5) - (6).

$$H(Y) = -\sum_{y \in Y} p(y)\log_2 p(y) \quad (5)$$

$$H(Y \mid X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y \mid x)\log_2 p(y \mid x) \quad (6)$$

2) Feature Selection based on Mutual Information (MI)

In 1948, Shannon developed the MI, a measure that describes the number of data two random variables carry about each other. The MI is symmetric and detects the variable's non-linear relationships. This latter attribute has made MI a well-known parameter for feature selection because other broadly utilized parameters like the correlation coefficient work only with linear dependencies.

$$I(x;y) = \sum_{i=1}^{n}\sum_{j=1}^{n} p(x(i), y(j)).\log\left(\frac{p(x(i), y(j))}{p(x(i)).p(y(i))}\right) \quad (7)$$

Wherein, when x and y are statistically independent, that is, $p(x(i), y(i)) = p(x(i)).p(y(j))$, the MI becomes zero.

The robustness to transformation and noise are the major benefits of MI. After feature selection, the top 100 features select from the 384 features extracted from Wavelets. Certain features employ to train and check the classifiers described in the following section.

Classifiers

To evaluate the proposed framework RF, ANN, CNN, and Proposed Wavelet based CNN are evaluated.

1) Random Forest (RF)

A "troupe of unproved classification trees" is fundamentally referred to as Random Forest [RF]. The RF provides outstanding performance on numerous functional problems, mainly because it is not prone to overfitting and it is not delicate to agitation in the collection of information. It is a combination of the predictions of individually trained numerous trees (Khan et al., 2020). The key attributes generation is followed by the assembly of many trees. The error rate is evaluated to select the tree. This order demonstrates deduced by choices tree-based classifiers. RF classifier utilizes three major key parameters:

1. Number of Predictors Sampled: At each split, the key factor is the number of tested predictors, which impacts the performance of the random forests.
2. Number of Trees: A decent decision is about 500 trees.
3. Node Size: Unlike in the decision trees, there can be a low number of nodes. The purpose is the expansion of trees with minimum bias.

This ensemble classifier has selected a class of input data (Chaudhary et al., 2016). It implies that a distinct classifier's decision merges to become a collective classifier. Each tree in the RF was trained to utilize a random vector independently sampled from a training set having the same distribution. A majority of votes will decide a class vector from the trees which are inside the ensemble. Thus, RF is a set of m classifier as per Equation (8):

$$R = \{\hbar(x, \Omega_k),\ \text{k=1,...m}\} \tag{8}$$

Wherein, for a random vector X, an n-dimensional random sample is denoted by x. The autonomously selected random vectors are denoted by Ω_k which is taken from the training set. Based on input variables, which are randomly selected from training set, trees are built. The random vector k Ω is for the kth tree which is created utilizing random vectors, $\Omega_1, \Omega_2 \ldots, \Omega_{k-1}$ to classify the input data.

The advantages of random forest classifiers are: highly accurate, runs efficiently on large databases. It allows an experimental method for identifying variable interactions.

2) Artificial Neural Network (ANN)

ANN comprises neurons, which are a collection of interconnected information processing units (Islam & Zhang 2018). The neurons are arranged into three layers: input layer, single or multiple hidden, and output layer. The ANN layers are not dependent on each other, the bias node (perpetually set to be equivalent to one). The bias nodes are comparable to offset in linear regression provided as *y ¼ ax þ b*, wherein, if "*a*" is independent's coefficient "*x*," and then, the slope is expressed by "*b*." The main purpose is to give a constant and trainable value to the node (Wang et al., 2019); (Valladares-Rodríguez et al., 2019). Essentially, a bias value allows the activation function to be moved either to the left or the right, which is analytical for the success of ANN training. When NN is employed as classifier, during training input features and output classes are matched by adjusting the neurons.

The biological neural network's organization and functions are obtained using artificial neural network models. Artificial neurons are the ANN's fundamental building units. Each input is multiplied with a weight by neuron's, the inputs are weighted. In the internal neuron, a sum function totals all bias and weighted inputs. While, at output neuron, case of prior weighted participation and bias are carried over an activation function. The ANN's product with K elements are implemented as per Equation (9):

$$y(x) = \sum_{i=1}^{k} w_i y_i(x) \tag{9}$$

Where, at a layer i, the output is y_i and the weight is w_i .

The input layer neurons utilize synaptic weights to transfer the input signal to the initial hidden layer. An activation function is applied for deciding if the value should transfer to the proceeding layer. Therefore, through weights adjustment, the neural network will continue its learning procedure. For the computation and adjustment of the weights, a very popularly utilized technique is back-propagation algorithm.

Advantages of Artificial Neural Network are: powerful, it can model difficult functions, ease to use.

3) Convolutional Neural Network (CNN)

Deep learning methods are one of the many machine learning models utilized for the classification task. An issue of the classification task is that the mobility data can display complicated pattern sequences of varying lengths over time. As a result, deep learning techniques are the best

classifiers for this data type as they can exploit the sequence's internal structure. CNN's, which are models from the deep learning family with rising renown, can identify pattern hierarchies from smaller sequences without expanding the complexity of the model (Wang et al., 2018).
Input, output, and many hidden layers enclosed in an artificial neural network structure make up the CNN. The linear operation between input and filter or kernel which functions as feature detector, is defined as convolution. The CNN's hidden layers are convolution-based. The CNN's hidden layers are convolution-based, and patterns are recognized by running CNN internally using sliding windows to recognize local patterns. The best representative can be discerned.
Linear rectification is used as the activation function by each convolutional layer. From the input data, the model can draw out patterns using this layer set. Last convolution's output is then flattened and utilized as last two layers' input: Softmax layer with three neurons and a Fully Connected (FC) layer. This network section's efficiency is to carry out the classification by assessing the probability that the provided input is part of a particular label.
The advantage of Convolutional Neural Network is fast to train model.

4) Wavelet-Convolutional Neural Network (Wavelet-CNN)

The two-dimensional Wavelet Transform is employed in applications like noise removal, image enhancement, data compression, and so on in medicine. The CNN's inputs are the whole image's wavelet coefficients.
Spectral analysis is suitable for the capture of scale-invariant features. Whereas CNNs are frequently fit to capture spatial features. Compared to the traditional CNNs, wavelet CNNs have competitive or better accuracy in classification accuracy and have a distinctly lower number of trainable criteria. A CNN is a neural network variation that employs a deep network that is connected sparsely. All inputs are connected to every unit in the next layer within a regular neural network model. Using fully connected layer and activation function, CNNs utilize convolution/pooling layers that will connect only with the local neighborhoods around each input.
CNN can decrease the parameter numbers and accomplish translation invariance in the image space through parameter sharing. In CNN's, the convolution layers will generally utilize diverse weight sets for similar input and will give a concatenated vector as the output. The pooling layers are utilized right after the convolution layers. Various CNN applications utilize max pooling.
In Wavelet CNNs, the input image's size must be fixed. This is because the Wavelet CNNs work on an energy layer with a similar size as the layer's input. Whereas, in traditional CNNs it has been proved that, just the low-frequency portions are utilized while the entire high-frequency portions are removed.

Results and Discussion

The proposed Wavelet CNN is compared with Random Forest, ANN, and CNN. For experiments, 690 AD and 540 normal images are considered. This work used 10 fold cross validation and 15 runs were carried out using the Hadoop framework in a E5 Xeon Core processor using VMware and one master 4 slaves configuration. Table 1 to table 4 and figure 1 to figure 7 illustrates Sensitivity, Classification Accuracy, Positive Predictive Value and F Measure, respectively. Figure 8 shows the 90
best fitness value.
Different terms are used with specificity, sensitivity and accuracy description.

- Accuracy = (TN + TP)/(TN+TP+FN+FP)
- Sensitivity = TP/(TP + FN)
- Specificity = TN/(TN + FP)
- Positive Predictive Value (PPV)=[TP/(TP+ TN)]×100
- F-Measure = (2 * Precision * Recall) / (Precision + Recall)

Where TP represents true positive, TN represents true negative, FN represents false negative, and FP represents false-positive.

Table 1. Classification Accuracy for CFS-RF, CFS-ANN, MI-RF, MI-ANN, CNN and Wavelet-CNN

Techniques Used	Classification Accuracy
CFS-Random Forest	85.37
CFS-ANN	86.42
MI-Random forest	87.97
MI-ANN	88.46
CNN	90.24
Wavelet-CNN	91.87

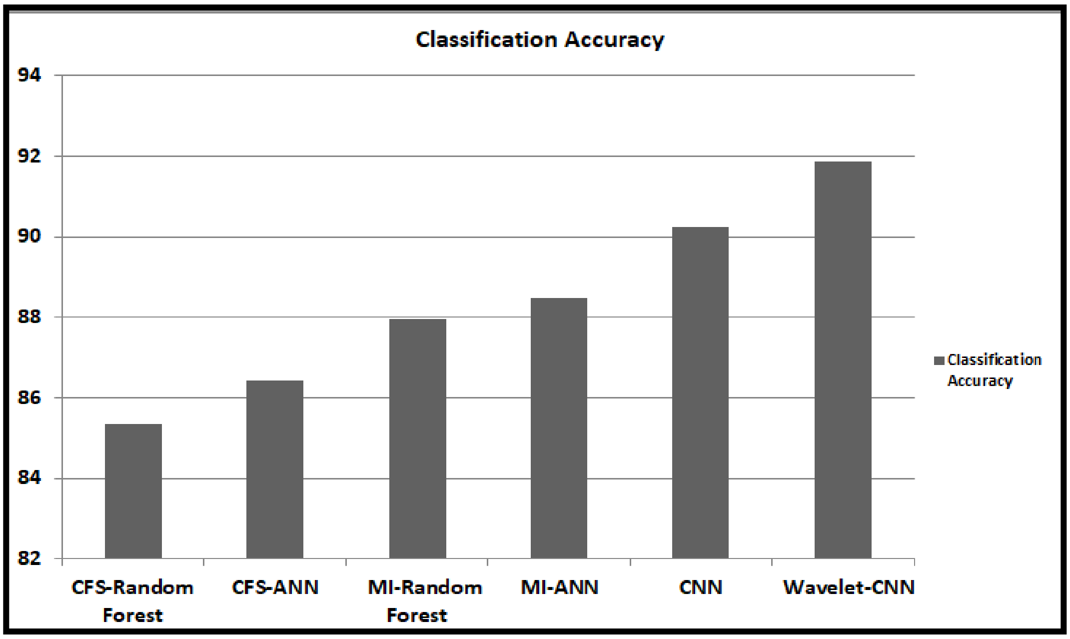


Figure 4. Graphical Representation of Classification Accuracy of CFS-RF, CFS-ANN, MI-RF, MI-ANN, CNN and Wavelet-CNN

Table 1 and figure 4 and table 1 indicates that proposed Wavelet-CNN's classification accuracy is better by 7.33%, by 6.11%, by 4.34%, by 3.78% and by 1.79% than CFS-Random Forest, CFS-ANN, MF-Random Forest, MI- ANN, CNN respectively.

Table 2. Sensitivity for CFS-RF, CFS-ANN, MI-RF, MI-ANN, CNN and Wavelet-CNN

Techniques Used	Sensitivity for Normal Brain	Sensitivity for AD affected brain
CFS-Random Forest	0.858	0.8481
CFS-ANN	0.8681	0.8593
MI-Random forest	0.8899	0.8667
MI-ANN	0.9014	0.863
CNN	0.9261	0.8722
Wavelet-CNN	0.942	0.8889

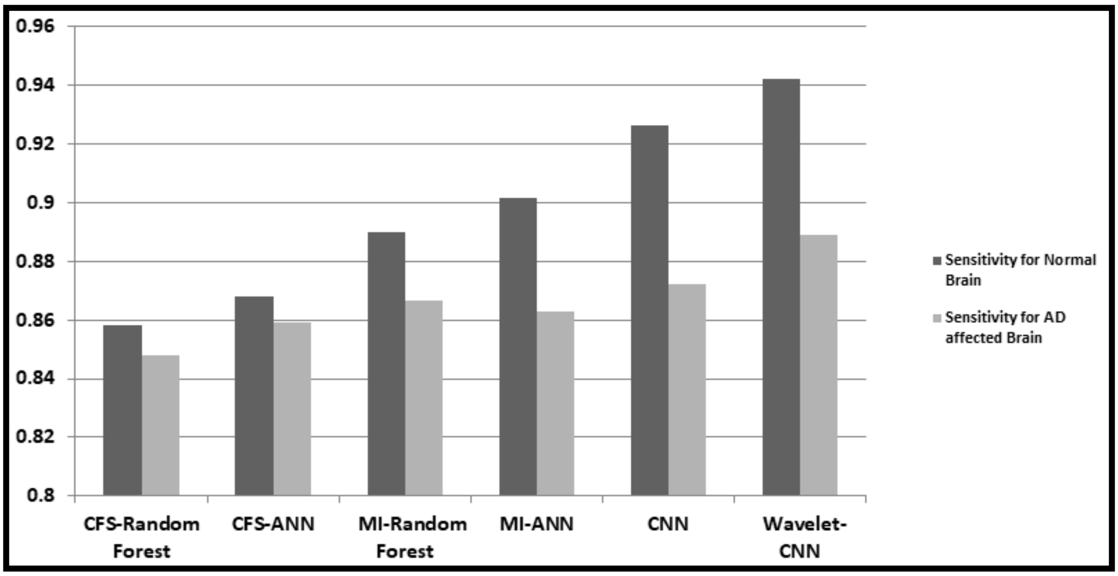


Figure 5. Graphical Representation of Sensitivity of CFS-RF, CFS-ANN, MI-RF, MI-ANN, CNN and Wavelet-CNN

Table 2 and figure 5 and table 2 indicates that proposed Wavelet-CNN 's sensitivity is better by 9.33%, by 8.2%, by 5.7%, by 4.4% and by 1.7% than CFS-Random Forest, CFS-ANN, MF-Random Forest, MI- ANN, CNN respectively for normal. The Sensitivity of proposed Wavelet-CNN performs better by 4.7%, by 3.4%, by 2.5%, by 2.96% and by 1.89% than CFS-Random Forest, CFS-ANN, MF-Random Forest, MI- ANN, CNN respectively for AD.

Table 3. Positive Predictive Value for CFS-RF, CFS-ANN, MI-RF, MI-ANN, CNN and Wavelet-CNN

Techniques Used	Positive predictive value for Normal Brain	Positive predictive value for AD affected brain
CFS-Random Forest	0.8783	0.8237
CFS-ANN	0.8874	0.836
MI-Random forest	0.895	0.8603
MI-ANN	0.8937	0.8727
CNN	0.9025	0.9023
Wavelet-CNN	0.9155	0.9231

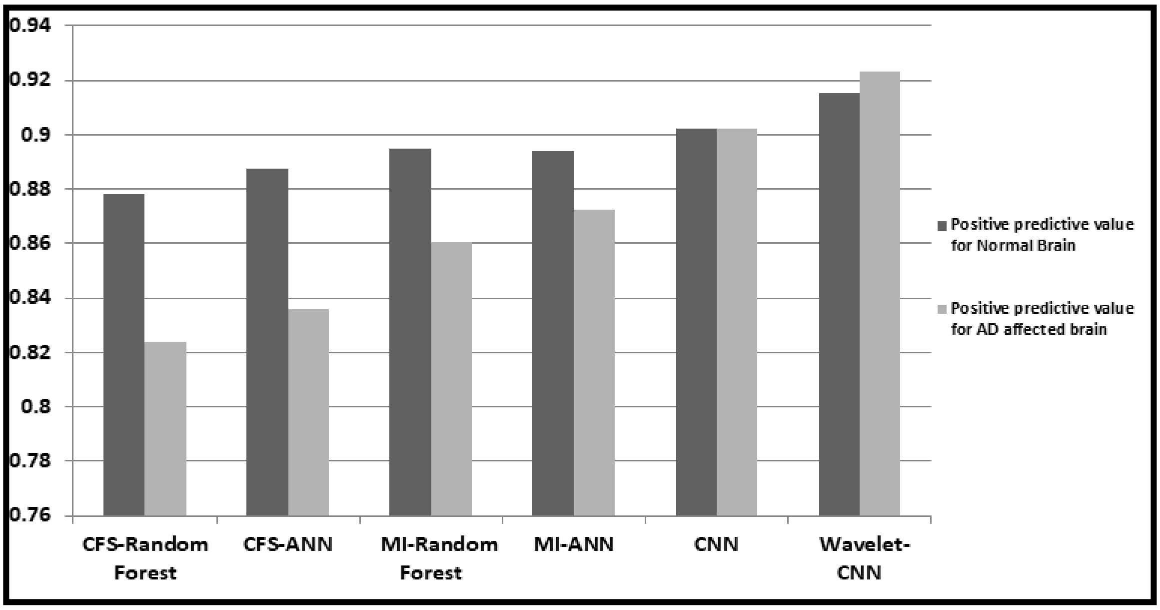


Figure 6. Graphical Representation of Positive Predictive Value of CFS-RF, CFS-ANN, MI-RF, MI-ANN, CNN and Wavelet-CNN

Table 3 and figure 6 shows that the Positive Predictive Value of proposed Wavelet-CNN performs better by 4.15%, by 3.12%, by 2.26%, by 2.41% and by 1.43% than CFS-Random Forest, CFS-ANN, MF-Random Forest, MI- ANN, CNN respectively for normal. The Positive Predictive Value of proposed Wavelet-CNN performs better by 11.4%, by 9.9%, by 7%, by 5.6% and by 2.3% than CFS-Random Forest, CFS-ANN, MF-Random Forest, MI- ANN, CNN respectively for AD.

Table 4 and figure 7 shows that the F Measure of proposed Wavelet-CNN performs better by 6.75%, by 5.65%, by 3.98%, by 3.41% and by 1.57% than CFS-Random Forest, CFS-ANN, MF-Random Forest, MI- ANN, CNN respectively for normal. The F Measure of proposed Wavelet-CNN performs better by 8.04%, by 6.64%, by 4.8%, by 4.3% and by 2.1% than CFS-Random Forest, CFS-ANN, MF-Random Forest, MI- ANN, CNN respectively for AD.

Table 4. F-Measure for CFS-RF, CFS-ANN, MI-RF, MI-ANN, CNN and Wavelet-CNN

Techniques Used	F-Measure for Normal Brain	F-measure for AD affected brain
CFS-Random Forest	0.868	0.8357
CFS-ANN	0.8776	0.8475
MI-Random forest	0.8924	0.8635
MI-ANN	0.8975	0.8678
CNN	0.9141	0.887
Wavelet-CNN	0.9286	0.9057

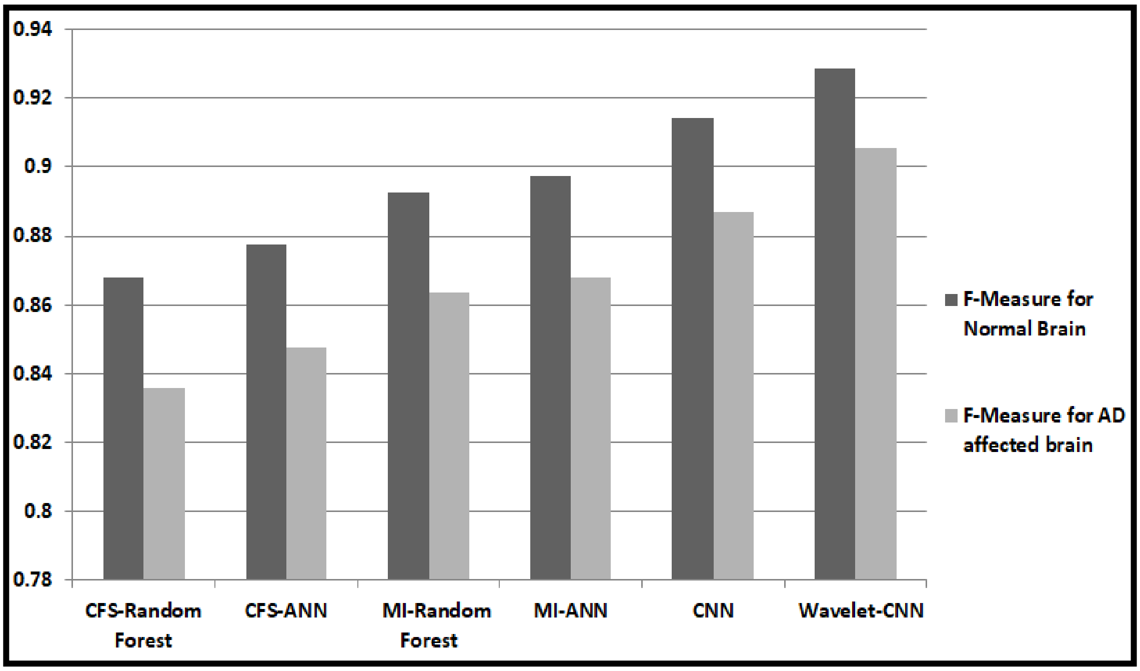


Figure 7. Graphical Representation of F-Measure of CFS-RF, CFS-ANN, MI-RF, MI-ANN, CNN and Wavelet-CNN

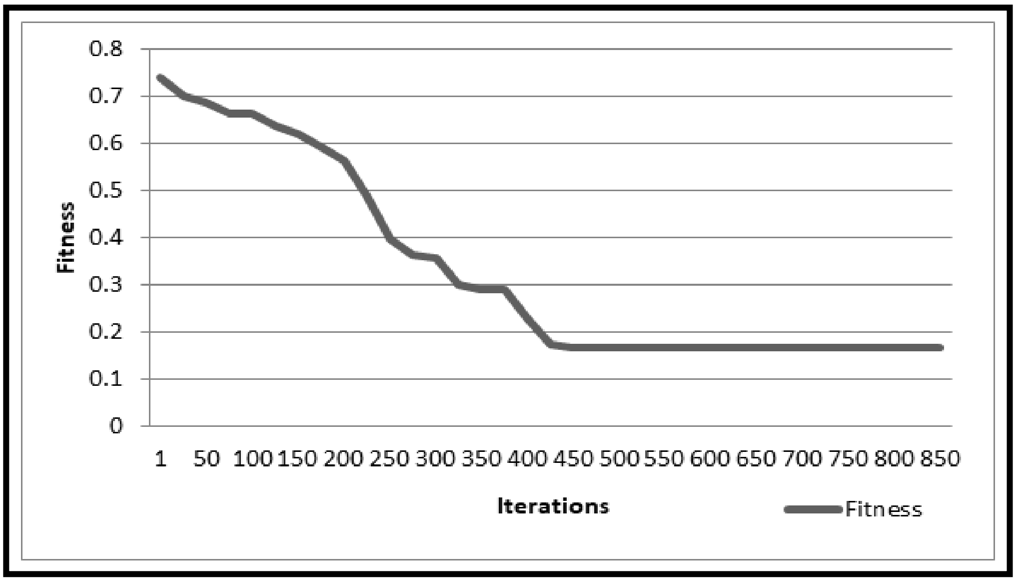


Figure 8. Graphical Representation of Best Fitness Where Convergence Occurs at Iteration Number 450

Figure 8 shows the Best Fitness value. Here the convergence occurred at iteration number 450.

Conclusion

Most patients cannot detect the warning signs of Alzheimer's Disease (AD) by themselves, as it is essential to diagnose before it reaches the critical stage. MRI brain images are popularly used to identify AD. Automatic AD detection frameworks based on image processing methods and machine learning algorithms are widely used. Feature selection techniques based on MI and CFS are used for decreasing the dimensionality of feature sets obtained from Wavelet Transform. Few feature subsets are chosen depending on relevancy factors. Thus selected features are subjected to classification methods like Random Forest, Artificial Neural Network, due to their exceptional attributes of non-linearity, fault tolerance, adaptivity, self-learning, and improvement in input to an output mapping, CNN which is a very well-known deep learning algorithm and Wavelet-CNN. It is shown from the results that the proposed Wavelet-CNN's Classification Accuracy is better by 7.33% when compared to CFS-Random Forest, by 6.11% when compared to CFS-ANN, by 4.34% when compared to MF-Random Forest, by 3.78% when compared to MI- ANN and by 1.79% when compared to CNN. The AD stage's classification accuracy can be further improved in future work. For evaluation, larger dataset may be utilized. The efficacy of the proposed method can be evaluated for other medical conditions like brain tumor, breast cancer, lung nodules.

References

Patterson C, The state of the art of dementia research: New frontiers. *World Alzheimer Report* 2018: 1-46.

Soliman SA, Mohsen H, El-Dahshan ESA, Salem ABM. Exploiting of Machine Learning Paradigms in Alzheimers Disease. *International Journal of Psychiatry and Psychotherapy* 2019: 1-11.

Beheshti I, Demirel H, Farokhian F, Yang C, Matsuda H, Alzheimer's Disease Neuroimaging Initiative. Structural MRI-based detection of Alzheimer's disease using feature ranking and classification error. *Computer methods and programs in biomedicine* 2016; 137: 177-193.

Matsuda H, Asada T, Tokumaru AM. (Eds.). *Neuroimaging diagnosis for Alzheimer's disease and other dementias.* Tokyo: Springer Japan 2017.

Jahnavi M. *Introduction to Neural Networks, Advantages and Applications, towards Data Science* 2017.

Tamije Selvy P, Dharani VP, Indhuja A. A Novel Approach for Brain Tumor Detection using Deep Learning Technique. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 2019; 5(2): 169-175.

Kim M, Yun J, Cho Y, Shin K, Jang R, Bae HJ, Kim N. Deep learning in medical imaging. *Neurospine* 2019; 16(4): 657-668.

Masoumi M, Toews M, Lombaert H. Wavelet Brain: Characterization of human brain via spectral graph wavelets, arXiv preprint arXiv:1906.06158, 2019: 1-9.

Acharya UR, Fernandes SL, Wei Koh JE, Ciaccio EJ, Fabell MKM, Tanik UJ, Yeong CH. Automated detection of Alzheimer's disease using brain MRI images–a study with various feature extraction techniques. *Journal of Medical Systems* 2019; 43(9): 1-14.

Jha D, Alam S, Pyun JY, Lee KH, Kwon GR. Alzheimer's disease detection using extreme learning machine, complex dual tree wavelet principal coefficients and linear discriminant analysis. *Journal of Medical Imaging and Health Informatics* 2018; 8(5): 881-890.

Long X, Chen L, Jiang C, Zhang L, Alzheimer's Disease Neuroimaging Initiative. Prediction and classification of Alzheimer disease based on quantification of MRI deformation. *PloS one* 2017; 12(3): 1-19.

Tanveer M, Richhariya B, Khan RU, Rashid AH, Khanna P, Prasad M, Lin CT. Machine learning techniques for the diagnosis of Alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 2020; 16(1s): 1-35.

Khagi B, Kwon GR, Lama R. Comparative analysis of Alzheimer's disease classification by CDR level using CNN, feature selection, and machine-learning techniques. *International Journal of Imaging Systems and Technology* 2019; 29(3): 297-310.

Bringas S, Salomón S, Duque R, Montaña JL, Lage C. A Convolutional Neural Network-Based Method for Human Movement Patterns Classification in Alzheimer's Disease. *In Multidisciplinary Digital Publishing Institute Proceedings* 2019; 31(1): 1-9.

Wang SH, Zhang Y, Li YJ, Jia WJ, Liu FY, Yang MM, Zhang YD. Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization. *Multimedia Tools and Applications* 2018; 77(9): 10393-10417.

Fujieda S, Takayama K, Hachisuka T. *Wavelet convolutional neural networks.* arXiv preprint arXiv:1805.08620, 2018: 1-10.

Jack Jr CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Weiner MW. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 2008; 27(4): 685-691.

Hampel H, O'Bryant SE, Durrleman S, Younesi E, Rojkova K, Escott-Price V, Alzheimer Precision Medicine Initiative. A precision medicine initiative for Alzheimer's disease: the road ahead to biomarker-guided integrative disease modeling. *Climacteric* 2017; 20(2): 107-118.

Ravi A, Giriprasad MN, Naganjaneyulu PV. SAR Images Denoising Using Invasive Weed Optimization Scheme. *International Journal of Pure and Applied Mathematics* 2018; 119(15): 2371-2389.

Karpagam V, Rangarajan R. A simple and competent system for content based retrieval of images using color indexed image histogram combined with discrete wavelet decomposition. *European Journal of Scientific Research* 2012; 73(2): 278-190.

Udhaya Suriya TS, Rangarajan P. Brain tumour detection using discrete wavelet transform based medical image fusion. *Biomedical Research* 2017; 28(2): 684-688.

Karegowda AG, Manjunath AS, Jayaram MA. Comparative study of attribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management* 2010; 2(2): 271-277.

Vergara JR, Estévez PA. A review of feature selection methods based on mutual information. *Neural computing and applications* 2014; 24(1): 175-186.

Khan Z, Gul A, Perperoglou A, Miftahuddin M, Mahmoud O, Adler W, Lausen B. Ensemble of optimal trees, random forest and random projection ensemble classification. *Advances in Data Analysis and Classification* 2020; 14(1): 97-116.

Chaudhary A, Kolhe S, Kamal R. An improved random forest classifier for multi-class classification. *Information Processing in Agriculture* 2016; 3(4): 215-222.

Wang N, Chen J, Xiao H, Wu L, Jiang H, Zhou Y. Application of artificial neural network model in diagnosis of Alzheimer's disease. *BMC neurology* 2019; 19(1): 1-8.

Valladares-Rodríguez S, Anido-Rifón L, Fernández-Iglesias MJ, Facal-Mayo D. A Machine Learning Approach to the Early

Diagnosis of Alzheimer's Disease Based on an Ensemble of Classifiers. *In International Conference on Computational Science and Its Applications* 2019: 383-396.

Islam J, Zhang Y. Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain informatics* 2018; 5(2): 1-14.

Wang SH, Phillips P, Sui Y, Liu B, Yang M, Cheng H. Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *Journal of medical systems* 2018; 42(5): 1-11.

Xie Z, Cao X, Liu Y, Wu X. Decisions on investing social funds in technology research and development based on cognitive psychology. *NeuroQuantology* 2018; 16(6): 397-404.