Thesis Final -modified - plag.pdf

by NEHA GARG

Submission date: 05-Dec-2025 07:05PM (UTC+0530)
Submission ID: 2836579513
File name: Thesis_Final_-modified_-_plag.pdf (10.5M)
Word count: 49403
Character count: 240030

Chapter 1

Introduction - Alzheimer's Disease Detection

This chapter introduces the background of Alzheimer's Disease, its stages, symptoms, structural changes in the brain, biomarkers, and neuroimaging studies. This chapter throws light on the significance and challenges of early detection of Alzheimer's Disease. Furthermore, motivation, problem statement, and thesis organization are discussed.

1.1. Background

Alzheimer's disease (AD) is a chronic degenerative brain disorder marked by the deteriorations of cognitive, behavioural, and communication skills.AD manifests through a range of mental, behavioural, and psychological symptoms that progressively worsen over time [1]. AD symptoms affect a person's ability to function independently. As the world's population grows older, the prevalence of Alzheimer's Disease is predicted to rise significantly, creating an increasing burden on healthcare systems, families, and society.

With 60–80% of cases, AD is the primary contributor to dementia in older persons. One of the disease's early symptoms is forgetting recent events. An individual affected by Alzheimer's has severe memory loss and becomes unable to perform everyday ordinary tasks as the condition worsens [2]. The report [3] states that more than 6.1 million people in the USA already have AD with projections estimating 13.8 million by 2060. The disease is not just affecting the people of the USA, but it is spreading worldwide with matching intensity. At present, AD is not fully treatable, though its progression can be controlled and mitigated. So, early AD detection is very crucial.

Neuropsychological tests, clinical assessments, and neuroimaging testing tools have been used traditionally to diagnose AD [4]. However, these methods often detect changes in the brain only after significant cognitive impairment has already occurred. Often, substantial irreversible damage has been done by this point, which reduces the efficacy of interventions. The current research focuses on early AD detection, biomarker discovery, and the development of disease-modifying medicines. Studies are underway to leverage machine learning and artificial intelligence in order to increase the precision of diagnosis and predict the course of disease. Early detection and intervention continue to be crucial priorities to combat Alzheimer's disease and enhance the living standards of patients with their loved ones.

1.2. Mild Cognitive Impairment Stage

Identifying AD in its initial cycle particularly during Mild Cognitive Impairment (MCI) stage, is vital for initiating appropriate treatments. MCI is a pivotal phase in comprehending and treating the course of cognitive deterioration. MCI [5] is the intermediate stage between the severe condition of Alzheimer’s disease and normal cognition level. Patients with MCI can perform their daily normal tasks despite having a lower cognitive capacity and a less functional brain. This reduced cognitive level includes memory, attention, language, or executive function deficits. MCI patients face difficulty in problem-solving and decision-making. There are two forms of MCI: progressive MCI and stable MCI. Individuals with MCI who convert into AD are referred to have progressive MCI and those who do not convert into AD and revert to normal cognitive function are considered stable MCI.

1.3. Symptoms of Alzheimer’s Disease

AD is a slowly advancing neurological deteriorative syndrome that considerably worsens intellectual and functional abilities. MCI is regarded as a transitional phase, where cognitive decline surpasses normal aging but is not severe enough to significantly obstruct everyday functioning. Alzheimer’s disease is frequently mistaken for dementia and normal aging. AD [6] is characterized by severe memory loss, which is not a sign of aging normally. In normal aging, bone density may decrease, and skin may become more brittle. Hearing and vision may also deteriorate along with a drop-in metabolic rate. Although a small memory loss, such as a slower recall of knowledge, is frequent, cognitive decline that interferes with day-to-day functioning is not a standard part of normal aging.

Dementia refers to a significant cognitive decline that is severe enough to impair social functioning. It may be the consequence of several diseases that harm brain tissue. Dementia can take many different forms, each with unique symptoms and causes. Vascular dementia, for instance, results from reduced blood supply to a particular brain site, such as after a stroke. Frontotemporal dementia includes personality changes in the early stages and memory deficits in advanced stages. In this dementia, the primarily affected brain regions are the frontal and temporal lobes which are responsible for the personality and behaviour of a person. Patients with hydrocephalus and Parkinson's disease may also have dementia. Alzheimer’s disease is a special case that targets the medial and temporal region of brain in early stages while Dementia is like an umbrella that includes many similar brain disorders.

1.4. Structural changes in Alzheimer’s Disease brain

Alzheimer’s Disease (AD) is the most frequently diagnosed form of Dementia arising due the formation of tau proteins and beta-amyloids in the brain. This makes the brain to lose synapses. As a result, several brain regions start to shrink, neuronal connections may decline, and numerous neurons around the brain become harmed and die. The entorhinal cortex [7] and hippocampus [8] undergo structural atrophies in the preclinical stage. In later stages, the overall volume of the brain decreases considerably due to the variations in various lobes of the brain. The atrophies in the parietal and temporal lobes impact body orientation, language, and memory abilities [9]. The atrophies in the frontal lobe led to impairments in decision-making, reasoning, and social behaviour. Significant loss of grey matter occurs in areas associated with memory, language, and executive functions. The heavily impacted areas are the posterior cingulate cortex, medial temporal lobe, hippocampus, and amygdala. The posterior cingulate cortex looks for memory retrieval and attention. The other structural atrophies include cortical surface reduction and ventricle expansion. This brain degeneration becomes extensive by the final phase of AD and results in a notable loss of brain tissue. Fig. 1.1 depicts the atrophies in the brain due to AD.

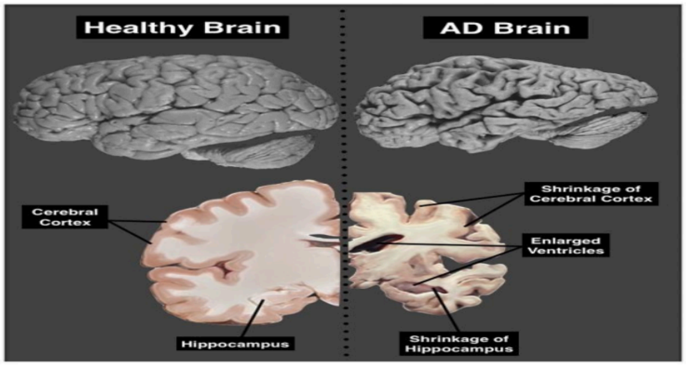


Fig. 1.1 Brain affected with Alzheimer’s Disease

Fig.1.1 depicts the normal and AD brain. The structural and grey matter changes at MCI stage are very minute which cannot be detected by the standard tests like Positron Emission Tomography ligands and Cerebral Spinal Fluid assays. Early AD detection requires the detection of these subtle structural and grey matter atrophies in the brain. With the advancement of the disease, these structural and grey matter changes also increase. Thus, a feature extraction technique is required that can detect these subtle and localized structural changes along with grey matter density in the AD brain and may provide effective classification among MCI, AD, and NC groups, thereby facilitating early diagnosis of AD.

1.5. Significance of Early Alzheimer's Disease Detection

Although AD is not completely curable at this moment, early diagnosis and intervention can mitigate symptoms and improve the well-being of people who are affected [10-12]. Research is still being done to learn more about the disease's underlying processes, create better treatments, and possibly even avoid it altogether. Drugs can either temporarily alleviate AD symptoms or halt its progression. Early AD detection helps in the following ways.

- Early detection allows for the timely initiation of treatments.
- The quality of life is improved by early AD detection. Early awareness makes it possible to maintain a sense of control over one's life, engage in worthwhile activities, and strengthen social ties.
- Patients who are detected early are more likely to be included in clinical trials, which improves the understanding of how diseases proceed and helps in the creation of new treatments.
- The preclinical and prodromal stages of AD are identified with the aid of early diagnosis, which is crucial for comprehending the disease process and developing preventative therapies.

1.6. Modalities Used for Alzheimer's Disease

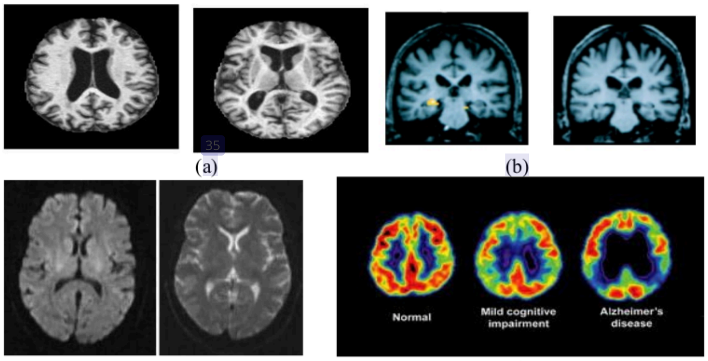
Neuroimaging is essential for the early AD detection because it provides details on the structural, functional and metabolic changes in the brain. There are different image modalities available in the literature that have been used for AD detection. These are Magnetic Resonance Imaging (MRI), Functional MRI, Diffusion Tensor Imaging (DTI), and Positron Emission Tomography (PET).

MRI [13] is a commonly accessible non-invasive diagnostic tool capable of detecting minute structural abnormalities in the brain. MRI offers high resolution and soft tissue contrast, which help to capture structural details. MRI highlights hippocampal and cortical atrophy, ventricular enlargement, and grey matter atrophies in the brain. This nature of MRI makes it strong enough for detecting AD in the early stages as well as monitoring its progression in subsequent stages.

Functional Magnetic Resonance Imaging (fMRI) [14-15] measure the activity of the brain by analysing blood flow dynamics. It provides insight into the functioning of different brain regions and how these regions get affected by AD. The fMRI is not generally used as a single diagnostical tool, rather it is used in multimodal approaches.

DTI [16] measures the white matter density in the brain by analyzing the water molecules movement in the brain. The white matter density is the pathways that connect the different brain regions. White matter pathways frequently suffer degeneration in Alzheimer's disease, which impairs communication between various brain regions.

PET [17] is a highly effective diagnostic tool for Alzheimer’s disease detection. PET can capture metabolic alterations, tau tangles, and amyloid plaques and thus provides important information for disease diagnosis, early detection, and tracking. Fig. 1.2 depicts the MRI, fMRI, DTI, and PET modalities.



(a) (b)

(c) (d)

Fig.1.2 Image Modalities a) MRI b) fMRI c) DTI and d) PET

Through the utilization of these neuroimaging techniques, clinicians and researchers can achieve better understanding of AD, facilitating better diagnostic accuracy, early detection, and personalized treatment strategies.

1.7. Limitations of Neuroimaging Studies for Early AD Detection

Although neuroimaging studies have greatly contributed to the detection of AD, they have a few drawbacks [18]. These are

- **High Cost:** The accessibility and cost of neuroimaging tests, particularly PET scans, limit their widespread use in standard clinical practice.
- **Limited Availability:** There is a lack of availability of advanced imaging facilities in regions of low-resource environments which makes the early AD detection difficult.
- **Invasiveness and Safety Concerns:** MRI is non-invasive while PET imaging is invasive in nature. PET includes the injection of radioactive tracers, whose repeated use raises safety issues.
- **Variability in Interpretation:** Diagnostic accuracy may be impacted by variations in radiologists' interpretations of imaging results.
- **Sensitivity vs. Specificity:** Despite its sensitivity in detecting changes in the brain, neuroimaging may not necessarily be Alzheimer's disease specific. White matter abnormalities and other imaging findings can overlap with other neurological disorders or with normal aging.

42

1.8. Biomarkers of Alzheimer's Disease

The biomarkers are the clinical markers that are used to assess the existence and evolution of a disease [19]. These are helpful in early AD detection. There are three bioindicators used for AD detection. These are Amyloid beta, Tau Protein, and Neurodegeneration indicators. These biomarkers are available in several bodily fluids, tissues, or imaging data and help in early AD detection, tracking, and assessment of AD as well as the efficacy of various treatments.

In AD, the brain undergoes substantial structural changes driven by widespread neuronal death. Neuroimaging techniques like MRI [20] provide information about the internal structures of the human brain. MRI can assess the shape, size, and volume of the hippocampus, amygdala, and cortical surface of the brain. MRI is also a good indicator of grey matter density variation in AD brain. Different types of MRI scans are utilized in studies to study cerebral anatomy and functional dynamics in normal aging and Alzheimer's progression. Structural MRI [21] is used to assess the structural changes, and functional MRI is used to assess the functional activity of the brain.

The CSF [22] is a clear and colorless fluid that covers and insulates the brain and spinal cord. Besides, CSF includes different nutrients like water, proteins, lipids, electrolytes, sugars, and other chemicals. These substances support the health of brain cells. Any neurological issue creates a disturbance in these substances and the measurement of variations can help in the detection of neurological issues. According to the research, Early-stage AD may modify the measurements of biomarkers such as tau and beta-amyloid in cerebrospinal fluid. CSF biomarkers follow the low amyloid beta and high levels of tau proteins in Alzheimer’s disease. CSF testing helps in detecting AD in early stages with high accuracy. This helps in monitoring disease progression and differentiating AD from other dementias. CSF testing is invasive and requires a lumbar puncture.

1.9. Challenges in Early AD Detection

Early AD detection is required for early treatment and threptic measures, leading to more effective symptom control. Early AD detection [23-24] faces several challenges in many directions like lack of standardization in clinics, biological complexity, data complexity, heterogeneity of the disease, proper biomarker identification, undermining the link between AD and normal aging, dearth of techniques for asymptomatic AD detection. The pattern of AD progression is different from one brain to another so the same treatment cannot be given to two different individuals. The symptoms in the initial stage of AD are very mild and often confused with normal aging. Biomarkers like amyloid plaques or tau tangles are linked to AD but their presence is difficult to detect in the asymptomatic AD brain. An Asymptomatic person does not show a cognitive decline in the initial stages, but the level of these biomarkers continues to change as the disease progresses.

These challenges can be mitigated by utilizing advanced imaging techniques like PET, MRI, and fMRI with improved machine learning methods. Such an approach facilitates the detection of early brain tissue atrophies related with Alzheimer's disease. Fig. 1.3 indicates the challenges for AD detection.

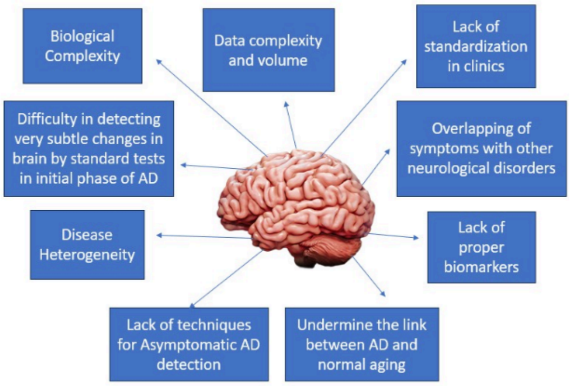


Fig. 1.3 Challenges in early AD detection

Even with improvements in AD diagnostics, early detection is still difficult because of several shortcomings in current techniques. The sensitivity, expense, accessibility, and invasiveness of existing methods are some of these drawbacks. These are explained in the following subsections.

1.9.1. Limited sensitivity to early AD detection

Limited sensitivity refers to the inability of AD detection methods to classify persons having early-stage AD accurately. Some individuals in the initial phase of the disease may be left for detection because their symptoms or biomarkers are not distinct enough from normal aging or other conditions. The factors contributing to the limited sensitivity of early AD detection methods are as.

- **Overlap with Normal Aging:**

The symptoms of early AD detection are overlapped with normal aging. This makes it challenging to differentiate between healthy individuals and those in the preclinical stages of AD.

- **Non-Specific Biomarkers:**

Many biomarkers like as amyloid-beta and tau levels are not disease-specific and may also be elevated in other neurodegenerative disorders. The presence of these biomarkers does not always correlate with clinical symptoms, leading to potential misdiagnosis.

- **Delayed Detection:**

The disease may have progressed beyond the early stages by the time structural alterations (such as brain atrophy) can be seen through imaging. Early AD detection has subtle structural changes and is generally mistaken for other neurological diseases, which results in limited sensitivity to early detection of AD.

1.9.2. High cost of advanced techniques

The cost of neuroimaging studies in Alzheimer's disease detection is very high which imposes a major challenge against AD patients. The high cost of advanced techniques makes them less accessible to the common man and contributes to delays in diagnosing AD. The imaging tools such as MRI, fMRI, PET, and CT scans are expensive due to the high cost of equipment, maintenance, and specialized personnel. Biochemical Assays are used to examine the AD possibility. Analysing cerebrospinal fluid (CSF) for tau and amyloid proteins requires specialized procedures that not all patients can afford or use.

1.9.3. Difficult to manage high dimensional data

Neuroimaging studies are high-dimensional due to the complex and detailed nature of brain imaging. The high-dimensional data generated by neuroimaging and other diagnostic tools require sophisticated analysis techniques, which can be computationally expensive and difficult to interpret. The high-dimensional data can lead to overfitting problems. Longitudinal studies are good for tracking the disease, but they need large storage requirements. In multimodal techniques, different modalities are used which increases data complexity and dimensionality.

1.9.4. Extraction of relevant features

Identifying relevant features from complex and noisy data is crucial but challenging. Even advanced imaging techniques like MRI, PET, and CT scans may not capture the earliest microstructural atrophies in the brain, limiting the capability to diagnose AD in initial phase. Extraction of relevant features helps in improving the precision, computational performance, and interpretability of machine learning models. Irrelevant features introduce noise and reduce the classification accuracy.

1.10. Role of Machine Learning for AD Detection

Machine learning (ML) [25] plays an important role in AD detection. ML brings transformation in early AD detection by analyzing subtle structural changes in brain scans which helps in early AD detection, increasing classification accuracy, and enabling individualized treatment methods. Normally, the use of clinical assessments, cognitive tests like mini-mental state examinations, and neuroimaging are in practice to detect AD. These methods take time and do not guarantee true results. These methods are subjective, and subject to variation. Alternatively, machine learning can deal with large amount of complex data and provide effective data-driven solutions.

Different neuroimaging methods like PET, MRI, and Functional MRI (fMRI) offer essential details regarding structural and functional atrophies in the brain, supporting prompt detection of AD. ML techniques can detect subtle changes through the analysis of structural, functional, and metabolic abnormalities in the brain. Machine Learning techniques enable the automated extraction of features from these images, enhancing the ability to diagnose AD in initial phase, predict the trajectory of the disease, and distinguish between AD, MCI, and normal aging.

Furthermore, machine learning [26] makes it easier to combine imaging data with additional biomarkers, like blood-based biomarkers and cerebrospinal fluid (CSF) studies, to create comprehensive predictive models. Machine learning (ML) can integrate information through several sources, such as MRI, PET, genetic information, and cognitive scores to develop more thorough and precise models for early AD detection. ML models can detect subtle changes in speech patterns, such as word-finding difficulty, sentence structure abnormalities, or reduced speech fluency, which may indicate early signs of AD.

1.11. Motivation and Problem Statement

Alzheimer's Disease (AD) is an irretrievable neurological syndrome marked by progressive cognitive decline, memory loss, and behavioural disturbances, causing a major drop in the quality of life. AD creates an enormous burden on patients, caregivers, and healthcare systems around the globe. The AD cases are anticipated to increase substantially with the world’s population age, making early detection and treatments more critical than ever.

The effectiveness of current treatments is limited because AD is frequently detected at advanced stages when substantial neuronal damage has already happened. Thus, there is a compelling need for early and accurate detection methods that can identify the disease in its preclinical stages. Early detection can provide well-timed interventions that can decelerate disease advancement, improv health results, and provide better planning and support for patients and their families. Late diagnosis is the main problem of this disease, and it is due to the lack of the ability of standard tests for diagnosing very minute structural changes in the initial stages of AD brain. There are very minute differences in structure and grey matter density details at various stages of AD. That's why existing techniques are giving good outcomes for AD/NC classification but with AD/MCI and MCI/NC detection, it still needs improvement. The main problems against existing techniques are high dimensionality of features, inaccurate segmentation of atrophied regions, imperfect registration of images over templates, lack of standardized data, heterogeneity in Magnetic Resonance Texture Analysis, and expertise in different software tools like FreeSurfer. Thus, this highlights the need of feature extraction techniques that can capture both biomarkers i.e., structural changes as well as grey matter density fluctuation without facing the above-mentioned problems and can give good classification results with AD/MCI and MCI/NC.

In recent years, advancements in imaging technologies and computational methods have opened new avenues for identifying subtle changes in the brain associated with early stages of AD. This thesis focuses on leveraging such techniques, particularly wavelet-driven feature extraction using machine learning, to upgrade the preciseness and credibility of early AD diagnosis. By exploring these methodologies, this research seeks to promote the development of reliable diagnostic solutions that can facilitate early interventions and upgrade the quality of life for those at risk. This thesis introduces seven novel feature extraction methodologies designed to overcome the constraints of existing advanced methods particularly addressing challenges like high dimensionality and detection of subtle structural and grey matter density changes across various stages of AD.

1.12. Contribution of the Proposed Research Work

The contribution of the proposed research work for early and accurate AD detection can be summarized as:

- Feature extraction technique using multiple MR slices is used which can capture the structural atrophies in large regions of the brain effectively in comparison to a single slice. This technique is better in a case when the disease spreads in different parts of the brain regions.
- A novel feature extraction algorithm is used to address the problem of scalability. The proposed technique i.e. wavelet-based shifted elliptical local binary pattern can be automated, allowing for the rapid processing of large volumes of MR images.
- A novel feature extraction algorithm is used to provide early AD detection. The proposed wavelet-based shifted circular-elliptical local binary descriptor can capture localized directional information from the MR scans and provide exceptional results for early AD detection.
- A novel feature extraction algorithm is used to mitigate the challenges of high dimensional data space. The proposed mean energy-based feature extraction method provides the maximum results with a minimum number of features (160) without using feature dimensionality reduction or feature selection methods. Thus, this technique is the simplest one, yet it provides good classification results, minimizes the scalability problem, and can even provide good results for early AD detection.
- The proposed feature extraction techniques like direct coefficient-based and GLCM-based feature extraction techniques evaluate various feature selection algorithms like PSO, ACO, GA, and t-test for AD classification. The Genetic Algorithm gives the best results in both techniques.

1.13. Thesis Organization

This thesis is divided into seven chapters, each accompanied by a concise summary presented below.

- **Chapter 1** provides concise overview of the background of Alzheimer’s Disease, its biomarkers, and neuro-imaging tests used for AD detection. It also gives the motivation for this thesis and challenges for early AD detection; the inherent problems faced in early AD detection and defines the problem statement. This chapter also elaborates on the core contribution presented in the thesis.
- **Chapter 2** focuses on the literature review, which includes the existing advanced methods related to feature extraction and classification for early AD detection. The strengths and limitations of Alzheimer’s disease detection have been identified by analyzing the various

existing methods. Based on the limitations of existing advanced methods, the research gaps are noted, and the objectives of this research work are identified. Further, the publicly available datasets are also discussed which are used in this research work for experimental validation of the proposed methods.

- **Chapter 3** presents the direct coefficient-based method for early AD detection. In this technique, a 3-D MR image is divided into 32 2-D slices and then DD-DTCWT is applied on the 32 2-D MR center slices, and fifth-level detail coefficients (2048 coefficients) have been extracted. These features are used for AD classification. This work provides the most optimized wavelet-based feature extraction method for true 3-D Alzheimer's Disease data. Multiple brain slices are used in the proposed work which provides more consistent, accurate, and reliable information which may not be inspected by a single slice. This work also investigates the performance of conventional statistical methods namely, t-test and evolutionary algorithms viz. PSO, ACO, and GA for feature selection in AD classification.
- **Chapter 4** presents the histogram feature extraction technique for AD detection. This technique operates with lower dimensionality, making it more efficient relative to the direct coefficient-based feature extraction technique. Two histogram feature extraction-based methods are proposed in this chapter. In the first method, the 2-D MR images undergo a level 1 DD-DTCWT, resulting in decomposition into sixteen sub-images. Subsequently, the Edge-Weighted Local Binary descriptor is applied to sixteen sub-images and histogram features have been obtained for AD classification. In the second method instead of Edge-weighted LBP, sixteen sub-band images are subjected to SCELD, and histogram features are obtained for AD classification. In the second method, the performance of different versions like Mean WSCELD, Median WSCELD, Energy WSCELD, and Variance WSCELD has been investigated for histogram features.
- **Chapter 5** presents the statistical feature extraction techniques for AD classification. A total of three methods based on statistical features are proposed in this chapter. In all three methods, the 2-D MR images undergo a level 1 DD-DTCWT, resulting in decomposition into sixteen sub-images. In the first method, sixteen sub-images are subjected to shifted ELBP to obtain multiple micro and macro patterns. The statistical features obtained from these patterns are used for AD classification. In the second method, GLCM has been computed for sixteen sub-images from which statistical texture features were extracted and subsequently utilized for Alzheimer's disease classification. In the third method, sixteen sub-images are subjected to SCELD to obtain multiple micro and macro patterns.

Subsequently, statistical features derived from these patterns are utilized for Alzheimer's disease classification.

- **Chapter 6** presents a Mean energy-based feature extraction technique for early AD detection. This technique is quite simple but more efficient in comparison to the previous techniques as this technique does not require any dimensionality reduction or feature selection technique. The MR images are decomposed up to the fifth level using DD-DTCWT. Mean Energy features are obtained from real, and imaginary sub-bands up to fifth level and different combinations of mean energy have been investigated. The results with this technique are outstanding for all AD/ NC, AD/MCI, and MCI/NC classifications.

Finally, in **Chapter 7**, the conclusions inferred from this research work are summarized and potential future work in this area is highlighted.

Chapter 2

Literature Review

This chapter discusses the various methods of feature extraction and machine learning models presently employed for detecting Alzheimer's Disease. Their advantages and disadvantages have been discussed also. An intense review of existing techniques revealed the limitations and research voids in early Alzheimer's Disease detection. Subsequently, the research objectives are developed in response to the identified gaps, and they are thoroughly addressed in this thesis. Finally, the datasets and performance metrics have been discussed.

2.1. Introduction

This chapter includes the prominent studies available in the literature for Alzheimer's Disease detection. This survey includes studies based on various feature extraction methods with dimensionality reduction and feature optimization techniques that can be applied for the early AD detection. The main criteria for a literature review can be listed as.

- Early AD detection is crucially required for providing timely intervention to the patients, so Literature Survey is carried out to find the causes that create problems in early AD detection.
- The studies based on Magnetic Resonance Imaging are included as MR images capture the structural and grey matter changes in the brain nicely.
- A Literature Survey is conducted to analyze the existing techniques for feature extraction for AD detection.
- A Literature Survey is also conducted for the machine learning algorithms that are commonly utilized for AD detection.
- A Literature Survey is also conducted to find the maximum results for classifying AD with NC, AD with MCI, and MCI with NC in the existing studies.

Research shows that a lot of work is available on the feature extraction process of AD detection but there is less work involved in designing a novel classifier that is specifically capable of handling neuroimaging data.Fig.2.1 represents the different feature extraction techniques available in Literature for AD detection.

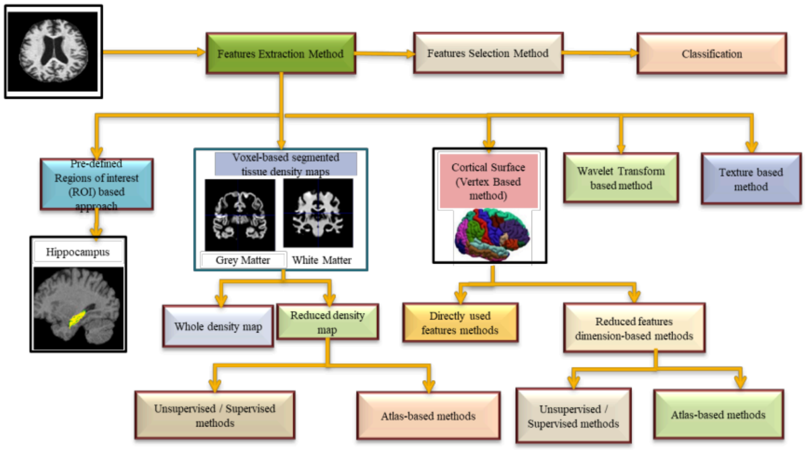


Fig. 2.1 Different Feature Extraction Methods

The following techniques are preferably used for extracting features for AD classification:

- Pre-defined Region of Interest (ROIs) - based approaches
- Voxel-Based Morphometry (VBM), or Voxel-based segmented tissue density maps
- Cortical surface or Vertex-based method
- Wavelet Transform-based method
- Texture-based method

2.2. Predefined ROI-Based Approaches

ROI technique includes the segmentation of that region of the brain which is strongly influenced by AD. The affected parts are the hippocampus, amygdala, and cortical surface which can be extracted manually and automatically. The neuronal loss results in the shrinking of the hippocampus, enlargement of ventricles, and cortical thinning. This technique requires prior information about the disease progression and pathology [27].

Uysal et al. [28] used Hippocampal atrophies as features for AD detection. The authors conducted the research using ITK-SNAP software [29,30] and observed that the group containing the features like left and right hippocampal volume, age, and gender, produced significant results. Sorensen et al. [31] used features like shape and texture of hippocampus, cortical thickness and volumetric measurement for classifying different stages of AD. The authors employed Freesurfer volumetry [32] to measure the volumetric and cortical thickness. The left and right hippocampi were independently segmented using a non-local patch-based multi-atlas approach. The segmented hippocampus was marked with 30 surface markers to calculate the leftward and dextral hippocampus shape labels. Gaussian filter response was used to generate the histograms of the hippocampus texture features from both hippocampi. This study highlights the hippocampus's texture as the foremost feature alongside parietal lobe thickness, the ventricles' and the hippocampus's volume.

Wang et al. [33] used hippocampal features for early AD detection. In this study, the hippocampal surface was divided into three regions using a hippocampal template viz. a superior region which includes the gyrus dentatus and CA2, CA3, and CA4 subfields, a lateral region having CA1 subfield, and an inferior medial region having the subiculum. They observed that variations in the subfields of CA1 and subiculum are more potential biomarkers than those in the subfields of CA2, CA3, and CA4. This work provided 81.1 % accuracy for AD/NC classification. Aghaei et al. [34] extracted ROI’s through heatmap of images. This work was not focussed on early AD detection. The study [35] applied a fusion approach of classifiers for AD classification using hippocampus size as a feature. In this work, the performance of different features like volumes of CSF, grey matter, white matter, combined volume of all three (grey matter, white matter, CSF), area of left hippocampus, right hippocampus and combination of features have been evaluated. The authors observed the best results with the area of the left hippocampus. This work provided 93.8 % accuracy for AD/NC classification. Ben et al. [36] combined the two biomarkers: hippocampal visual features and CSF for AD classification. This work provided 87% accuracy for AD/NC classification. D. Baskar et al. [37] extracted 447 texture and 10 shape features from the posterior cingulate cortex and hippocampus. Then nineteen features were selected for AD classification using a multiple-criterion feature selection method. The authors used the ensemble mode by combining two classifiers viz. Back-Propagation Artificial Neural Network (BPANN) and Kernel Fuzzy C-Means Clustering (KFCM) to boost the predictive performance. This method provided 98.63%, 96.54%, and 95% accuracy for distinguishing AD from NC, AD from MCI, and MCI from NC respectively. Table 2.1 includes the ROI-based studies available in the Literature for AD classification.

Table 2.1 Prominent studies for AD classification with ROI-based methods

Author	**Data Set**	**Database**	**Feature Type**	**Classifier**	**Performance**	**Observations**
(D. Bhaskar et al., 2018) [37]	AD=137 NC=162 MCI=210	Alzheimer's Disease Neuroimaging Initiative (ADNI)	Shape and textural features	BPANN + KFCM	**AD / NC** (Acc=98.63, Sen=79.56, Spe=85.63) **MCI/ NC** (Acc=95, Sen=78, Spe=82) **AD/MCI** (Acc=96.4, Sen=76, Spe=85)	Shape features have been used which are high in dimensionality. This makes them noisy, and not reliable in nature [38].
(Aghaei et al.,2023) [34]	sMCI=88 pMCI =87	ADNI	ROI's obtained through heatmap of image	3D-CNN	**sMCI /pMCI** (Acc=98.63)	This work does not focus on early AD detection.
(Farhan et al.,2014) [35]	AD=37 NC =48	Open Access Series of Imaging Studies (OASIS)	Left hippocampus area	Ensemble (SVM, MLP, Decision Tree)	**AD/NC** (Acc=93.75, Sen=87.5, Spe=100)	A manual segmentation process is used to prepare the ROI masks [40]. The dataset is very small.
(Uysal et al.,2020) [33]	AD=18 NC =159 MCI=217	ADNI	Hippocampus volume + age+ gender	KNN	**AD/NC** (Acc=98)	Study observed improvement in results when biomarkers are combined.
(Ben et al.,2015) [36]	AD=35 NC =72 MCI=111	ADNI "Bordeaux dataset	CSF Volume and visual features	SVM	**AD/NC** (Acc=87, Sen=100, Spe=75.5) **MCI/ NC** (Acc=78.22, Sen=83.34, Spe=70.73) **AD/MCI** (Acc=72.23, Sen=70, Spe=75) **AD/NC** (Acc=85, Sen=81, Spe=76)	The authors used Atlas-based segmentation for extracting hippocampi. This segmentation approach does not provide boundary details [41]; thus, ROI segmentation is not accurate in this approach.

Accuracy=Acc, Sensitivity=Sen, Specificity=Spe

Table 2.1 shows that hippocampal texture and volumetric features yield more promising AD classification results compared to features associated with hippocampus shape and size [42]. Shape features are unreliable due to their high dimensionality and noise [38]. These features yield noticeably better outcomes by reducing their dimensions. It has been noted that the ROI-based approaches face challenges while precisely extracting the hippocampal features from the complex brain structure [43–45].

2.3. Voxel-Based Morphometry (VBM)

VBM [46] technique provides the grey matter analysis of the brain. It helps in comparing the brains of different subjects by superimposing them on templates. The VBM process consists of the following steps.

- Creating a template using structural magnetic resonance images with high resolution.

- Spatial normalization [47] of MR images is performed and then they are segmented into the grey and white matter. The segmented images are smoothed with the help of a consistent Gaussian kernel.
- At last, the grey and white matter concentrations of different brains are compared voxel-wise.

VBM technique is better than the ROI-based method as it does not require segmentation [48-49]. VBM methods fall into two groups. In the first group, features are directly used for classification. These methods are called whole-density maps and suffer from high dimensionality. In the second group, feature dimensionality is decreased prior to classification. These methods are called reduced density maps. Dimensionality reduction is achieved through supervised methods, unsupervised methods, and an atlas.

Kloppel et al. [50] extracted GM density maps from the whole brain and used them as features. The authors obtained 95.6% accuracy for AD/NC classification and observed that the hippocampal regions of AD patients had a comparatively low GM density. Moller et al. [51] also acquired GM density maps from the entire brain for SVM-based AD/NC classification. Salvoitre et al. [52] and Gupta et al. [53] used PCA to reduce feature dimensions. This method comes under the unsupervised category. Unsupervised techniques do not require labelled data. The algorithm searches for the structure and underlying patterns to classify any object. Behashti et al. [54,55] used a supervised technique for reducing feature dimensions. These supervised techniques for feature selection are the Probability distribution function [54], statistical test, and genetic algorithm [55]. These studies obtained 89.7% [54] and 93% [55] accuracy respectively.

Cuingnet et al. [56] used a single atlas to perform five techniques using VBM, three surface-based, and two ROI-based techniques and compared their performance. The performance of VBM-based and surface-based studies was outstanding. Liu et al. [57] divided the brain into different patches using a single atlas. The authors captured the spatial relationship between different patches. Later in 2014, Liu et al. [58] solved the dimensionality problem by using a single atlas. The whole brain image was divided into different patches by using a single atlas. The two types of features were extracted i) Grey Matter (GM), and ii) spatial-correlation features. These features were then transformed into high-level features by two low-level classifiers, which were then fed to multiple high-level classifiers. A conclusion was made based on the results of each high-level classifier. Because a single atlas [59] is typically biased toward a certain class, it might not be able to adequately represent all the individuals by illustrating the subtle differences between the brains of AD patients and healthy participants. Consequently, most researchers employed multiple atlases.

Multiple atlases eliminate the impact of registration errors and provide complementary information for AD classification. Multiple atlases were utilized by Lepore et al. [60] to register images. The data gathered from multiple atlases was then averaged by the authors. Lepore et al. [60] and Koikkalainen et al. [61] extracted features by using multiple atlases. These atlases had the drawback that they were nonlinearly confined to a common space and thus were not able to capture the anatomical differences in different atlases. To solve this issue, Min et al. [62] selected multiple atlases from the different groups of patients by using the Affinity Propagation Algorithm [63]. These atlases were not confined to the common space and thus extracted more distinguished features. This method suffered from the problem of information redundancy due to the concatenation of features generated from different atlases. To solve this, Liu et al. [64] 2015 used multi-atlases and kept all the atlases in their original space. The features were extracted from the main atlas along with complementary information from other atlases.

In 2016, Liu et al. [65] examined the relationships between patients in the same atlas and between multiple atlases. The authors proposed the Relationship Induced Sparse Feature Selection approach (RISFS) which extracted uncorrelated features in each atlas space after registering the image onto multiple atlases. In 2016, Liu et al. [66] also presented a novel feature selection method based on structural information from several patients in the same class. The class may be divided into encoded subclasses that can be utilized for classification based on these structural details.

Huang et al. [67] in 2023 used VBM for analysing morphological changes in AD brain. They proposed a newly developed deep CNN model for identifying patients with MCI based on cerebral GM changes. The proposed method provided an accuracy of 80.9%, a sensitivity of 88.9%, and a specificity of 75%. Table 2.2 includes the prominent studies related to VBM-based feature extraction in Literature for AD/NC, AD/MCI, and MCI/NC classifications.

Table 2.2 Prominent studies for AD classification using voxel-based methods

Author	**Dataset**	**Method**	**Database**	**Feature**	**Classifier**	**Performance**	**Observations**
(Salvatore et al.,2015) [52]	AD=137 NC =162 sMCI=134 pMCI=76	Unsupervised (PCA)	ADNI	VOXEL WISE	SVM	**AD/NC** (Acc=76) **pMCI /sMCI** (Acc=66) **MCI/NC** (Acc=72)	PCA is utilized to lower the number of feature dimensions which results in spatial information loss [68]
(Gupta et al. ,2019) [53]	AD=81 NC =171 sMCI=35 pMCI=39	Unsupervised (PCA)	National Research Center for Dementia (NRCD)	Hippocampal Volume, Voxel-Based Morphometry, Cortical and Subcortical segmented region	k-Nearest Neighbour (k-NN)	**AD/NC** (Acc=89.47, Sen=93.75, Spe=88.33) **pMCI / sMCI** (Acc=73.91, Sen=70.58, Spe=83.33) **MCI/NC** (Acc=88.88, Sen=94.1)	PCA is utilized to lower the number of feature dimensions which results in spatial information loss [68]
					Random Forest (RF)	**AD/NC** (Acc=88.15, Sen=78.26, Spe=92.45) **pMCI/sMCI** (Acc=78.26, Sen=76.92, Spe=80) **MCI / NC** (Acc=87.3, Sen=93.87)	
					SVM	**AD/NC** (Acc=93.06, Sen=87.87, Spe=95.58) **pMCI / sMCI** (Acc=86.95, Sen=77.77, Spe=92.85) **MCI/NC** (Acc=95.23, Sen=95.77)	
(Beheshti et al.,2015) [54]	AD=130 NC =130	Supervised (Probability distribution function)	ADNI	VOXEL WISE	Linear SVM	**AD / NC** (Acc=89.65, SEN=87.73, SPY=91.57)	Image registration is done by DARTEL registration technique that does not create a deformations field for variable velocities. Feature extraction is done by using Partial Least Squares (PLS)method which overrates and gives redundant features [69].
(Beheshti et al.,2017) [55]	AD=160 NC =160 sMCI=65 pMCI=71	Supervised (test +GA)	ADNI	VOXEL WISE	SVM	**AD/NC** (Acc=93.01, Sen=89.13, Spe=96.8) **pMCI / sMCI** (ACC=75, Sen=76.92, Spe=73.23)	Features are selected through a t-test, which is prone to false positives (Type I errors) [70].
(Liu et al.,2014) [58]	AD=198 NC =229 MCI=225	Single atlas	ADNI	Patch-Based GM	SVM ensemble	**AD/NC** (Acc=92, Sen=91, Spe=93) **MCI/NC** (Acc=85.3, Sen=82.3)	Features are selected through a t-test, which is prone to false positives (Type I errors) [70].
Huang, Huaidong et a.,2023[67]	NC =73 MCI=111	Single atlas	ADNI	Grey Matter volumes	CNN	**MCI / NC** (Acc=80.9, Sen=88.9, Spec=75)	Compared to subjects with normal cognition, participants with MCI had grey matter atrophy mainly in the entorhinal cortex, frontal cortex, and bilateral frontotemporal lobes.

pMCI=progressive MCI sMCI=stable MCI

Table 2.2 indicates the methods for dimensionality reduction. One drawback of the unsupervised dimensionality reduction method is that it doesn't examine spatial information in images [68]. The single atlas-based method typically yields inaccurate results because it is biased towards a particular class [59]. Multiple atlas-based techniques have recently drawn attention from researchers due to their ability to examine subcategories within a class [71-72]. In addition, the multiple atlases are less biased toward particular groups of classes.

2.4. Cortical Surface-Based Methods

Surface-based morphometry is used to assess the cortical surface measurements like cortical thickness, cortical volume, density, etc. Alzheimer's disease shows noticeable cortical atrophies particularly in the temporal and parietal regions [73,74]. The cortical features that are directly used for classification come under directly used feature extraction methods and the cortical features used for classification after reducing their dimensions come under reduced features dimension-based methods.

Li et al. [75] combined six cortical features such as cortical surface area, cortical thickness, sulcal depth, GM volume, metric distortion, and cortical curvature. These features have been obtained from the pial surface’s vertices. The authors investigated the grey matter density in MCI patients and found the significance of all these cortical features for early AD detection. The right hemisphere features gave 80% accuracy and left hemisphere features gave 76.0% accuracy. Yubraj et al. [76] used FreeSurfer tool to extract features like cortical thickness and subcortical volume and obtained 98% accuracy with SVM classifier for AD/NC classification. In this study, PCA has been used to reduce feature dimensions. Using a manifold harmonic transform, Cho et al. [77] transformed cortical thickness data into its spatial frequency components. The authors employed Linear Discriminant Analysis (LDA) to reduce feature dimensions by eliminating high-frequency (noise) components. Park et al. in 2012 [38] extracted cortical thickness and sulcal depth features for AD/NC classification. The authors used PCA for feature dimensionality reduction and secured 90.0% accuracy in SVM -based AD/NC classification. Park et al. in 2013 [78] performed AD/MCI classification using identical features from longitudinal MRI and obtained 83% accuracy. The prominent cortical surface-based studies for AD detection are shown in Table 2.3.

Table 2.3 Prominent studies for AD classification using cortical surface-based methods

Author	**Dataset**	**Database**	**Feature**	**Classifier**	**Performance**	**Observations**
(Li et al., 2014) [75]	MCI=24 NC =26	Xuan Wu Hospital (CHINA)	All vertices (right hemisphere)	SVM	**sMCI / NC** (Acc=80)	Different types of features have been used in this approach. The small sized dataset negatively impacts the performance of the classifier [79].
(Yubraj et al.,2019) [76]	AD=81 NC =171 pMCI=39 sMCI=35	National Research Center for Dementia (NRCD)	Reduced vertices	SVM	**AD / NC** (Acc=98.02, Sen=97.87, Spe=98.05) **pMCI/sMCI** (Acc=97.77, Sen=100, Spe=95.23) **pMCI/NC** (Acc=99.2, Sen=99.02, Spe=100)	The FreeSurfer toolbox has been used for cortical thickness measurement. It shows biases while doing measurements. These biases need to be statistically adjusted [80].
(Cho et al., 2012) [77]	AD=128 NC =160 pMCI=72 sMCI=131	ADNI	Reduced Vertices	Linear Discriminant Analysis (LDA)	**AD/NC** (Acc=88.33, Sen=82, Spe=93) **pMCI/ sMCI** (Acc=71.21, Sen=63, Spe=76) **pMCI/NC** (Acc=82, Sen=66, Spe=89)	The single template is utilized for registering all individuals. This template must not be inclined for particular class. The influence of noise on cortical thickness is cancelled by removing high-frequency components.
(Park et al., 2012) [38]	AD=25 MCI=25 NC =50	OASIS	Reduced Vertices	SVM	**AD/NC** (Acc=90) **pMCI/NC** (Acc=86) **AD/MCI**(Acc=90)	The dimensionality is reduced by using PCA which is scale-invariant [81]. The authors used Sulcal depth for AD/MCI classification.
(Park et al., 2013) [78]	MCI=30 pMCI=12 NC =30	ADNI	Reduced Vertices	SVM	**pMCI/NC** (Acc=90) **AD/MCI**(Acc=83)	The dimensionality is reduced by using PCA which is scale-invariant [81]. Longitudinal studies have been used for AD prediction.

Accuracy=Acc, Sensitivity=Sen, Specificity=Spe

Table 2.3 lists several methods for reducing the size of features extracted from the cortical surface. PCA has generally been utilized by authors to reduce the dimensionality of features. The fact that PCA is scale-invariant and can handle Gaussian data is one of its drawbacks [81]. The selection of an appropriate atlas and the correct alignment of images over atlases are prerequisites for atlas-based approaches [82].

2.5. Wavelet Transforms-Based Methods

Research shows that different wavelet transforms have been used for feature generation from MR images. The imaginary coefficients in complex wavelets capture the structure details nicely and the detail wavelet coefficients provide the grey matter density fluctuation due to AD [83-85]. Wavelets provide the time and frequency information enabling analysis of a signal at different scales.

Studies reveal that atrophy spreads to various parts of the brain in the advanced stages of AD instead of being confined to a single area. Therefore, brain region segmentation is not necessary as it is in ROI-based techniques. Above all, the whole brain analysis views the brain as a single ROI. The whole brain analysis is completed in two steps i.e., feature extraction and classification. However, it has issues with feature dimensionality as it requires scanning every voxel in the brain, [86-87].

Some researchers used detail coefficients directly as features in their research, while some of them extracted features from coefficients. Chaplot et al. [88] obtained coefficients after Db4 wavelet decomposition and used these coefficients as features for AD classification. This study achieved 98% accuracy with SVM. Jha et al. [89] utilized the Dual Tree Complex Wavelet Transform (DTCWT) to obtain wavelet coefficients from MRI scans, achieving a 90.06% accuracy rate in distinguishing AD from NC. Subsequently, Alam et al. [90] applied PCA and LDA to reduce the dimensionality of the extracted wavelet features. The authors obtained 92.65% accuracy with a twin support vector machine using the ADNI dataset. Bhasin et.al [91] extracted LBP-20 features from MR images by applying 3D-DWT and obtained 90.3% accuracy for MCI/NC classification. Acharya et al. [92] classified AD with 94.54% accuracy for MR images using the shearlet transform in conjunction with the K-Nearest Neighbour classifier. Wang et al. [93] employed 3D DWT for extracting coefficients from the 3D MR image. Jha et al. [94] employed contourlet transform on MR images for extracting features and obtained 91.14% accuracy with Random Under Sampling Boosted Trees. Recently Topannavar et al. [95] used DT-CWT and Rotated Complex Wavelet Filter (RCWF) for feature extraction. The authors extracted entropy and energy features in twelve different directions and used them for AD classification. Yang et al. [96] used wavelet energy-based features and obtained 97.78% accuracy with an SVM classifier using the Biogeography Based Optimization (BBO) technique. Feng et al. [97] combined ROI-based approach with the wavelet transform-based approach to obtain features in the frequency domain. They segmented 90 ROIs by using an Automated Anatomical Labelling (AAL) atlas and then used shearlet transform to obtain features. Feng et al. [98,99] in another work used the VBM technique to obtain grey matter density maps and then applied contourlet transform [97] and nonsubsampled contourlet transform [98] to obtain the sub-images. Subsequently, they obtained features viz. energy features [97] and mean, variance, and origin moment [98] for AD classification. In these works [97-99], Feng et al. used frequency domain features for AD classification, but they also included the ROI's and VBM-based approach for selection of appropriate atrophied regions and grey matter density maps respectively. The prominent studies using wavelet transform-based methods are detailed in Table 2.4.

Table 2.4 Prominent wavelet transform-based studies for AD classification

Author	**Feature extraction**	**Database**	**Dataset**	**Classifier**	**Performance**	**Observations**
(Jha et al.,2017) [89]	DTCWT+ PCA	OASIS	AD=28 NC =98	FNN	**AD/NC** (Acc=90.06, Sen=92, Spe=87.78)	The selection of slices is done manually. Feature dimensionality is reduced by using PCA which leads to spatial information loss [100].
(Alam et al.,2017) [90]	DTCWT+ PCA + TSVM+LDA	OASIS	NC =44 AD=51	TSVM	**AD/NC** (Acc=96.68, Sen=97.72, Spe=95.61)	The selection of slices is done manually. Feature dimensionality is reduced by using PCA which leads to spatial information loss [100].
		ADNI	NC =86 AD=86	TSVM	**AD/NC** (Acc=92.65, Sen=93.11, Spe=92.19)	
(Bhasin et al.,2020) [91]	3-D DWT+LBP20	ADNI	sMCI=112 pMCI=75 NC =89	SVM	**pMCI / sMCI** (Acc=88.7, Sen=90.16, Spe=89.16) **MCI vs NC** (Acc=90.31, Sen=90.22, Spe=90.2)	DWT has been used for extracting features. DWT does not show shift invariance and high directional sensitivity.
(Acharya etal.,2019) [92]	ST	Clinical Brain MRI Dataset	NC =110 AD=55	k-NN	**AD/NC** (Acc=98.48, Sen=96.97, Spe=100)	Features are selected through a t-test, which is prone to false positives (Type I errors) [70].

Accuracy=Acc, Sensitivity=Sen, Specificity=Spe

Table 2.4 lists the prominent studies using wavelet transform for feature extraction. Wavelet Transform-based methods generally suffer from high dimensionality as it is applied to the whole brain. PCA is used by the authors to address high dimensionality. PCA results in spatial information loss which is its major problem.

2.6. Texture-Based Methods

Texture [101] gives information about shape, density, pixel value, etc in an image by measuring small variations in intensity at regular intervals. The texture can be characterised using structural and statistical approaches. The structural approach uses Local Binary Pattern (LBP), Fourier coefficients, and wavelet coefficients and the statistical approach uses histogram, Grey-Level Co-occurrence Matrix (GLCM), and Grey Level Run Length Matrix (GLRLM). The elements of a matrix called GLCM indicate how frequently different sets of pixel grey level intensity occur in an image [102]. The total grey level runs at different run lengths are calculated using a matrix called the GLRLM [103]. The line of pixels with equal intensity values in a specific direction is represented by a grey level run; the count of corresponding pixels is represented by the grey level run length; and the count of incidents is represented by the run-length value. GLCM is the best among all texture features.

Simoes et al. [104] considered those brain regions that get influenced during the initial phase of AD. The 3-D texture features i.e. LBP-TOP (Three Orthogonal Projections) were extracted from the specified locations in a 3-D MRI. LBP-TOP examines the brain's axial, coronal, and sagittal sections to obtain LBP histograms from these areas. This study observed that as AD advances, more areas of the brain are impacted, starting with the hippocampus and amygdala. Feng et al. [105] worked with intensity features and textural features which were obtained through GLCM and GLRLM. They concluded the right hippocampus was the most atrophied region in the AD brain. Joana Silva et al. [106] obtained textural features. These features were obtained through GLCM. Gao et al. [107] obtained textural features from the hippocampus of 299 individuals by applying contourlet transform and GLCM. The study concluded that classification was more effective when textural features were included in the multivariate model. Salunkhe et al. [108] applied GLCM on the hippocampus to obtain textural features for detecting its slight structural changes due to AD. Table 2.5 lists the prominent AD classification studies using Textural features.

Table 2.5 Prominent studies for AD classification using Texture-based method

Author	**Feature extraction**	**Dataset**	**Classifier**	**Performance**	**Observations**
Simoes R et al. [104]	LBP TOP from the Whole brain	NC =66 AD=70	SVM	This study focuses on finding the most discriminative regions in early AD by analyzing LBP patches	LBP-TOP uses three orthogonal projections to obtain the edge features. This study does not require segmentation, prior details, and image registration as LBP-TOP has been extracted from the whole brain [109].
Feng et al. [105]	Histogram, GLCM, Wavelet, GLRLM from the hippocampus	NC =45 AD=38 aMCI=33	SVM	**AD / NC** (Acc=86.75, Sen=84.21, Spe=88.89) **sMCI / NC** (Acc=70.51, Sen=57.58, Spe=80) **sMCI / AD** Acc=59.15, Sen=54.55, Spe=63.16)	Different subregions are obtained from Hippocampus by using atlases. The identical shape features obtained in different subregions are not included in the study. Atlas-based segmentation creates problems in extracting Radiomic features viz., intensity features, and texture features of GLCM and GLRLM [110-111].
Joana Silva et al. [106].	Texture statistical measures (contrast, correlation, energy, homogeneity, entropy, variance, and standard deviation)	NC =24 AD=21 MCI=44	CNN	AD / NC (Acc=82.2) AD / MCI (Acc=75.4) MCI / CN (Acc=83.8	Dataset is small. Textural features have been obtained from GLCM that have been applied on subband images of three planes MR slices.
Gao et al. [107]	GLCM	NC =7 sMCI=73 pMCI=62	Gaussian process models	**pMCI/NC** (Acc=86.75) **sMCI / NC** (Acc=79.5) **MCI to NC** (Acc=85.7)	The small dataset is used. Only the hippocampus is utilized to extract the texture features. It may increase the accuracy if other regions are included.
			Partial least squares	**pMCI / NC** (Acc=85.5) **sMCI / NC** (Acc=83.6) **MCI to NC** (Acc=100)	

Accuracy=Acc, Sensitivity=Sen, Specificity=Spe

Table 2.5 lists the prominent studies based on textural features. Proper segmentation is required to obtain textural information from the region of interest [112]. Texture feature extraction relies heavily on the discretization of image intensities and the acquisition of parameters from MR images. Its improper execution leads to heterogeneity in Magnetic Resonance Texture Analysis (MRTA) [113].

2.7. A Comparison of Feature Extraction Methods

ROI approach works well for early AD detection. In addition to examining grey and white matter volume at the voxel level and group differences, VBM supports the analysis of how developmental abnormalities, aging, sex, and genetic traits affect the structure of the brain [114]. ROI can identify moderate to severe abnormalities in the grey matter, whereas VBM can detect minute changes in the grey matter [115]. VBM demands that all the images on the template be perfectly registered. The cortical surface-based method [116] emphases on surface registration, which is superior to volume registration. The VBM technique concentrates on volume registration. Table 2.6 outlines the various feature extraction techniques.

Table 2.6 Evaluation of various feature extraction techniques

Methods	Advantages	Disadvantages
Predefined ROI-Based Approaches	In ROI-based techniques, that part is extracted from the brain which is highly influenced like hippocampal and amygdala. The extracted features are shape, volume, and size of hippocampus or amygdala [45]. ROI-based approach is efficient technique in case of early AD detection.	Understanding the size and spatial distribution of AD is crucial for precise ROI segmentation, and this depends on previous data and expert knowledge. Selecting an ROI is a major issue. It takes more time to manually segment and choose ROIs in several slices. Variability between and within observers must be investigated using statistical techniques. Because it is challenging to replicate ROI from one image to another, ROI techniques are challenging to apply in longitudinal investigations.
Voxel-Based Morphometry (VBM)	VBM is highly useful in extracting voxel-based information with the application of statistical parametric mapping (SPM). It is used to explore subtle grey matter changes in brain structure [117]. VBM is used for volume measurement of brains of different topologies and sizes.	VBM requires perfect image registration. Misalignment errors may give false results. Motion artifacts in MRI may affect tissue segmentation and create differences in results. All the images must be scanned through the same scanners for better outcomes.
Cortical Surface-Based Methods	The measurement of cortical thickness requires fewer operators compared to the measurements related with hippocampus volume. This is good for extracting localised cortical surface features [118].	These methods require expertise in the FreeSurfer tool. These methods show limitations with clinical data for AD vs MCI detection. The efficiency of the method depends upon the perfect alignment of images over templates.
Wavelet transform Based	Wavelet transform provides multiresolution analysis for non-stationary signals. They are good at capturing structural changes and grey matter fluctuations. It does not require segmentation [84,85].	It gives less directional sensitivity in the 3-D image [119]. These methods generally suffer from high dimensionality. These methods may be computationally high thus a balance is maintained between detail preservation and the computation process.
Texture-Based Methods	MRIs are not able to capture deposition of unwanted proteins in the brain in initial phase of AD. However, their deposition on the brain surface changes the pixel intensity of MR images. The analysis of such surfaces can be done by using texture analysis (TA) [113].	There is no standardization in MR acquisition methods and discretization of image intensities which results in challenges faced by Texture-based methods for their successful implementation in clinics. MRTA software heterogeneity is another major issue. Texture-based methods are highly influenced by noises.

Table 2.6 highlights different feature extraction techniques with their limitations and advantages. ROI-based methods and wavelet-based methods are efficient for early AD detection. VBM methods are automated in nature but less effective for cortical folding. Cortical surface-based methods provide finer cortical details but are computationally intensive. Texture-based methods capture microstructural details but are sensitive to artifacts. The wavelet-based method does not require segmentation and covers the entire brain. High-speed computers and feature dimension reduction techniques can be used to address the high dimensionality, which is the only issue with the wavelet-based method [119].

2.8. Machine Learning Algorithms used in AD Detection

Classification refers to the act of finding the class of unseen input data. A machine learning algorithm that sorts input data into labelled classes is called a classifier [120]. Two sets of data are required for this: one used to train the classifier and the other to test its performance. AD classification means classifying input data into the AD, MCI, or NC classes. MR scans of healthy individuals and AD patients at various stages of the disease are included in the training image set, whereas MR images that have not yet been classified as normal or AD are included in the test image set. Features are extracted from the MR scans and using these features different classes are classified by machine learning algorithms. The subsequent subsections present the details of these classifiers used for AD classification.

2.8.1. Support Vector Machine (SVM)

Research shows that SVM is widely applied for classifying Alzheimer's disease. The support vector machine (SVM) makes a hyperplane connecting the training samples of both classes and then chooses the optimal hyperplane to keep the margin between them at maximum limit. SVM are used with different kernels. Kloppel et al. [50] used Linear SVM and Zhang et al. used polynomial SVM in the AD classification framework. Kheder et al. [121] observed that, in comparison to polynomial kernels, linear kernels produce excellent classification accuracy of 89.0% for large dimensional data. On the other hand, other research uses a polynomial kernel. Using a polynomial kernel, Lahmiri et al. [122] observed that the multiclass classification of NC, MCI, and AD had good classification accuracy. Zhang et al. [123] used PCA with a polynomial kernel to attain 92.4% accuracy.

2.8.2. Decision Tree (DT)

The structure of a decision tree classifier is like a tree with internal nodes, branches, leaves, and root nodes. Branches show the decision result, internal nodes show the decision at the next level, leaf nodes show the class label, and the root node shows the initial decision [124]. The expected output label of the test sample is represented as a leaf on the tree, and it is obtained by moving the test sample through the tree. The study [125] proposed a hyperparameters tuning-based decision tree classifier for early AD detection. This study obtained 99.1% classification accuracy on the Open Access Imaging Studies Series (OASIS) dataset.

2.8.3. Random Forest (RF)

The Random Forest (RF) classifier achieves the greatest accuracy when using multimodal neuroimaging data. Random Forest minimizes overfitting issues and easily work with non-linear input [126]. The Random Forest algorithm is a collection of several Decision Trees. The training is given to each tree based on randomly selected portion of the training dataset. At every split, a randomly selected portion of the features is evaluated, reducing the risk of overfitting. Each tree is constructed independently which makes it effective for parallel computation. According to Saraica et al. [126], the Random Forest classifier minimizes overfitting issues and provides good accuracy level when handling multimodal neuroimaging data. Lebedev [127] and Gupta et al. [53] used a Random Forest classifier in their work. Lebedev et al. [127] used volumetric and cortical thickness features and obtained 88.6% sensitivity and 92.0% specificity by using Random Forest.

2.8.4. K-Nearest Neighbour (KNN)

The test data is categorized by a KNN using distance or similarity metrics between the test and training samples. The K-Nearest Neighbours classifier identifies the K closest training instances to a test sample based on distance metric and then classifies the test data [128]. Krishnkumar et al. [129] used classifiers viz. RF, Linear SVM, and KNN in their work for AD/NC classification. They extracted textural features from 3-D MR Images and used them for classification. With KNN, RF, and Linear SVM, the authors achieved 87.4%, 87.6%, and 82.6% accuracy for classifying AD with NC respectively. Khasanah et al. [130] used a KNN classifier to evaluate the performance of an imbalanced dataset. This study obtained 45.86%

accuracy which highlights the impact of an imbalanced dataset on classification performance. Gupta et al. [53] obtained 88.1% classification accuracy using KNN for AD/NC classification. The study [131] used fuzzy KNN with an improved binary Salp swarm algorithm for early AD detection. This study obtained respective accuracies of 95.37%,100%, and 93.95% for classifying AD with MCI, AD with NC, and MCI with NC.

2.8.5. Naive Bayes (NB)

The Naive Bayes (NB) [132] is a probabilistic machine learning model. It is referred to as "naive" because it assumes that every feature is unrelated to every other feature, which is rarely the case in actual data. It works well in a variety of applications despite this presumption. Kruthika et al. [133] integrated the expertise of professional radiologists with the automatic classification technique to improve classification accuracy. The authors employed a multistage classifier, using Naive Bayes for binary classification in the first stage. In the second stage, SVM and KNN were used for binary and multiclass classifications respectively. The feature selection was done using Particle Swarm Optimization. Compared to individual classifiers like Naive Bayes, KNN, and SVM, the multistage classifier performed well for AD detection.

2.8.6. Extreme Learning Machine (ELM)

ELM is a feed-forward network with single layer and needs less time for processing compared to ANN because it assigns hidden nodes at random, whereas ANN assigns hidden nodes continually [134]. ELM shows good generalization performance and fast execution speed. Jha et al. [135] achieved 90.3 ± 1.8% accuracy with the ADNI dataset and 95.7 ± 1.5% accuracy with the OASIS dataset by utilizing an ELM classifier. Gao et al. [136] used ELM for classifying statistical features obtained from MR images through GLCM.

2.9. Selection Criteria of Different Machine Learning Algorithms

SVM is effective for predicting high-dimensional data since it has few parameters and works well with datasets that have a distinct margin of separation between the classes. SVM is capable of handling multiclass problems because of its kernel nature. SVM is less effective with overlapping targets [137]. A KNN performs well when features are standardized, and the dataset is normalised and labelled [128]. KNN gives good results with multiclass classification. KNN is non-parametric and thus can handle non-linear data. KNN is sensitive to noise or outliers. When features are independent of one another, NB performs better. It performs fast with categorical data. A decision tree is well suited for structured data with clear feature importance. It does not require scaling of features and can handle non-linear relationships. It may suffer from an overfitting problem. Random Forest can handle missing data and perform well with mixed data types. RF is computationally expensive and less interpretable. The neural network requires large datasets and can learn complex patterns. Table 2.7 shows the selection criteria for different classifiers in prominent AD classification studies.

Table 2.7 Selection Criteria of classifiers in prominent AD classification studies.

Author	**Method**	**Database**	**Classifier**	**Performance**	**Selection Criteria of Classifier**
(Kloppel et al., 2008) [50]	VBM	From multiple scanners	SVM (linear)	**AD/NC** (Acc=95.6, Sen=97.1, Spe=94.1)	The linear kernel is suitable for linear separable data and not suitable for non-linear data [137].
(Zhang et al., 2015) [93]	PCA	OASIS	SVM (polynomial)	**AD/NC** (Acc=92.36, Sen=83.48, Spe=94.9)	The polynomial kernel works efficiently with linearly inseparable data. It is suitable when a smaller number of features are present [138].
(Lebedev et al.,2014) [127]	41 ROI	ADNI	RF	**AD/NC** (Sen=88.6, Spe=92) **pMCI/ sMCI** (Sen=81.3, Spe=83.3)	Random Forest [139] is used with large datasets. It is good when interpretability is not a concern. It can handle high-dimensional features
(Gupta et al.,2019) [53]	HV+ CSC+ VBM	NRCD	k-NN	**AD/NC** (Acc=89.47, Sen=93.75, Spe=88.3) **pMCI/ sMCI** (Acc=73.91, Sen=70.58, Spe=83.33)	The KNN works well when normalized or scaled dataset is used. It can be used for multiclass classification with small datasets. The datasets should be noise free and labelled.
			RF	**AD/NC** (Acc=88.15, Sen=78.26, Spe=92.45) **pMCI / sMCI** (Acc=79.99, Sen=78.26, Spe=76.92)	Random Forest is used with a large dataset. It is good when interpretability is not a concern. It can handle high-dimensional features.
			SVM	**AD/NC** (Acc=93.06, Sen=87.87, Spe=95.58) **pMCI / sMCI** (Acc=87.08, Sen=86.95, Spe=77.77)	Different SVM Kernels can deal with nonlinear data also. SVM does not perform well with overlapping targets [138].
(Krishnakumar et al.,2019) [129]	Fisher score+Elastic net+Recursive Feature Elimination	ADNI	k-NN	**AD/NC** (Acc=87.39, Sen=89.58, Spe=85.82) **pMCI / sMCI** (Acc=66.38, Sen=60.24, Spe=69.8)	The KNN works well when normalized or scaled dataset is used. It can be used for multiclass classification with small datasets. The datasets should be noise free and labelled.
			RF	**AD/NC** (Acc=87.39, Sen=85.42, Spe=88.81) **pMCI/ sMCI** (Acc=61.64, Sen=65.06, Spe=59.73)	Random Forest is used with a large dataset. It is good when interpretability is not a concern. It can handle high-dimensional features [139].
			Linear SVM	**AD/NC** (Acc=82.61, Sen=78.13, Spe=85.82) **pMCI / sMCI** (Acc=64.66, Sen=56.63, Spe=69.13)	The linear kernel is suitable for linearly separable data and not suitable for non-linear data [137].

Accuracy=Acc, Sensitivity=Sen, Specificity=Spe

Table 2.7 demonstrates that different classifiers perform differently with the same dataset. Machine learning algorithms are selected based on linearity, noise, outliers, normality of distribution, missing values, dimensionality, speed, and preprocessing requirements of training data.

2.10. Recent Trends in AD Detection

The recent trend in AD detection includes multimodal studies and deep-learning-based studies. Multimodal techniques include different types of modalities like Positron Emission Tomography (PET), Single-Photon Emission Computed Tomography (SPECT), MRI, Computerized Axial Tomography (CAT), etc. These techniques extract features from different modalities thus complete information are obtained. Dukart et al. [140] extracted volumetric details from MRI and fluorodeoxyglucose PET. They used these details as features for AD/NC classification. This study provided 100% classification results using SVM as a classifier. Kong et al. [141] proposed an image fusion strategy that includes PET and MRI. The authors proposed a 3-D convolutional network for detecting features obtained from multi-modal fused images. Multimodal studies capture micro and macro-structural details in the image like studies [142,143] that used MRI for macro-detail features and DTI for micro-detail features. Multimodal studies lack proper alignment and mutual interaction of multimodal data for better results and are at high risk of data overfitting [144].

The deep learning-based approach is currently the second most used method. Deep Learning helps in extracting features that are hidden in deep layers of data. Deep learning enhances Alzheimer's disease detection by automating feature extraction, improving prediction accuracy, handling multimodal data, and enabling early detection. Liu et al. [145] performed hippocampal segmentation using a multi-task deep CNN model. The features are extracted from the 3-D patches which have been segmented from the hippocampus. These features were used to train 3-D Densely Connected Convolutional Networks (3-D DenseNet). AD classification is done based on the combined results obtained through the training of CNN and DenseNet models through these features. Chen et al. [146] proposed an ensemble model of deep learning that combines Soft Non-Maximum Suppression (NMS) with Faster R–CNN architecture to enhance classification accuracy. Further, features are extracted using an improved ResNet50 network. The improved Faster R-CNN is used to produce the final classification results.

2.11. Research Gaps

- To diagnose Alzheimer's disease at an early stage is still challenging due to symptoms overlapping with aging and the complexity of brain changes.
- The good classification accuracy between Alzheimer's disease and normal controls is already achieved but still, there is scope for improvement in accuracy for MCI/NC and AD/MCI classifications, that would help in timely detection of Alzheimer's disease.
- AD heterogeneity is a major problem which creates misclassification issues. The development of tools that can deal with heterogeneous imaging patterns is important.
- Another significant problem is identifying the appropriate biomarkers for AD. Similar structural changes in the brain might occasionally be caused by other neurological conditions, and using those data in training or testing sets may result in an incorrect model result indicating 100% correctness.

2.12. Objectives

- To introduce feature optimization techniques in the novel feature set so that redundancy in information may be removed, and classification accuracy may be improved.
- To develop a novel feature extraction method so that the problem of scalability may be minimized.
- To use novel features for AD classification by using wavelet transform feature extraction technique which will retain the spatial information and improve the classification accuracy.
- To develop a feature extraction model for early diagnosis of Alzheimer's disease by getting good classification accuracy between MCI/NC and MCI/AD.

2.13. Dataset

Two types of datasets are used in the research work. These are described in subsequent subsections.

2.13.1. Open Access Series of Imaging Studies (OASIS) dataset

OASIS includes four datasets: OASIS-1, OASIS-2, OASIS-3 and OASIS-4. OASIS-1 is widely benchmarked in literature. OASIS-1 consists of cross-sectional single visit T1-

weighted MRI data. This dataset provides clear diagnostic labels and demographic information. It does not include longitudinal complexity which makes it good for early AD detection. OASIS-2 includes repeated MRI scans over several years which adds variability and scanning inconsistencies. OASIS-3 is large and heterogenous. It combines data from multiple scanners, multiple modalities and longitudinal sessions. This makes it more suitable for monitoring disease progression rather than for early-stage Alzheimer's detection. OASIS-4 includes the younger cohort with minimal structural degeneration, making it not suitable for early AD detection.

OASIS-1 dataset has been used in the proposed work which includes 416 cross-sectional studies in an age span of 18 to 96 years. Among 416 subjects,316 subjects belong to the non-demented category, and 100 subjects belong to the demented category. Among these 100 subjects,70 subjects belong to very mild dementia,28 subjects belong to mild dementia, and 2 subjects belong to moderate dementia. This dataset [147] is freely publicly available. In its compressed state, the dataset is 15.6 GB; uncompressed, it requires 50 GB of storage. The 3-D images have been downloaded in analyze 7.5 format. There are both men and women among the subjects, and they are all right-handed. Table 2.8 depicts the demographic statistics of the participants in the OASIS-1 dataset.

Table 2.8 Statistical details of the OASIS-1 dataset

		Non-Demented				**Demented**				
Age Group	**N**	**N**	**Mean**	**Male**	**Female**	**N**	**Mean**	**Male**	**female**	**CDR 0.5/1/2**
<20	19	19	18.53	10	9	0		0	0	0/0/0
20s	119	119	22.82	51	68	0		0	0	0/0/0
30s	16	16	33.38	11	5	0		0	0	0/0/0
40s	31	31	45.58	10	21	0		0	0	0/0/0
50s	33	33	54.36	11	22	0		0	0	0/0/0
60s	40	25	64.88	7	18	15	66.1	6	9	12/3/0
70	83	35	73.37	10	25	48	74.4	20	28	32/15/1
80s	62	30	84.07	8	22	32	82.9	13	19	22/9/1
≥90	13	8	91	1	7	5	92	2	3	4/1/0
Total	**416**	**316**		**119**	**197**	**100**		**41**	**59**	**70/28/2**

Table 2.8 shows that a total of 416 subjects are in the OASIS-1 dataset. Out of which,316 subjects belong to the non-demented category or Normal Controls class. A total of 70 subjects belong to very mild dementia i.e., MCI class, 28 subjects belong to mild dementia means the AD category, and 2 subjects belong to moderate dementia means a severe stage of AD.

2.13.2. Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset

The Alzheimer's Disease Neuroimaging Initiative (ADNI) [148] is a data house which contains imaging, genetic, and clinical data for understanding AD progression. The scientific community uses ADNI data extensively to create biomarkers and to enhance the knowledge about disease. The proposed work is implemented by using T1-weighted structural MR images obtained at 1.5T scans in baseline visit or screening from the ADNI-1 dataset. A total of 818 subjects are in the 1.5 T screening dataset of which 229 belong to the Normal class,401 belong to MCI class and 188 belong to AD class. The subject-age group belongs to 55 to 90 years in 1.5T screening dataset. There are five standard analysis datasets in ADNI-1. Different time points were chosen for inclusion in these five datasets. Table 2.9 illustrates the demographic statistics of ADNI-1 dataset.

Table 2.9 Demographic statistics of the ADNI-1 dataset

Screening Visit		
NC=229	MCI=401	AD=188
Complete 1-year visits (screening, 6 months, 12 months)		
NC=195	MCI=311	AD=133
Complete annual year 2 visits (screening, 12 months, 24 months)		
NC=169	MCI=234	AD=101
Complete 2-year visits (screening, 6 months, 12 months, 18 months*, 24 months)		
NC=168	MCI=212	AD=99
Complete 3 years visits (screening, 6 months, 12 months,18*months,24 months,36months)**		
NC=135	MCI=148	AD=99

From Table 2.9, screening visit scans are a total of 818 in numbers. The complete 1-year visit scans include screening scans, six-month scans, and twelve-month scans. The same scanner was used for the acquisition of these scans. Complete two years of annual visitors include screening, twelve-month and twenty-four months scans. Complete two-year visitors include screening, 6, 12, 18 (MCI only), and 24 months scans. Complete three years include screening ,6,12,18 (MCI only), 24-, and 36-months scans (only NC and MCI).

2.14. Performance Evaluation Parameter

The parameters like accuracy, sensitivity, and specificity are used to assess the predictive capability of the models under classification. The Receiver Operating Characteristics (ROC) curve is used to graphically evaluate the classification models.

2.14.1. Quantitative Measures

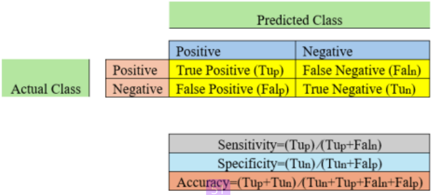
Accuracy determines the efficiency of a classifier to identify the correct true positive and true negative cases. Sensitivity determines the efficiency of a classifier in identifying the correct true positive cases, and specificity determines the efficiency of a classifier in finding the correct true negative cases. Mathematically, they are formulated as

$$\text{Accuracy} = (\text{Tu}_\text{p} + \text{Tu}_\text{n})/(\text{Tu}_\text{n} + \text{Tu}_\text{p} + \text{Fal}_\text{n} + \text{Fal}_\text{p}) \qquad (2.1)$$

$$\text{Sensitivity} = (\text{Tu}_\text{p})/(\text{Tu}_\text{p} + \text{Fal}_\text{n}) \qquad (2.2)$$

$$\text{Specificity} = (\text{Tu}_\text{n})/(\text{Tu}_\text{n} + \text{Fal}_\text{p}) \qquad (2.3)$$

Tu_p be a symbol of True positive. It implies the right identification of an AD person as AD. Tu_n be a symbol of True negative. It signifies the right identification of healthy person as healthy. Fal_p be a symbol of False Positive. It implies the wrong identification of a healthy person as AD. Fal_n be a symbol of False Negative. It signifies the wrong identification of an AD person as healthy. Fig. 2.2 depicts the Confusion Matrix.



		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (Tu_p)	False Negative (Fal_n)
	Negative	False Positive (Fal_p)	True Negative (Tu_n)

Sensitivity=(Tu_p)/(Tu_p+Fal_n)

Specificity=(Tu_n)/(Tu_n+Fal_p)

Accuracy=(Tu_p+Tu_n)/(Tu_n+Tu_p+Fal_n+Fal_p)

Fig. 2.2 Confusion Matrix

Confusion Matrix shows the performance of a classification model by indicating the correct and incorrect predicted instances. With the help of Confusion Matrix, accuracy, sensitivity, and specificity can be calculated by using formulas mentioned in the Fig.2.2.

2.14.2. Receiver Operating Characteristic (ROC) Curve

ROC graph plots sensitivity (true positive rate) against 1- specificity (false positive rate) at different threshold values. The true positive rate is kept at the vertical axis and False positive rate is kept at the horizontal axis. This plot is used to assess the performance of a binary classifier. The curve should ascend swiftly in the direction of the plot's upper-left corner. This shows that the classifier maintains a low false positive rate (FPR) while achieving a high true positive rate (TPR) ie. (sensitivity). The 100% sensitivity and 0% false positive rate would be represented by the ideal ROC curve, which would cross the upper-left corner of the plot (0, 1). The classifier is better if the curve is closer to this corner.

2.15. Summary

In this chapter, extensive Literature review is presented on AD detection using various techniques for feature extraction and machine learning algorithms. The different feature extraction techniques include ROI-based methods, VBM-based methods, Cortical surface-based methods, Wavelet-based methods and Texture-based methods. ROI-based methods need segmentation of the specific brain region which is highly influenced due to AD.VBM methods require perfect registration of images over templates to extract grey matter maps. Cortical surface-based methods require expertise in FreeSurfer tool to extract cortical features. Texture-based methods face challenges due to the lack of protocols for standardization of MR acquisition. The results with wavelet-based methods are outstanding in comparison to existing methods but these techniques generally suffer from high dimensionality. Principal Component Analysis is often used in wavelet-based methods to reduce the dimensionality, but this results in spatial information loss. This chapter also includes the comparison between different feature extraction methods. Each feature extraction technique has its own limitation, yet it helps in extracting different features for AD detection. Recent trends include the multimodalities and deep learning techniques. Selection criteria of different machine learning techniques viz. KNN, SVM, ELM, DT, RF has been included. Lastly, research gaps, objectives, datasets and performance metrics have been discussed.

The publication related to this chapter

Neha Garg, M. S. Choudhry, & R. M. Bodade, (2023). A review on Alzheimer's disease classification from normal controls and mild cognitive impairment using structural MR images. **Journal of neuroscience methods**, 384, 109745. https://doi.org/10.1016/j.jneumeth.2022.109745 **I.F. 2.3**

Chapter 3

Alzheimer’s Disease Detection using Direct Coefficient-Based Feature Extraction Technique

This chapter presents an effective technique for early Alzheimer’s Disease detection. This technique utilises multiple 2-D slices of magnetic resonance imaging which helps in capturing a broader spatial and anatomical context, potentially improving the sensitivity of early detection. Early AD includes subtle and localized structural changes in the regions like hippocampus and entorhinal cortex which cannot be clearly represented by single slice only. The proposed technique is the direct coefficient-based feature extraction technique where wavelet-coefficients obtained from detail sub-bands are directly employed for Alzheimer’s Disease classification. Further, the feature selection method has been used to elevate the model accuracy. The performance of the proposed method is demonstrated and verified through comprehensive experimentation on benchmark data and a detailed comparison with current existing approaches.

3.1. Introduction

This chapter introduces the Direct-coefficient-based feature extraction method for extracting features for AD classification. This work provides the most optimised wavelet-based feature extraction method for true 3-D Alzheimer’s Disease data. In this work,32 centre axial slices have been obtained from 3-D MR image. Further, features are extracted from these slices instead of using a single slice. The multiple brain slices provide more consistent, accurate, and reliable information which may not be inspected by a single slice. Alzheimer's disease often affects multiple brain regions, resulting in spatially distributed structural changes. Multiple slices provide extensive insight about the structural changes than a single slice, which may overlook the important pathological features. In advanced stages of AD, disease spreads to the different regions of the brain which can be captured by multiple slices nicely.

The proposed method uses Double Density Dual Tree Complex Wavelet Transform (DD-DTCWT); a complex wavelet adapts at capturing structural details. This method employs fifth-level detail coefficients as feature inputs as they are being high frequency coefficients, capture high frequency information i.e. grey matter density variation in AD brain. In addition to the main approach, assessments were performed using DWT, DTCWT, and DD-DWT wavelet transforms. The Double-Density Dual-Tree Complex Wavelet Transform (DD-DTCWT) demonstrated the most favourable outcomes among them. This work also investigates the performance of conventional statistical methods namely, t-test and evolutionary algorithms viz. Particle Swarm Optimization (PSO), Genetic Algorithm (GA) and Ant Colony Optimization (ACO), for selecting appropriate features in AD classification.

3.2. Wavelet Transform

The wavelet transform is a mathematical method that breaks down a signal into its constituent frequencies at different scales, such as an image, audio, or time series data. It decomposes a signal into "wavelets"—localized, tiny waves that are shifted and scaled variants of a selected base function. The wavelets enable multi-resolution analysis, thus very helpful for capturing information at different scales, encompassing both coarse and fine information. Wavelet Transform encompasses two categories: Discrete Wavelet Transform (DWT), and Continuous Wavelet Transform (CWT). CWT is computationally expensive in comparison of DWT as CWT decomposes the signal at all possible shifts and scales while DWT performs the transformation at specific scales and shifts. DWT has wide applications in image processing, signal compression, and noise reduction tasks. Complex wavelets are the extension of DWT which adds a complex component, providing more information about the signal phase. The DTCWT, and DD-DTCWT are complex wavelets that offer improved directionality and shift-invariance, making them valuable in applications like medical imaging. Different wavelets used in the proposed work are addressed in the subsequent subsections.

3.2.1. Discrete Wavelet Transform (DWT)

Discrete Wavelet Transform (DWT) [149] consists of two filters - Low pass and high pass filters during both decomposition and reconstruction stages. Low pass filter extracts the approximation coefficients, which inherits the signal’s low-frequency components. It essentially captures the general trend or smooth variations in the signal. High pass filter is utilized to extract the detail coefficients, which shows the signal's high-frequency components. It captures the finer details, such as edges or rapid changes in the signal. Fig. 3.1 represents the implementation of 2-D DWT.

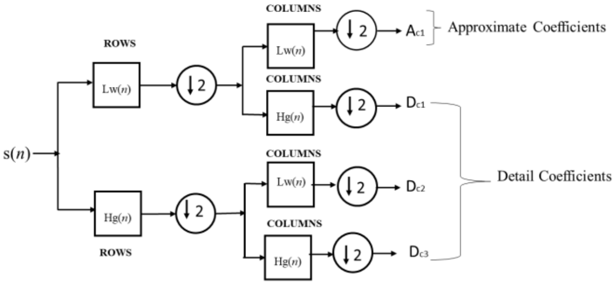


Fig. 3.1 2-D Discrete Wavelet Transform

In Fig. 3.1, the signal s(*n*) undergoes decomposition using a low pass filter $L_w(n)$ and a high pass filter $H_g(n)$ resulting in A_{c1} as approximate coefficients and D_{c1}, D_{c2}, and D_{c3} as detail coefficients in first level. To implement 2-D DWT, two stage filter banks are used. In the first stage, rows of 2-D signal are convolved with $L_w(n)$ and $H_g(n)$ filters and then down sampled by 2. In second stage, columns of the 2-D signal are convolved with $L_w(n)$ and $H_g(n)$ filters and then down sampled by 2. This process decomposes the 2-D signal into four sub-bands in which one is giving approximate coefficients and three are giving detail coefficients.

3.2.2. Double Density Discrete Wavelet Transform (DD-DWT)

The Double Density Discrete Wavelet Transform (DD-DWT) [150] extends the capability of Discrete Wavelet Transform (DWT) in terms of improved directional sensitivity. The one-dimensional DD-DWT is designed by using one low pass and two high pass filters for both decomposition and reconstruction purposes. The DD-DWT is applied on both rows and columns of an image alternatively. This results in nine sub-bands, of which one is an approximate (low-frequency) sub-band, and the remaining eight are detailed (high-frequency) sub-bands. Fig. 3.2 illustrates the structure of 2-D DD-DWT.

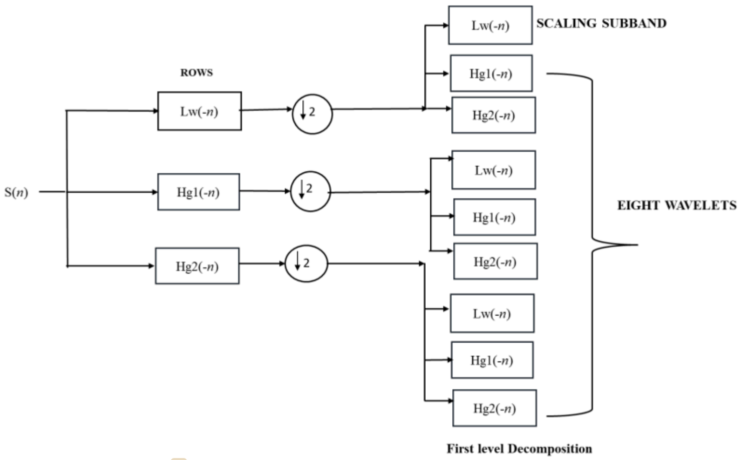


Fig. 3.2 2-D Double Density Discrete Wavelet Transform

Fig.3.2 depicts the first level implementation of 2-D DD-DWT.The s(n) is the signal to be decomposed. The $H_{g1}(-n)$ and $H_{g2}(-n)$ are the high pass filters and $L_{w}(-n)$ is the low pass filter. For perfect reconstruction of the signal, the integer translates of one wavelet lie halfway between the integer translates of another wavelet. 2-D DD-DWT shows reduced shift sensitivity and reduced rectangular artifacts.

3.2.3. Dual Tree Complex Wavelet Transform (DTCWT)

DTCWT [151] extracts feature in six directions of $\pm 15^\circ, \pm 45^\circ, \pm 75^\circ$. It does not show shift variance and has high directional selectivity. It has two DWT trees, operating in parallel. Tree 'a' filters are the real part and Tree 'b' filters are the imaginary part of the complex wavelet.Fig.3.3 illustrates the implementation of 2-D DTCWT up to the second level of decomposition.

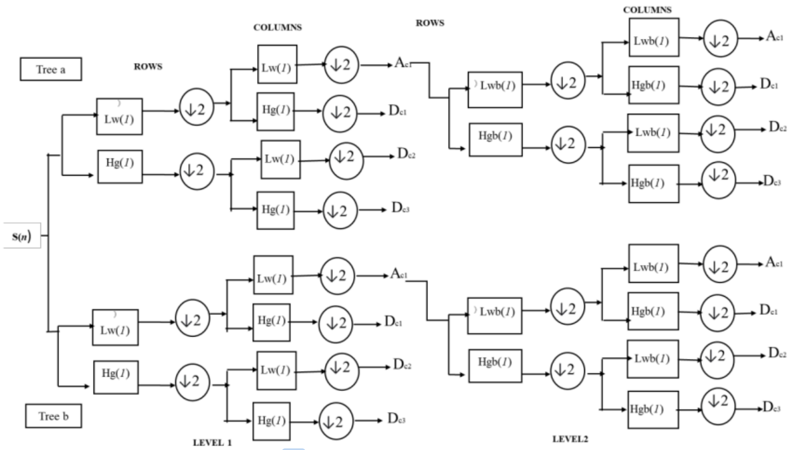


Fig. 3.3 2-D Dual Tree Complex Wavelet Transform

Analysis and synthesis filter banks in both DWT trees are selected in a manner by which the wavelet of Tree 'b' would be the Hilbert Transform of the wavelet of Tree 'a' i.e. $\varphi_b(t) = H\{\varphi_a(t)\}$.Here $\varphi_b(t)$ is the wavelet function of tree 'b' given by $\varphi_b(t) = \sqrt{2}\sum_n \chi_b H_g(n)(2t\text{-}n)$, $\varphi_a(t)$ is the wavelet function of tree 'a' given by $\varphi_a(t) = \sqrt{2}\sum_n \chi_a H_g(n)(2t\text{-}n)$, $\chi_a(t)$ is the scaling function of tree 'a' given by $\chi_a(t) = \sqrt{2}\sum_n \chi_a L_w(n)(2t\text{-}n)$ and $\chi_b(t)$ is the scaling function of tree 'b' given by $\chi_b(t) = \sqrt{2}\sum_n \chi_b L_w(n)(2t\text{-}n)$. The $L_w(n)$ and $H_g(n)$ functions work as the low pass filter and the high pass filter. The analysis filter in the first stage and subsequent stages are different in both trees to achieve perfect reconstruction conditions.

3.2.4. Double Density Dual Tree Complex Wavelet Transform (DD-DTCWT)

DD-DTCWT [152] is the combination of DTCWT and DD-DWT. DD-DTCWT implementation requires four different wavelet functions ($\psi_{p1}(t),\psi_{p2}(t),\psi_{r1}(t),\psi_{r2}(t)$ and two different scaling functions $(\phi_p(t),\phi_r(t))$.The wavelet functions $\psi_{r1}(t),\psi_{r2}(t)$ and scaling function $\phi_r(t)$ are used for the real part of complex wavelet. The wavelet functions $\psi_{p1}(t),\psi_{p2}(t)$ and scaling function $\phi_p(t)$ are used for the imaginary part of the complex wavelet. The wavelet functions are designed in a manner so that the $\psi_{p1}(t)$ is half sample delay of $\psi_{p2}(t)$ and $\psi_{r1}(t)$ is half sample delay of $\psi_{r2}(t)$ as given

$$\psi_{p1}(t) = \psi_{p2}(t\text{-}0.5) \quad \text{and} \quad \psi_{r1}(t) = \psi_{r2}(t\text{-}0.5) \tag{3.1}$$

And both wavelets are Hilbert transform of each other.

$$\psi_{p1}(t) = H\{\psi_{r1}(t)\} \quad \text{and} \quad \psi_{p2}(t) = H\{\psi_{r2}(t)\} \tag{3.2}$$

Fig.3.4 illustrates the structure of 2-D DD-DT-CWT up to the first level.

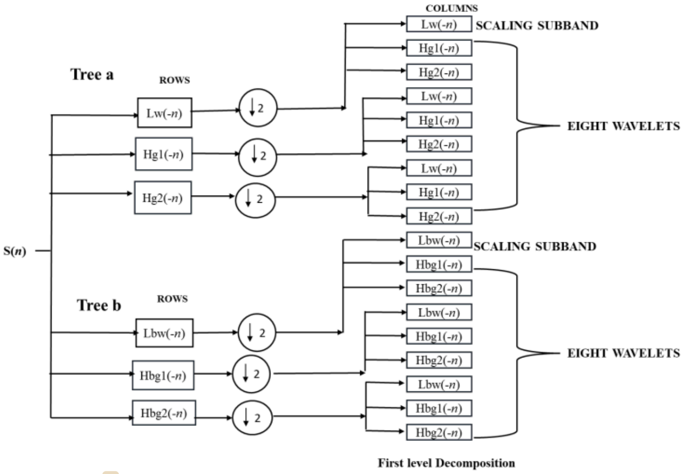
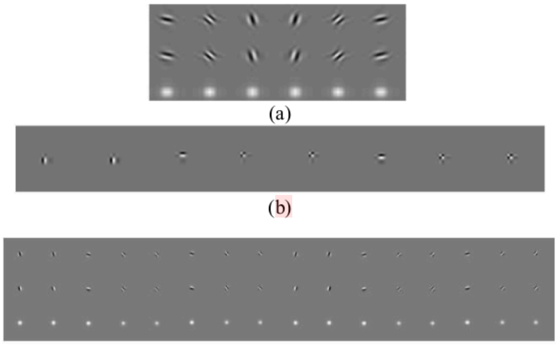


Fig. 3.4 2-D Double Density Dual-Tree Complex Wavelet Transform

Tree ‘a’ shows the real part and Tree ‘b’ shows the imaginary part of the complex wavelet. Fig 3.4 shows the two scaling functions as $L_w(-n)$ and $Lb_w(-n)$ of tree ‘a’ and tree ‘b’ respectively and four wavelet functions as $H_{g1}(-n)$, $H_{g2}(-n)$ and $Hb_{g1}(-n)$, $Hb_{g2}(-n)$ of tree ‘a’ and tree ‘b’ respectively. For 2-D implementation, transform is applied on both rows and columns of an image alternatively. This results in eighteen sub-bands, of which two are an approximate (low-frequency) sub-band, and the rest sixteen are detailed (high-frequency) sub-bands. The DT-CWT, DD-DWT, and DD-DTCWT wavelets are illustrated in Fig.3.5.



(a) (b) (c)

Fig. 3.5 Wavelets (a) 2-D DTCWT (b) 2-D DD-DWT (c) 2-D DD-DTCWT

2-D DTCWT provides six complex high sub-bands in six directions $\pm 15^\circ, \pm 45^\circ, \pm 75^\circ$ and one low pass sub-band at each decomposition level. Fig. 3.5(a) shows the six wavelets for real part in first line, six wavelets for imaginary part in second line and the magnitude of the complex wavelet in third line. Fig. 3.5 (b) shows eight wavelets provided by 2-D DD-DWT in eight orientations. The first two wavelet lies in the vertical orientation; the horizontal direction is occupied by the third and sixth wavelet, while the remaining four wavelets merge the two diagonal directions thus giving rise to the checkerboard effect. 2-D DD-DTCWT has 16 complex wavelets in 16 directions with 16 real wavelets in first line and 16 imaginary wavelets in second line in Fig.3.5 (c). Thus, DD-DTCWT provides 32 wavelets oriented at 16 different angles.

3.3. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) lowers the dimensionality of data while keeping variance. This low dimensional space includes the principal components which are perpendicular to each other. Principal Components are linear transformations of the original data variables that maximize the variance of all the variables. First Principal component shows the maximum variance. The PCA includes the following steps:

(i) The initial data is standardised so that all input variables may be at same scale. This is done by arranging the input data into matrix I (M×L), where M and L are the number of rows and column representing samples and features respectively. Calculate the mean of each column of input matrix and standardize the input data by deducting mean from each feature in the column. Put the result in a form of matrix M_{mn} having dimensions M×L.

(ii) The covariance matrix $C_{ov}=\frac{1}{M}M_{mn}^{T}.M_{mn}$ is computed.

(iii) The relation $V_e^{-1}C_{ov}V_e=D_{gl}$ with V_e as eigenvector matrix and D_{gl} as the diagonal matrix is used to find the eigenvalue and the eigenvector of a covariance matrix. Eigenvector represents the principal component as well as the eigenvalue corresponds to the variances.

(iv) The input data matrix is projected onto the axes defined by principal components.

3.4. Feature Optimization Methods

Feature optimization is crucial for enhancing the classification accuracy of machine learning models. The performance of the following Feature optimization methods has been evaluated in the proposed work.

3.4.1. Particle Swarm Optimization (PSO)

The Particle Swarm Optimization (PSO) [153] follows the intelligence and movement of swarms that they use while foraging. The swarms can find the food place due to their cooperation, intelligence, and continuous updating of their position and velocity. PSO constitutes many particles as swarms that are traveling in the optimization problem search space to determine the optimized solution. Each particle dynamically modifies its position and velocity, to determine the global minimum solution in the search space. The velocity and location of a particle are modified according to the flying experiences of its own and other particles in the group. Particles update their location and velocity according to their fitness value which is evaluated by fitness function. The location X_p^t of a p particle in t iteration can be defined in N dimensional space as

$$X_p^t = \left[x_{0,p}^t, x_{1,p}^t, x_{2,p}^t, x_{3,p}^t \dots \dots, x_{N,p}^t\right] \tag{3.3}$$

where $x_{N,p}^t$ is the value of the N^{th} co-ordinate of the p particle in t iteration

The particles are in continuous motion with a velocity Vx_p^t in search space. With this velocity they can change their location to find the global minimum solution.

$$Vx_p^t = \left[vx_{0,p}^t, vx_{1,p}^t, vx_{2,p}^t, vx_{3,p}^t \dots \dots, vx_{N,p}^t\right] \tag{3.4}$$

where $vx_{N,p}^t$ is velocity of p particle in N^{th} dimension at t iteration

Each particle modifies its location $\left(X_p^{t+1}\right)$ and velocity $\left(Vx_p^{t+1}\right)$ according to its personal best location ($pbest$) and global best location ($gbest$). Global best represents the best location acquired by a particle among all particles.

$$X_p^{t+1} = X_p^t + Vx_p^{t+1} \tag{3.5}$$

$$Vx_p^{t+1} = wVx_p^t + c_1 r_1\left(pbest_p^t - X_p^t\right) + c_2 r_2\left(gbest_p^t - X_p^t\right) \tag{3.6}$$

In Eqn. (3.6), $pbest_p^t$ and $gbest_p^t$ represent the personal best and global best location of the p particle at t iteration. Here, w represents the inertia weight constant and provides balance between the process of exploration and exploitation. The c_1 and c_2 are the acceleration coefficients which maintain the convergence ability of PSO. The r_1 and r_2 are the random numbers. `

3.4.2. Ant Colony Optimization (ACO)

ACO is a metaheuristic algorithm which is introduced by Dorigo and Caro [154]. This algorithm is based on the foraging strategy of ants. Under this strategy, ants search the shortest route between their home and the food source without any direct communication among them. Besides this, Pheromone levels act as a communication tool for searching the path between nest and food. The path having a higher level of pheromone deposits shows a greater probability of being the appropriate and shortest path. The amount of pheromone across the path between any two arbitrary points 'm' and 'n' for k^{th} ant is

$$\Delta\tau_{m,n}^k = \begin{cases} {}^{1}/{}_{l_k} & when\ k^{th}\ ant\ travel\ along\ m\ to\ n\ path \\ 0 & otherwise \end{cases} \tag{3.7`}$$

Where 'l_k' is the length of the path followed by k^{th} ant. Total number of ants along path 'm' to 'n' is 'N' so summation of pheromone ($\tau_{m,n}^k$) along 'm' to 'n' path is

$$\tau_{m,n}^k = \sum_{k=1}^{N} \Delta\tau_{m,n}^k \tag{3.8}$$

This amount is without vaporization. This equation can be written as considering vaporization.

$$\tau_{m,n+1}^k = (1-\rho)\tau_{m,n}^k + \sum_{k=1}^{N} \Delta\tau_{m,n}^k \tag{3.9}$$

Where ρ is constant and depicts the vaporization rate.

3.4.3. Genetic Algorithm (GA)

Genetic Algorithm is inspired by Charles Darwin's theory of Natural evolution [155]. According to this theory, GA functions on the idea of reproduction and survival of the fittest.GA handles many parameters and is robust. It provides multiple solutions rather than a single solution. GA uses the fitness function for evaluation. In GA, a single solution is called an individual, and a group of individuals exploring the search space is referred to population. In GA, the population of randomly generated solutions (individuals) to an optimization problem evolved toward better solutions. Each candid solution in the population is called a chromosome. Each chromosome is represented by a string of binary values, "0" s and "1" s. These binary values indicate the state of attributes in the chromosome. The fitness of each chromosome is then checked by using a fitness function. Further chromosomes are sorted according to their fitness values. Now two chromosomes are selected based on their positions on a roulette wheel and then passed through genetic operations viz. crossover and mutation. Now the new offspring are evaluated and if they are found strong enough, they are passed into a new generation This process iterates until the termination criteria are fulfilled. Fig.3.6 shows the complete flow of process in Genetic Algorithm.

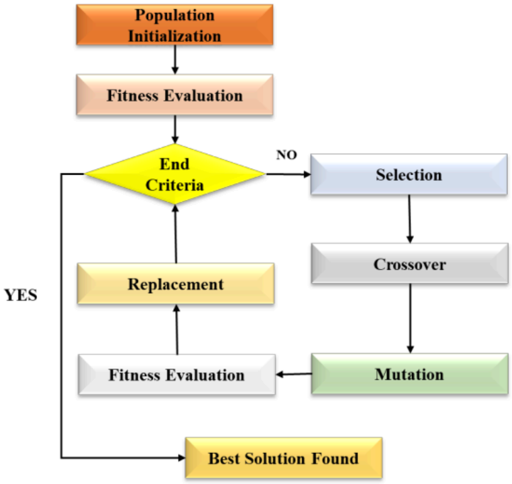


Fig. 3.6 Flow diagram: Genetic Algorithm

According to Fig. 3.6, GA completes in the given steps:

1) The population of individuals is randomly generated initially.
2) The fitness number of each individual is calculated by evaluating them with a fitness function.
3) Examine whether the individuals fulfil the termination criteria or not.
4) The best individual is selected for reproduction purposes if termination criteria are not met. The algorithm comes to an end after the termination criteria are fulfilled.
5) Selected individuals undergo genetic operations viz. Crossover and mutation to produce offspring set.

Now the fitness value of new off-spring is calculated and the whole process starts from step 3. The following points can be considered while setting termination criteria:

(1) If the model has achieved maximum possible accuracy.
(2) If successive generations achieved the same results in terms of accuracy.
(3) If there has been a maximum number of generations.

3.5. Statistical Methods

Statistical-based feature selection methods select those input variables which are highly related to the target variables with the help of statistics. The following types of t-tests [156] are used as statistical methods for feature selection.

3.5.1. Independent t-test (student's t-test)

Independent t-test is used when two independent groups are to be compared. This method helps in finding the features with significant mean differences. This method includes two different kinds of hypotheses- Null hypothesis (H0) and Alternative hypothesis (H1). The null hypothesis is true when the two groups have equal means. The alternative hypothesis is that the two groups have different means.

3.5.2. Paired t-test

This t-test is used when the same subjects are tested under different conditions (e.g., before and after treatment). This test is less common in feature selection.

3.5.3. Welch's t-test

This t-test is used when the two groups have unequal variances. This t-test is more robust than the standard t-test.

The performance of the student's t-test has been evaluated for feature selection in the proposed work. The t-test compares the means of two groups and determines whether the differences between the means are statistically significant or not. If the means have a significant difference, then that feature is more likely relevant for classification. The value of 't' can be found as

$$t = \frac{\bar{x}1 - \bar{x}2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{3.10}$$

where $\bar{x}1$ and $\bar{x}2$ represent sample means, s_1 and s_2 represent sample variances and n_1 and n_2 represent sample sizes. The value of 't' is equal to the p-value. The following steps are used to perform the t-test

1. The t-test is performed between the two different groups for each feature like in AD/NC classification, one group is AD and another group is NC.
2. The p-value is obtained by calculating 't' value in Eqn. (3.10) and if $p<0.05$, it indicates that the means of the two groups are significantly different.
3. The features with statistically significant differences are selected for classification task.

3.6. Dataset

For implementation, 150 subjects have been downloaded from the OASIS-1 dataset. These subjects include 50 ADs, 50 NCs, and 50 MCIs. Skull stripping and intensity modulation have been performed on these 3-D images. A set of thirty-two two-dimensional slices, selected for their high informational content, were extracted using XMedCon software and saved in JPEG format for feature extraction purposes. These slices are 176 × 208 in size and further resized into 256 × 256 to achieve adequate resolution in the fifth-level detail coefficients. The sample images for AD, NC, and MCI classes from the OASIS dataset are depicted in Fig. 3.7.

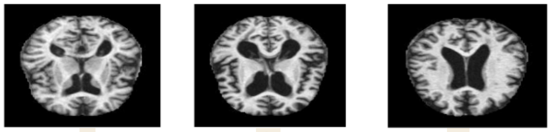

(a) (b) (c)

Fig. 3.7 Demonstrative 2-D MR scan (a) AD (b) MCI (c) NC

3.7. Proposed Method

In the proposed work, 3-D images have been used for feature extraction. They are converted into 2-D slices.

- Let the input 3-D MRI

$$\mathrm{I}\,(\mathrm{x},\mathrm{y},\mathrm{z}) \text{ of size } N \times M \times L \tag{3.11}$$

- Set of 32 axial (x-y plane) slices obtained from 3-D MRI is shown by $\sum_{K=1}^{32} I_k$ (x, y) at z_k , where z_k is the thickness of the slice along the z-axis and N × M is the size of each 2-D slice.
- Each 2-D slice is resized into 256 × 256.
- The decomposition of image at n^{th} level in approximation and detail sub-bands using DD-DTCWT can be represented as

$$DD - DTCWT_{(n)} = \{AL_{n,1}, AL_{n,2}, S_{n,m}\} \tag{3.12}$$

where $AL_{n,1}$ and $AL_{n,2}$ represent the approximation sub-bands and $S_{n,m}$ represents the m^{th} detail sub-band at the n^{th} decomposition level.

- Fifth level DD-DTCWT is applied on 2D slices, which gives two approximate bands and sixteen detail sub-bands.

$$DD - DTCWT_{(5)} = \{AL_{5,1}, AL_{5,2}, S_{5,1} \ldots\ldots\ldots\ldots S_{5,16}\} \tag{3.13}$$

The complex coefficients, $C_{i,n,m}$ at fifth level in the detail sub-bands are represented as given in Eq. (3.14),

$$C_{i,5,m} = R\{C_{i,5,m}\} + jI_m\{C_{i,5,m}\} \tag{3.14}$$

where, $R\{C_{i,5,m}\}$ denote real coefficients and $I_m\{C_{i,5,m}\}$ denote imaginary coefficients of i^{th} complex wavelet coefficients of 'm^{th}' sub-band at 'n^{th}' decomposition level. These detail coefficients at fifth level are used for classification task.

- Each 3-D MR volume constitutes 32 2-D centre MR slices, so total number of detail coefficients will be 32×2048 equal to 65536 for each 3-D MRI at fifth level of decomposition. Fig.3.8 indicates the total of 2048 detail coefficients at fifth level DD-DTCWT from single 2-D MR slice of size 256 × 256.

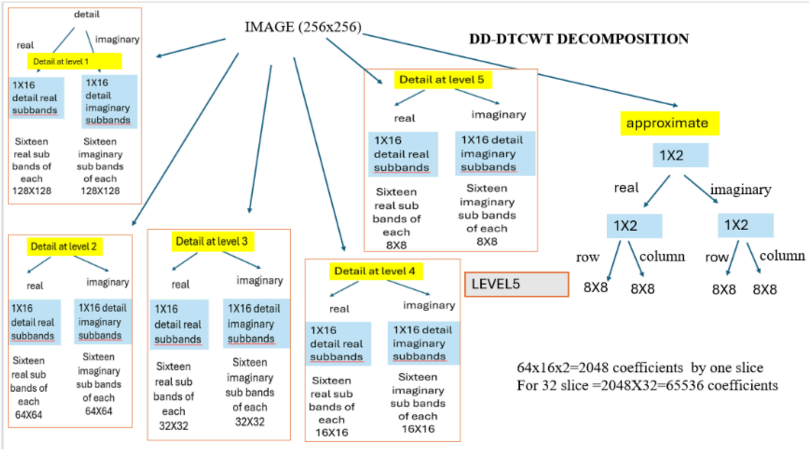


Fig. 3.8 Detail coefficients at fifth level DD-DTCWT decomposition

Fig. 3.8 depicts how 32 slices of size 256×256 provide 65536 coefficients. The dimensionality of the extracted coefficients was reduced through the application of both dimensionality reduction and feature selection techniques. These refined features were subsequently utilized for AD classification. Fig. 3.9 illustrates the workflow of the proposed technique through a pipeline model.

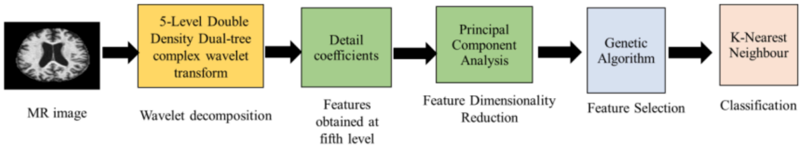


Fig.3.9 Pipeline model: Proposed Method

Fig. 3.9 includes the main steps in the proposed method. A detailed process of feature extraction is shown in Fig. 3.10. PCA has been used for reducing feature dimensions and GA has been used for selecting optimum features.

Fig. 3.10 Feature Extraction Process

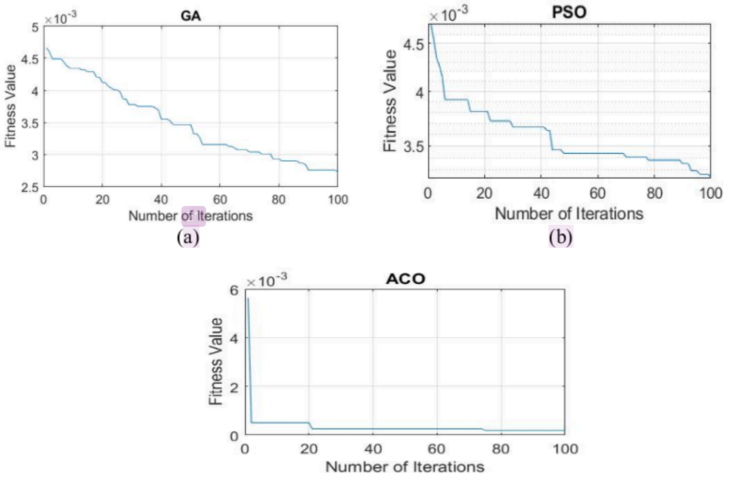
In Fig. 3.10, a row vector of 1 × 65536 has been formed from the matrix 32 × 2048. Here one represents the single 3-D MRI and 65536 signifies the count of detail coefficients. Further, a matrix for 100 rows (including 50 AD and 50 NC subjects) has been formed with 65536 number of columns (detail coefficients). This matrix suffers from high dimensionality thus at the second stage, PCA has been applied to reduce dimensionality. The smallest number of components that explain 95% of the total variance are selected. PCA provides a score matrix of dimensions 100 × 99 where 100 is the total count of subjects and 99 is the number of principal components. A feature selection algorithm i.e. GA has been introduced in the third stage for enhancing the classification results. A total of 99 features are fed to GA, and a total of 35 features are chosen by GA. These 35 features are further used for AD classification.

The proposed method is executed using three different wavelets also viz. DWT, DTCWT and DD-DWT. The performance of t-test and feature selection algorithms viz. PSO and ACO has also been evaluated. A total of 24 models have been investigated and the best results have been obtained with the model DD-DTCWT+PCA+GA. Four feature selection methods have been tested with all four wavelets. Four classifiers viz. Decision Tree (DT), Naive-Bayes (NB), Support Vector Machine (SVM) and KNN have been employed for classification purposes. Table 3.1 indicates the parameters of feature selection algorithms used in the implementation of the proposed work.

Table 3.1 Parameters

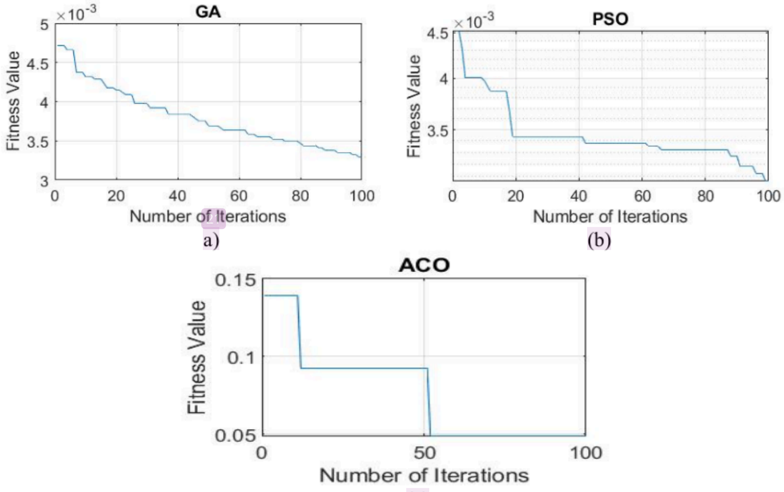
PSO		ACO		GA	
Parameter	**Value**	**Parameter**	**Value**	**Parameter**	**Value**
Number of solutions	30	Number of solutions	30	Number of solutions	30
Number of maximum iterations	100	Number of maximum iterations	100	Number of maximum iterations	100
c1	2	Pheromone value (τ)	1	Cross-over rate	0.8
c2	2	heuristic desirability (eta)	1	Mutation rate	0.01
W	0.9	Pheromone factor (α)	1		
		control heuristic (beta)	0.1		
		pheromone evaporation rate (ρ)	0.2		

Table 3.1 states the optimum values of the number of solutions per generation, number of maximum iterations, c1, c2, w, pheromone value, pheromone evaporation rate, heuristic desirability, cross-over rate, mutation rate for feature selection algorithms. Figs. 3.11, 3.12, and 3.13 show the convergence plots of feature selection algorithms with the proposed model i.e., DD-DTCWT+PCA+GA+KNN for AD classification with NC, AD classification with MCI, and MCI classification with NC respectively. The convergence plot illustrates the correlation between the number of iterations (or generations) and the fitness (or cost) value. It tells how the quality of the chosen feature subset has improved over time. It helps in evaluating the algorithm's convergence behaviour and efficiency.



(a) (b) (c)

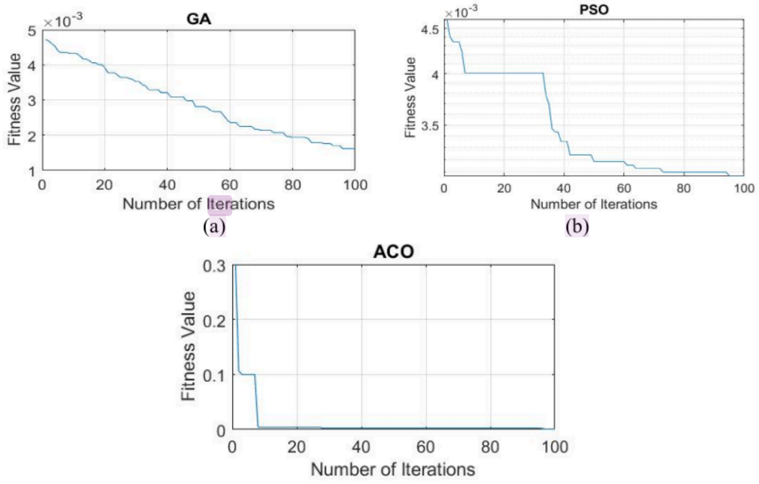
Fig. 3.11 Convergence plots (a) GA (b) PSO (c) ACO with proposed method (AD/NC)



a) (b)

(c)

Fig. 3.12 Convergence plots (a) GA (b) PSO (c) ACO with proposed method (AD/MCI)



(a) (b)

(c)

Fig. 3.13 Convergence plots (a) GA (b) PSO (c) ACO with proposed method (MCI/NC)

In the proposed work, the fitness value for the feature selection algorithm has been obtained from the cost function given as

Cost function $=\alpha$ * error + β * (number of selected features/total number of features)

Where error = 1 - Accuracy   and $\alpha = 0.99$, $\beta = 0.01$

The subset of features that have a minimum value of cost function will be the fittest one.

3.8. Experimental Results

The proposed work has been implemented on MATLAB R2019 A. The ten-fold cross validation technique has been used to divide the dataset into training and testing sets. In this technique whole dataset is divided into ten equal parts. Nine parts are used for training, and one part is used for testing. This process is repeated 10 times, each time using a different part as the test set.

The proposed work i.e. DD-DTCWT-based direct coefficients feature extraction technique provides an accuracy of 96.9%,92.9%,90.6%, a sensitivity of 97.6%,90%,91%, and specificity of 95.8%,94.1%,90% for distinguishing AD from NC, AD from MCI, and MCI from NC respectively. This work is assessed with four wavelets i.e. DWT, DTCWT, DD-DWT, and DD-DTCWT. The number of wavelet coefficients in detail sub-bands provided by DWT, DTCWT, DD-DWT, and DD-DTCWT are 192,512,768 and 2048 respectively at the fifth level of decomposition with a single 2-D MR slice, and with 32 slices, these detail coefficients are 6144,16384,24576 and 65536 respectively. These detail coefficients have been used as features as they highlight the structural and edge information. A total of 24 models are implemented in the proposed work. The first four models use coefficients from DWT, DTCWT, DD-DWT, and DD-DTCWT for AD classification. The models from 5th to 8th apply PCA to reduce dimensionality and the rest 9th to 24th models use PCA+ feature selection algorithms (ACO, PSO, GA, t-test) for AD classification.

3.8.1. AD/NC Classification Results

Table 3.2 indicates the results of models from 9 to 24 with different classifiers like SVM, KNN, NB, and DT. The other models (1 to 8) are not included in Table 3.2, due to their low reported classification accuracy. The best results are observed with the KNN classifier.

Table 3.2 Models (9 to 24) for AD classification with NC using different classifiers

Feature Selection Method	**Classifier**	**DWT+PCA**			**DTCWT+PCA**			**DD-DWT+PCA**			**DD-DTCWT+PCA**		
		Acc	**Sen**	**Spe**	**Acc**	**Sen**	**Spe**	**Acc**	**Sen**	**Spe**	**Acc**	**Sen**	**Spe**
GA	**KNN**	91.1	94.3	87.9	95.6	98	90.4	92.9	95.1	90.9	96.9	97.6	95.8
	DT	88.6	88.4	88.8	90.5	92.7	88.4	87.9	87	88.9	89.2	91.4	87
	SVM	87.6	92.8	82.7	88.2	92	84.6	89.9	96	84.2	92.7	90.8	94.5
	NB	91.2	93.4	89.3	93.1	95.4	90.8	90.9	92.6	89	93.5	94.5	92.5
PSO	KNN	89.2	91.1	87.1	94.4	96.5	91.6	87.8	92.6	83	95	94	93
	DT	83.2	80.6	85.9	87.4	90.3	84.5	79.1	78.6	80	82.6	87.3	78
	SVM	87.4	88.5	86.4	87.9	90.2	85.8	87.3	90.8	83.6	86.3	87.9	84.6
	NB	83.1	88.4	77.7	87.8	86.5	89.2	87.1	89.7	84.5	89.2	89.9	88.4
ACO	KNN	84.9	84.8	85.2	87.2	91.3	83	82.2	87.6	76.8	84.5	85.1	83.9
	DT	74.1	71.8	75.8	85.6	87.4	84.4	78.7	78.8	78.4	80.6	80.9	80.4
	SVM	82.6	79.8	85.5	85.1	87.6	82.6	83.8	88.3	79.2	85.7	87.5	83.9
	NB	79.2	82.8	75.3	82.6	82.3	82.8	78.9	81.9	76.1	82.4	83.2	81.6
t- Test	KNN	69.7	78.5	60.9	72.2	73.0	71.0	71.4	79.0	63.0	75.4	76.0	75.0
	DT	66.7	66.5	67.1	69.8	69.0	71	66.7	70.0	63.0	70.6	68.0	73
	SVM	71.8	73.9	69.8	75.4	73	78	72.2	75.0	70.0	75.4	78.0	73.0
	NB	60.8	64.2	57	71.4	71.0	71.0	69.8	70.0	70.0	72.2	70.0	75.0

From Table 3.2, the KNN classifier performs better in comparison to other classifiers in almost all the models. The reason is that KNN performs well with nonlinear data and does not make assumptions with the dataset and adjusts parameters based on the neighbours. The performance of all 24 models using KNN classifier is depicted in Table 3.3.

Table 3.3 Models implemented for AD/NC classification using KNN classifier

Model No.	**Models**	**Acc/Sen/Spe**	**Model No.**	**Models**	**Acc/Sen/Spe**
1	DWT	70.0/77.1/62.8	13	DTCWT+PCA+PSO	94.4/96.5/91.6
2	DTCWT	71.2/77.8/64.5	14	DTCWT+PCA+ACO	87.2/91.3/83.0
3	DD DWT	68.6/75.5/61.8	15	DTCWT+PCA+GA	95.6/98.0/90.4
4	DD-DTCWT	67.0/76.8/57.3	16	DTCWT+PCA+ t-Test	72.2/73.0/71.`0
5	DWT+PCA	72.7/80.2/65.1	17	DD DWT+PCA+PSO	87.8/92.6/83
6	DTCWT+PCA	73.6/79.6/69.6	18	DD DWT+PCA+ACO	82.2/87.6/76.8
7	DD DWT+PCA	72.4/82.9/62.2	19	DD DWT+PCA+GA	92.9/95.1/90.9
8	DD-DTCWT+PCA	74.3/71.8/76.9	20	DD DWT+PCA+ t-Test	71.4/79.0/63.0
9	DWT+PCA+PSO	89.2/91.1/87.1	21	DD-DTCWT+PCA+PSO	95.0/94.0/93.0
10	DWT+PCA+ACO	84.9/84.8/85.2	22	DD-DTCWT+PCA+ACO	84.5/85.1/83.9
11	DWT+PCA+GA	91.1/94.3/87.9	**23**	**DD-DTCWT+PCA+GA [PROPOSED]**	**96.9/97.6/95.8**
12	DWT+PCA+ t-Test	69.7/78.5/60.9	24	DD-DTCWT+PCA+ t-Test	75.4/76.0/75.0

From Table 3.3, the models1,2,3 and 4 suffer from the problem of high dimensionality as in these models, coefficients are directly fed to the classifier. The classification performance of these groups is not good due to high dimensionality. DWT (model 1) gives an accuracy of 70.0%, DTCWT (model 2) gives an accuracy of 71.2 %, DD-DWT (model 3) gives an accuracy of 68.6%, DD-DTCWT (model 4) gives an accuracy of 67.0% using KNN classifier. The results with model 4 i.e., DD-DTCWT are not good. This model has dimensionality and overfitting problems as input to the KNN classifier is 65536 features which is quite high in number. The classification is also done using other classifiers viz. DT, NB, SVM but their performance is not good. The models 5,6,7 and 8 included PCA for reducing dimensionality and for finding uncorrelated features. The dimensionality reduction by PCA enhances the classification accuracy. The classification accuracy with groups DWT+PCA, DTCWT+PCA, DD-DWT+PCA and DD-DTCWT+PCA are 72.7%, 73.6%, 72.4% and 74.3% respectively. Among these models, DD-DTCWT+PCA is the best performer as it extracts features in sixteen different directions. From model 9 onwards, the feature selection stage is introduced which further improves the classification accuracy by selecting appropriate features. Three evolutionary algorithms viz. PSO, ACO, GA, and statistical tests namely, t-test are used independently for the feature selection process. GA provides better results with all wavelets compared to all other feature selection algorithms like ACO, PSO and t-test.

The GA algorithm takes longer to compute in MATLAB than ACO and PSO, even if it produces the best results when using the proposed technique. It takes 47.82 seconds to compute using the DD-DTCWT wavelet. ACO algorithm is faster when compared to PSO and GA. With DD-DTCWT, the PSO takes 41.62 seconds, and the ACO takes 22.96 seconds.

The performance of models DD-DTCWT+PSO, DD-DTCWT+ACO, and DD-DTCWT+GA have also been examined. In these models, the feature selection algorithm receives 65536 features directly, which makes them suffer from high dimensionality. The overfitting problem causes MATLAB's computation time to increase significantly, leading to poor classification results. Tables 3.2 and 3.3 do not feature the results of these models. The model DD-DTCWT+PCA+GA has been proposed as it provides 96.9% prediction accuracy, 97.6% sensitivity, and 95.8% specificity.

Fig.3.14 shows the ROC curves of the implemented models while classifying AD and NC using KNN classifier.

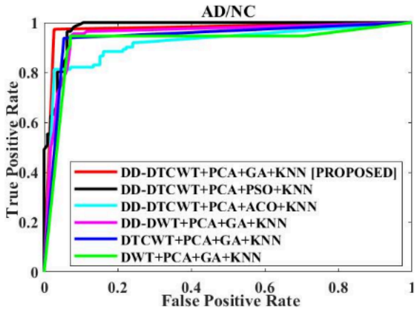


Fig. 3.14 ROC curves of implemented models (AD/NC)

The proposed model in Fig. 3.14, shows the steep rise in ROC curve towards the left-top quadrant. In the proposed work, GA gives the maximum results. GA is probabilistic, requires less information, and has global search capability. It also offers a bigger search field. With high dimensional datasets, PSO and ACO are more likely to become caught in local optima. GA makes use of both mutation and crossover operators. These operators increase the diverse population thus expanding the search space. As a result, GA is less likely to be trapped in a local optimum. GA works well in situations involving several constraints. GA offers a range of excellent outcomes from several generations. Fig. 3.15 illustrates the comparative average performance of all feature selection algorithms (ACO, PSO, GA, t-Test) with models DWT+PCA, DD-DWT+PCA, DTCWT+PCA and DD-DTCWT+PCA.

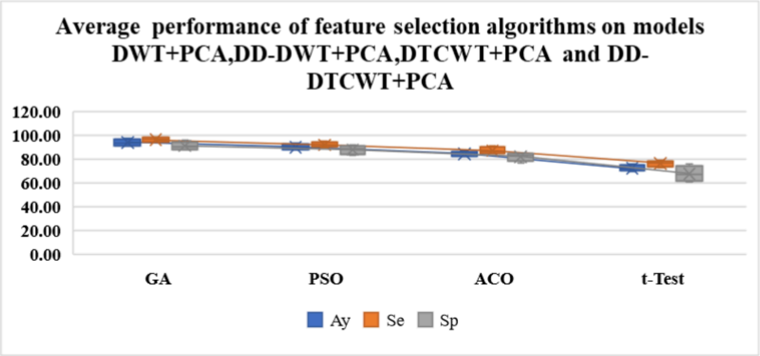


Fig. 3.15 Average performance of feature selection algorithms with all wavelets

It is clear from Fig.3.15 that the average performance of GA with DWT, DD-DWT, DTCWT, and DD-DTCWT is better than other feature selection algorithms. Fig.3.16 demonstrates the performance of GA with all four wavelets.

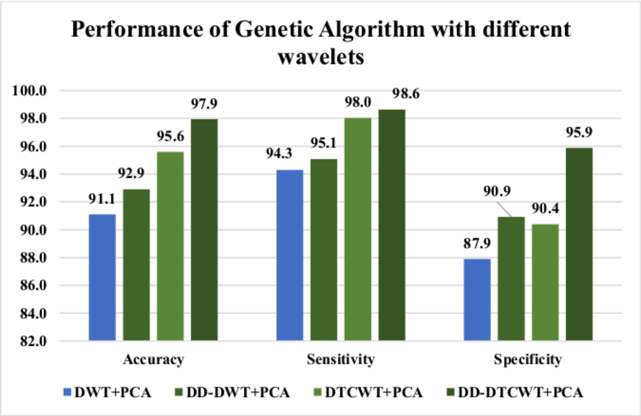


Fig. 3.16 Performance of GA with different wavelets

Fig. 3.16 shows the good performance of GA with different wavelets. GA efficiently balances exploration and exploitation through selection, cross-over, and mutation. GA shows parallel search capability which prevents GA from trapping in local optima.

3.8.2. Experimental Results: AD with MCI and MCI with NC

Additionally, the proposed method has been evaluated to distinguish between AD and MCI as well as between MCI and NC. The model DD-DTCWT+PCA+GA yields the best results in distinguishing MCI from AD and NC.DD-DTCWT extracts real and imaginary coefficients in sixteen directions which act as an important biomarker for AD detection. The dimension of these coefficients is very high which is further reduced by PCA and even the features selected by GA improve the classification accuracy further. GA takes respective 45.76 and 46.6 seconds for classifying AD from MCI and MCI from NC to execute with DD-DTCWT wavelet while PSO takes respective 40.50 and 41.1 seconds, and the ACO takes respective 21.86 and 22.2 seconds for these binary classifications. ACO algorithm is faster in execution in comparison to PSO and GA. Tables 3.4 and 3.5 present the performance of the models 9 to 24 using distinct classifiers for AD/MCI and MCI/NC classifications, in the given sequence.

Table 3.4 Models (9 to 24) for AD/MCI classification using different classifiers

Feature Selection Method	Classifier	DWT+PCA	DTCWT+PCA	DD-DWT+PCA	DD-DTCWT+PCA
		Acc/Sen/Spe	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**
GA	**KNN**	79.2/77/81	87.7/85/88	82.1/83/82	92.9/90/94.1
	DT	75.1/72/78	81.2/80/82	77.1/73/80	86.5/87/85
	SVM	76.2/73/79.1	85.1/85.2/84	80.4/78/82	89.5/87/91
	NB	74.2/73.1/75	79.1/78.1/80	75/78/72	84.5/84/86
PSO	KNN	73.7/76/71	76.4/76/77	74.1/75/73	86.6/85/88
	DT	71.3/72/69	74.2/73/75	72.2/73/71	82.5/81/83
	SVM	72.2/73/71	75.2/76/73	73.1/74/72	84.5/83/85
	NB	70/71.1/69.1	72.1/71/73	70.1/71/69	81.5/80/82
ACO	KNN	72.9/70/74	73.8/72/74	71.9/71/73	83.9/87/81
	DT	70.7/71/69	71.4/70/72	70.5/70/69	79.7/81/77
	SVM	71.2/72/68.9	72.9/71/73	71/69/72	80.2/82/78
	NB	69.8/67/70	70.1/72/69	69.1/71/68	76.8/75/77
t- Test	KNN	67.5/67/68	69.9/69/70	68.3/64/72	68.3/64/72
	DT	65.2/66/64	67.5/66/68	65.5/66/64	61.7/60/62
	SVM	66.5/65/67	68.2/69/67	67.2/69/66	65.2/67/63
	NB	63.2/64/61	65.2/66/64	65.2/66/63	60.8/61/60

Table 3.5 Models (9 to 24) for MCI/NC classification using different classifiers

Feature Selection Method	Classifier	DWT+PCA	DTCWT+PCA	DD-DWT+PCA	DD-DTCWT+PCA
		Acc/Sen/Spe	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**
GA	**KNN**	77.9/76/78	85.6/84/86	80.8/76/86	90.6/91/90
	DT	73.1/72/75	80.5/81/79	78.5/76/80	85.1/87/83
	SVM	74.4/73/75.1	82/81/83	79.5/77/81	89.2/87/91
	NB	70/71/69	80.7/81/79	77.5/76/78	83.3/81/85
PSO	KNN	73.7/76/71	77.9/78/78	76.6/77/75	85.3/80/90
	DT	70.5/71/69	74.5/77/71	73.2/71/75	82.5/81/83
	SVM	72.7/70/73	76.5/77/75	75.5/71/79	84.5/81/87
	NB	69.8/66/71	72.4/71/73	71.1/72/70	81.5/80/82
ACO	KNN	72.9/70/76	74.2/76/73	74.7/73/77	81.3/83/79
	DT	71.2/70/72	73.4/71/74	72.4/70/74	79.4/78/80
	SVM	73.5/71.1/74	74.9/71/76	73.9/71/74	80.9/81/79
	NB	70.4/71/69	71.1/72/70	70.1/72/68	76.1/75/77
t- Test	KNN	67.5/65/70	68.2/66/70	69.6/70/69	76.8/78/74
	DT	64.5/63/65	63.5/61/67	67.7/66/68	74.5/71/77
	SVM	66.5/67/64	65.5/66/64	68.2/66/70	75.5/76/74
	NB	61.5/60/62	61.5/60/62	65.2/67/63	71.5/72/70

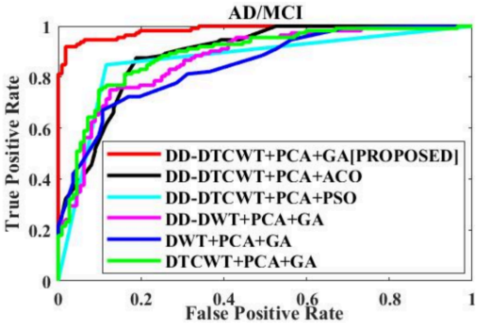
It can be seen from Tables 3.4 and 3.5 that the KNN classifier is the best performer among all classifiers in almost all the models. The performance of all 24 models using the KNN classifier is depicted in Table 3.6.

Table 3.6 Models implemented for AD/MCI and MCI/NC classifications with KNN classifier

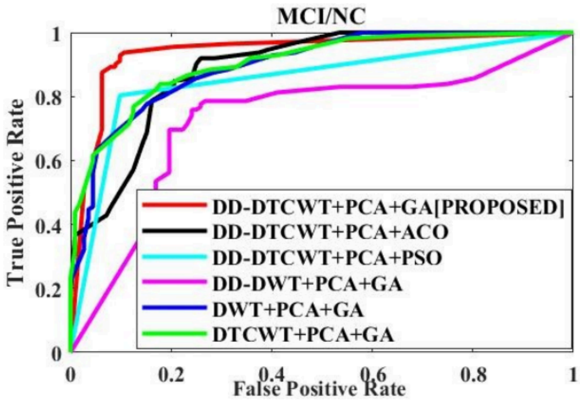
S. No	Models	AD/ MCI Acc/Sen/Spe	MCI / NC Acc/Sen/Spe	S. No	Models	AD/ MCI Acc/Sen/Spe	MCI / NC Acc/Sen/Spe
1	DWT	53/58/48	50/52/48	13	DTCWT+PCA+PSO	76.4/76/77	77.9/78/78
2	DTCWT	58.5/53/64	57.3/58/56	14	DTCWT+PCA+ACO	73.8/72/76	74.2/76/73
3	DD DWT	56.3/64/48	55.1/56/54	15	DTCWT+PCA+GA	81.7/82/81	82.6/84/81
4	DD-DTCWT	55.1/57/53	53.5/55.1/51.5	16	DTCWT+PCA+ t-Test	69.9/69/70	68.2/66/70
5	DWT+PCA	57.1/59/55	53.8/52/56	17	DD DWT+PCA+PSO	74.1/75/73	73.7/76/71
6	DTCWT+PCA	62.1/63/61	63.7/66/61	18	DD DWT+PCA+ACO	71.9/71/73	72.9/70/76
7	DD DWT+PCA	`60.3/57/63	58.7/58/59	19	DD DWT+PCA+GA	78.1/77/79	75.9/76/76
8	DD-DTCWT+PCA	66/67/65	65.3/61/69	20	DD DWT+PCA+ t-Test	68.3/64/72	67.5/65/70
9	DWT+PCA+PSO	73.7/76/71	76.6/77/75	21	DD-DTCWT+PCA+PSO	86.6/85/88	85.3/80/90
10	DWT+PCA+ACO	72.9/70/76	74.7/75/77	22	DD-DTCWT+PCA+ACO	83.9/87/81	81.3/83/79
11	DWT+PCA+GA	77.2/71/84	80.8/76/86	**23**	**DD-DTCWT+PCA+GA [PROPOSED]**	92.9/90/94.1	90.6/91/90
12	DWT+PCA+ t-Test	67.5/67/68	69.6/70/69	24	DD-DTCWT+PCA+ t-Test	79.5/84/75	76.8/80/73

Table 3.6 shows the outstanding performance of the proposed approach. The proposed model i.e. DD-DTCWT+PCA+GA provides 92.9% accuracy for classifying AD against MCI and 90.6% accuracy for classifying MCI against NC. The experimental outcomes verify that the proposed approach is applicable for both early and advanced stages of Alzheimer's Disease detection.

Fig. 3.17 indicates the ROC plots of implemented models for binary classifications i.e. AD/MCI and MCI/NC using OASIS dataset.



(a)



(b)

Fig. 3.17 ROC curves of implemented models (a)AD/MCI (b)MCI/NC

The proposed model in Fig. 3.17 shows the steep rise in ROC curve towards the left-top section of the plot. This shows the excellent outcome of the proposed model for early AD detection. The proposed method leverages detail coefficients to handle the image's high frequency details, which clearly illustrates the frequent variations in grey matter density.

3.9. Proposed Method vs Existing Methods

The proposed method has been benchmarked against twelve existing approaches, as detailed in Table 3.7. The existing techniques fall into different categories of feature extraction. These algorithms are executed on MATLAB-19.

Wavelet offers multiresolution signal analysis and helps in the extraction of data that is localized in both frequency and time. This helps to capture the directional and minute structural details in the AD brain. The current methods concentrate on both structural changes and variations in grey matter density while in the existing technique, the extracted features either provide information on structural changes [36] or variations in grey matter density [57]. The proposed method based on a wavelet transform-based approach, produces exceptional classification results by offering more directed and informative features. Table 3.7 presents the assessment of the proposed and existing algorithms for binary classifications using the OASIS dataset.

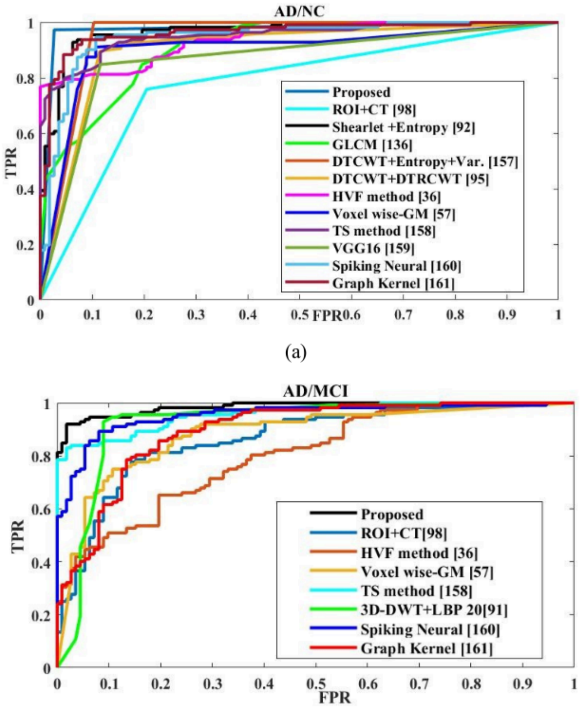
Table 3.7 Proposed vs Existing Methods

		AD/NC			AD/MCI			MCI/NC		
Method	**Classifier**	**Acc**	**Sen**	**Spe**	**Acc**	**Sen**	**Spe**	**Acc**	**Sen**	**Spe**
[PROPOSED]	KNN	**96.9**	**97.6**	**95.8**	**92.9**	**90**	**94.1**	**90.6**	**91**	**90**
ROI+ Contourlet Transform (CT) [98] [2022]	SVM	81.1	83.3	79.2	78.1	83.1	73.2	80.4	82	79
Shearlet + Entropy [92] [2019]	KNN	94.1	95.3	93.4						
GLCM [136] [2021]	ELM	93	92	94	-	-	-	-	-	-
DTCWT+ Entropy+ Variance [157] [2016]	TSVM	94.1	93.2	95.1	-	-	-	-	-	-
DTCWT+ DTRCWT [95] [2021]	FNN	90.3	91.4	92.5	-	-	-	-	-	-
Hippocampus Visual Feature (HVF) [36] [2015]	SVM	85.4	77.1	94	73.9	74.6	72.8	78.1	77	79
Voxel wise GM [57] [2012]	SVM	89.3	87	92	81.8	85.9	77.9	84	87.1	78.7
Tissue-based Segmented (TS) method [158] [2015]	SVM	88.5	89	86	86.3	87.1	86.2	85.4	86.1	85.2
VGG16 [159] [2022]	Neural Network	86.7	85.2	87.3						
Spiking Neural Network [160] [2021]	CNN	90.5	92	89	89.7	89	90	86.2	85	88
3D-DWT+LBP-20 [91] [2020]	SVM				90	89.5	91	88.3	83	94
Graph kernel-based approach [161] [2023]	SVM	92	90	94	82.6	88	77	81.3	82	80

In Table 3.7, the proposed method is evaluated against currently available methods. The study [91] applied 3D-DWT to extract Local Binary Pattern-20 features for AD Classification. Acharya et al. [92] applied shearlet transform to extract entropy-based features. Toppanavar et al. [95] used Dual Tree Rotated Complex Wavelet (DTRCWT) and DTCWT to extract textural information. This study captures intricate directional and textural features to improve the diagnostic accuracy between Alzheimer's disease and normal controls. The study [98] used contourlet transform to extract energy features from Grey Matter (GM) density maps. Wang et al. [157] extracted variance and entropy by using DTCWT. Gao et al. [136] obtained GLCM from MR images and extracted statistical features viz. entropy, energy, and correlation from GLCMs and used these features for AD classification. The VBM method was used in the study [57] to choose the relevant grey density maps from MR images. The authors employed ensemble mode for AD classification and sparse representation for feature extraction. The studies [36] and [158] are part of the ROI-based technique since they use segmented hippocampal [36] and grey matter [158] tissues to extract features for AD detection. The feature extraction capabilities of Visual Geometry Group 16 (VGG 16) Convolutional Neural Network architecture are examined in a recent study [159]. This study gives an accuracy of 86.7% for AD/NC classification on the OASIS dataset. The study [160] utilizes spiking neural networks to classify AD, aiming to enhance diagnostic accuracy by integrating deep learning techniques with biologically inspired neural processing. This study offers 90% accuracy. A graph-based model is used in the study [161] for AD detection utilizing SVM as a classifier. The authors used MR images to extract 3-D Texture characteristics. These textural properties are used in the construction of the graphs. Nodes in graphs represent different brain areas, whereas edges show the relationships between these regions. This study uses graph kernels with SVM to achieve 92% classification accuracy.

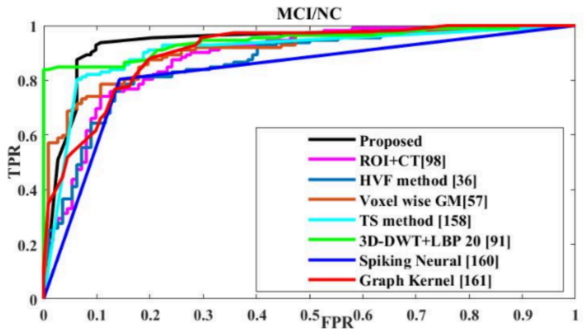
The proposed technique offers a respective increment in accuracy by 2.9%, 3.2%, and 2.5% for classifying AD against NC, AD against MCI, and MCI against NC respectively in comparison with the best performing existing technique. These excellent results validate the efficacy of the wavelet transform-based method for MR image analysis.

Fig. 3.18 shows the ROC plots of the proposed and existing techniques for all three binary classifications.



(a)

(b)



(c)

Fig. 3.18 ROC Plots: Proposed vs Existing methods (a)AD/NC (b)AD/MCI (c) MCI/NC

The proposed approach in Fig. 3.18 shows steep rise of the ROC curve towards the top-left quadrant. This signifies that the proposed method attains high TPR with low FPR value producing good classification results.

3.10. Summary

This chapter presents the direct coefficient-based feature extraction method for AD detection. In the proposed method, wavelet coefficients are obtained at the fifth level of decomposition from multiple MR slices. These coefficients have been used for AD/NC, AD/MCI, and MCI/NC classifications. The detail coefficients from multiple slices provide the high frequency variation in grey matter density and reflect the structural changes in the brain which is essential for early AD detection. The proposed approach shows the effectiveness of multiple slices for early AD detection. This research work also investigates the feature optimization algorithms i.e. GA, ACO, and PSO for feature selection in AD detection. The best results are obtained with DD-DTCWT and GA. Achieving impressive metrics, the method reports 96.9% accuracy, 97.6% sensitivity, and 95.8% specificity for AD/NC; 92.9%, 90%, and 94.1% for AD/MCI; and 90.6%, 91%, and 90% for MCI/NC classification.

The publication related to this chapter

- Neha Garg, M. S. Choudhry and R. M. Bodade, "Alzheimer's Disease Detection using Particle Swarm optimization," 2022 8th International Conference on Signal Processing and

Communication (ICSC), Noida, India, 2022, pp. 271-276, doi: 10.1109/ICSC56524.2022.10009510.

- Neha Garg, M. S. Choudhry and R. M. Bodade, “Alzheimer’s Disease Detection using Double Density Dual-Tree Complex Wavelet Transform based on multiple MR slices.” **Evolving Systems,** SCIE, Impact factor 2.7**. (Communicated)**

Chapter 4

Alzheimer’s Disease Detection using Histogram-Based Feature Extraction Technique

---

This chapter presents an effective method for early AD detection by using a histogram-based feature extraction technique. Two histogram-based techniques are proposed in this chapter. Both techniques decompose the magnetic resonance images into high frequency sub-bands. Further, edge-weighted local binary pattern in the first technique and shifted circular-elliptical local binary pattern in the second technique are applied to the sub-bands for obtaining the histograms of pixel intensity. The proposed method’s performance is demonstrated and verified through comprehensive experimentation using benchmark data and a detailed comparison with current existing approaches.

4.1. Introduction

This chapter presents the histogram-based feature extraction technique. As AD worsens, MRI scan intensity values can fluctuate due to tissue loss, fluid buildup, and other structural atrophies. The pixel intensity histograms can capture these fluctuations by summarizing the frequency of intensity values. In AD, structural and grey matter changes occur in the hippocampus and cortical surface of the brain due to noticeable shrinkage. Histograms can reveal shifts in intensity or texture distributions in these areas, helping to capture AD-specific biomarkers. Histograms are less noise-sensitive than the technique where wavelet coefficients are directly used as features for AD classification. In histograms, wavelet coefficients are aggregated into bins making the representation less sensitive to noise and variations.

Histogram graphically represents of the data distribution. It is created by measuring how many data points fall into each bin after splitting the data range into intervals known as bins. A histogram gives insights into brightness, contrast, and pixel intensity distribution in an image. Histograms are crucial for identifying patterns linked with AD by extracting, analysing, and utilizing texture and intensity changes in MRIs.

4.2. Circular Local Binary Pattern (CLBP)

Local Binary Pattern provides textural details from the images. Circular Local Binary Pattern (CLBP) [162] is formed by neighbouring pixels which fall on a circumference of ‘r’ radius circle with a centre pixel Bc as shown in Fig. 4.1.

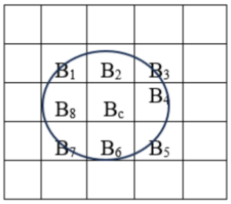


Fig. 4.1 Eight neighbours

The CLBP descriptor for centre pixel Bc(X_c,Y_c) having N number of neighbours can be obtained as in Eqn. 4.1.

$$\text{CLBP}\ (X_c,Y_c) = \sum_{n=1}^{N} \text{Mag}(Y)2^{n-1} \qquad (4.1)$$

here, $Y = I_n(r) - I_c$

The intensity values I_n and I_c for neighbour pixel Bn and centre pixel Bc respectively are used to assign the binary value (0,1) to Mag(Y) using Eqn. 4.2.

$$\text{Mag} = \begin{cases} 1, Y \geq 0 \\ 0, Y < 0 \end{cases} \tag{4.2}$$

The coordinates of neighbour pixel Bn(X_n,Y_n) can be obtained as shown in Eqn. 4.3 and Eqn. 4.4

$$X_n = X_c + r\cos\left({}^{2\pi}/_{N}\right) \tag{4.3}$$

$$Y_n = Y_c + r\sin\left({}^{2\pi}/_{N}\right) \tag{4.4}$$

4.3. Elliptical Local Binary Pattern (ELBP)

ELBP [163] consists of neighbouring pixels lying on an ellipse of r1 horizontal radius and r2 vertical radius around the centre pixel. The ellipse may be oriented in three directions diagonal, horizontal, and vertical. Fig. 4.2 shows the different orientations of ELBP.

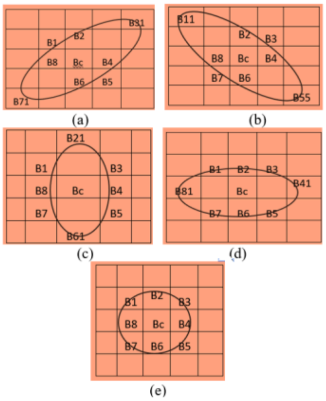


Fig. 4.2 a) Right- Diagonal ELBP b) Left- Diagonal ELBP c) V-ELBP d) H-ELBP e) CLBP

The ELBP descriptor of a pixel Bc(X_c,Y_c) having N number of neighbours present on the ellipse of $r1$ horizontal radius and $r2$ vertical radius can be obtained as shown in Eqn. 4.5.

$$\text{ELBP}\ (X_c,Y_c) = \sum_{\text{n}=1}^{\text{N}} \text{Mag}(Y)2^{\text{n}-1} \tag{4.5}$$

where $Y = I_n(r1, r2) - I_c$

The intensity values I_n and I_c for neighbour pixel Bn and centre pixel Bc respectively are used to assign the binary value (0,1) to Mag(Y) using Eqn. 4.6. The neighbour pixel Bn is at the circumference of ellipse having $r1$ and $r2$ horizontal and vertical radii respectively.

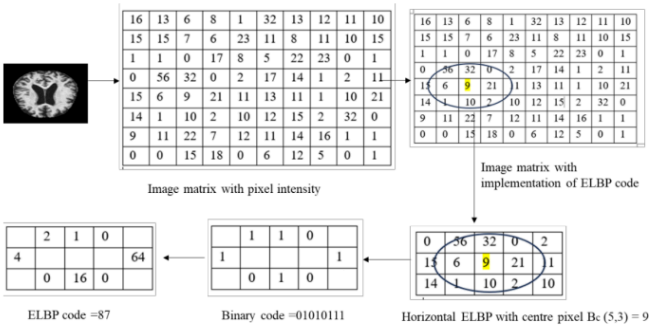
$$\text{Mag} = \begin{cases} 1, Y \geq 0 \\ 0, Y < 0 \end{cases} \tag{4.6}$$

The X_n,Y_n coordinates of Bn neighbour can be obtained through Eqn. 4.7 and Eqn. 4.8. Here N represents the total number of neighbours.

$$X_n = X_c + r1\, cos\left({}^{2\pi}/_{N}\right) \tag{4.7}$$

$$Y_n = Y_c + r2\, sin\left({}^{2\pi}/_{N}\right) \tag{4.8}$$

Implementation of ELBP on MR images requires setting the value of horizontal radius ($r1$), vertical radius ($r2$), and count of neighbourhood pixels (N). The large number of neighbourhood pixels increases the computational complexity, yet it offers more detailed texture patterns. Each pixel in the MR image is considered the centre pixel of an elliptical neighbourhood. Now the intensity of the centre pixel and neighbourhood pixels are compared. If neighbouring pixel intensity is equal to or greater than that of the center pixel, then allocate a binary value of 1; otherwise, allocate 0, as demonstrated in Figure 4.3. Finally, binary values are converted into decimal numbers which represent an ELBP code for the central pixel, encapsulating the local texture pattern.



16	13	6	8	1	32	13	12	11	10
15	15	7	6	23	11	8	11	10	15
1	1	0	17	8	5	22	23	0	1
0	56	32	0	2	17	14	1	2	11
15	6	9	21	11	13	11	1	10	21
14	1	10	2	10	12	15	2	32	0
9	11	22	7	12	11	14	16	1	1
0	0	15	18	0	6	12	5	0	1

Image matrix with pixel intensity

16	13	6	8	1	32	13	12	11	10
15	15	7	6	23	11	8	11	10	15
1	1	0	17	8	5	22	23	0	1
0	56	32	0	2	17	14	1	2	11
15	6	9	21	11	13	11	1	10	21
14	1	10	2	10	12	15	2	32	0
9	11	22	7	12	11	14	16	1	1
0	0	15	18	0	6	12	5	0	1

Image matrix with implementation of ELBP code

0	56	32	0	2
15	6	9	21	11
14	1	10	2	10

Horizontal ELBP with centre pixel Bc (5,3) = 9

	1	1	0	
1				1
	0	1	0	

Binary code =01010111

	2	1	0	
4				64
	0	16	0	

ELBP code =87

Fig. 4.3 Implementation of ELBP on MR scans

Fig. 4.3 indicates the ELBP execution on MR scans. Here eight neighbours are B_1=0, B_2=11, B_3=2, B_4=10, B_5=1, B_6=15, B_7=56, B_8=32 with centre pixel Bc(5,3) = 9. Binary code ‘01010111’ is generated by comparing the intensity of the neighbour pixel and centre pixel.

This binary code yields an ELBP value of 87. Fig. 4.4 shows the ELBP pattern obtained from the sub-band image through DD-DTCWT implementation.

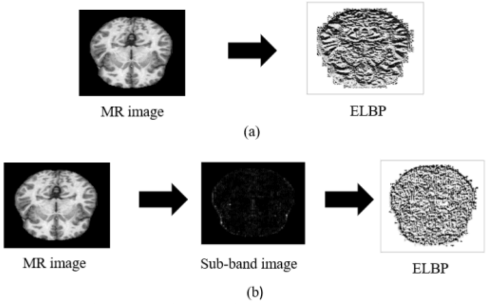


Fig. 4.4 (a) ELBP pattern from MR image (b) ELBP pattern from sub-band of MR image

The ELBP pattern obtained from sub-band images captures the directional texture details.DD-DTCWT provides sixteen sub-bands on the first level of decomposition thus, sixteen patterns will be obtained through ELBP. Histogram features are obtained from these patterns and subsequently utilized for AD classification. In this chapter, edge-weighted LBP and shifted circular-elliptical local descriptors are used in place of ELBP.

4.4. Histogram

A histogram captures the frequency of patterns in an image. It can be constructed in the following way:

1. **Pattern Encoding**: Every pixel in the image is applied with LBP, which creates a distinct binary pattern by comparing the intensities of the central pixel and its neighbours. After that, a decimal value is generated from each binary pattern.
2. **Histogram Creation**: The occurrence of each distinct LBP code throughout the image is counted to create a histogram. All the possible LBP codes are represented by different bins in the histogram, and the value of each bin indicates how frequently that code appears in the image.
3. **Feature Array**: The histogram is a feature array that provides an overview of the image's local texture. This histogram then offers a compact representation of the texture patterns in the image, which can be used for classification.

These LBP histograms can reveal textural differences between healthy and diseased brain tissue, thus helping in early AD detection.

4.5. Dataset

In this work, both datasets, i.e., OASIS and ADNI have been used. The OASIS-1 dataset provided 28 3-D MR images for each group: AD, MCI, and NC. The four centre 2-D MR slices from each 3-D MRI have been extracted with the help of XMedCon software. In this way, 112 2D MR slices have been obtained for each class. Similarly, the same dataset has been prepared from the 3-D MR images downloaded from the ADNI-1 dataset. Fig.4.5 shows the sample of each class of 2-D MR slices obtained from the OASIS-1 dataset.

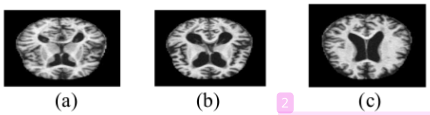

(a) (b) (c)

Fig. 4.5 Demonstrative 2-D MR scan (a) AD (b) MCI (c) NC

4.6. Proposed Methods

In this chapter, two histogram-based feature extraction methods have been proposed. The first method uses Edge weighted Local Binary Pattern (EWLBP), while the second method uses a Wavelet-based Shifted Circular-Elliptical Local Descriptor (WSCELD) for extracting histograms of pixel intensities. EWLBP includes the edge information while encoding the binary patterns, and WSCELD combines the properties of Circular and Elliptical LBP and presents their shifted version. The histograms obtained from EWLBP and WSCELD have been used as features for AD classification with NC and MCI. The second method also investigates the performance of different versions of WSCELD, viz. energy WSCELD, variance WSCELD, mean WSCELD, and median WSCELD. Both techniques yield outstanding performance in classifying MCI against AD and NC, rendering them highly effective for early detection of Alzheimer's disease.

4.6.1. First Proposed Method

This method proposes an EWLBP-based histogram feature extraction technique. Fig. 4.6 illustrates the workflow of the proposed technique through a pipeline model.

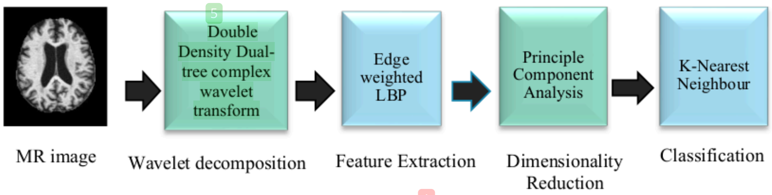


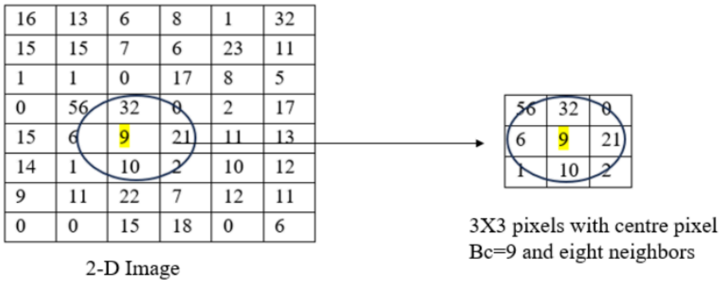
Fig. 4.6 Pipeline model: Proposed method

In the proposed work, DD-DTCWT has been applied to 2-D MR slices to produce sixteen sub-band images. EWLBP is implemented on sub-band images to generate texture patterns with prominent edges. Histograms obtained from these patterns are used as features. For eight neighbour pixels, EWLBP provides 256 histogram bins. EWLBP has been applied on sixteen sub-band images, so the feature vector length of histogram features will be 4096, which is quite high in dimensionality. The study implemented PCA for dimensionality reduction, subsequently using the transformed features to identify individuals with Alzheimer's disease

According to the experimental findings, EWLBP performs better than traditional local binary patterns because EWLBP provides information about local gradient variations. DD-DTCWT helps in extracting features in sixteen directions. The proposed method incorporates edge information into the LBP encoding process and thus provides improved features that enable it to distinguish across several texture patterns with distinct edge structures. This makes the proposed method, able to detect AD at its different stages.

4.6.1.1. Edge Weighted Local Binary Pattern (EWLBP)

EWLBP consists of a gradient operator called Sobel. The Sobel operator extracts the gradient magnitude of the central pixel located in 3×3 neighbourhoods. The gradient magnitude of the central pixels causes the edge information to be included in the binary pattern computation. In pattern computing, greater weight is assigned to the central pixels with high gradient magnitudes since they include a greater amount of edge information. Then, EWLBP is computed using these weighted pixels like the original LBP. Fig. 4.7 indicates the eight neighbouring pixels with centre pixel Bc in a 3 × 3 neighbourhood.



16	13	6	8	1	32
15	15	7	6	23	11
1	1	0	17	8	5
0	56	32	0	2	17
15	6	9	21	11	13
14	1	10	2	10	12
9	11	22	7	12	11
0	0	15	18	0	6

56	32	0
6	9	21
1	10	2

Fig. 4.7 3 X 3 Neighbourhood pixels

The Sobel operator in EWLBP computes the gradient variations. This operator utilizes two kernels to compute the gradient of pixel intensity in horizontal ‘x’ and vertical ‘y’ directions. The magnitude of the gradient is determined by calculating the square root of the sum of the squared gradients of x and y directional pixels. Now, the weights for that 3×3 window are determined by the gradient magnitude of the central pixel. The steps for computing EWLBP on sixteen sub-band images are given in Eqn. 4.9 to Eqn. 4.14.

- Let the input MRI slice be represented as a 2D image:

$$\mathrm{I}\ (x, y) \quad (x, y) \in \mathrm{R}^2 \tag{4.9}$$

  where I (x, y) is the pixel intensity at co-ordinates (x, y).

- First level DD-DTCWT is applied on MR image, which gives one approximate band and sixteen detail sub-bands.

$$DD - DTCWT(1) = \{AL, S_{1,1} \ldots, S_{1,16}\} \tag{4.10}$$

  where, $S_{n,m}$ represents the m^{th} sub-band at the n^{th} decomposition level.

- For each sub-band $S_{n,m}$, the edge weights (gradient magnitude) using the Sobel operator are calculated for each centre pixel Bc(X_c,Y_c) in 3 × 3 neighbourhood.

$$\mathrm{G}(x, y) = \sqrt{\frac{\partial S_{n,m}}{\partial x}^2 + \frac{\partial S_{n,m}}{\partial y}^2} \tag{4.11}$$

- Now, Local Binary Pattern is calculated for a centre pixel Bc(X_c,Y_c) in $S_{n,m}$ with its N number of neighbours (Bn)

$$\mathrm{LBP}\ (X_c, Y_c) = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{Mag}(Y) 2^{\mathrm{n}-1} \tag{4.12}$$

  Where, $Y = I_n(r1, r2) - I_c$

  I_n symbolizes for neighbour pixel intensity (Bn) and I_c symbolizes for centre pixel intensity Bc and the value of Mag(Y) can be assigned 0 and 1 using Eqn. 4.13.

$$\text{Mag} = \begin{cases} 1, Y \geq 0 \\ 0, Y < 0 \end{cases} \tag{4.13}$$

- Now, weight the Local Binary Pattern with gradients

$$\text{EWLBP}(X_c, Y_c) = \sum_{n=1}^{N} \text{G}(\text{x}, \text{y})\text{Mag}(Y).\, 2^{n-1} \tag{4.14}$$

- Now histograms have been computed from sixteen sub-bands using EWLBP. The feature vector size will be 4096, which is quite high in dimensionality. Further PCA has been introduced to reduce the dimensionality. Final features are used for AD/NC, AD/MCI, and MCI/NC classifications.

4.6.1.2. Results of proposed method for binary classifications:AD/NC,AD/MCI,MCI/NC

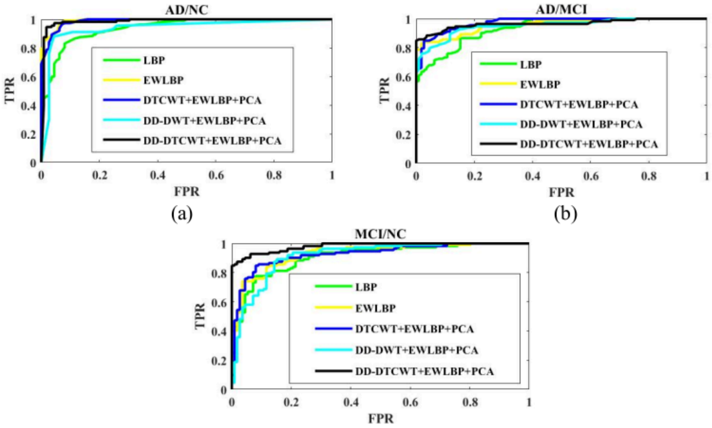
A total of ten models have been implemented in this work for all three binary classifications with KNN classifier. These results are summarized in Table 4.1. The performance of other classifiers like SVM, NB, and DT has been evaluated also but the best results are with KNN and are thus included here. The model DD-DTCWT+EWLBP+PCA has been proposed due to its outstanding performance.

Table 4.1 Different implemented models for binary classifications

S.N.	METHODS	AD/ NC			AD/ MCI			MCI / NC		
		Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
1	LBP	87.1	85	90	82.1	77	88	81.7	81	82
2	DWT+LBP+PCA	88.05	82.3	93.9	83.9	87	81	82.1	81	83
3	DTCWT+LBP+PCA	89.1	88	91	84.7	88	82	83.9	88	80
4	DD-DWT+LBP+PCA	88.05	85.29	90.9	83.5	84	83	84.8	85	85
5	DD-DTCWT+LBP+PCA	89.55	84.84	94.11	84.4	88	81	83.1	82	86
6	EWLBP	92.1	90	94.1	86.2	88	84	85.7	88	84
7	DWT+EWLBP+PCA	93.8	92.4	96.2	87.5	89	86	87.1	91	83
8	DTCWT+EWLBP+PCA	94.28	96	93	89.7	89	90	88.4	85	92
9	DD-DWT+EWLBP+PCA	94	94.9	92.1	87.1	88	87	86.2	89	93
10	DD-DTCWT+EWLBP+PCA	95.5	94	96	93	92	94	91	92.5	89

The first model includes the application of LBP on 2-D MR images. This model utilizes histogram features for classification. The count of histogram features is 256 with eight neighbouring pixels. This model gives a classification accuracy of 87.1%. In the second to fifth models, different wavelets have been used to obtain sub-band images, and subsequently, LBP is employed to sub-band images for extracting histogram features. The second model uses DWT as a wavelet, the third model uses DTCWT as a wavelet, the fourth model uses

DD-DWT as a wavelet, and the fifth model uses DD-DTCWT as a wavelet. The second, third, fourth, and fifth models give 256 × 4, 256 × 6, 256 × 8, and 256 × 16 histogram features, respectively. These features are quite high in dimensionality and are reduced by applying PCA. The classification accuracy from models 2 to 5 is given by 88.05%,89.1%,88.05%, and 89.55%, respectively. In the sixth model, EWLBP has been introduced in place of LBP, and a total of 256 histogram features have been obtained. The classification accuracy with this model is 92.1%. In models 7th to 10th, wavelets have been combined with EWLBP. The 7th model includes DWT and provides 93.8% accuracy. The 8th model includes DTCWT and provides 94.28% accuracy. The 9th model includes DD-DWT and provides 94% accuracy. The 10th model includes DD-DTCWT and provides 95.5% accuracy. These models from 7th to 10th give 256 × 4, 256 × 6, 256 × 8, and 256 × 16 histogram features respectively. The high dimensionality is reduced by using PCA. The best results are with the proposed model, i.e., DD-DTCWT+EWLBP+PCA. The proposed technique provides outstanding results for classifying AD against MCI and NC and thus applicable for early AD detection. Fig. 4.8 shows the ROC plots for binary classifications.



(a) (b)

(c)

Fig. 4.8 ROC Plots of Implemented Models (a) AD/NC (b) AD/MCI (c) MCI/NC

The ROC graph in Fig. 4.8 indicates high TPR and low FPR values, highlighting the outstanding efficacy of the proposed classification model, i.e., DD-DTCWT+EWLBP-PCA.

4.6.1.3. Evaluation of the proposed framework against existing methods

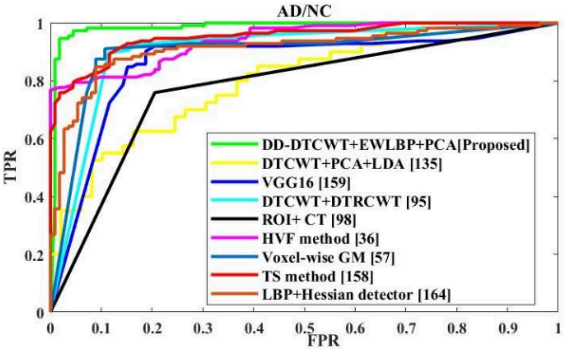
Table 4.2 listed the performance of the proposed method and nine existing methods. The algorithms underwent ten iterations in MATLAB-19 and their average accuracy has been noted using the OASIS-1 dataset. Ten-fold cross-validation technique partitioned dataset into training and testing subsets. The proposed method not only provides good accuracy for AD/NC classification but also gives outstanding performance for AD/MCI and MCI/NC classifications in comparison to existing methods.

Table 4.2 Proposed vs Existing Advanced Methods

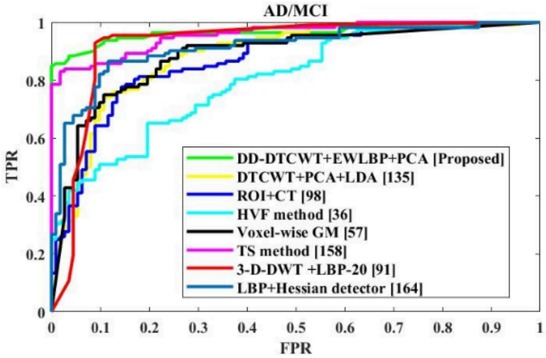
			AD/ NC			**AD/ MCI**			**MCI / NC**	
Methods	**Classifier**	**Acc**	**Sen**	**Spe**	**Acc**	**Sen**	**Spe**	**Acc**	**Sen**	**Spe**
Proposed	KNN	95.5	94	96	93	92	94	91	92.5	89
DT-CWT+PCA+LDA [135] [2018]	ELM	88.2	86.5	84.4	80.2	78.4	76.3	82.1	76.3	85
LBP+ Hessian detector [164] [2021]	CNN	87.5	89	86	86.2	83	89	71.9	72	71
ROI +CT [98] [2022]	SVM	81.1	83.3	79.2	78.1	83	73	80.4	82	79
HVF [36] [2015]	SVM	85.3	77	94	74	75	73	78.1	77	79
Voxel wise GM [57] [2012]	SVM	89.3	87	92	81.7	86	78	83.5	88	79
TS method [158] [2015]	SVM	88.5	89	86	86.2	87	86	85.3	86	85
VGG16 [159] [2022]	NN	86.7	85.2	87.3	-	-	-	-	-	-
DTCWT+DTRCWT [95] [2022]	FNN	90.7	91.2	92.4	-	-	-	-	-	-
3D-DWT+LBP-20 [91] [2020]	SVM	-	-	-	91.5	92	91	89.3	84	95

The existing methods in Table 4.2 belong to the different categories of feature extraction methods. The study[164] used the combination of LBP and hessian detector to extract the textural details at the key points selected by the hessian detector. The proposed method demonstrates enhanced performance over existing approaches [98,36,158] primarily, because it eliminates the need for segmentation, which is a prerequisite in these studies.The proposed method forgoes the image registration step inherent in VBM methods [57]. The proposed method operates with low dimensionality relative to studies [135,95].

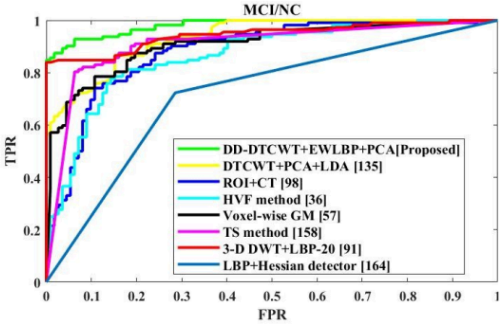
Fig. 4.9 presents the ROC plots of the proposed and current methods.


AD/NC
TPR
FPR
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
DD-DTCWT+EWLBP+PCA[Proposed]
DTCWT+PCA+LDA [135]
VGG16 [159]
DTCWT+DTRCWT [95]
ROI+ CT [98]
HVF method [36]
Voxel-wise GM [57]
TS method [158]
LBP+Hessian detector [164]

(a)


AD/MCI
TPR
FPR
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
DD-DTCWT+EWLBP+PCA [Proposed]
DTCWT+PCA+LDA [135]
ROI+CT [98]
HVF method [36]
Voxel-wise GM [57]
TS method [158]
3-D-DWT +LBP-20 [91]
LBP+Hessian detector [164]

(b)


MCI/NC
TPR
FPR
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
DD-DTCWT+EWLBP+PCA[Proposed]
DTCWT+PCA+LDA [135]
ROI+CT [98]
HVF method [36]
Voxel-wise GM [57]
TS method [158]
3-D DWT+LBP-20 [91]
LBP+Hessian detector [164]

(c)

Fig. 4.9 ROC Plots: Proposed vs Existing methods (a)AD/NC (b)AD/MCI (c) MCI/NC

The proposed approach in Fig. 4.9 shows the ROC curve that climbs sharply towards the upper-left section of the plot. This demonstrates that the proposed method achieves high TPR and low FPR, indicating excellent sensitivity and specificity for early AD detection

4.6.2. Second Proposed Method

In this work, a Wavelet-based shifted Circular-Elliptical Local Descriptor (WSCELD) technique has been proposed for early detection of AD. The proposed WSCELD extracts directional and structural features by combining the strength of DD-DTCWT with the shifted Circular and Elliptical Local Descriptor (SCELD). Fig. 4.10 illustrates the workflow of the proposed technique through a pipeline model.

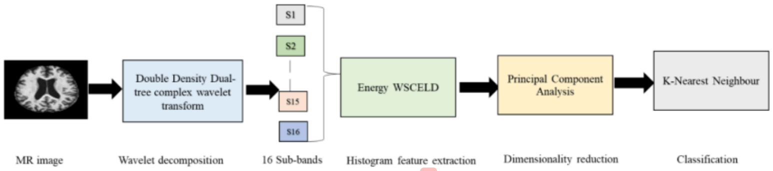


Fig. 4.10 Pipeline model: proposed method

In the proposed work, DD-DTCWT has been applied to 2-D MR images, and sixteen sub-bands have been obtained. Various micro and macro patterns are generated by applying the WSCELD technique on these sub-bands. These patterns provide histogram features that are used for AD classification.

The different versions of WSCELD like Median, Mean, Energy, and Variance have been examined using wavelets like DD-DWT, DTCWT, and DD-DTCWT. The experiments conducted reveal that Energy WSCELD, in conjunction with DD-DTCWT, yields the best results.

The Circular Elliptical Local Descriptor (CELD) possesses the properties of both circular and elliptical LBP's. Elliptical LBP yields anisotropic information, while circular LBP yields isotropic information. Traditionally, to capture both isotropic and anisotropic features in the image, the histograms of CLBP and ELBP had to be concatenated [165]. The feature vector size increases as a result. The proposed method makes use of CELD, which reduces the feature vector size by half while still providing circular and elliptical LBP properties. The proposed method can detect various stages of Alzheimer's disease as it detects atrophies in the brain's multi-structural variations and fluctuating grey matter density as the disease progresses. Grey matter density fluctuation is revealed by detail sub-bands, while imaginary coefficients capture structural information. The subsequent subsections present a comprehensive description of the proposed method.

4.6.2.1. Circular Elliptical Local Descriptor (CELD)

CELD combines the Circular LBP and Elliptical LBP and thus extracts isotropic and anisotropic information from the image. "The feature vector produced by CELD is smaller in size compared to the feature vector obtained from the traditional combination of CLBP and ELBP. CELD provides distinct codes by comparing the intensities of surrounding pixels lying on the periphery of circular LBP, horizontal, vertical, and diagonal ELBPs with the intensity of centre pixel Bc(X_c,Y_c) . The 8-neighbouring points in 3 × 3 CELD are obtained by taking the average of intensities of two pixels (B_6 and B_{61}) at bottom, (B_2 and B_{21}) at top, (B_8 and B_{81}) at left, (B_4 and B_{41}) at right, and four pair of two diagonal pixels (B_1 and B_{11}, B_5 and B_{51}, B_3 and B_{31}, B_7 and B_{71}). This is depicted in Fig. 4.11.

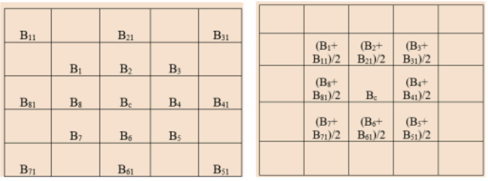


Fig. 4.11 CELD sampling points

In Fig. 4.11, eight neighbours in CELD can be obtained by using Eqns. from (4.15) to (4.18)

$$B_1 = int\ (B_1 + B_{11})/2 \qquad B_2 = int\ (B_2 + B_{21})/2 \tag{4.15}$$

$$B_3 = int\ (B_3 + B_{31})/2 \qquad B_4 = int\ (B_4 + B_{41})/2 \tag{4.16}$$

$$B_5 = int\ (B_5 + B_{51})/2 \qquad B_6 = int\ (B_6 + B_{61})/2 \tag{4.17}$$

$$B_7 = int\ (B_7 + B_{71})/2 \qquad B_8 = int\ (B_8 + B_{81})/2 \tag{4.18}$$

Using the above equations, CELD can be expressed as in Eqn. (4.19)

$$CELD\ \mathrm{N, r1, r2}(X_c,Y_c) = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{Mag}(Y)2^{\mathrm{n}-1} \tag{4.19}$$

where, $Y = I_n(r1, r2) - I_c$

I_n represents the neighbour pixel intensity (Bn) and I_c represents the centre pixel intensity (Bc) and the value of Mag(Y) can be assigned 0 and 1 using Eqn. 4.20.

$$\text{Mag} = \begin{cases} 1, Y \geq 0 \\ 0, Y < 0 \end{cases} \tag{4.20}$$

In Eqn. 4.19, r1 belongs to the radius of CLBP. It is also equal to the horizontal radius of Vertical ELBP and the vertical radius of Horizontal ELBP. The r2 belongs to the horizontal radius of Horizontal ELBP and the vertical radius of Vertical ELBP.

4.6.2.2. Shifted Circular Elliptical Local Descriptor (SCELD)

SCELD is the shifted form of CELD as shown in Fig. 4.11. A shifted CELD provides different micro and macro patterns that show minute and large-scale information from the brain images respectively. Micropatterns capture small-scale textures, focusing on local, fine details within an image. This can assist in identifying cellular or subcellular atrophies, which include alterations in synapse density and neuronal architecture or the presence of microscopic lesions. Macro patterns capture larger, more global texture information, incorporating broader structural or spatial relationships in the image such as ventricles enlargement or thinning of cortical surface. These patterns provide high structural and good textural details resulting in improved classification performance. Figs. 4.12 and 4.13 show the CELD with eight neighbours and the multiple shifted patterns respectively.

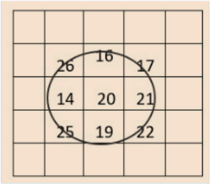
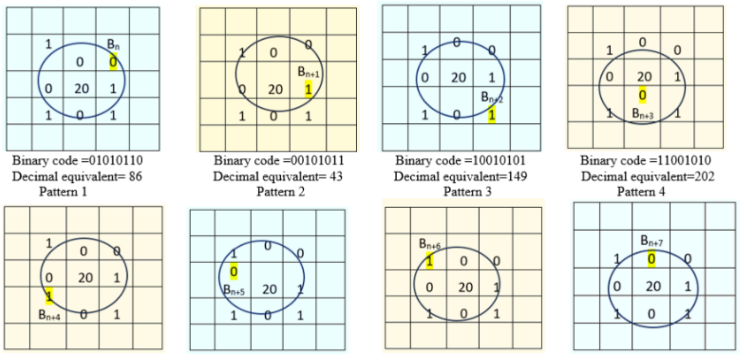


Fig. 4.12 CELD with eight neighbours



Binary code =01010110
Decimal equivalent= 86
Pattern 1

Binary code =00101011
Decimal equivalent= 43
Pattern 2

Binary code =10010101
Decimal equivalent=149
Pattern 3

Binary code =11001010
Decimal equivalent=202
Pattern 4

Binary code =01100101
Decimal equivalent= 101
Pattern 5

Binary code =10110010
Decimal equivalent=178
Pattern 6

Binary code =01011001
Decimal equivalent= 89
Pattern 7

Binary code =10101100
Decimal equivalent= 172
Pattern 8

Fig. 4.13 Shifted eight patterns in shifted CELD

Fig.4.13 shows the shifted version of the CELD shown in Fig.4.12. Shifted CELD is obtained by considering each neighbour pixel (Bn) of centre pixel (Bc) as a first pixel. In shifted CELD, different binary codes are generated by comparing the intensities of the neighbour's pixels with centre pixels. These different binary codes are shown in Fig. 4.13 representing different texture patterns. In the proposed work, SCELD includes the average CELD value of all eight patterns as shown in Eqn. 4.21. This is done to reduce the computation burden as histograms are used as features in the proposed work. The Eqns. from 4.22 to 4.29 help in computing the eight CELD patterns generated by considering each neighbour (Bn) of centre pixel (Bc) as a first pixel.

$$SCELD\ N, r1, r2, \mathrm{Bc}(X_c, Y_c) = \sum_{i=1}^{8} \frac{\mathrm{CELD}_{\mathrm{Pat(i)}}}{8} \tag{4.21}$$

$$\mathrm{CELD}_{\mathrm{Pat(1)}} = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{Mag}(Y1)2^{\mathrm{n}-1} \tag{4.22}$$

$$\mathrm{CELD}_{\mathrm{Pat(2)}} = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{Mag}(Y2)2^{\mathrm{n}-1} \tag{4.23}$$

$$\mathrm{CELD}_{\mathrm{Pat(3)}} = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{Mag}(Y3)2^{\mathrm{n}-1} \tag{4.24}$$

$$\mathrm{CELD}_{\mathrm{Pat(4)}} = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{Mag}(Y4)2^{\mathrm{n}-1} \tag{4.25}$$

$$\mathrm{CELD}_{\mathrm{Pat(5)}} = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{Mag}(Y5)2^{\mathrm{n}-1} \tag{4.26}$$

$$\mathrm{CELD}_{\mathrm{Pat(6)}} = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{Mag}(Y6)2^{\mathrm{n}-1} \tag{4.27}$$

$$\mathrm{CELD}_{\mathrm{Pat(7)}} = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{Mag}(Y7)2^{\mathrm{n}-1} \tag{4.28}$$

$$\mathrm{CELD}_{\mathrm{Pat(8)}} = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{Mag}(Y8)2^{\mathrm{n}-1} \tag{4.29}$$

where $\mathrm{Mag}(Y1) = \begin{cases} 1, Y1 \geq 0 \\ 0, Y1 < 0 \end{cases}$ $\mathrm{Mag}(Y2) = \begin{cases} 1, Y2 \geq 0 \\ 0, Y2 < 0 \end{cases}$ $\mathrm{Mag}(Y3) = \begin{cases} 1, Y3 \geq 0 \\ 0, Y3 < 0 \end{cases}$

$$\mathrm{Mag}(Y4) = \begin{cases} 1, Y4 \geq 0 \\ 0, Y4 < 0 \end{cases} \quad \mathrm{Mag}(Y5) = \begin{cases} 1, Y5 \geq 0 \\ 0, Y5 < 0 \end{cases} \quad \mathrm{Mag}(Y6) = \begin{cases} 1, Y6 \geq 0 \\ 0, Y6 < 0 \end{cases}$$

$$\mathrm{Mag}(Y7) = \begin{cases} 1, Y7 \geq 0 \\ 0, Y7 < 0 \end{cases} \quad \mathrm{Mag}(Y8) = \begin{cases} 1, Y8 \geq 0 \\ 0, Y8 < 0 \end{cases}$$

And $Y1 = I_n(r1, r2) - I_c$ $Y2 = I_{n+1}(r1, r2) - I_c$ $Y3 = I_{n+2}(r1, r2) - I_c$

$$Y4 = I_{n+3}(r1, r2) - I_c \quad Y5 = I_{n+4}(r1, r2) - I_c \quad Y6 = I_{n+5}(r1, r2) - I_c$$

$$Y7 = I_{n+6}(r1, r2) - I_c \quad Y8 = I_{n+7}(r1, r2) - I_c$$

4.6.2.3. Wavelet-based Shifted Circular Elliptical Local Descriptor (WSCELD)

The SCELD with wavelet results in wavelet-based Shifted Circular Elliptical Local Descriptor (WSCELD). WSCELD provides the directional micro, and macro patterns and thus improves the classification results. Wavelet transform decomposes the signal into many sub-bands; each related to a specific direction and scale. These sub-bands provide the directional information. SCELD is employed independently to each sub-band preserving local texture patterns within each frequency component. The proposed algorithm provides a multi-scale representation by combining wavelet sub-bands with SCELD and allows the detection of structural alterations, variations in grey matter density, and texture details at varying levels of scale.

4.6.2.4. Different Versions of WSCELD

In the proposed work, DD-DTCWT is used to obtain sixteen sub-bands, and then SCELD is applied to these sixteen sub-bands to obtain the histograms that are used as features for AD classification. The proposed work investigates the performance of various variants of WSCELD. These versions are Mean WSCELD, Median WSCELD, Variance WSCELD, Energy WSCELD, and Entropy WSCELD as shown in Fig. 4.14.

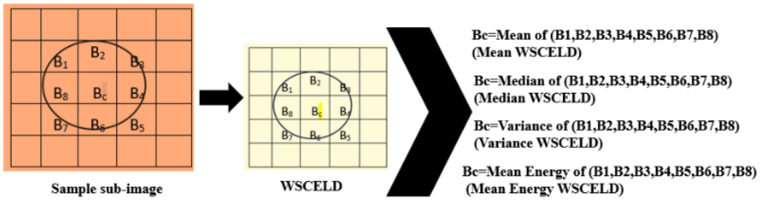


Fig. 4.14 Different Versions of WSCELD

These versions are obtained by substituting the centre pixel value with the mean, median, variance, and energy values of the neighbourhood pixels in Mean WSCELD, Median WSCELD, Variance WSCELD, and Energy WSCELD, respectively. The binary codes are generated by comparing the intensity of surrounding pixels with the intensity of that substituted centre pixel.

The energy version of WSCELD performs better than the other versions since it accounts for the energy input from nearby pixels. Energy is the most promising feature for detecting high frequency variations in grey matter density and structural changes. Edge detection helps in observing structural changes in the brain. There is maximum high frequency energy at edges because of the quick changes in pixel values in the orthogonal direction [166]. Edges give important information about the location and morphological structure of objects in images by illustrating the change between rough and smoother regions. The energy version captures the edge details and thus provides the structural information. The other versions, such as variance, mean, and median, record the average value of nearby pixels, the centre value of neighbouring pixels, and the local contrast information respectively. These versions fail to capture fine structural details, leading to reduced performance.

4.6.2.5. Experimental Results for binary classifications

A total of 24 models have been implemented in this work. All these models have been implemented on MATLAB 19 using the OASIS-1 dataset. Table 4.3 shows the performance of different models using a decision tree classifier.

Table 4.3 Performance of different implemented models using OASIS dataset

S.No	Models	AD/NC			AD/MCI			MCI/NC		
		Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe

1	CLBP	84.2±1.4	85.6±1.2	83.1±1.1	83.2±1.4	84.1±1.2	82.9±1.1	81.5±1.2	80.1±1.1	82.1±0.9
2	H-V-ELBP	87.1±1.3	88.1±1.1	86.5±1.2	85.2±0.8	87.0±1.1	83.0±1.1	83.1±1.2	81.5±1.3	85.5±1.5
3	SCLD	86.3±1.8	87.3±1.4	85.5±1.1	84.7±1.4	82.1±1.2	86.2±1.6	83.5±1.5	84.0±1.3	82.2±1.3
4	S-H-V-ELBP	87.4±1.5	88.9±1.2	86.5±1.1	85.1±1.2	85.1±1.4	85.5±1.2	83.3±1.3	85.1±1.2	81.9±1.1
5	SCELD	89.7±1.6	88.1±1.2	90.7±1.2	86.7±1.3	84.1±0.9	88.0±1.2	84.5±1.7	88.3±1.5	80.0±1.4
6	Mean-SCELD	89.3±1.7	88.5±1.3	91.1±1.1	86.3±1.2	86.2±1.2	85.0±1.3	84.8±1.7	85.5±1.4	83.5±1.5
7	Median-SCELD	86.4±1.1	86.1±1.1	87.7±1.2	82.6±1.1	81.2±0.8	83.9±1.3	81.3±1.1	79.4±1.2	83.5±1.2
8	Energy-SCELD	91.6±1.7	92.2±1.4	89.9±1.2	88.0±1.2	91.1±1.1	85.2±0.9	86.3±1.4	88.1±1.1	84.3±0.9
9	Variance-SCELD	85.4±1.1	87.5±1.2	83.2±1.5	80.5±1.2	80.0±1.4	80.1±1.3	79.2±1.1	78.2±1.2	80.3±1.3
10	DD-DWT+SCELD+PCA	91.2±1.2	92.5±1.3	90.5±1.1	88.3±1.3	85.1±1.2	91.2±0.8	86.1±1.7	88.5±1.2	84.6±1.1
11	DD-DWT+ Mean-SCELD+PCA	91.9±1.7	92.1±1.1	90.1±1.2	87.6±1.3	87.1±1.1	88.1±1.2	85.1±1.3	88.2±1.1	82.2±1.2
12	DD-DWT+ Median-SCELD+PCA	90.5±0.8	92.5±1.1	88.5±1.2	85.1±1.3	87.1±1.2	83.2±0.9	84.0±1.4	85.1±1.3	83.2±1.3
13	DD-DWT+ Energy-SCELD+PCA	93.5±1.5	90.1±1.2	96.5±1.5	90.0±1.3	89.1±0.9	90.1±1.1	87.6±0.9	84.2±1.1	91.1±1.2
14	DD-DWT+ Variance-SCELD+PCA	89.1±1.2	88.2±0.9	91.2±1.1	85.2±0.8	81.1±1.1	89.2±1.2	83.9±1.3	82.1±1.1	84.2±1.2
15	DTCWT+SCELD+PCA	94.0±1.3	93.1±1.1	95.5±1.2	89.5±1.1	88.5±1.2	91.5±1.1	88.7±1.3	89.5±1.1	87.1±1.2
16	DTCWT+ Mean-SCELD+PCA	94.2±1.4	95.9±1.2	93.1±1.1	89.2±1.1	90.1±1.1	88.9±0.9	88.3±1.2	87.1±1.1	89.5±1.2
17	DTCWT+ Median-SCELD+PCA	92.7±1.7	94.2±1.5	91.0±1.2	89.1±1.7	90.1±1.2	89.9±1.1	87.7±1.3	89.5±1.4	85.1±1.1
18	DTCWT+ Energy-SCELD+PCA	95.4±0.8	92.2±1.2	98.5±1.1	91.6±1.3	92.5±1.1	90.5±0.9	90.6±1.3	91.9±1.1	89.5±1.2
19	DTCWT+ Variance-SCELD+ PCA	91.1±1.3	90.1±1.1	89.5±1.2	88.7±1.8	89.5±1.2	87.5±1.1	86.2±1.3	85.5±1.4	87.5±1.1
20	DD-DTCWT+SCELD+PCA	95.2±1.3	96.2±1.4	94.3±1.5	92.1±0.9	93.1±1.3	91.5±1.2	92.0±1.3	91.1±1.2	93.5±1.4
21	DD-DTCWT+ Mean-SCELD+PCA	95.0±1.2	94.4±1.2	96.2±1.1	92.6±0.9	91.6±1.1	90.4±1.1	92.5±1.2	94.2±1.1	90.1±1.2
22	DD-DTCWT+ Median-SCELD+PCA	94.3±1.2	94.1±1.1	93.9±1.1	92.1±0.9	94.2±1.1	90.1±1.1	92.0±1.1	96.5±1.4	88.5±1.2
23	**DD-DTCWT+ Energy-SCELD+PCA [PROPOSED]**	**97.3**±1.6	**97.1**±1.2	**97.2**±1.1	**94.6**±1.1	**96.1**±1.2	**93.1**±1.1	**93.8**±1.4	**92.4**±1.5	**96.2**±1.2
24	DD-DTCWT+ Variance-SCELD+PCA	93.4±1.2	94.1±1.1	92.2±1.2	91.1±0.9	90.2±1.1	92.1±1.5	89.7±1.3	90.5±1.2	88.5±1.2

As presented in Table 4.3, the Circular Local Binary Pattern (CLBP) attains an accuracy of 84.2±1.4% for AD classification with NC, 83.2±1.4% for AD classification with MCI, and 81.5±1.2% for MCI classification with NC.The classification results are enhanced by using Horizontal and Vertical Elliptical Local Binary Patterns (H-V-ELBP) as ELBP provides directional details also.

In the proposed work, a t-test has been performed to check whether the improvement between different models is significant or not. Table 4.4 indicates the p-value for accuracy with different groups of implemented models. The proposed model DD-DTCWT+ Energy-SCELD+ PCA provides exceptional results. The p-value for the combination of proposed and implemented models is less than 0.05 in each case for AD classification. The group of models such as H-V-ELBP/SCLD, H-V-ELBP/S-H-V-ELBP, and SCELD/Mean-SCELD do not show significant improvement. The p-value of these groups is greater than 0.05 and is marked by * in Table 4.4.

Table 4.4 p-values for different combinations of implemented models

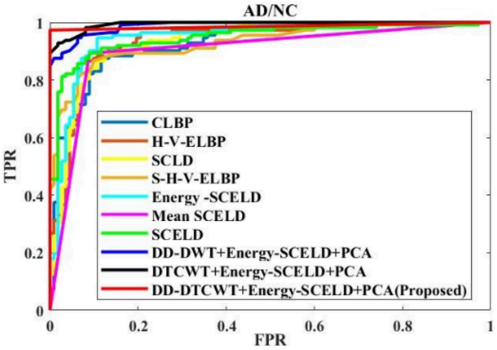
S.N.	**GROUPS**	**AD/NC**	**AD/MCI**	**MCI/NC**
1	CLBP/H-V-ELBP	2.60E-04	2.22E-04	0.015
2	CLBP/SCLD	0.012	0.024	0.005
3	H-V-ELBP/SCLD*	0.340*	0.075*	0.465*
4	H-V-ELBP/S-H-V-ELBP*	0.608*	0.330*	0.919*
5	S-H-V-ELBP/SCELD	0.004	0.025	0.048
6	SCELD/Mean-SCELD*	0.599*	0.848*	0.705*
7	SCELD/Energy-SCELD	0.017	0.042	0.022
8	Mean-SCELD/Energy-SCELD	0.007	0.024	0.049
9	Median-SCELD/Energy-SCELD	7.25E-07	2.37E-08	1.89E-07
10	Variance-SCELD/Energy-SCELD	7.63E-08	1.96E-10	1.22E-09
11	SCELD/DD-DWT-SCELD+PCA	2.74E-02	0.014	0.030
12	DD-DWT-SCELD+PCA/DD-DWT+ Energy-SCELD+ PCA	2.00E-03	0.014	0.007
13	Energy-SCELD/DD-DWT+ Energy- SCELD+PCA	2.03E-02	0.008	0.035
14	DD-DWT+ Mean-SCELD+PCA/DD-DWT+ Energy-SCELD+PCA	4.52E-02	0.008	2.15E-04
15	DD-DWT+ Median-SCELD/DD-DWT+ Energy-SCELD+ PCA	1.29E-04	0.008	7.04E-06
16	DD-DWT+ Variance-SCELD/DD-DWT+ Energy-SCELD+PCA	2.50E-06	8.61E-07	1.78E-06
17	DD-DWT-SCELD+PCA/DTCWT-SCELD+PCA	1.36E-04	0.041	3.49E-04
18	DTCWT-SCELD+PCA/DTCWT+ Energy-SCELD+PCA	0.014	0.002	0.003
19	DD-DWT+ Energy-SCELD+ PCA/DTCWT+ Energy-SCELD+PCA	0.005	0.027	2.12E-05
20	DTCWT+ Mean-SCELD+PCA/DTCWT+ Energy-SCELD+PCA	0.039	0.041	8.08E-04
21	DTCWT+ Median-SCELD+PCA/DTCWT+ Energy-SCELD+PCA	0.0006	0.002	1.04E-04
22	DTCWT+ Variance-SCELD+PCA/DTCWT+ Energy-SCELD+PCA	4.21E-07	8.92E-04	7.86E-07
23	DTCWT-SCELD+PCA/DD-DTCWT-SCELD+PCA	3.94E-02	3.46E-05	1.69E-05
24	DTCWT+ Energy-SCELD+PCA/DD-DTCWT+ Energy-SCELD+PCA	0.007	7.06E-05	5.82E-05
25	DD-DTCWT+ Variance-SCELD+PCA/DD-DTCWT+ Energy-SCELD+PCA	1.75E-05	2.01E-06	3.42E-06
26	DD-DTCWT+ Mean-SCELD+PCA/DD-DTCWT+ Energy-SCELD+PCA	0.003	0.004	0.050
27	DD-DTCWT+ Median-SCELD+PCA/DD-DTCWT+ Energy-SCELD+PCA	3.20E-04	2.28E-04	0.007

From Table 4.4, the p-value for group CLBP/H-V-ELBP is less than 0.05 which indicates the significant improvement in performance by H-V-ELBP over CLBP. Shifted Circular Local Descriptors (SCLD) is the shifted version of CLBP so results get improved due to the inclusion of different shifted patterns. The group CLBP/SCLD has a p-value of 0.012. SCELD include shifted circular and elliptical patterns. Mean-SCELD, Median-SCELD, Energy-SCELD, and Variance SCELD are the different variants of SCELD. The Energy-SCELD is the best performer as the p-value for Energy SCELD vs other versions(groups from 7 to 10) in Table 4.4 is less than 0.05. These models use histograms as features which are 256 in number.In Table 4.3, Models 10 to 14 use DD-DWT with SCELD and test the performance of different versions. These models (10 to 14) provide 8 detail sub-bands on first-level DD-DWT decomposition. The application of SCELD on each detail sub-band results in 256 textural features. Thus, the total count of features is 256 × 8= 2048. PCA is subsequently applied to minimize feature dimensionality. This model provides 91.2±1.2%, 88.3±1.3%, and 86.1±1.7% accuracy for AD/NC, AD/MCI, and MCI/NC classifications respectively. The performance of version DD-DWT+SCELD +Energy is outstanding. The groups from 12 to 16 in Table 4.4 have a p-value of less than 0.05 which indicates the significant improvement by DD-DWT+Energy SCELD. The proposed method has been examined with three wavelets namely DD-DWT, DTCWT, and DD-DTCWT. In Table 4.3,Models 15 to 19 use DTCWT which provides 6 high-frequency sub-bands at first decomposition stage and thus give 256 × 6=1536 histograms as features. Models 20 to 24 use DD-DTCWT which provides 256 × 16=4096 histograms as features. Now, PCA has been applied to reduce dimensionality. The version DD-DTCWT+Energy-SCELD+PCA shows the best performance. This model gives 97.3±1.6%, 94.6±1.1%, and 93.8±1.4% accuracy for AD/NC, AD/MCI, and MCI/NC classifications respectively. The groups 24 to 27 in Table 4.4 show the p-value which is less than 0.05. This shows the significance of improvement by DD-DTCWT+Energy-SCELD+PCA for AD classification at all stages. The proposed work is evaluated with other classifiers as demonstrated in Table 4.5.

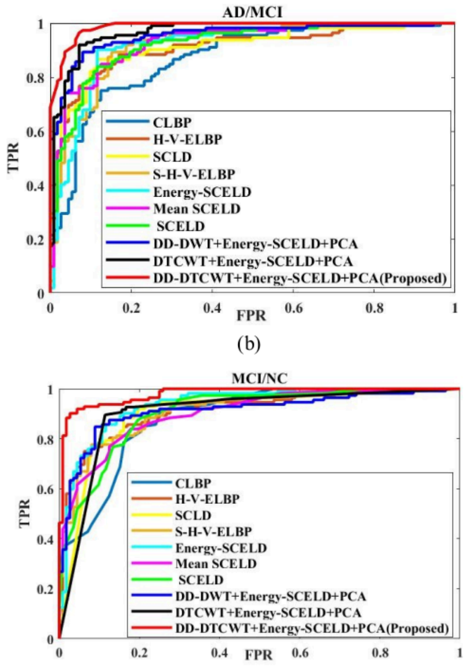
Table 4.5 Classifiers performance for AD/NC classification

Classifiers	Acc	Sen	Spe
DT	97.3	97	97
KNN	94.1	96	92
NB	87.1	88	87
LSVM	95.1	92	98

Table 4.5 indicate the good performance of Decision Tree among other classifiers. Decision Tree can work with missing data and does not require normalisation of features. The ROC plots of different models with decision tree classifier are demonstrated in Fig. 4.15 for all three binary classifications.



(a)



(b)

(c)

Fig. 4.15 ROC plots (a) AD/NC (b) AD/MCI (c) MCI/NC

The proposed method in Fig. 4.15 shows the steep rise in the ROC curve towards the upper left quadrant. This indicates the high TPR and low FPR values showing the significance of the proposed approach for early AD detection.

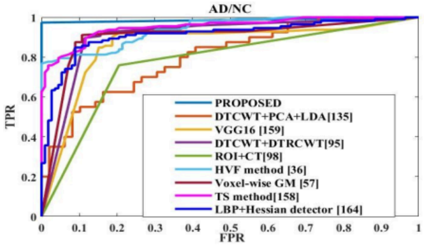
4.6.2.6. Evaluation of the proposed method against existing state-of-art-methods

The proposed method has been benchmarked against nine existing approaches, as detailed in Table 4.6. The existing techniques fall into different categories of the feature extraction. The algorithms have been executed on MATLAB-19 by 10 times and their average accuracy has been noted using OASIS-1 dataset. Ten-fold cross-validation technique have been used to divide the dataset into training and testing dataset. The proposed method not only provides good accuracy for AD/NC classification but also shows outstanding performance for classifying MCI stage against AD and NC in comparison to existing methods.

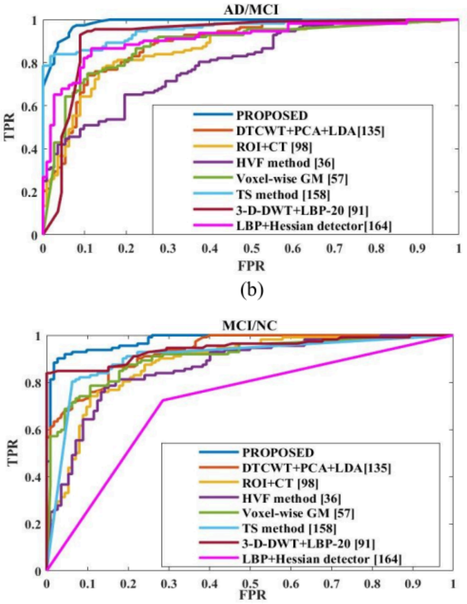
Table 4.6 Proposed vs Existing Methods

Methods	Classifier	AD/NC			AD/MCI			MCI/NC		
		Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
Proposed	**KNN**	**97.3±1.6**	**97.1±1.2**	**97.2±1.1**	**94.6±1.1**	**96.1±1.2**	**93.1±1.1**	**93.8±1.4**	**92.4±1.5**	**96.2±1.2**
DTCWT+PCA+LDA [135] [2018]	ELM	88.2±1.1	86.5±1.1	84.4±1.2	80.2±1.4	78.4±1.2	82.3±1.1	82.0±1.6	79.3±1.5	85.1±1.1
LBP+ Hessian detector [164] [2021]	CNN	87.5±1.3	89.0±1.0	86.0±1.1	86.3±1.7	83.1±1.5	89.5±1.5	71.5±1.8	72.2±1.4	71.0±1.5
ROI+CT [98] [2022]	SVM	81.3±1.1	83.3±0.9	79.2±1.2	78.8±1.4	83.0±1.5	73.0±1.4	80.4±1.2	82.1±1.1	79.1±1.1
HVF [36] [2015]	SVM	85.3±1.3	77.1±1.3	94.1±1.4	74.2±1.8	75.1±1.5	73.0±1.5	78.1±1.0	77.2±1.2	79.5±1.2
Voxel wise GM [57] [2012]	SVM	89.3±1.3	87.0±1.2	91.2±1.1	81.7±1.0	84.0±1.1	78.1±1.2	83.5±1.2	88.2±1.1	79.2±1.1
TS method [158] [2015]	SVM	88.5±1.2	89.5±1.1	87.0±1.2	86.2±1.6	87.1±1.1	86.0±0.9	85.3±1.3	86.2±1.2	85.2±1.1
VGG16 [159] [2022]	NN	86.7±1.2	85.2±1.1	87.3±1.0	83.5±1.6	83.0±1.1	84.1±1.2	82.4±1.0	84.1±1.2	80.4±1.0
DTCWT+DTRCWT. [95] [2022]	FNN	90.4±1.1	91.2±1.1	89.4±1.0	89.7±0.9	88.2±1.2	91.0±1.1	85.3±1.1	86.1±0.9	84.0±1.1
3D-DWT+LBP-20 [91] [2020]	SVM	92.8±1.7	94.0±1.0	90.0±1.1	91.5±1.3	92.1±0.9	91.1±1.5	89.3±1.3	84.2±1.2	95.1±1.0

The study [135] used DTCWT wavelet coefficients at fifth level as features. The rotated version of DTCWT in the study [95] improves the classification accuracy up to 90.4±1.1%, 89.7±0.9%, and 85.3±1.1% for classifying AD against NC, AD against MCI, and MCI against NC. These studies [135,95] suffer from high dimensionality issues. The study [57] used a single atlas for extracting grey matter density maps. The problem with a single atlas is that it may be biased towards a particular class. The studies [36] and [158] used features extracted from the hippocampus [36] and grey matter tissues [158]. These studies [36,158] need a high level of accuracy in segmentation. The recent study [159] is based on a deep learning approach and provides 86.7±1.2%, 83.5±1.6%, and 82.4±1.0% accuracy for classifying AD against NC, AD against MCI, and MCI against NC respectively. The proposed method provides accuracy, sensitivity, and specificity of 97.3±1.6 %, 97.1±1.2%, and 97.2±1.1% for AD/NC,94.6±1.1%, 96.1±1.2%, and 93.1±1.1% for AD/MCI and 93.8±1.4%,92.4±1.5% and 96.2±1.2% for MCI/NC classifications respectively. Fig. 4.16 shows the ROC plots of the proposed and existing algorithms for all three binary classifications.



(a)



(b)

(c)

Fig. 4.16 ROC Plots: Proposed vs Existing methods (a)AD/NC, (b)AD/MCI (c) MCI/NC

The proposed method in Fig. 4.16, shows steep rise in a ROC curve towards the upper-left quadrant. This indicates that the proposed method attains high TPR and low FPR values. Table 4.7 shows the p-values for the different groups having proposed and existing techniques.

Table 4.7 P-values for proposed vs existing methods

Group	Existing Methods/Proposed	AD/NC	AD/MCI	MCI/NC
1	DTCWT+PCA+LDA [135]/Proposed	1.07E-10	1.12E-14	1.52E-12
2	LBP-Hessian detector [164]/Proposed	3.61E-11	1.48E-09	6.20E-16
3	ROI+CT [98]//Proposed	1.28E-14	8.90E-16	2.25E-14
4	HVF [36]/Proposed	1.39E-12	3.64E-15	2.48E-15
5	Voxel wise GM [57]/Proposed	8.14E-10	3.70E-15	1.88E-12
6	TS method [158]/Proposed	1.53E-10	3.82E-10	4.88E-11
7	VGG16 [159]/Proposed	8.22E-12	5.37E-12	3.54E-13
8	DTCWT+DTRCWT [95]/Proposed	6.46E-09	2.22E-08	3.56E-11
9	3D-DWT+LBP-20 [91]/Proposed	1.73E-05	4.95E-05	9.58E-07

Table 4.7 shows the significant improvement made by the proposed method over existing algorithms. The p-value in each group for all three binary classifications is less than 0.05

which shows the significant improvement of proposed method over other existing methods. The proposed work is useful in early AD detection due to its exceptional results in all stages.

4.7. Summary

This chapter introduces two methods for early AD detection using a histogram-based feature extraction technique. The first approach focuses on capturing edge details using edge-weighted LBP. Edge weighted LBP consists of the Sobel operator which is a gradient-based algorithm to capture the edge information. In this method, the centre pixel of 3×3 neighbourhood which has a high value of gradient magnitude is given more weightage while calculating binary patterns. Greater gradient magnitude values indicate a higher amount of edge information captured at the center pixel. This method provides accuracy, sensitivity, and specificity of 95.5% and 94% and 96% for AD/NC ,93%,92% and 94% for AD/MCI and 91% ,92.5% and 89 % for MCI/NC classifications respectively.

The second approach focuses on capturing micro and macro structural details through different directional textural patterns. This method proposes a Wavelet-based Shifted Circular Elliptical Local Descriptor (WSCELD) which is the combination of circular and elliptical local descriptors. WSCELD can extract isotropic and anisotropic structural information in terms of multiple micro and macro textural patterns. This approach can detect the subtle structural changes in the brain which is essential for early AD detection. This approach also investigates the performance of various versions of WSCELD and suggests Energy WSCELD as an efficient feature extraction technique for early AD detection. Energy WSCELD has been used for extracting histogram features and provides exceptional results for early AD detection.This method provides accuracy, sensitivity, and specificity of 97.3±1.6 %, 97.1±1.2% and 97.2%±1.1 for AD/NC ,94.6±1.1%, 96.1±1.2% and 93.1±1.1% for AD/MCI and 93.8±1.4% ,92.4±1.5% and 96.2±1.2% for MCI/NC classifications respectively.

The publication related to this chapter

- Neha Garg, M. S. Choudhry, & R. M. Bodade ,(2024). Alzheimer's disease classification using wavelet-based image features. **Traitement du Signal**, Vol. 41, No. 4, pp. 1899-1910. https://doi.org/10.18280/ts.410420 **IF. 1.2**
- Neha Garg, M. S. Choudhry, and R. M. Bodade. "Wavelet-based Alzheimer's Disease Detection using Edge-Weighted Local Binary Pattern." 2023 9th International Conference

on Signal Processing and Communication (ICSC). IEEE, 2023. [10.1109/ICSC60394.2023.10441017](https://doi.org/10.1109/ICSC60394.2023.10441017)

- Neha Garg, M. S. Choudhry and R. M. Bodade, “Early Alzheimer’s Disease Detection using wavelets with Edge-Weighted Local Binary Pattern." , **IETE journal of research**, SCIE, Impact factor 1.3 (**Communicated)**

Chapter 5

Alzheimer’s Disease Detection using Statistical Feature Extraction Technique

This chapter focuses on the performance of statistical features for the classification of Alzheimer’s Disease from Normal Controls and Mild Cognitive Impairment. Statistical features have been obtained in three different ways which directs to three methods proposed in this chapter. These three methods include, shifted elliptical local binary descriptor-based feature extraction technique, grey level co-occurrence matrix-based feature extraction technique, and shifted circular elliptical local descriptor-based feature extraction technique. These methods address the problem of scalability also. The proposed methods give good classification results on diverse populations and show exceptional performance on binary as well as on multi-class classification. The proposed method is demonstrated and verified through comprehensive experimentation using benchmark data and a detailed comparison with current existing approaches.

5.1. Introduction

This chapter is based on statistical feature extraction techniques. Statistical features are quantitative measurements that are obtained from the spatial relationships between the pixels and are utilized to characterize texture, contrast, randomness, and other aspects of the image. First-order statistical attributes are derived directly from the pixel intensity values without taking spatial relationships into account. Examples include variance, mean, skewness, entropy and standard deviation. Second order statistical measures consider the spatial relationship between the pixels such as their separation and directional alignment.

Alzheimer's disease frequently causes tissue deterioration, which modifies MR imaging intensity values. Statistical features like entropy, energy, mean, variance, etc can capture these changes in the MR images. Mean detects general atrophy in afflicted areas by reflecting average intensity. Variance captures heterogeneity, which is frequently increased by structural degradation. Skewness and Kurtosis draw attention to asymmetry and sharpness of distribution, which are susceptible to certain pathological changes. Entropy is the measure of unpredictability that is frequently elevated in areas that have degraded. Energy represents uniformity, useful for identifying regions with consistent structural loss.

Statistical features [167] are better than histogram-based features as they provide a compact representation (e.g., mean, variance, skewness, entropy) that significantly reduces dimensionality while preserving key information. Statistical features are less sensitive to noise as they focus on the overall data distribution rather than individual bins. Statistical features are interpretable as histograms might have dozens or hundreds of bins, making it challenging to analyze directly while statistical features condense this into a small number of descriptors (e.g., mean, standard deviation, entropy).

The techniques proposed in this chapter give exceptionally good results on two datasets for binary as well as for multiclass classification. These techniques can address the issue of scalability as these techniques can classify diverse and large populations with good accuracy. An algorithm is called scalable if it can handle increased demand without compromising performance.

5.2. Dataset

In this chapter, three methods are proposed. The OASIS-1 and ADNI-1 datasets have been used to implement all the three methods. The four central 2-D MR slices have been extracted from each 3-D MRI volume using XMedCon software. In this way, 112 2D MR slices have been obtained for each class. Similarly same dataset has been prepared for ADNI-1 dataset. Fig.5.1 shows the sample of each class of 2-D MR slices obtained from OASIS-1 dataset.

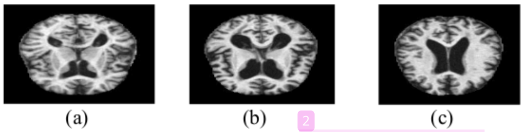
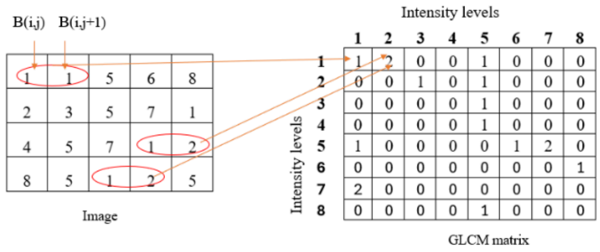


Fig. 5.1 Demonstrative 2-D MR scan (a) AD (b) MCI (c) NC

5.3. Grey-Level Co-occurrence Matrix (GLCM)

The Grey-Level Co-occurrence Matrix (GLCM) [168] considers the spatial relationships between the pixels to analyze textures. The GLCM quantifies the occurrence of pixels pairs with grey level intensity and spatial relationship in an image. Statistical texture features are then computed from this matrix to characterize the image's surface patterns. GLCM of an image can be calculated in the following way as depicted in Fig. 5.2.



1	1	5	6	8
2	3	5	7	1
4	5	7	1	2
8	5	1	2	5

	1	2	3	4	5	6	7	8
1	1	2	0	0	1	0	0	0
2	0	0	1	0	1	0	0	0
3	0	0	0	0	1	0	0	0
4	0	0	0	0	1	0	0	0
5	1	0	0	0	0	1	2	0
6	0	0	0	0	0	0	0	1
7	2	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0

Fig. 5.2 GLCM

In Fig.5.2, two adjacent pixels $B_{(i,j)}$ and $B_{(i,j+1)}$ having intensity (1,1) are coming one time in the horizontal direction in the whole image, thus in the GLCM matrix, 1 is written at (1,1) pixel location. Here i and j symbolize for rows' count and columns' count respectively. In Fig. 5.2, the distance between adjacent pixels is taken as d=1 and the directions are taken as horizontal i.e. k=0°. The adjacency between the pixels can be observed at four angles 0°,45°,90°, and 135° as demonstrated in Fig. 5.3.

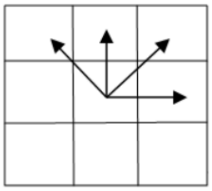


Fig. 5.3 Different orientations in GLCM

Fig.5.3 indicates the different directions in which spatial relationships between the pixels can be calculated. Multiple orientations provide comprehensive texture representation.

5.4. Proposed Methods

This chapter includes three methods based on statistical feature extraction techniques. The first method employs a shifted version of ELBP on sixteen sub-band images of 2-D slices. The second method uses the GLCM in place of shifted ELBP. The third method uses Shifted Circular-Elliptical Local Descriptors (SCELD) in place of GLCM. The shifted ELBP and SCELD-based techniques provide first-order statistical features while the GLCM-based technique provides second-order statistical features. These features are used for AD classification with MCI and NC.

5.4.1. First Method

In this work, a Wavelet-based Shifted Elliptical Local Binary Pattern (WSELBP) feature extraction technique is proposed. This technique provides directional features from multiple micro and macro patterns. These patterns are generated by integrating complex wavelets with shifted elliptical local binary patterns. Fig. 5.4 illustrates the workflow of the proposed technique through a pipeline model.

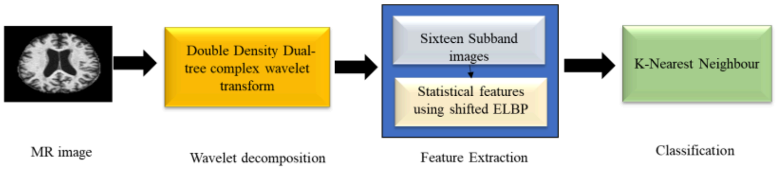


Fig. 5.4 Pipeline model: Proposed method

In Fig.5.4, DD-DTCWT is applied on 2-D MR images which decompose images into sixteen sub-bands. In the next stage, shifted ELBP is applied on the sixteen sub-bands to capture directional structural details in terms of multiple micro and macro patterns. The eight-neighbour shifted ELBP provides 8 patterns from one sub-band. So, sixteen sub-bands will provide 128 (16×8) patterns. The different statistical features like energy, entropy, mean, skewness, kurtosis, and variance have been obtained from these multiple micro and macro patterns and have been used for early AD diagnosis. Shifted ELBP enhances the classification accuracy by concatenating the features obtained from different patterns. Experiments conducted show the outstanding performance of the shifted ELBP over conventional LBP. Fig.5.5 shows the complete process of the proposed work.

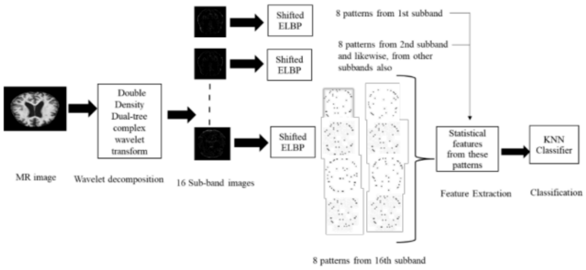


Fig. 5.5 Detailed Feature Extraction Process

In Fig. 5.5, statistical features have been obtained from sixteen sub-bands through shifted ELBP. These features are defined as

1. **Energy:** The signal's strength and properties are represented by its energy. Energy is a measure of how much an image, or a specific area varies in intensity. It is represented as

$$\text{Energy} = \sum_{i=1}^{N} |s_x(i)^2| \tag{5.1}$$

here $s_x(i)$ represents the i^{th} element of shifted ELBP and N represents the total count of elements.

2. **Mean:** The mean shows the middle point of the data distribution. It can be applied to characterize an image's distribution of intensity. It can be represented as

$$\text{Mean} = \frac{1}{N}\sum_{i=1}^{N} s_x(i) \tag{5.2}$$

3. **Variance:** The data's deviation from its mean value is represented by variance. It captures the variety of information and highlights the data's unpredictability. It can be represented as

$$\text{Variance} = \frac{1}{N}\sum_{i=1}^{N}(s_x(i) - \text{Mean})^2 \tag{5.3}$$

4. **Entropy:** Entropy measures the degree of disorder in an image. When the texture of an image is not uniform, the entropy is large. Entropy is calculated as

$$\text{Entropy} = -\sum_{i=1}^{N} \frac{s_x(i)}{\sqrt{\sum_{i=1}^{N} s_x(i)^2}} \ln\left(\frac{s_x(i)}{\sqrt{\sum_{i=1}^{N} s_x(i)^2}}\right) \tag{5.4}$$

5. **Kurtosis**: A statistical measure called kurtosis is used to describe how the tails of a distribution resemble a normal distribution. It helps to determine if the tails are heavier or lighter than the distribution. High kurtosis indicates sharp edges, minute details, or abrupt transitions (sharper peaks), while low kurtosis indicates smoother, more uniform regions. It can be represented as

$$\text{Kurtosis} = \frac{\frac{1}{N}\sum_{i=1}^{N}(s_x(i) - \text{Mean})^4}{Variance^2} \tag{5.5}$$

6. **Skewness**: A statistical measure called skewness shows how asymmetrical a distribution is around its mean. It offers details on the asymmetry of pixel intensity values, which is a useful feature for classification issues. It can be represented as

$$\text{Skewness} = \frac{N}{(N-1)(N-2)}\sum_{i=1}^{N}\left(\frac{s_x(i) - Mean}{\sqrt{Variance}}\right)^3 \tag{5.6}$$

5.4.1.1. Shifted ELBP

The Shifted ELBP is the modified version of normal ELBP. Shifted ELBP provides distinct binary codes that correspond to various texture patterns by treating each neighbour of the centre pixel as the first pixel of the binary pattern. Multiple micro and macro patterns can be captured in an image due to the shifted and anisotropic nature of shifted ELBP. Macro Patterns capture large-scale features like changes in shapes and sizes of the hippocampus and cortical surface. Micro Patterns capture the atrophies at a cellular or subcellular level. The Eqns. from (5.7) to (5.14) show the multiple patterns obtained from the shifted version of ELBP.

$$\text{ELBP Pattern 1} = \textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{Mag}(Y1)2^{\mathrm{n}-1} \tag{5.7}$$

$$\text{ELBP Pattern 2} = \textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{Mag}(Y2)2^{\mathrm{n}-1} \tag{5.8}$$

$$\text{ELBP Pattern 3} = \textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{Mag}(Y3)2^{\mathrm{n}-1} \tag{5.9}$$

$$\text{ELBP Pattern 4} = \textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{Mag}(Y4)2^{\mathrm{n}-1} \tag{5.10}$$

$$\text{ELBP Pattern 5} = \textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{Mag}(Y5)2^{\mathrm{n}-1} \tag{5.11}$$

$$\text{ELBP Pattern 6} = \textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{Mag}(Y6)2^{\mathrm{n}-1} \tag{5.12}$$

$$\text{ELBP Pattern 7} = \textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{Mag}(Y7)2^{\mathrm{n}-1} \tag{5.13}$$

$$\text{ELBP Pattern 8} = \textstyle\sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{Mag}(Y8)2^{\mathrm{n}-1} \tag{5.14}$$

where $\mathrm{Mag}(Y1) = \begin{cases}1, Y1 \geq 0\\ 0, Y1 < 0\end{cases}$ $\mathrm{Mag}(Y2) = \begin{cases}1, Y2 \geq 0\\ 0, Y2 < 0\end{cases}$ $\mathrm{Mag}(Y3) = \begin{cases}1, Y3 \geq 0\\ 0, Y3 < 0\end{cases}$

$\mathrm{Mag}(Y4) = \begin{cases}1, Y4 \geq 0\\ 0, Y4 < 0\end{cases}$ $\mathrm{Mag}(Y5) = \begin{cases}1, Y5 \geq 0\\ 0, Y5 < 0\end{cases}$ $\mathrm{Mag}(Y6) = \begin{cases}1, Y6 \geq 0\\ 0, Y6 < 0\end{cases}$

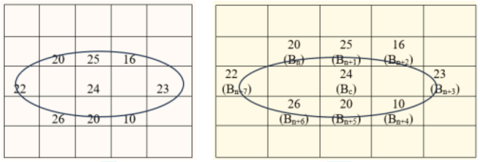
$\mathrm{Mag}(Y7) = \begin{cases}1, Y7 \geq 0\\ 0, Y7 < 0\end{cases}$ $\mathrm{Mag}(Y8) = \begin{cases}1, Y8 \geq 0\\ 0, Y8 < 0\end{cases}$

And $Y1 = I_n(r1, r2) - I_c$ $Y2 = I_{n+1}(r1, r2) - I_c$ $Y3 = I_{n+2}(r1, r2) - I_c$

$Y4 = I_{n+3}(r1, r2) - I_c$ $Y5 = I_{n+4}(r1, r2) - I_c$ $Y6 = I_{n+5}(r1, r2) - I_c$

$Y7 = I_{n+6}(r1, r2) - I_c$ $Y8 = I_{n+7}(r1, r2) - I_c$

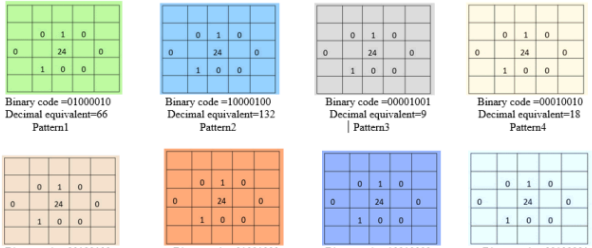
$I_n, I_{n+1}, I_{n+2}, I_{n+3}, I_{n+4}, I_{n+5}, I_{n+6}, I_{n+7}$ are the intensities of eight neighbours $\mathrm{B_n, B_{n+1}, B_{n+2}, B_{n+3}, B_{n+4}, B_{n+5}, B_{n+6}, B_{n+7}}$ respectively around the centre pixel $\mathrm{B_C}$ as depicted in Fig 5.6. A sample sub-image of dimensions 5×5 is shown in Fig 5.6 (a). Neighbour locations are demonstrated in Fig. 5.6 (b).



(a) (b)

Fig. 5.6 (a) Sample sub image (b) Sub image with neighbours location

Fig.5.7 shows the shifted version of the horizontal ELBP shown in Fig.5.6(b). Shifted ELBP is obtained by considering each neighbour pixel (Bn) of centre pixel (Bc) as a first pixel. In shifted ELBP, different binary codes are generated by comparing the intensities of neighbour's pixels with centre pixels.



Binary code =01000010
Decimal equivalent=66
Pattern1

Binary code =10000100
Decimal equivalent=132
Pattern2

Binary code =00001001
Decimal equivalent=9
Pattern3

Binary code =00010010
Decimal equivalent=18
Pattern4

Binary code =00100100
Decimal equivalent=36
Pattern5

Binary code =01001000
Decimal equivalent=72
Pattern6

Binary code =10010000
Decimal equivalent=144
Pattern7

Binary code =00100001
Decimal equivalent=33
Pattern8

Fig. 5.7 Different patterns from horizontal ELBP

The different binary codes are shown in Fig. 5.7 representing different texture patterns.

5.4.1.2. Experimental Results: Binary Classification

The proposed work evaluates the performance of nine combinations. These combinations are 1) pattern 1 statistical features across sixteen sub-bands, 2) pattern 2 statistical features across sixteen sub-bands, 3) pattern 3 statistical features across sixteen sub-bands, 4) pattern 4 statistical features across sixteen sub-bands, 5) pattern 5 statistical features across sixteen sub-bands, 6) pattern 6 statistical features across sixteen sub-bands, 7) pattern 7 statistical features across sixteen sub-bands, 8) pattern 8 statistical features across sixteen sub-bands, and 9) all 8 patterns statistical features across sixteen sub-bands. The ninth combination i.e. statistical features obtained from all 8 patterns of sixteen sub-bands gives the maximum results. Since six statistical features (mean, variance, energy, entropy, kurtosis, skewness) have been obtained from each sub-band therefore the combinations from 1 to 8 provide 96 statistical features and 9th combination provide 768 statistical features.

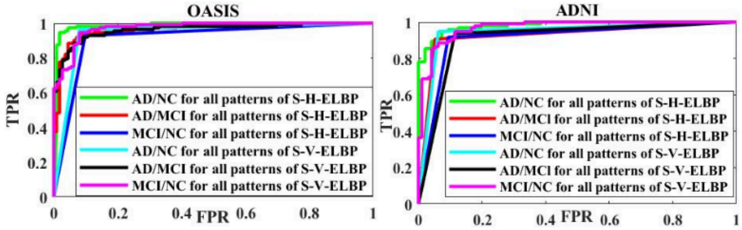
Table 5.1 illustrates the performance of different local patterns obtained from shifted horizontal and vertical ELBPs for three binary classifications. The execution of the proposed work is performed on MATLAB-19. The dataset is divided into training and testing parts using the ten-fold cross-validation approach. The classification is done using different classifiers but results with K-Nearest Neighbour (KNN) classifier are outstanding and have been listed in Table 5.1.

Table 5.1 Analysis with different patterns

		OASIS DATASET			**ADNI DATASET**		
		AD/NC	**AD/MCI**	**MCI/NC**	**AD/NC**	**AD/MCI**	**MCI/NC**
Approach	**Patterns considered**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**
DD-DTCWT+S-H-ELBP	Pattern 1 across 16 sub bands	91.5/93/90	88.4/85/92	87.5/88/86	91.1/94/89	88/86/90	87.5/92/83
	Pattern 2 across 16 sub bands	90.2/89/91	84.8/86/84	85.3/86/85	89.1/86/92	83.1/82/84	84.6/84/86
	Pattern 3 across 16 sub bands	85.7/84/88	80.8/81/80	85.7/88/84	84.4/86/82	80.2/78/82	85.9/88/84
	Pattern 4 across 16 sub bands	89.7/88/91	87.1/83/91	84.8/87/83	87.5/89/86	87.5/88/88	83.9/84/83
	Pattern 5 across 16 sub bands	82.1/81/83	83.9/82/86	86.2/89/93	81.8/83/80	82.6/87/79	86.2/87/86
	Pattern 6 across 16 sub bands	85.3/82/88	82.6/78/88	87.1/91/83	85.4/84/86	82.1/81/83	87/90/84
	Pattern 7 across 16 sub bands	86.2/88/84	81.6/77/88	84.8/88/82	85.3/88/83	80.4/82/79	83/81/85
	Pattern 8 across 16 sub bands	90.6/92/89	80.4/75/86	82.6/87/79	89.1/91/88	80.7/76/85	82.8/83/82
	All 8 patterns across 16 sub bands	**95.1/95/96**	**92/90/94**	**91.5/93/90**	**94.8/94/95**	**91.9/90/92**	**91.1/92/91**
DD-DTCWT+S-V-ELBP	Pattern 1 across 16 sub bands	91.1/92/90	86.2/83/89	87.5/89/86	89.6/91/89	85.7/85/87	86.5/88/85
	Pattern 2 across 16 sub bands	89.3/88/91	83.5/84/83	86.2/88/85	89.1/93/85	83.9/84/83	86/85/87
	Pattern 3 across 16 sub bands	85.3/77/94	79.9/79/80	83.5/85/82	84.6/85/85	80.4/82/79	83/80/86
	Pattern 4 across 16 sub bands	89.3/87/92	87.5/84/91	87.1/89/85	87.5/88/88	85.7/85/86	85.4/84/86
	Pattern 5 across 16 sub bands	83.5/89/78	85.3/81/89	85.7/85/87	83.9/88/80	84.8/83/86	84.8/88/82
	Pattern 6 across 16 sub bands	84.4/83/86	82.1/81/83	83.9/87/81	83.9/88/80	81.8/83/80	82.8/83/82
	Pattern 7 across 16 sub bands	85.3/85/86	84.4/81/88	85.7/88/89	85.3/82/88	83.5/84/83	85.7/79/92
	Pattern 8 across 16 sub bands	89.3/89/89	79.5/74/85	82.6/88/77	87/88/86	79.7/74/85	81.4/83/80
	All 8 patterns across 16 sub bands	**94.2/97/91**	**91.1/89/93**	**90.2/94/87**	**93.8/95/93**	**91/94/90**	**90.1/92/89**
DD-DTCWT+S-H-V-ELBP	All 8 patterns through 16 sub bands	92.9/90/96	91.5/92/91	91.5/90/87	91.1/90/93	91.1/94/89	91.1/92/90

Table 5.1 depicts the pattern-wise evaluation of the proposed work. The pattern 1 gives the maximum results among all eight patterns with shifted horizontal and vertical ELBPs both. The results are further enhanced by combining the features of all eight patterns from all sixteen sub-bands. This performance enhancement can be seen as 3.9% and 3.4% with Shifted Horizontal Elliptical Local Binary Pattern (S-H-ELBP) and Shifted Vertical Elliptical Local

Binary Pattern (S-V-ELBP) respectively for AD/ NC classification using OASIS dataset. These figures are 4.07% and 5.6% respectively for AD/MCI classification and 4.5% and 3.5% respectively for MCI/NC classification using OASIS dataset. The ADNI dataset provides enhancement accuracy of 4.0% and 4.6% for AD/NC classification, 4.4% and 6.1% for AD/MCI classification,4.1% and 4.1% for MCI/NC classification with S-H-ELBP and S-V-ELBP respectively. These eight patterns retain micro and macro structural details. They provide directional statistical features since they are extracted by applying S- H-ELBP or S-V-ELBP on 16 directional sub-bands. The improved classification performance is due to the micro and macro structural details available in the various patterns. The ROC plots of different combination of patterns implemented in the Table 5.1 for all three binary classifications is shown Fig. 5.8.



(a) (b)

Fig. 5.8 ROC Plots (a) OASIS (b) ADNI

The ROC plots in Fig. 5.8 show the good performance of models using S-H-ELBP for all three binary classifications with both datasets.

5.4.1.3. Performance of the proposed method with different wavelets

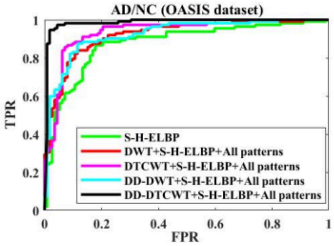
The proposed method has been examined using other three wavelets also. These are DWT, DTCWT and DD-DWT. Table 5.2 represents the classification analysis of the proposed method with DWT, DTCWT, DD-DWT and DD-DTCWT.

Table 5.2 Performance of different wavelets using statistical features

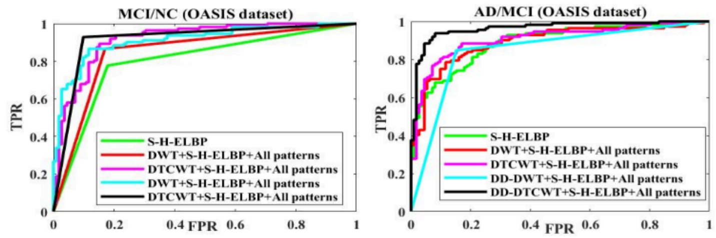
			OASIS DATASET			**ADNI DATASET**		
			AD/NC	**AD/MCI**	**MCI/NC**	**AD/NC**	**AD/MCI**	**MCI/NC**
Wavelet	**ELBP type**	**Patterns considered**	**Acc/ Sen/ Spe**	**Acc/ Sen/ Spe**	**Acc/ Sen/ Spe**	**Acc/ Sen/ Spe**	**Acc/ Sen/ Spe**	**Acc/ Sen/ Spe**

_	S-H-ELBP	All 8 patterns	82.1/81/83	79/77/81	79.9/78/82	81.8/83/80	79.2/79/79	75.5/75/75
DWT		All 8 patterns through 4 subbands	85.7/84/88	82.6/78/88	84.8/87/83	84.4/85/84	82.3/81/83	83.5/88/79
DTCWT		All 8 patterns through 6 subbands	87.5/88/87	84.8/85/85	88.4/85/92	87/90/84	85.4/85/86	87.5/88/87
DD-DWT		All 8 patterns through 8 subbands	86.2/88/85	83.9/82/86	86.2/89/93	85.9/85/86	83.9/84/83	85.4/85/86
DD-DTCWT		All 8 patterns through 16 subbands	**95.1/95/96**	**92/90/94**	**91.5/93/90**	**94.8/94/95**	**91.9/90/92**	**91.1/92/91**
_	S-V-ELBP	All 8 patterns	78.1/77/79	75.4/76/75	76.3/71/82	77.7/76/79	74.1/75/73	75.4/77/74
DWT		All 8 patterns through 4 subbands	81.7/86/78	79/77/81	81.7/81/82	80.4/82/79	78.1/83/73	81.3/79/84
DTCWT		All 8 patterns through 6 subbands	83.5/88/79	82.5/81.1/83	85.3/84/87	83/80/86	82.1/78/87	85.1/87/83
DD-DWT		All 8 patterns through 8 subbands	82.1/81/83	80.4/82/79	84.8/85/85	82.6/88/77	79.5/85/74	83.9/88/80
DD-DTCWT		All 8 patterns through 16 subbands	**94.2/97/91**	**91.1/89/93**	**90.2/94/87**	**93.8/95/93**	**91/94/90**	**90.1/92/89**
_	S-H-V-ELBP	All 8 patterns	77.2/83/71	73.7/76/71	71.9/72/71	75.9/73/79	71.4/77/66	70,1/67/73
DWT		All 8 patterns through 4 subbands	78.9/76/80.1	78.1/76/72	74/75/73	77.7/76/79	75.9/73/79	73.7/76/71
DTCWT		All 8 patterns through 6 subbands	82.4/84.1/80.4	80.8/81/80	83.9/82/86	81.7/83/80	81.7/81/82	84.6/82//87
DD-DWT		All 8 patterns through 8 subbands	81.5/83.7/80	79/77/81	82.1/81/83	80.4/80/80	78.1/75/81	80.8/83/79
DD-DTCWT		All 8 patterns through 16 subbands	**92.9/90/96**	**91.5/92/91**	**91.5/90/87**	**91.1/90/93**	**91.1/94/89**	**91.1/92/90**

Table 5.2 presents the remarkable results achieved by the proposed method using DD-DTCWT. The performance of complex wavelets (DD-DTCWT and DTCWT) is better than the performance of real wavelets (DWT and DD-DWT). The reason is that complex wavelets provide both real and imaginary coefficients. The imaginary coefficients provide structural, and edge information thus enhance the classification accuracy in various patterns. Figs. 5.9 and 5.10 show the ROC plots for different combinations implemented in Table 5.2.

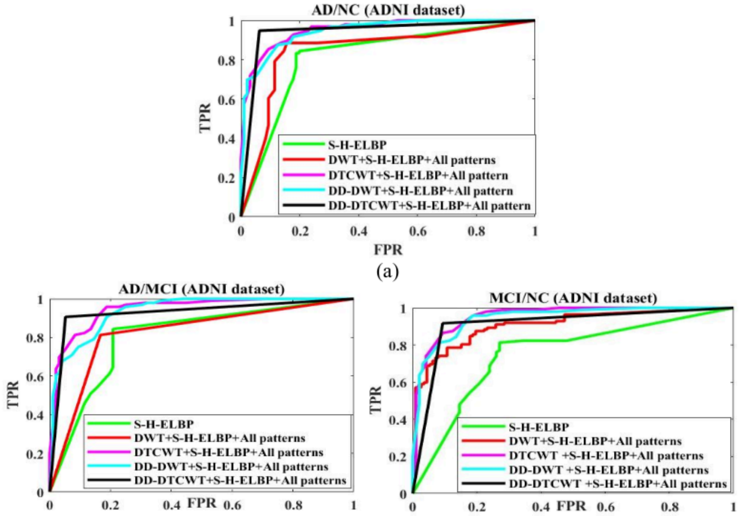


(a)



(b) (c)

Fig. 5.9 ROC Plots on OASIS dataset (a) AD/NC (b)AD/MCI (c) MCI/NC



(a)

(b) (c)

Fig. 5.10 ROC Plots on ADNI dataset (a) AD/NC (b)AD/MCI (c) MCI/NC

The ROC curves depicted in Figs. 5.9 and 5.10 exhibit a steep incline toward the upper-left quadrant, highlighting the effectiveness of the proposed approach. As illustrated in the plot, the proposed approach yields elevated TPR and FPR values, demonstrating its effectiveness in terms of both sensitivity and specificity.

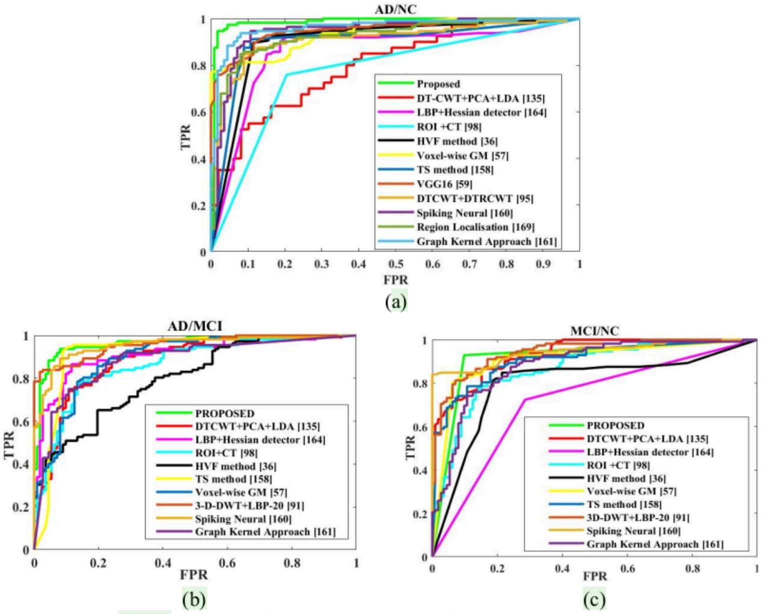
5.4.1.4. Evaluation of the proposed method against existing advanced methods for Binary Classification

The proposed method has been benchmarked against twelve existing approaches, as detailed in Table 5.3. The existing techniques fall into different categories of the feature extraction. These are wavelet-based methods [91,95,135], texture-based methods [164], VBM-based methods [57], ROI-based methods [36,98,158]. The study [169] uses deep learning for locating pathological regions in structural MR images and targets at improving the classification accuracy for early AD detection. All state-of-the-art techniques have been implemented on MATLAB-19, using the same dataset as the proposed work.

Table 5.3 Proposed vs existing methods for binary classifications

Methods	Classifier	OASIS			ADNI		
		AD/NC **Acc/ Sen/ Spe**	**AD/MCI** **Acc/ Sen/ Spe**	**MCI/NC** **Acc/ Sen/ Spe**	**AD/NC** **Acc/ Sen/ Spe**	**AD/MCI** **Acc/ Sen/ Spe**	**MCI/NC** **Acc/ Sen/ Spe**
Proposed	**KNN**	**95.1/95/96**	**92/90/94**	**91.5/93/90**	**94.8/94/95**	**91.9/90/92**	**91.1/92/91**
DT-CWT+PCA+LDA [135] [2018]	ELM	88.2/86.5/84.4	80.2/78.4/76.3	82.1/76.3/85	85.1/86/84	78.4/75.2/79.8	81.4/75.4/84.2
LBP+ Hessian detector [164] [2021]	CNN	87.5/89/86	86.2/83/89	71.9/72/71	86.2/87/86	85.9/85/86	72.8/70/76
ROI+ Contourlet Transform (CT) [98] [2022]	SVM	81.1/83.3/79.2	78.1/83/73	80.4/82/79	85.5/86/84	82.7/84.6/80	83.1/87.8/76.6
HVF [36] [2015]	SVM	85.3/77/94	74/75/73	78.1/77/79	85.7/77.6/91.2	71.5/72.9/70.2	76.2/72.3/81.5
Voxel wise GM [57] [2012]	SRC Ensemble	89.3/87/92	81.7/86/78	83.5/88/79	90.8/86.3/94.8	81.5/83.7/80.1	81.7/87.0/77.0
TS method [158] [2015]	SVM	88.5/89/86	86.2/87/86	85.3/86/85	89.9/92.3/86.9	83.5/89.0/78.0	82.4/84.1/80.5
VGG16 [159] [2022]	NN	86.7/85.2/87.3	-	-	85.4/86.5/84.6		
DTCWT+DTRCWT. [95] [2022]	FNN	90.7/91.2/89.4	-	-	90.2/91.7/88.3		
Spiking Neural Network [160] [2021]		90.5/92/89	89.7/89/90	86.2/85/88	90.1/94/87	87.1/89/85	83.9/80/86
Region localisation-based method [169] [2023]		88.4/85/92			87.5/85/90		
Graph kernel -based approach [161] [2023]	SVM	92/90/94	82.6/88/77	81.3/82/80	90.2/94/87	83.5/85/82	82.1/78/87
3D-DWT+LBP-20 [91] [2020]	SVM	-	91.5/92/91	89.3/84/95		91.0/89/91	89.1/90.2/88

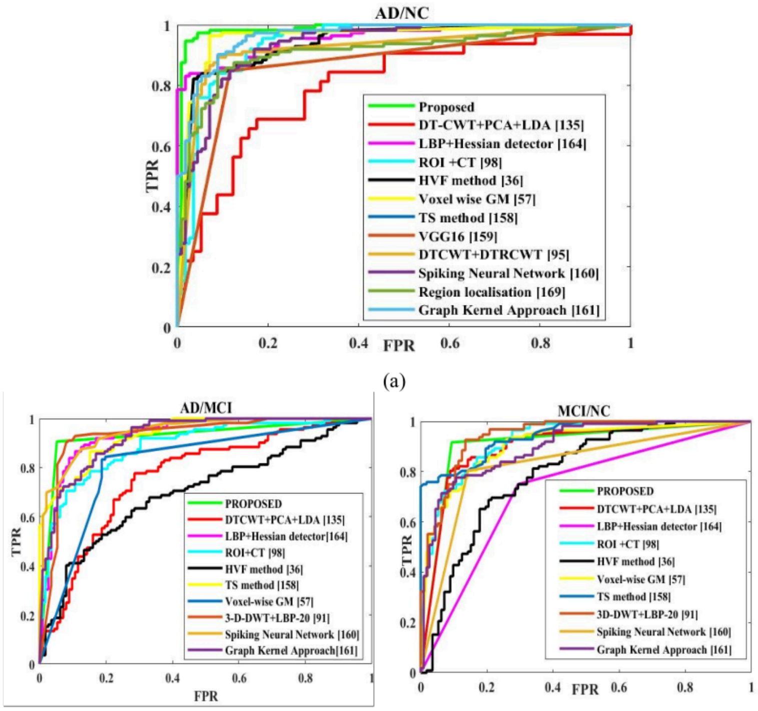
The Table 5.3 indicates outstanding performance of the proposed technique in terms of accuracy, sensitivity, and specificity. On the OASIS dataset, the proposed method achieves higher accuracy than the top-performing method by 3% for AD and NC classification, 0.54% for AD and MCI classification, and 2.4% for MCI and NC classification. With ADNI dataset, these figures are 4.4%,0.98%, and 2.0% for respective classifications. This data shows the significance of the proposed technique over existing techniques. A key contribution of the proposed technique is that it includes the strength of the complex wavelets and shifted ELBP for feature extraction. Unlike spiking neural network-based methods, it does not necessitate the intricate process of encoding data into spikes. In the proposed work, statistical features have been obtained from the shifted ELBPs' histograms. The proposed method is more resilient to noise as it concentrates on the intrinsic textural characteristics of the data and extracts statistical information for classification. However, because region localization-based techniques rely on precise region localization, which can be impacted by image noise or small variations, they might be more vulnerable to noise and variations. The ROC plots for both the proposed and existing algorithms, covering all three binary classification scenarios, are illustrated in Figs. 5.11 and 5.12.



(a)

(b) (c)

Fig. 5.11 ROC Plots (Proposed vs Existing) on OASIS dataset
(a) AD/NC (b)AD/ MCI (c) MCI/NC



(a)

(b) (c)

Fig. 5.12 ROC Plots (Proposed vs Existing) on ADNI dataset
(a) AD/NC (b)AD/MCI (c) MCI/NC

The ROC curves depicted in Figs. 5.11 and 5.12 exhibit a steep incline toward the upper-left area, highlighting the proficiency of the proposed approach. The plot shows that the proposed approach attains high TPR value and low FPR value thus producing good sensitivity with specificity on both datasets.

5.4.1.5. Experimental Results: Multiclass-Classification

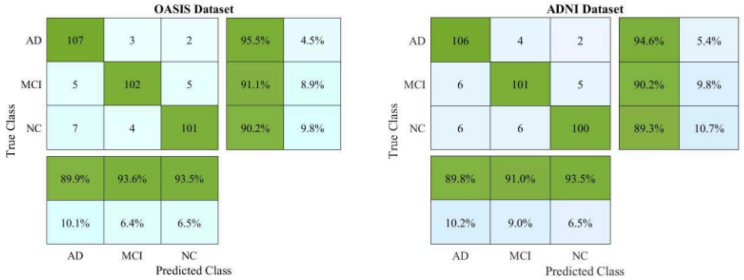
The proposed method shows outstanding results with multiclass classification also. The multi-class classification task includes three categories: AD, NC, and MCI. The effectiveness of the proposed method has been evaluated with three classifiers namely SVM, KNN, and Decision Tree. The results with KNN classifiers are outstanding. The proposed approach has been evaluated for multiclass classification with four wavelets namely DWT, DTCWT, DD-

DWT and DD-DTCWT. Maximum results are observed with DD-DTCWT. Table 5.4 depicts the results of KNN classifier with different wavelets for multi-class classification.

Table 5.4 Analysis with different wavelets for multi-class classification

				OASISI DATASET AD/MCI/NC	**ADNI DATASET AD/MCI/NC**
Wavelet	**ELBP type**	**Sub-bands**	**Patterns used**	**Acc/Sen/ Spe**	**Acc/Sen/ Spe**
_	S-H-ELBP	Zero	Complete 8 patterns	80.1/82/79	81.2/80/82
DWT		4	A total of 8 patterns obtained across 4 sub-bands	84.1/85.2/84	83.2/81/85
DTCWT		6	A total of 8 patterns obtained across 6 sub-bands	87.5/88/87	84.4/85/84
DD-DWT		8	A total of 8 patterns obtained across 8 sub-bands	85.4/84.1/86	82.9/84/81
DD-DTCWT		16	A total of 8 patterns obtained across 16 sub-bands	**94.9/92.2/96**	**94.2/91.3/95.6**
_	S-V-ELBP	Zero	Complete 8 patterns	79.6/78/80	78.2/76/77
DWT		4	A total of 8 patterns obtained across 4 sub-bands	82.4/81.4/83	81.2/83/73
DTCWT		6	A total of 8 patterns obtained across 6 sub-bands	85.7/84/86	86.2/88/84
DD-DWT		8	A total of 8 patterns obtained across 8 sub-bands	84.6/85/83	84.5/85/83
DD-DTCWT		16	A total of 8 patterns obtained across 16 sub-bands	**92.7/92/93**	**91.7/9290**
_	S-H-V-ELBP	Zero	Complete 8 patterns	77.1/79/76	77.2/78/76
DWT		4	A total of 8 patterns obtained across 4 sub-bands	80.5/81/79	81.5/82/80
DTCWT		6	A total of 8 patterns obtained across 6 sub-bands	83.5/84/82	84.5/85/83
DD-DWT		8	A total of 8 patterns obtained across 8 sub-bands	82.6/83/82	83.1/84/83
DD-DTCWT		16	A total of 8 patterns obtained across 16 sub-bands	**89.2/88/91**	**90.1/91/89**

The confusion matrices of the proposed approach using the KNN classifier and DD-DTCWT + S-H-ELBP on both datasets are illustrated in Fig. 5.13. The multiclass classification results are 94.9% and 94.2% accuracy, 92.2% and 91.3% sensitivity, and 96% and 95.6% specificity respectively with the OASIS and ADNI datasets.



a) (b)

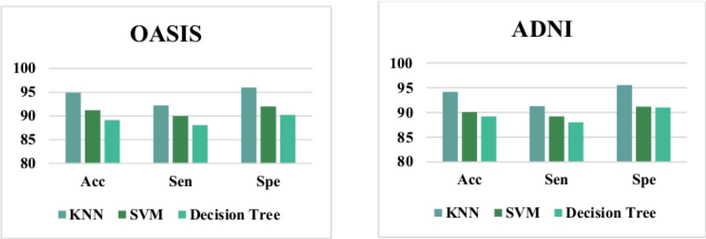
Fig. 5.13 Confusion Matrix (a) OASIS (b)ADNI

The proposed method is also examined with different classifiers like SVM, Decision Tree, and KNN. The classification outcomes are stated in Table 5.5.

Table 5.5 Various classifiers used in multi-class classification

	OASIS DATASET	ADNI DATASET
Classifier	Acc/Sen/ Spe	Acc/Sen/ Spe
KNN	94.9/92.2/96.0	94.2/91.3/95.6
SVM	91.2/90.0/92.0	90.1/89.2/91.2
Decision Tree	89.1/88.1/90.2	89.2/88.0/91.0

Fig. 5.14 illustrates a bar graph demonstrating the performance of several classifiers for multiclass classification.



(a) (b)

Fig. 5.14 Multiclass Classification (a) OASIS, (b)ADNI

The proposed method yields excellent results for AD classification. Nonetheless, both the proposed method and existing approaches exhibit a decline in performance when classifying MCI/NC compared to AD/NC. This is due to the variation in the progression of the disease from one person to another. The minute structural alteration in the brain and the signs linked with MCI make the early AD detection less obvious than the more severe changes seen in AD. Unlike, AD, MCI is more challenging to detect using biomarkers or imaging, as pathological changes such as tau tangles and amyloid-beta plaques are typically in their early stages.

5.4.1.6. Evaluation of the proposed method against existing state-of-the-art-methods for Multiclass Classification

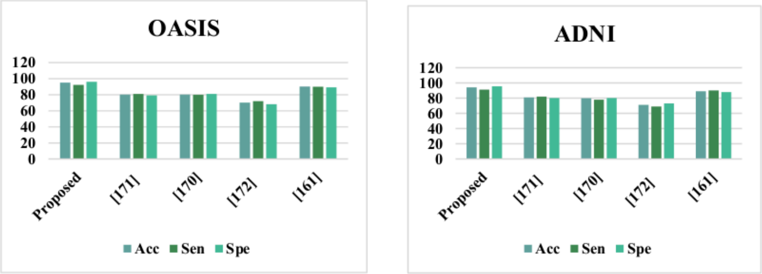
The proposed method has been benchmarked against four existing approaches for multiclass classification, as detailed in Table 5.6. These methods have been implemented on MATLAB -19 with the same datasets as the proposed approach.

Table 5.6 Proposed Method vs Existing Methods for multi-class classification

			OASIS DATSET	ADNI DATASET
			AD/MCI/NC	AD/ MCI/NC
Methods	**Extracted Feature**	**Classifier**	**Acc/Sen/Spe**	**Acc/Sen/Spe**
Proposed	**Statistical features**	**KNN**	**94.9/92.2/96**	**94.2/91.3/95.6**
Hybrid Features [171 [2017]	GLCM texture feature	SVM	80.2/81/79	81/82/80
Bag of visual words [170] [2018]	Clinical+ texture features	KNN	80.1/80/81	79.7/78/80
LDA scoring method [172] [2020]	Genetic features	Decision tree	70.1/72/68	71/69/73
Graph-kernel Approach [161] [2023]	3-D Texture Features	SVM	90.1/90/89	89.1/90/88

The study [170] achieves an accuracy of 80.1% utilizing OASIS dataset by combining clinical data and textural features. Among other textural features, this study emphasizes the significance of the GLCM for AD detection. The study [171] classified Alzheimer's disease into several classes using GLCM texture features as well as clinical features. This method yields 80.2% accuracy when implemented with OASIS dataset. The recent study [172] combined multimodal data using a linear discriminant analysis score approach. This method yields 70.1% accuracy when implemented with OASIS dataset. For multiclass classification, a recent work [161] has also been investigated. This study yields 90.1% when implemented with OASIS dataset. The multiclass classification using the proposed technique provides 94.9% and 94.2 % accuracy on OASIS and ADNI datasets respectively.

The bar chart in Fig. 5.15 contrasts the proposed method and existing advanced techniques.



(a) (b)

Fig. 5.15 Bar Plots (a) OASIS (b) ADNI

The outcomes confirm that the proposed method performs exceptionally good for both binary and multi-class diagnosis. Multi-class classification is more significant since it classifies the data into one of the three groups directly, eliminating the need for repeated binary classifications.

5.4.2. Second Method

In this method, GLCM-based feature extraction technique is proposed. GLCM-based features capture complementary information to wavelet coefficients, focusing on texture properties and spatial relationships that may not be fully captured by frequency-domain representations alone. GLCM offers configurable parameters like distance, angle, and grey levels count to enable texture-specific feature extraction. Fig.5.16 illustrates the workflow of the proposed technique through a pipeline model.

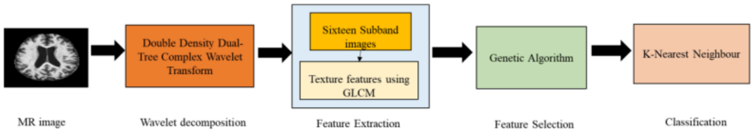


Fig. 5.16 Pipeline Model: Proposed Method

In the proposed work, GLCM [168] has been computed from the sixteen sub-bands obtained after first level DD-DTCWT decomposition. Further, the statistical features have been obtained from computed GLCMs. These features are used for AD classification. This can be understood by using following steps.

- Let the input MRI slice be represented as a 2D image:

$$I(x,y) \qquad (x,y) \in R^2 \tag{5.15}$$

where $I(x,y)$ is the pixel intensity at co-ordinates (x,y)

- First level DD-DTCWT is applied on MR image, which gives one approximate band and sixteen detail sub-bands.

$$\text{DD} - \text{DTCWT}(1) = \{AL, S_{1,1} \ldots\ldots S_{1,16}\} \tag{5.16}$$

Where $S_{n,m}$ represents the m^{th} sub-band at the n^{th} decomposition level.

- Now, GLCM has been computed for these sixteen sub-bands and statistical features have been obtained through computed matrices. The GLCM $P(i,j,\Delta x,\Delta y)$ at a distance $\Delta x, \Delta y$ is defined as:

$$P(i,j,\Delta x,\Delta y) = \sum_{x=1}^{M} \sum_{y=1}^{N} \delta(I(x,y),i).\delta(I(x+\Delta x, y+\Delta y),j) \tag{5.17}$$

where:

- M and N are the dimensions of the image (height and width).
- *i* and *j* are possible gray level values (e.g., from 0 to G−1 for G gray levels).

The total 22 statistical features have been obtained from each computed matrix of all sub-band images and these statistical features have been fed to feature selection algorithm i.e. Genetic Algorithm for selecting optimum features and thus improving classification accuracy.

The 22 Harlick texture features are represented in Table 5.7. There is a wide application of these features in texture analysis.

Table 5.7 GLCM Statistical Features

S.No.	Features	Description
1	Angular second Moment	It shows homogeneous patterns in the image.
2	Joint Entropy	It shows variation in next pixel intensity.
3	Dissimilarity	It defines the distance between the interested pair of pixels in image.
4	Contrast / Inertia	It measures the difference in between the pixel and its neighbour intensity.
5	Homogeneity1 / Inverse difference	It shows the local homogeneity of an image
6	Correlation	It measures how a pixel is correlated with its neighbour pixel. For maximum correlation, count is 1 and for minimum correlation, count is -1.
7	Homogeneity 2/ Inverse difference moment	It shows the local homogeneity of an image.
8	Autocorrelation	It quantifies the coarseness of an image.
9	Cluster Shade	It quantifies the skewness of the GLCM matrix
10	Cluster Prominence	It measures the changes in local intensity of pixels. Its high value indicates the asymmetry of the image.
11	Maximum probability	It shows occurrences of the most predominant pair of adjacent intensity alues.
12	Sum of Squares or variance	It measures the dispersion of grey intensity level around the mean value.
13	Sum Average	It measures the average of the sum of all `possible grey-level intensities of the image.
14	Sum Variance	It measures the dispersion of the average of the sum of all `possible grey-level intensities of the image around the mean value.
15	Sum Entropy	It measures the clutter regarding the sum of grey-level intensities of the mage.
16	Difference variance	It is a measure of dispersion of the variation of grey level spread of the image with respect to the mean.
17	Difference entropy	It quantifies the clutter regarding the grey-level variation spread in the image
18	Information measures of correlation (1)	It computes the strength of correlation, and the complexity of texture.
19	Information measures of correlation (2)	It computes the strength of correlation, and the complexity of texture.
20	Maximal correlation coefficient	It shows texture complexity.
21	Inverse difference normalized (INN)	It shows local homogeneity of image.
22	Inverse difference moment normalized (IDN)	IDN standardizes the difference between the intensity of neighbouring pixels.

These features offer a detailed description of the spatial distribution and relationship of pixel intensities within an image. Features like contrast, energy, and correlation capture subtle textural variations, making them ideal for fine-grained analysis. The proposed method extracts a total of 352 statistical features.

DD-DTCWT provides high directional sensitivity which is useful for extracting more informative features in sixteen different directions. Besides this, Gray Level Co-occurrence Matrix directly interacts with image intensity, it provides information on mutually occurring patterns. The Genetic Algorithm enhances classification accuracy by extracting optimal features.

5.4.2.1. Experimental results for binary classifications

This method captures global and local texture features which enhances the classification accuracy. DD-DTCWT provides global features by capturing multiscale and directional information, ensuring that the overall texture patterns and structural elements are well-represented. GLCM provides local features by analyzing pixel-pair relationships within the decomposed sub-bands, capturing fine details and local texture variations. The combination of DD-DTCWT and GLCM results in a feature set that is rich in both global and local information, enhancing the robustness and accuracy of AD classification model. Tables 5.8, 5.9 and 5.10 show the classification results of three binary tasks with all implemented models using OASIS dataset.

Table 5.8 AD/NC classification using OASIS dataset

Models	**Method**	**AD/NC**			
		0º	**45º**	**90º**	**145º**
1	GLCM	83/85/81	83/82/85	82/84/80	84.4/85.1/83
2	DTCWT+GLCM	89/88/90	89/88.8/90	90/92/89	90/89/91
3	DTCWT+GLCM+PCA	92/93/91	92/90/94	92.3/93/91	91.9/93/89.1
4	DTCWT+GLCM+GA	96/97/95	96.0/94.0/98	96/96/96	96.1/98.1/94
5	DD-DWT+GLCM+GA	95/94/97	95/95/96	95.1/93/97	95.1/95/96
6	DD-DTCWT +GLCM +PCA	94.2/96/93	94/94/96	94.6/92/97	94.2/94/95
7	DD-DTCWT +GLCM +PSO	96.9/98/94	96/96/96	96.9/97/96	96.4/99/94
8	DD-DTCWT +GLCM +ACO	95.5/95/96	95/92/98	95.1/94/96	95.5/93/98
9	**DD-DTCWT +GLCM +GA**	**97.9/97/99**	**97.8/98/97**	**97.8/96/99**	**97.3/98/96**

Table 5.9 AD/MCI classification on OASIS dataset

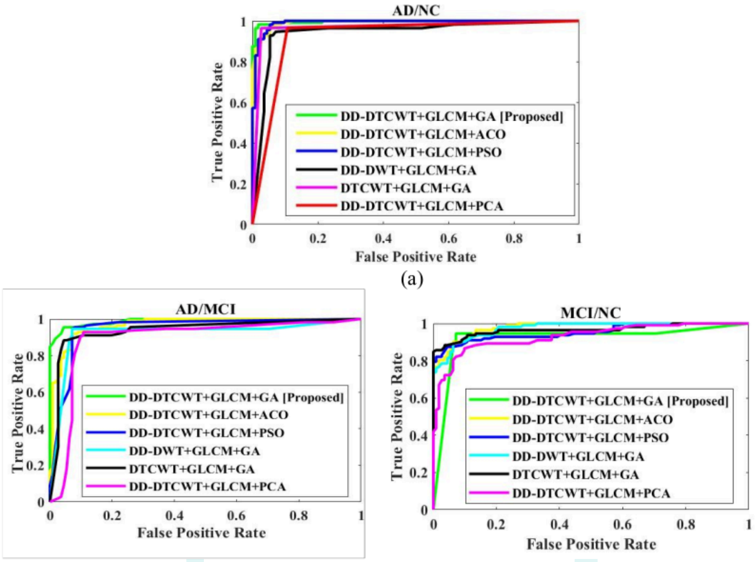
Models	**Method**	**AD/MCI**			
		0º	**45º**	**90º**	**145º**
1	GLCM	81/83/80	80.4/82/79	81/81/82	80.4/82/79
2	DTCWT+GLCM	87.1/82/82	87.5/85/90	87/88/87	86.2/88/84
3	DTCWT+GLCM+PCA	90.6/88/94	90.2/91/89	90/90/90	90.6/96/86
4	DTCWT+GLCM+GA	92/91/93	92.4/91/94	92/91/94	92.9/94/92
5	DD-DWT+GLCM+GA	91.1/89/93	91/87/94	91/92/90	91.5/93/90
6	DD-DTCWT +GLCM +PCA	90.2/92/88	90.2/93/86	90/89/91	90/90.1/90.1
7	DD-DTCWT +GLCM +PSO	92/90/94	92.4/95/90	92/89/95	92.4/91/94
8	DD-DTCWT +GLCM +ACO	91.1/88/94	91.5/92/91	91/94/89	91.1/96/87
9	DD-DTCWT +GLCM +GA	93.9/94/94	93.3/97/89	92.9/90/96	93.8/95/93

Table 5.10 MCI/NC classification on OASIS dataset

Models	**Method**	**MCI/NC**			
		0º	**45º**	**90º**	**145º**
1	GLCM	79.9/78/82	79/77/81	78.1/76/72	79.9/78/82
2	DTCWT+GLCM	85.7/87/85	85.7/88/84	84.8/87/83	84.8/88/82
3	DTCWT+GLCM+PCA	87.5/89/86	87.9/88/88	87.5/85/90	87.5/89/86
4	DTCWT+GLCM+GA	90.2/94/87	90.6/92/89	90.1/91/89	90.2/90/90
5	DD-DWT+GLCM+GA	89.5/87.8/91.1	89.3/92/87	89.7/94/86	89.3/94/85
6	DD-DTCWT +GLCM +PCA	87.5/89/86	87.9/88/88	87.9/86/90	87.5/85/90
7	DD-DTCWT +GLCM +PSO	91.1/88/94	91.5/91/92	91.5/89/94	91.1/87/96
8	DD-DTCWT +GLCM +ACO	90.6/89/92	90.2/89/90	90.2/94/87	90.2/92/88
9	DD-DTCWT +GLCM +GA	91.4/89/93	91.5/90/92	91.1/90/92	91.5/92/91

In Tables 5.8,5.9 and 5.10, the performance of nine models have been investigated on OASIS dataset. In first model, GLCM is directly applied on MR images and statistical features have been obtained. The number of these features are 22. This model gives 83%, 81% and 79% accuracy for classifying AD from NC, AD from MCI and MCI from NC respectively using d=1.In second model DTCWT is applied on the MR images and then six sub-bands have been obtained. Further, GLCM is applied to six sub-bands resulting in 132 (22×6) statistical features. The classification accuracy is improved by 7.2%,7.5% and 7.2% in AD classification with NC, AD classification with MCI and MCI classification with NC respectively. This is due to the local and global information captured by the combination of DTCWT+GLCM. In third model, PCA is introduced after extracting 132 statistical features. PCA results in score matrix with 224 rows and 223 columns. This model provides respective accuracy of 92%,90.6%, and 87.5 % in AD classification with NC, AD classification with MCI and MCI classification with NC using d=1.In fourth model PCA is replaced with feature optimization algorithm i.e. Genetic Algorithm. In fifth model DTCWT is replaced with DD-DWT.DD-DWT provides eight sub-bands. GLCM has been applied to these eight sub-bands and statistical features have been obtained through computed matrices. The total number of features is 176 (22×8). Further, GA has been applied to select appropriate features. This model provides 95%,91%, and 89.5 % accuracy in AD classification with NC, AD classification with MCI and MCI classification with NC respectively with d=1.In 6th,7th,8th and 9th models, DD-DTCWT has been applied on MR images, and sixteen sub-bands have been obtained. Further, GLCM has been applied on these sixteen sub-bands and statistical features have been obtained through computed matrices. The numer of features is 352 (22×16). Now, PCA, PSO, ACO, and GA have been applied on models from six to nine and optimum features have been obtained for AD/NC, AD/MCI and MCI/NC classifications. The best results are with ninth model i.e., DD-DTCWT+GLCM+GA. The proposed model provides 97.9% accuracy, 96.0% sensitivity and 99% specificity for AD/NC classification ,93.9 % accuracy ,94.0% sensitivity, and 94.0% specificity for AD/MCI classification ,91.4% accuracy,89.0% sensitivity, and 93.0% specificity for MCI/NC classification on OASIS dataset with d=1.The proposed method has been tested for effectiveness with d=2 as well. The results are almost same and thus not included.

The ROC plots of the implemented models in Tables 5.8,5.9 and 5.10 are demonstrated in Fig. 5.17.



(a)

(b) (c)

Fig. 5.17 ROC Plots (a)AD/NC (b)AD/MCI (c)MCI/NC on OASIS dataset

In Fig. 5.17, the proposed model shows a steep rise in ROC curve towards the upper left quadrant. This indicates that the model attains high TPR value and FPR value thus producing good sensitivity with specificity.

5.4.2.2. *Feature Selection Algorithms and Convergence Plots*

The performance of the proposed work has been evaluated with three feature selection algorithms i.e. PSO [153], ACO [154] and GA [155]. The results with GA are outstanding due to its nature of having diverse populations. The crossover and mutation as genetic operators promote diversity within the population which prevents the algorithm to trap in local optima. While ACO and PSO tends to trap in local optima. The processing time of GA is 43 seconds with AD/NC, 42 seconds with AD/MCI and 41 seconds with MCI/NC classification. The ACO is the fastest with processing time nearly 20 seconds in each classification. PSO takes 35 seconds approximately in each classification. ACO uses pheromone levels to explicitly store a memory of the search history. This enables faster convergence by enabling the algorithm to swiftly adjust and improve solutions.GA depends on the population of the current generation and lacks a memory mechanism. PSO does not have a global memory of the search process, it just keeps track of the local and global best solutions. Table 5.11 shows the parameters used for different feature selection algorithms used in the proposed work.

Table 5.11 Parameters of feature selection algorithms

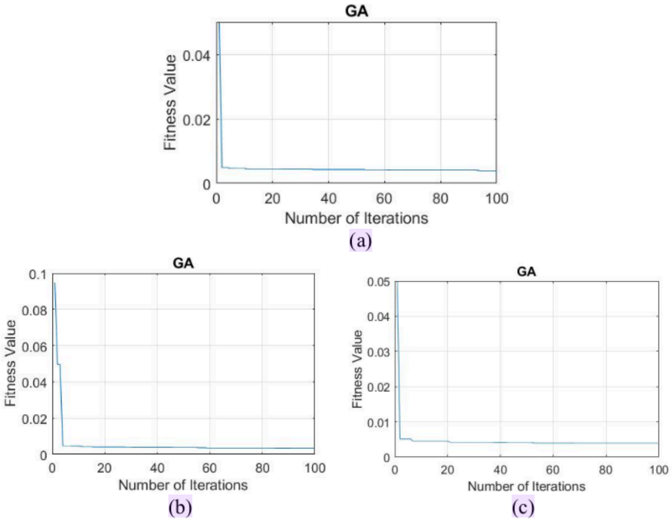
Algorithms					
PSO		ACO		GA	
Parameter	**Value**	**Parameter**	**Value**	**Parameter**	**Value**
Solutions count	30	Solutions count	30	Solutions count	30
Iterations count	100	Iterations count	100	Iterations count	100
c1	2	Pheromone value (τ)	1	Cross-over rate	0.8
c2	2	heuristic desirability (eta)	1	Mutation rate	0.01
W	0.9	Pheromone factor (α)	1		
		control heuristic (beta)	0.1		
		pheromone evaporation rate (ρ)	0.2		

Initial parameter settings of GA in proposed work includes crossover rate as 0.8, rate of mutation as 0.01, number of solutions as 30 and iterations count as 100.The performance of genetic algorithm has been evaluated at different values of genetic operators viz. mutation rate, crossover rate and their optimum value has been found experimentally. The performance has been evaluated at different number of population and iterations. Table 5.12 includes the experimental values of genetic operators where classification results are outstanding.

Table 5.12 Different experimental values for genetic operators

S.No.	**No of solution**	**No of iterations**	**Cross over rate**	**Mutation Rate**	**Accuracy**
1	50	100	0.6	0.001	89.1
2	30	100	0.6	0.001	92.2
3	80	100	0.45	0.01	93.1
4	30	100	0.45	0.01	94.2
5	40	100	0.8	0.005	93.9
6	50	100	0.9	0.03	92
7	30	100	0.95	0.01	95
8	50	100	0.8	0.01	94.1
9	50	100	1	0.03	95
10	30	200	0.8	0.01	96.2
11	30	100	0.8	0.01	96.2

All the feature selection algorithms have been run by 20 times and each time similar subset of features have been selected and classification results are approximately same in all the runs. The GA convergence plot with proposed approach is presented in Fig. 5.18.



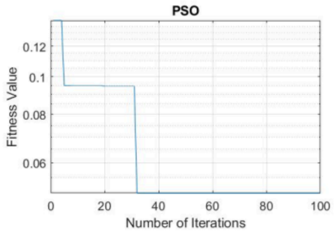
(a)

(b)

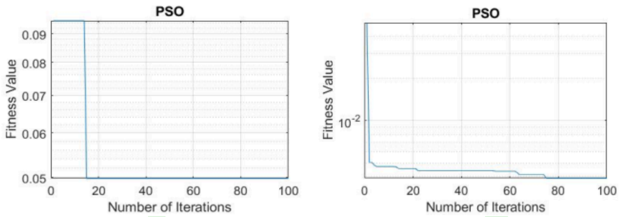
(c)

Fig. 5.18 GA Convergence Plots (a)AD/NC (b)AD/MCI (c) MCI/NC

Fig 5.18 shows the early convergence of GA. This is not the issue because GA has been checked for different genetic operators to avoid the local optima condition and the subset of features selected through GA provides optimum classification results. The PSO and ACO convergence plots with proposed approach are shown in Figs. 5.19 and 5.20.

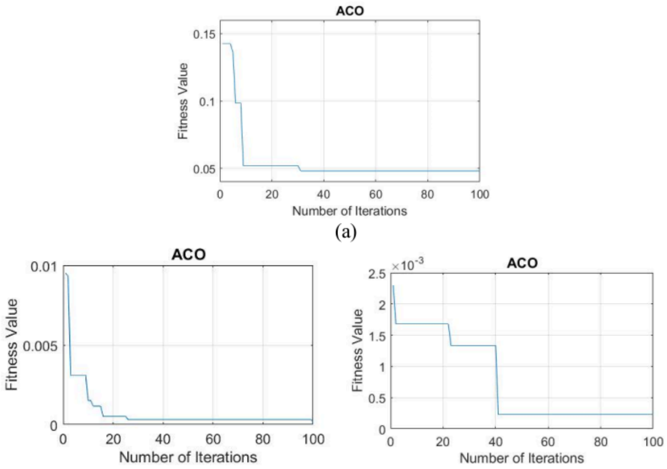


(a)



(b) (c)

Fig. 5.19 PSO Convergence Plots (a)AD/NC (b)AD/MCI (c) MCI/NC



(a)

(b) (c)

Fig. 5.20 ACO Convergence Plots (a)AD/NC (b)AD/MCI (c) MCI/NC

The fitness value for the feature selection algorithm is given as

Fitness value $=\alpha$ * error + β * (number of selected features/total number of features)

Where error = 1 - Accuracy and $\alpha = 0.99$, $\beta = 0.01$

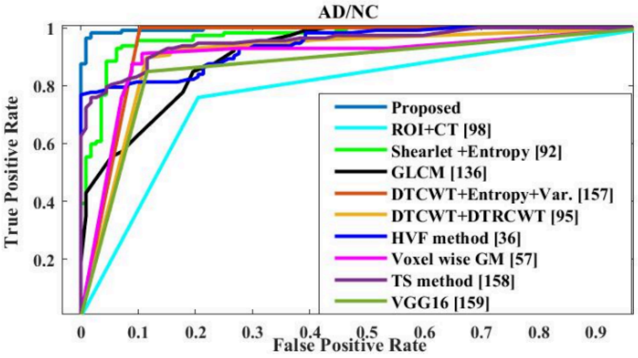
5.4.2.3. Evaluation of the proposed method against leading state-of-the-art-methods

The proposed method has been benchmarked against ten existing techniques of feature extraction. All these techniques have been implemented on MATLAB 19 with OASIS dataset employing ten-fold cross validation techniques. The mean results of ten iterations from each algorithm are compiled in Table 5.13.

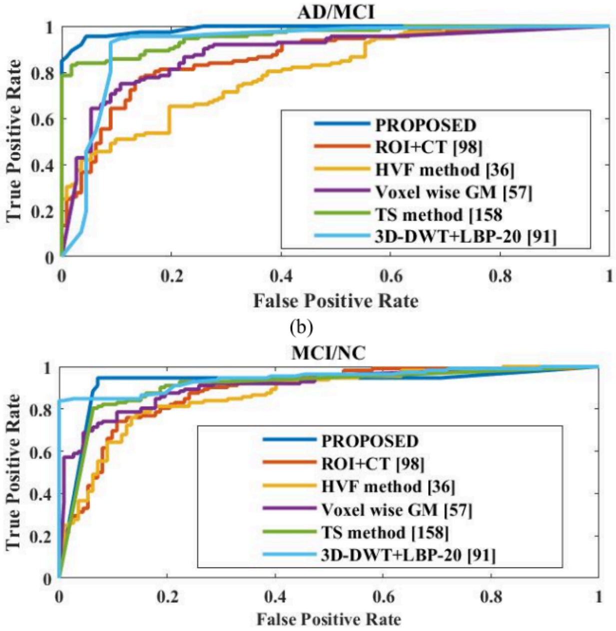
Table 5.13 Proposed vs Existing Methods

		AD/NC			**AD/MCI**			**MCI/NC**		
Method	**Classifier**	**Acc**	**Sen**	**Spe**	**Acc**	**Sen**	**Spe**	**Acc**	**Sen**	**Spe**
DD-DTCWT+GLCM+GA [PROPOSED]	KNN	97.9	96	99	93.9	94	94	91.4	89	93
ROI+ Contourlet Transform [98] [2022]	SVM	82.5	84.8	80.3	78	83.1	73.2	80.4	82	79
Shearlet + Entropy [92] [2019]	KNN	94.1	95.3	93.4	-	-	-	-	-	-
GLCM [136] [2021]	ELM	93	92	94	-	-	-	-	-	-
DTCWT+ Entropy+ Variance [157] [2016]	TSVM	94.1	93.2	95.1	-	-	-	-	-	-
DTCWT+ DTRCWT [95] [2021]	FNN	90.3	91.4	92.5	-	-	-	-	-	-
HVF method [36] [2015]	SVM	85.4	77.1	94	73.9	74.6	72.8	78.1	77	79
Voxel wise GM [57] [2012]	SVM	89.2	87.1	92.2	81.8	85.9	77.9	84	87.1	78.7
Tissue Segmented (TS) method [158] [2015]	SVM	88.5	89	86	86.3	87.1	86.2	85.4	86.1	85.2
VGG16 [159] [2022]	Neural Network	86.6	85.1	87.2	-	-	-	-	-	-
3D-DWT+LBP-20 [91] [2020]	SVM	-	-	-	90	89.5	91	88.3	83	94

Table 5.13 includes the existing advanced methods belonging to different categories of the feature extraction. The proposed method combines DD-DTCWT and GLCM and further use GA which enhances the classification accuracy. The proposed technique also investigates the efficacy of feature selection algorithms in detecting AD. The results of the proposed technique are 4.03%,4.3% and 3.17% higher than the best performing existing technique for classifying AD with NC, AD with MCI and MCI with NC respectively. The other techniques belong to the VBM-based technique, ROI-based technique, Deep learning-based techniques, Wavelet-based techniques and surface-based techniques for feature extraction. The proposed technique provides the local and global structural details, eliminates the need for segmentation, and avoids high-dimensional feature spaces. Fig.5.21 depicts the ROC plots of the proposed and existing techniques.



(a)



(b)

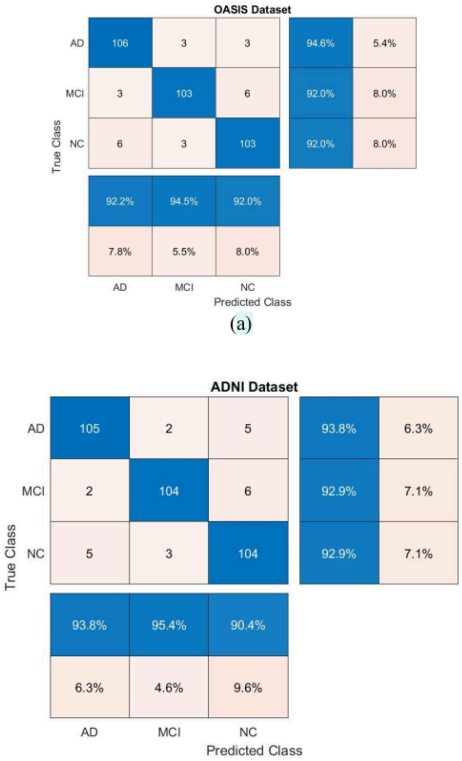
(c)

Fig. 5.21 ROC Plots (a)AD/NC (b)AD/MCI (c) MCI/NC on OASIS dataset

In Fig. 5.21, the proposed method shows a sharp rise in ROC curve towards the upper left quadrant. This suggests that the proposed approach attains high TPR value and FPR value thus producing good sensitivity with specificity along with an accuracy for all three binary classifications.

The proposed method has been assessed for binary classification with ADNI dataset also. The proposed approach provides 96.1% accuracy, 95.0% sensitivity and 97% specificity in classifying AD with NC ,93.0% accuracy ,94.0% sensitivity, and 95.0% specificity in AD classification with MCI ,90.1% accuracy, 88.0% sensitivity, and specificity of 92.0% in MCI classification with NC on ADNI dataset.

The proposed method yields exceptional results for multiclass classification also. The confusion matrices of multi-class classification for OASIS and ADNI datasets are demonstrated in Fig.5.22.



(a)

(b)

Fig. 5.22 Confusion Matrices (a) OASIS Dataset (b) ADNI Dataset

The proposed technique provides accuracy, sensitivity and specificity of 95.2 %, 92.8% and 96.3% on OASIS dataset and 95.4 %, 93.2% and 96.5% respectively on ADNI dataset for multiclass classification.

5.4.3. Third Method

In third method, Shifted Circular Elliptical Local Descriptor (SCELD) has been applied on sixteen sub-bands in place of GLCM and shifted ELBP and statistical features have been obtained. The shifted circular elliptical local descriptor combines the properties of Circular LBP and Elliptical LBP; thus, it can capture the isotropic and anisotropic features. The eight-neighbour shifted circular elliptical local descriptor provides eight patterns from single sub-band. These patterns capture the micro and macro textural details from the sub-band images. Six statistical features have been obtained from these patterns and used for AD classification. Fig. 5.23 illustrates the workflow of the proposed technique through a pipeline model.

.

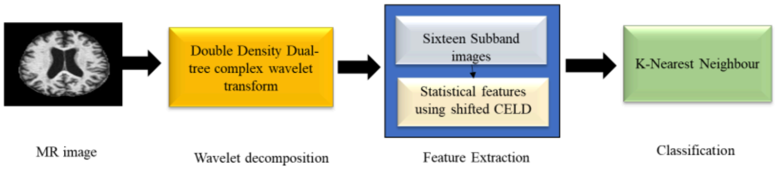


Fig. 5.23 Pipeline model: Proposed method

In Fig. 5.23, DD-DTCWT has been applied on 2-D MR images, and sixteen sub-bands have been obtained. Further statistical features like mean, variance, energy, entropy, kurtosis and skewness have been obtained from these sixteen sub-bands. The complete step in the proposed approach is depicted in Fig. 5.24.

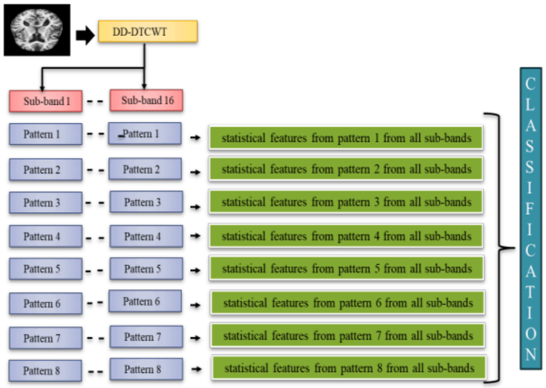


Fig. 5.24 Feature Extraction Process

In Fig. 5.24, statistical features have been extracted from pattern 1 of all sixteen sub-bands and have been used for AD classification. This is the first implemented model. Similarly, statistical features have been extracted from pattern 2 of all sixteen sub-bands and have been used for AD classification. This process is done with all 8 patterns. Thus, the total eight models have been implemented. In the ninth and the proposed model, statistical features from all eight patterns of sixteen sub-bands have been used for AD classification.

5.4.3.1. Experimental Results for binary classifications

The classification performance of all eight patterns has been examined and listed in Table 5.14 using OASIS dataset. Table 5.14 shows that the highest performance is achieved when statistical features from all eight patterns across all sixteen sub-bands are combined. The classification accuracy improves when statistical features from all eight patterns across sixteen sub-bands are concatenated, rather than when features are derived from a single pattern. These statistical features are directional in nature and provide micro and macro structural details for AD detection.

Table 5.14 Models implemented in the proposed work using OASIS dataset

		AD/NC	**AD/MCI**	**MCI/NC**
S.No	**Models**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**
1	DD-DTCWT+SCELD+ First pattern across 16 sub bands + features	94.5/95/93	91.4/90/92	90.5/88/92
2	DD-DTCWT+SCELD+ Second pattern across 16 sub bands + features	93.2/94/92	87.8/88/86	88.3/86/90
3	DD-DTCWT+SCELD+ Third pattern across 16 sub bands + features	88.7/89/87	83.8/85/82	88.7/88/89
4	DD-DTCWT+SCELD+ Fourth pattern across 16 sub bands + features	92.7/94/90	89.1/87/91	86.8/87/86
5	DD-DTCWT+SCELD+ Fifth pattern across 16 sub bands + features	85.1/86/84.2	86.9/87/86	89.2/90/88
6	DD-DTCWT+SCELD+ Sixth pattern across 16 sub bands + features	87.3/88/86	85.6/87/83	90.1/91/89
7	DD-DTCWT+SCELD+ Seventh pattern across 16 sub bands + features	90.2/91/89	84.6/85/83	87.8/88/87
8	DD-DTCWT+SCELD+ Eighth pattern across 16 sub bands + features	92.6/93/90.4	82.4/84/83	85.6/87/83
9	DD-DTCWT+SCELD+ All patterns across 16 sub bands + features	**98.1/97/98**	**94.8/93/95**	**94.2/93/95**

Table 5.14 indicates the total nine models implemented with DD-DTCWT. In first model, statistical features have been extracted from first pattern of all sixteen sub-bands. In second model, similar step is taken with second pattern of all sixteen sub-bands. The number of features with models from one to eight are 96 (16×6). In ninth i.e. proposed model, statistical features have been extracted from all eight patterns of sixteen sub-bands. The total number of features are 768 (16 × 6 × 8) with this model. The ninth model provides the maximum results due to the concatenation of features from all patterns. These results are 98.1% accuracy,98% sensitivity and 97% specificity for AD/NC classification, 94.8% accuracy,93% sensitivity and 95% specificity for AD/MCI classification and 94.2% accuracy,93% sensitivity and 95% specificity for MCI/NC classification on OASIS dataset.

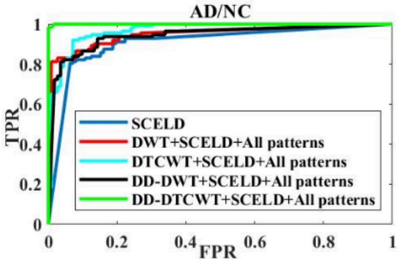
5.4.3.2. Assessment of the proposed framework with different wavelets

The proposed model has been tested with other wavelets also. These are DWT, DTCWT and DD-DWT. Table 5.15 indicates the comparative assessment of the proposed model with different wavelets using OASIS dataset.

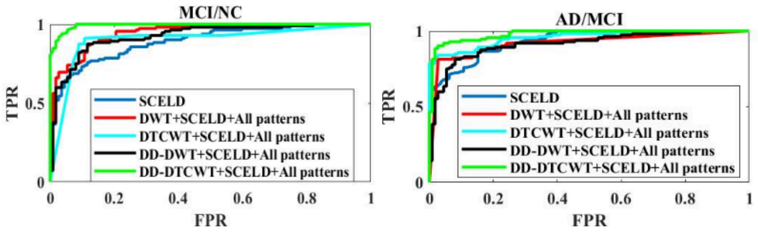
Table 5.15 Assessment of the proposed framework with different wavelets

Wavelet	**AD/NC**	**AD/MCI**	**MCI/NC**
	Acc/Sen/Spe	**Acc/Sen/Spe**	**Acc/Sen/Spe**
SCELD + All statistical features	85.1/86/84	82/83/81	81/83/79
DWT + SCELD +All patterns through 4 sub bands +features	87.7/89/85	84.6/85/83	85.8/87/83
DTCWT+ SCELD +All patterns through 6 sub bands + features	91.5/92/89.7	87.8/88/86	89.4/88/91
DD-DWT+ SCELD + All patterns through 8 sub bands + features	88.2/89/87	85.9/84/86	86.2/88/84
DD-DTCWT+ SCELD+ All patterns through 16 sub bands + features	**98.1/97/98**	**94.8/93/95**	**94.2/93/95**

From Table 5.15, it is clear the classification results get improved by including wavelet with SCELD. The group SCELD + statistical features provide respective accuracy of 85.1 % ,82% and 81% in AD classification with NC, AD classification with MCI and MCI classification with NC. DWT extracts feature in four directions. The group DWT+ SCELD+ statistical features provide the directional features thus classification results get better in this case. Similarly, DTCWT provides feature in six directions, DD-DWT provides features in eight directions while DD-DTCWT provides features in sixteen directions. DD-DTCWT is the best performer among all wavelets. Fig. 5.25 shows the ROC plots of SCELD with various wavelets for all three binary classifications on OASIS dataset.



(a)



(b) (c)

Fig. 5.25 ROC Plots (a) AD/NC, (b) AD/MCI (c) MCI/NC on OASIS dataset

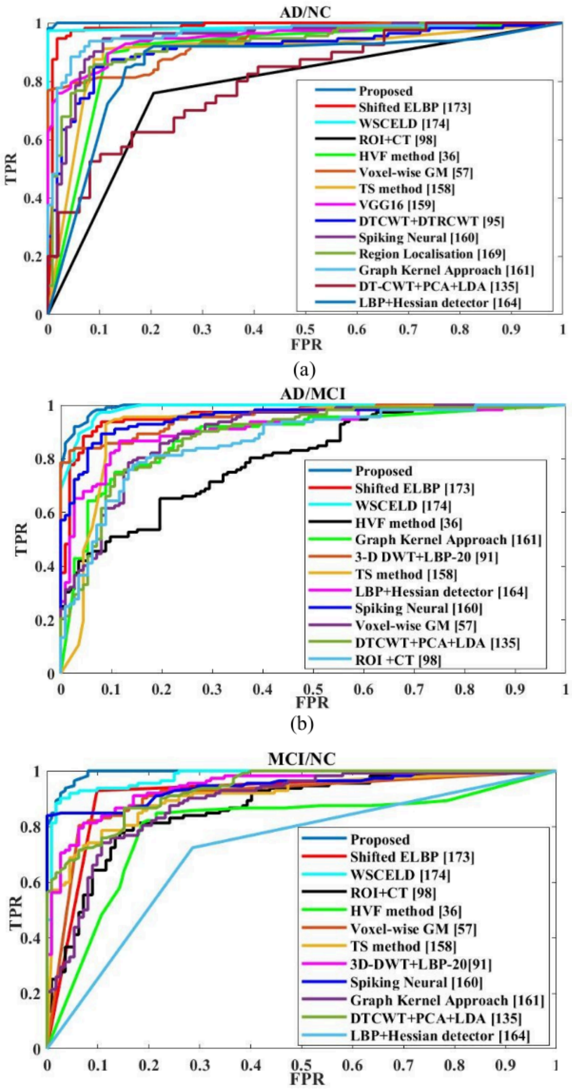
5.4.3.3. Evaluation of the proposed method against current state-of-the-art-methods

The proposed approach surpassed the eleven existing methods. All the methods are executed on MATLAB-19 applying ten-fold cross validation technique on OASIS dataset and the average outcomes over ten runs are detailed in Table 5.16.

Table 5.16 Proposed vs Existing Methods

		AD/ NC	**AD/ MCI**	**MCI / NC**
Methods	**Classifier**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**
Proposed	**KNN**	**98.1/97/98**	**94.8/93/95**	**94.2/93/95**
Shifted ELBP +statistical features [173]	KNN	95.1/95/96	92/90/90	91.5/93/90
WSCELD + histogram [174]	DT	97.3/97.2/94.6	94.6/96.1/93.1	93.8/92.4/96.2
DTCWT+PCA+LDA [135] [2018]	ELM	88.2/86.5/84.4	80.2/78.4/76.3	82.1/76.3/85
LBP+ Hessian detector [164] [2021]	CNN	87.5/89/86	86.2/83/89	71.9/72/71
ROI +CT [98] [2022]	SVM	81.1/83.3/79.2	78.1/83/73	80.4/82/79
HVF method [36] [2015]	SVM	85.3/77/94	74/75/73	78.1/77/79
Voxel-wise GM [57] [2012]	SVM	89.3/87/92	81.7/86/78	83.5/88/79
TS method [158] [2015]	SVM	88.5/89/86	86.2/87/86	85.3/86/85
VGG16 [159] [2022]	NN	86.7/85.2/87.3	-	-
DTCWT+DTRCWT [95] [2022]	FNN	90.7/91.2/92.4	-	-
3D-DWT+LBP-20 [91] [2020]	SVM	-	91.5/92/91	89.3/84/95

The proposed method outperforms existing techniques across all three binary classification tasks, making it highly suitable for early detection of Alzheimer's disease. The study [173] used statistical features through shifted ELBP, and the study [174] used histogram features through WSCELD for AD/NC classification. These studies also give outstanding results for AD detection. However, the proposed technique combines the statistical features with WSCELD thus provides more efficient results compared to the existing studies. Fig. 5.26 shows the ROC plots for the proposed and existing techniques for all three binary classifications on OASIS dataset.



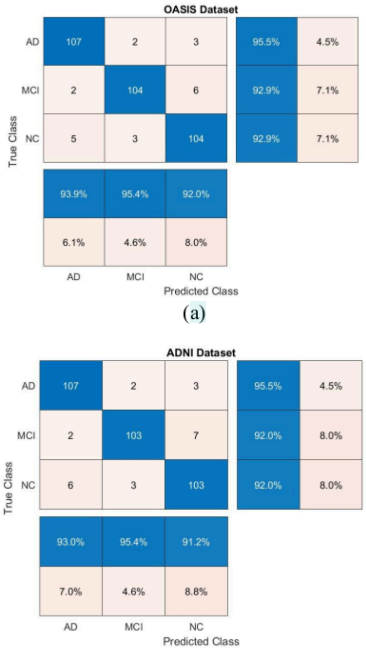
(a) (b) (c)

Fig. 5.26 Proposed vs Existing: ROC Plots (a) AD/NC (b)AD/MCI (c) MCI/NC (OASIS)

The proposed method in Fig. 5.26 shows the steep rise of the ROC curve towards the upper-left quadrant signifying high sensitivity (true positive rate) and low false positive rate, which collectively reflect strong classification performance.

The proposed method has been examined for binary classification with ADNI dataset also. The proposed method gives 98.1% accuracy, 96.1% sensitivity and 97.8% specificity for AD/NC classification ,94.0 % accuracy ,93.0% sensitivity, and 95.0% specificity for AD/MCI classification ,93.9% accuracy, 92.0% sensitivity, and specificity of 94.0% for MCI/NC classification on ADNI dataset.

The proposed method yields exceptional results for multiclass classification also. The corresponding confusion matrices for both datasets are demonstrated in Fig. 5.27.



(a)

(b)

Fig. 5.27 Confusion Matrices (a) OASIS Dataset (b) ADNI Dataset

The presented technique provides 95.8% accuracy ,93.7% sensitivity and 96.8% specificity for multiclass classification on OASIS dataset and with ADNI dataset, these figures are 95.4% ,93.1% and 96.5% respectively.

5.5. Summary

This chapter presents three statistical feature extraction techniques for early AD detection. Statistical features provide a compact representation (e.g., mean, variance, skewness, entropy) that significantly reduces dimensionality while preserving key information. The proposed techniques provide exceptional results on two datasets with binary as well as multi-class classification, thus minimize the problem of scalability in AD detection to a certain limit. In proposed techniques, DD-DTCWT has been applied to get sixteen sub-band images and then statistical features have been obtained from these sixteen sub-band images through different ways.

In first technique, shifted ELBP has been applied on sixteen sub-band images to obtain multiple micro and macro patterns. These micro and macro patterns reflect the fine textural details and large-scale features respectively. Statistical features have been extracted from these patterns and used for AD classification. The accuracy/sensitivity and specificity of proposed method are 95.1%/95% and 96% for AD/NC classification, 92%/90% and 94% for AD/MCI classification and 91.5%,93% and 90% for MCI/NC classification on OASIS dataset respectively. With ADNI dataset, these figures are 94.8%/94%and 95% for AD/NC classification ,91.9%/90% and 92% for AD/MCI classification ,91.1%/92% and 91% for MCI/NC classification respectively. With multiclass classification, respective results are 94.9%, 92.2% and 96% on OASIS dataset and 94.2 %, 91.3% and 95.6% on ADNI dataset.

In second technique, statistical features are obtained through GLCM. The GLCM of the sixteen sub-band images have been computed and then 22 statistical features have been extracted from the computed GLCM. The optimum features are selected by using genetic algorithm. These features are further used for AD classification. This research work also investigates the performance of different feature selection algorithms and proposes an optimum feature selection method for AD detection. The accuracy/sensitivity and specificity of proposed method are 97.9%/96% and 99% for AD/NC classification, 93.9%/94% and 94% for AD/MCI classification and 91.4%,89% and 93% for MCI/NC classification on OASIS dataset respectively. With ADNI dataset, these figures are 96.1%/ 95% and 97% for AD/NC classification ,93.0%/94.0% and 95.0% for AD/MCI classification ,90.1%/88.0% and 92.0% for MCI/NC classification respectively. With multiclass classification, respective results are 95.2%, 92.8% and 96.3% on OASIS dataset and 95.4 %, 93.2% and 96.5% on ADNI dataset.

In third technique, WSCELD has been applied on sixteen sub-band images to obtain multiple micro and macro pattern. WSCELD is better than shifted ELBP as it includes the properties of circular and elliptical LBP both while shifted ELBP includes the properties of ELBP only. The third technique gives exceptional results for early AD detection without using any feature dimensional reduction method or feature selection method. The accuracy/sensitivity and specificity of proposed method are 98.1%/97% and 98% for AD/NC classification, 94.8%/93% and 95% for AD/MCI classification and 94.2%,93% and 95% for MCI/NC classification on OASIS dataset respectively. With ADNI dataset, these figures are 98.1%/ 96.1% and 97.8% for AD/NC classification ,94.0%/93.0% and 95.0% for AD/MCI classification ,93.9%/92.0% and 94.0% for MCI/NC classification respectively. With multiclass classification, respective results are 95.8%, 93.7% and 96.8% on OASIS dataset and 95.4 %, 93.1% and 96.5% on ADNI dataset.

The publication related to this chapter

- Neha Garg, M. S. Choudhry, & R. M. Bodade, & Alzheimer's Disease Neuroimaging Initiative. (2025). Alzheimer's disease detection through wavelet-based shifted elliptical local binary pattern. **Biomedical Signal Processing and Control,** 100, 107067. https://doi.org/10.1016/j.bspc.2024.107067 **IF. 4.9**
- Neha Garg, M. S. Choudhry, and R. M. Bodade. "Wavelet Transform based Alzheimer's Disease detection using Grey Level Co-occurrence Matrix and Genetic Algorithm." 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN). IEEE, 2023, 10.1109/CICTN57981.2023.10141079
- Neha Garg, M. S. Choudhry and R. M. Bodade, "Early Alzheimer's Disease detection using wavelet and Grey Level Co-occurrence matrix". **Engineering research express. [ESCI] (Communicated)**
- Neha Garg, M. S Choudhry, & R. M. Bodade, "Alzheimer's disease classification with statistical features using wavelet based shifted Circular Elliptical Local Binary Descriptor." **Biomedical Signal Processing and Control**, SCIE Impact factor-4.9 (**Communicated)**

Chapter 6

Alzheimer’s Disease Detection using Mean Energy-Based Feature Extraction Technique

---

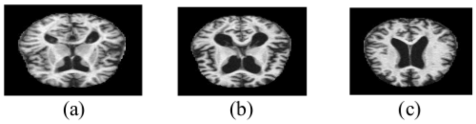
This chapter presents simple yet effective method for feature extraction in Alzheimer’s Disease detection. The proposed technique provides exceptional results on two datasets with low dimensionality and less computation time. The proposed technique uses novel feature i.e. mean energy of real and imaginary components taken independently for Alzheimer’s Disease detection. The proposed technique uses only 160 features and does not require feature dimension reduction methods and feature selection methods as these were used in the proposed techniques discussed in the earlier chapters. Feature dimensionality reduction methods create the loss of spatial information and feature selection methods unnecessary increase the computation burden while optimizing the features. The proposed method is assessed and verified by way of comprehensive experimentation on benchmark data and a detailed comparison with current existing approaches.

6.1. Introduction

This chapter proposes the Mean Energy-based Feature Extraction technique for early AD detection. This technique emphasises features in frequency domain which are more descriptive and directionally sensitive than those in the spatial domain, offering richer and more structured information. The existing wavelet-based techniques generally have the high dimensionality problem. The feature dimension reduction methods like PCA are used in the existing wavelet-based techniques. This results in spatial loss of information. This can blur or distort the local structures leading to a loss of the specific spatial patterns that may be important for AD detection. This method effectively captures edge features and grey matter density variations without relying on principal component analysis or feature selection techniques. The proposed method used mean energy features from real and imaginary sub-bands taken independently for AD classification. The mean energy features are less sensitive to noise in comparison to LBP-based and GLCM-based features. LBP patterns get disturbed due to the small local changes created by noise between the pixels. GLCM focuses on spatial relationship between the neighbouring pixels which also get affected due to the noise between the pixels. Mean energy-based method aggregates coefficients, inherently suppressing noise influence. In AD, the pathogenic abnormalities in the brain images are reflected as textural, structural, or intensity changes. The distribution of energy across frequency sub-bands is impacted by these changes, which makes mean energy a discriminative feature. The proposed method facilitates the detection of various stages of Alzheimer's disease by capturing both grey matter density fluctuations and structural brain changes associated with disease progression. The proposed work investigates the performance of different features viz. mean energy of real coefficients, mean energy of imaginary coefficients, mean energy of real and imaginary coefficients used independently, mean energy of complex coefficients, variance of magnitude of complex coefficients and entropy of complex coefficients.

6.2. Dataset

In this work, both datasets i.e. OASIS and ADNI have been used. The twenty-eight 3-D MR images of each class i.e., AD, MCI, NC have been downloaded from OASIS-1 dataset. The four central 2-D MR slices have been extracted from each 3-D MRI volume using XMedCon software. In this way, 112 2D MR slices have been obtained for each class. Similarly same dataset has been prepared for ADNI-1 dataset. Fig.6.1 shows the sample of each class of 2-D MR slices obtained from OASIS-1 dataset.



(a) (b) (c)

Fig. 6.1 Demonstrative 2-D MR scan (a) AD (b) MCI (c) NC

6.3. Proposed Method

In the proposed method, MR images are decomposed up-to fifth level using DD-DTCWT and complex coefficients through all sixteen sub-bands from every level are obtained. Mean Energy features are obtained from these coefficients from level 1 to level 5. The mean energy of wavelet coefficients at various stages of decomposition, can reflect the texture and strength of the frequency components in an image and that for a particular sub-band also. The lower and higher decomposition levels capture high-frequency details (fine textures, edges) and low-frequency, coarser features (smooth areas or broader shapes) respectively. Directional patterns in textures can be obtained by analyzing mean energy in wavelet sub-bands that correspond to specific orientations (horizontal, vertical, diagonal). The mean energy of sub-band reflects the small variations in brain tissue textures that correlate with Alzheimer's disease development. Mean energy [175] acts as a compact statistical feature which represents the distribution of textural and structural information in an image. Fig. 6.2 illustrates the workflow of the proposed technique through a pipeline model.

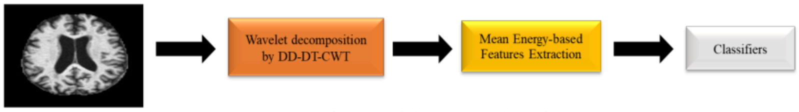


Fig. 6.2 Pipeline model: Proposed work

In Fig. 6.2, mean energy has been obtained from sixteen sub-bands from first level to fifth level of DD-DTCWT decomposition. DD-DTCWT produces sixteen sub-bands per level, so upto fifth level, 80 complex sub-bands will be there. Mean Energy features have been obtained from all 80 real and 80 imaginary sub-bands. The proposed work compares the effectiveness of different combinations of mean energy. These combinations are.

1. Mean Energy of real coefficients
2. Mean Energy of imaginary coefficients
3. Mean Energy of real and imaginary coefficients taken separately
4. Mean Energy of complex coefficients
5. Variance of magnitude of complex coefficients
6. Entropy of complex coefficients

Other than first four mean energy-based features, results have been investigated with entropy and variance also, so two more combinations of features are included as no.5 and 6.

Entropy is a metric that represents randomness in an image. High entropy indicates unpredictability of texture and lower entropy means smoother or uniform regions. The sub-bands provide spatial and frequency components after wavelet decomposition of a signal. The entropy from these sub-bands can represent textural and structural abnormalities, which can help in AD detection. Variance measures the deviation of pixel intensity from the average value. High variance means texture is more diverse and possibly less uniform. Low variance means texture is more uniform, with pixel intensities being similar to the mean. The intensity distribution in MR images is frequently changed by structural changes due to Alzheimer's disease, such as cortical thinning and hippocampal atrophies. These modifications are captured by variance in the form of intensity variations in sub-bands.

These features are calculated using Eqn. from (6.1) to Eqn. (6.6).

- The Mean Energy of real components at oth orientation and lth level of decomposition

$$M_{1,l,o} = \frac{1}{N}\sum_{i}(wr_{i,l,o})^2 \qquad (6.1)$$

where $wr_{i,l,o}$ is the real component of wth complex coefficient ($w_{i,l,o}$).

- The Mean Energy of imaginary components at oth orientation and lth level of decomposition

$$M_{2,l,o} = \frac{1}{N}\sum_{i}(wi_{i,l,o})^2 \qquad (6.2)$$

where $wi_{i,l,o}$ is the imaginary component of wth complex coefficient ($w_{i,l,o}$)

- The Mean Energy of real and imaginary components at oth orientation and lth level of decomposition

$$M_{3,l,o} = \frac{1}{N}\sum_{i}(wr_{i,l,o})^2 , \frac{1}{N}\sum_{i}(wi_{i,l,o})^2 \qquad (6.3)$$

- The Mean Energy of complex components at oth orientation and lth level of decomposition

$$M_{4,l,o} = \frac{1}{N}\sum_{i}[(wr_{i,l,o})^2 + \sum_{i}(wi_{i,l,o})^2] \qquad (6.4)$$

- The variance of complex coefficients

$$M_{5,l,o} = \frac{1}{N}\sum_{i}[\,|w_{i,l,o}| - \frac{1}{N}\sum_{i}|w_{i,l,o}|^2\,] \qquad (6.5)$$

- The Entropy of complex coefficients at oth complex sub-band

$$M_{6,l,o} = -\sum_{i=1}^{N} p_{i,l,o}\ln(p_{i,l,o}) \qquad (6.6)$$

Where $p_{i,l,o} = \frac{E_{l,o}}{E_{total}}$, $E_{l,o} = \frac{1}{N}\sum_{i=1}^{N}|w_{i,l,o}|^2$ and $E_{total} = \sum_{l}\sum_{o}\sum_{i=1}^{N}|w_{i,l,o}|^2$

and $p_{i,l,o}$ is relative wavelet energy, E_{total} is the total energy, $E_{l,o}$ is the wavelet energy at lth decomposition and oth sub-band. The complete steps for extracting features are outlined in Fig. 6.3.

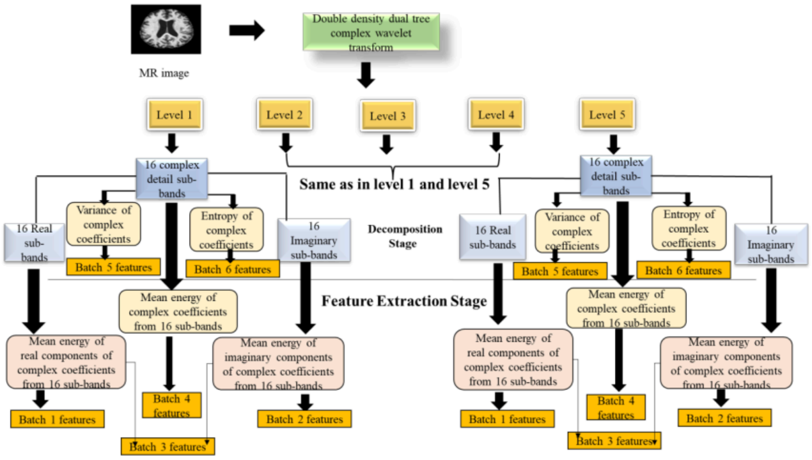


Fig. 6.3 Detailed process for extracting features

In Fig. 6.3, features for all six batches have been extracted from all five level of decompositions. These Batches are as

1. Batch 1 includes the mean energy of real components of complex coefficients. It includes 80 mean energy features as 80 real sub-bands will be there up to fifth level
2. Batch 2 includes the mean energy of imaginary components of complex coefficients. It includes 80 mean energy features as 80 imaginary sub-bands will be there up to fifth level
3. Batch 3 includes the mean energy of real and imaginary components used independently. It includes 160 (80+80) mean energy features as 80 imaginary and 80 real sub-bands will be there up to fifth level
4. Batch 4 includes the mean energy of complex coefficients. It includes 80 mean energy features as 80 complex sub-bands will be there up to fifth level
5. Batch 5 includes the variance of complex coefficients. It includes 80 features as 80 complex sub-bands will be there up to fifth level.
6. Batch 6 includes the entropy of complex coefficients. It includes 80 features as 80 complex sub-bands will be there up to fifth level

Each feature value from all batches is normalized to lie in the range [0,1]. In Eqn. (6.7), N_{min} and N_{max} are the minimum and maximum values of features within a particular batch and N_{lm} represents feature's normalized value (N) within the same batch.

$$N_{lm} = \frac{N - N_{min}}{N_{max} - N_{min}} \tag{6.7}$$

All features from Batch 1 to Batch 6 are thus, normalized by using Eqn. (6.7). These features have been used for AD classification. The normalized features are input to several independent classifiers viz. KNN, DT, Twin SVM, Cubic SVM, Linear SVM, LDA, FNN, ELM, and NB to determine the proposed approach's efficacy.

The proposed approach has also been assessed using alternative wavelets such as DTCWT and DD-DWT. Table 6.1. indicates the feature selection rule with all three wavelets for all six batches.

Table 6.1 Feature selection rule for all batches with different wavelets

	Batch 1			**Batch2**		
Wavelets	**Number of sub-bands upto 5th level**	**Features count upto fifth level**	**Feature vector**	**Number of sub-bands upto 5th level**	**Features count upto fifth level**	**Feature vector**
DT-CWT	30 (6 Ṙ subbands at each level)	30	Batch 1= {$M_{1,1,o}$ } l = 1, 2, . . .5, o = 1, 2, . . ., 6	30 (6 Î subbands at each level)	30	Batch 2= {$M_{2,1,o}$ } l = 1, 2, . . .5, o= 1, 2, . . ., 6
DD-DWT	40(8 Ṙ subbands at each level)	40	Batch 1= {$M_{1,1,0}$ } l = 1, 2, . . 5, o = 1, 2, . . ., 8	**Î sub-band does not exist**	NIL	NIL
DD-DT-CWT	80(16 Ṙ subbands at each level)	80	Batch 1= {$M_{1,1,o}$ } l = 1, 2, . . .5, o = 1, 2, . . ., 16	80(16 Î subbands at each level)	80	Batch 2= {$M_{2,1,o}$ } l = 1, 2, . . .5, o= 1, 2, . . ., 16
	Batch 3			**Batch4**		
DT-CWT	30 +30 (6 Ṙ and 6 Î sub-bands at each level)	30+30=60	Batch 3= {$M_{3,1,o}$ } l = 1, 2,. .5 o = 1, 2, . . ., 6	30 (6 complex sub-bands at each level)	30	Batch 4= {$M_{4,1,o}$ } l = 1, 2, . . .5 o = 1, 2, . . ., 6
DD-DWT	**Î sub-band does not exist**	NIL	NIL	**Î sub-band does not exist**	NIL	NIL
DD-DT-CWT	80+80(16 Ṙ and 16 Î sub-bands at each level)	80+80=160	**Batch 3 ={ $M_{3,1,o}$ }** **l = 1, 2, .5** **o = 1, 2, . . ., 16** **[Proposed]**	80(16 complex sub-bands at each level)	80	Batch 4= {$M_{4,1,o}$ } l = 1, 2,. .5 o = 1, 2, . . ., 16
	Batch 5			**Batch 6**		
DT-CWT	30 (6 complex sub-bands at each level)	30	Batch 5= {$M_{5,1,o}$ } l= 1, 2, . . .5 o = 1, 2, . . ., 6	30 (6 complex sub-bands at each level)	30	Batch 6= {$M_{6,1,o}$ } l= 1, 2, . . .5 o = 1, 2, . . ., 6
DD-DWT	**Î sub-band does not exist**	NIL	NIL	**Î sub-band does not exist**	NIL	NIL
DD-DT-CWT	80(16 complex sub-bands at each level)	80	Batch 5= {$M_{5,1,o}$ } l= 1, 2,. .5 o = 1, 2, . . ., 16	80(16 complex sub-bands at each level)	80	Batch 6= {$M_{6,1,o}$ } l = 1, 2, . ..5 o= 1, 2, . . ., 16

From Table 6.1, each wavelet DTCWT, DD-DWT and DD-DTCWT provide different number of features for all six batches depending upon the number of their real, imaginary and complex sub-bands up to fifth level of decomposition.

6.3.1. Results for Binary Classification (AD, MCI and NC)

The proposed work is implemented by using MATLAB 19. Classification is done using nine classifiers viz. CSVM, LSVM, KNN, DT, TSVM, NB, FNN, ELM, and LDA. The 10-

fold stratified cross validation approach partitioned the dataset into training and test data. After 10 runs of the proposed method, the average values of accuracy, sensitivity, and specificity for each classifier are determined. The results with KNN classifier are outstanding. Across all batches, the KNN consistently performs well and yields the best results with batch 3. This is so, because KNN continuously evolves and quickly adjusts to new training data during real-time applications. Its nature is non-parametric. KNN adapts well based on the parameters of its neighbours without requiring assumptions about the datasets. The KNN model works effectively in non-linear scenarios as well because it is flexible in nature. It depends only on the distance between data points and does not require any kernel for adapting itself according to any complex or nonlinear distributions [176].

Table 6.2 shows the results of all six batches with all three wavelets on OASIS and ADNI datasets. DD-DWT does not give imaginary coefficients thus it generates features for only one batch i.e., Batch 1. DTCWT and DD-DTCWT provide features for all six batches.

Table 6.2 Classifiers performance assessment for AD/NC classification.

		OASIS DATASET			**ADNI DATASET**		
Batch/ Feature Name	**Classifier**	**DT-CWT Acc/Sen/Spe**	**DD-DWT Acc/Sen/Spe**	**DD-DT-CWT Acc/Sen/Spe**	**DT-CWT Acc/Sen/Spe**	**DD-DWT Acc/Sen/Spe**	**DD-DT-CWT Acc/Sen/Spe**
Batch 1	k-NN	94.2/93/96	95.4/96/95	96.9/97.3/96.4	93.2/93/94	95.4/94/96.2	97/94.4/98.9
	TSVM	82/78.8/85.3	83.1/82.5/84.5	84.2/85.4/81.2	90.6/85.8/95.6	90.6/85.3/95.4	91.5/92.6/90.8
	CSVM	92/90/94	95.1/95/96	96/97.3/94.6	92.7/94/92	97.1/97/99	97/96.6/97.8
	LSVM	81.7/81/82	73.4/84/63	83.4/84.1/81.2	84.9/84/85	87.5/84/91	87.9/87.7/88
	NB	67/47/87	69.1/69/70	67.4/67/65	68.2/66/71	74.5/66/83	79.1/71.1/86.9
	FNN	95.5/94.7/96.4	96.1/95.1/97	96.8/95.8/97.1	89.7/93.8/84.6	92.7/91/94.6	93.2/92.8/93.7
	ELM	70.84/71.6/69.5	68.1/72.7/64.1	67.5/69/65	71.8/73.2/69.3	73.7/75.8/71.5	77.56/78.6/75.5
	LDA	81.3/85/78	70.8/72/78	84.8/88/81	93/94/93	88.5/89/89	94.3/95.5/93
	DT	71/73/69	71.9/77/67	76.3/76.7/75.8	74/75/73	77.4/76/79	77.5/77.7/77.1
Batch 2	k-NN	95.1/96/95	The Imaginary part does not exist	98.7/98/99	94.8/93/96	The Imaginary part does not exist	97.2/98/96
	TSVM	80.8/85.4/75.6		77.7/82.2/72.4	78.6/83.1/73.1		80.2/73.8/86.5
	CSVM	93.3/93/94		98.2/99/97	92.7/92/94		94.5/95.5/93.6
	LSVM	79.9/87/73		84.8/84/86	82.3/78/86		91.7/90/94
	NB	69.6/71/69		70.1/72/68	71.9/74/70		68.2/56/80
	FNN	95.1/93.2/97.2		97.1/98.2/96.2	94.3/93/96.7		95.2/95.7/93.9
	ELM	76.2/75.7/78.4		76.6/73.91/79.6	76.9/78.8/74.4		77.6/79.5/75.4
	LDA	76.3/80/72		89.3/87/92	78.6/76/81		93.8/93/95
	DT	76.8/83/71		77/78/76	74.5/77/72		77.1/76/78
Batch 3	k-NN	97.3/98.2/96.4	The Imaginary part does not exist	99.1/98.2/100	97.9/97/98	The Imaginary part does not exist	99.5/98.9/100
	TSVM	85.9/91.2/80.7		87.8/90.4/85.4	89/92.5/86.2		90.7/91.4/93.6
	CSVM	97.3/96/99		97.8/99.1/96.4	95.3/97/94		98.5/97.8/100

	LSVM	82.1/86/79		85.7/86.6/84.8	94.2/98/91		95.2/96.3/94.1
	NB	69.2/65/73		72.8/84/62	77.5/77.7/77.1		79.5/82/77
	ELM	73.3/71.1/75.2		73.6/71.7/75	77.7/79.3/75.3		78.7/79.1/76.4
	FNN	95.1/94.7/95.5		98.7/97.4/100	97.2/96.9/98.9		98.5/100/96
	LDA	81.7/83/80		87.1/87/88	87/86.4/88.1		92.2/94.1/90.2
	D T	75.4/74/77		79.5/80/79	81.8/80/83		90.7/91.1/90.2
Batch 4	k-NN	94.2/96.4/91.9	The Imaginary part does not exist	98.2/99.1/97	94.8/97/93	The Imaginary part does not exist	97.5/94.4/99.5
	TSVM	74.78/75.29/74.51		76/73.2/77.1	86.3/84.7/88.6		89.81/87.2/91.2
	CSVM	93.3/94.6/91.9		95.5/93/98	93.8/92/96		95.9/98.8/92.9
	LSVM	76.3/80.3/72.3		84.4/83/86	85.9/88/84		87.9/92.2/83.6
	ELM	74.5/76.2/72.4		74.7/76.5/72.5	73.4/75.3/71.3		74.6/75.6/72.3
	FNN	82.6/84.8/80.7		83/83.5/82.5	81.3/80.6/81.9		84.9/93.5/79.1
	NB	69.2/59.8/78.5		70.5/83/58	66.1/67/66		70.9/70/71.73
	LDA	76.8/75/78.5		90.6/92/89	91.7/93/91		96.7/97.7/95.6
	DT	75/75.8/74.1		80.4/84/77	77.6/77/78		80.2/78.8/81.5
Batch 5	k-NN	89.3/87/92	The Imaginary part does not exist	95/96/95	90.1/90/91	The Imaginary part does not exist	94.3/92/97
	TSVM	83.5/83/84		87.9/86/90	89.7/91/88		90.6/93.8/88.1
	CSVM	87.9/88/88		90.6/89/92	94.8/94/96		95.8/95/97
	LSVM	79/78/80		79.5/78/80	82.8/79/86		85.4/81/90
	ELM	71.2/73.6/69.5		76.8/78.4/74.6	70.6/71.5/69.7		74.2/76.3/72.7
	FNN	83.5/82.1/85		89.7/88/91.6	85.9/91.6/81.7		90.6/90.6/90.6
	NB	71/80/62		71.4/68/75	65.1/64/67		68.8/69.1/67.2
	LDA	73.7/75/74		78.1/77/79	91.5/92/91		94.3/95/94
	DT	79.5/84/75		80.4/82/79	76/71/81		81.3/77/85
Batch 6	k-NN	96/97/95	The Imaginary part does not exist	97.3/98/96	94.8/93/97	The Imaginary part does not exist	97.5/98/96
	TSVM	75.8/78.6/72.5		87.9/83.7/91.3	84.4/74.8/92.2		87.1/86/90
	CSVM	91.1/89/93		97.3/98/96	96.9/95/99		97/98/96
	LSVM	73.7/76/71		83.5/88/79	88.5/89/89		96.9/99/95
	ELM	68.9/70.1/67.5		70.8/69.1/71.4	78.3/79.3/77.4		83.2/85.1/81.3
	FNN	88.8/87.2/90.7		94.2/97.1/91.6	96.9/96.9/96.9		96.9/97.9/95.9
	NB	71/71/71		78.1/79/78	71.9/73/71		86.5/95/78
	LDA	79/77/81		91.5/92/91	90.6/91/91		96.4/98/95
	DT	72.3/72/72		75.9/79/73	81.3/82/80		91.1/92/91

Table 6.2 indicates that DD-DTCWT performs better than DTCWT and DD-DWT for all six batches and nearly with all nine classifiers as DD-DTCWT extracts features in 16 different directions. Batch 3 is the best performer among all batches. In Batch 1, KNN and FNN perform well with accuracy of 96.9% and 96.8% respectively for DD-DTCWT using OASIS dataset. In Batch 2, the best classification accuracy is 98.7% with KNN for DD-DTCWT. Additionally, in Batch 3, KNN outperforms FNN having 98.7% accuracy and KNN with

99.1% accuracy for DD-DTCWT. KNN is the best performer in batches 4, 5, and 6, with accuracy of 98.2%, 95%, and 97.3%, respectively for DD-DTCWT.

On the other hand, with ADNI dataset in Batch 1, KNN and Cubic SVM perform well with an accuracy of 97% using DD-DTCWT. Using same wavelet, KNN performs well in Batch 2, yielding an accuracy of 97.2% and performs even better in Batch 3 with 99.5% accuracy. FNN and Cubic SVM provide an accuracy of 98.5% in Batch 3 with DD-DTCWT. The best performance is shown by KNN in batches 4, 5, and 6, with accuracy of 97.5% ,94.3% and 97.5% respectively for DD-DTCWT. This data shows that KNN performs consistently in all batches and produces the most promising outcomes with Batch 3.

Finally, the proposed technique i.e. batch 3 yields classification accuracy, sensitivity, and specificity of 99.1%, 98.2%, and 100% respectively for AD and NC classification, 95.8%,96%, and 93% respectively for AD and MCI classification and 95.6%,96% and 94% respectively for MCI and NC classification using OASIS dataset. While on the ADNI dataset, these figures are 99.5%, 98.9%, and 100% for AD and NC classification,95.1%,94%, and 96% for AD and MCI classification, and 94.8%,95% and 93% for MCI and NC classification.

Fig.6.4 shows the accuracy of different classifiers on different batches with OASIS dataset.

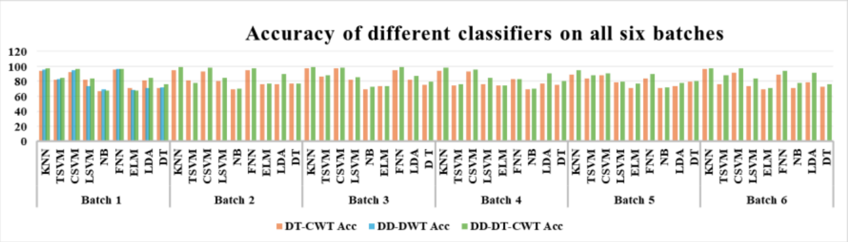
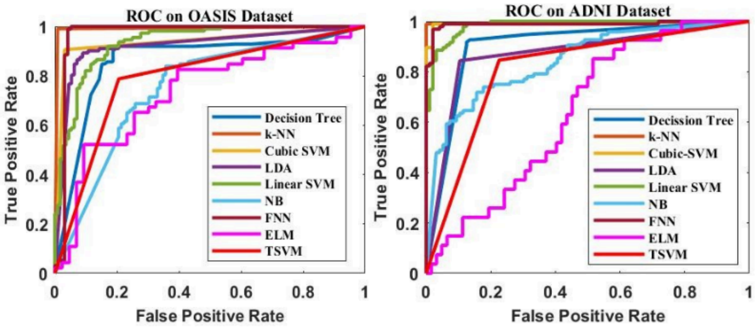


Fig. 6.4 Accuracy for different classifiers and batches on the OASIS dataset

In Fig. 6.4, DD-DWT does not provide imaginary coefficients so the accuracy for all classifiers with DD-DWT is dropped to zero beyond Batch 1. Best performance is shown by DD-DT-CWT, with peaks at k-NN classifier. Fig 6.5 illustrates the ROC plots of different classifiers with DD-DTCWT and batch 3 on the OASIS and ADNI Datasets.



(a) (b)

Fig. 6.5 ROC Plots (a) OASIS (b) ADNI

In Fig. 6.5, KNN performs exceptionally good on both datasets. The ROC curve steeply ascends towards the upper-left quadrant, demonstrating that the proposed batch gets high TPR value with low FPR value using KNN classifier. While FNN and Cubic SVM also exhibit strong performance, KNN consistently gives best performance in nine executions of the proposed algorithm. Tables 6.3 and 6.4 show the performance assessment across various batches for AD/MCI and MCI/NC classifications respectively on both datasets.

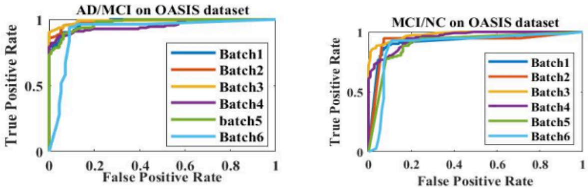
Table 6.3 Performance assessment across various batches for AD/MCI classification

		OASIS DATASET			ADNI DATASET		
Batch No.	**Classifier**	**DT-CWT**	**DD-DWT**	**DD-DT-WT**	**DT-CWT**	**DD-DWT**	**DD-DT-WT**
		Acc/Sen/Spe	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**
Batch 1	k-NN	85.7/87/85	87.9/88/88	90.2/92/88	89.3/94/85	91.5/88/96	92.4/91/94
Batch 2		87.1/88/87	The Imaginary part does not exist	92.4/90/95	90.2/94/87	The imaginary part does not exist	93.8/92/96
Batch 3		89.7/91/88		**95.8/96/93**	92.9/92/94		**95.1/94/96**
Batch 4		86.2/88/84		91.5/94/89	88/89/87		92.9/94/92
Batch 5		86/87/86		89.7/94/86	87.5/85/90		89.1/84/95.1
Batch 6		87.1/89/85		92/95/89	89.3/84/95		90.6/88/94

Table 6.4 Performance assessment for various batches for MCI/NC classification

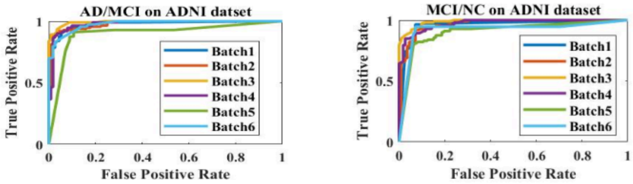
		OASIS DATASET			ADNI DATASET		
Batch No.	**Classifier**	**DT-CWT**	**DD-DWT**	**DD-DT-CWT**	**DT-CWT**	**DD-DWT**	**DD-DT-CWT**
		Acc/Sen/Spe	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**
Batch 1	k-NN	82.1.81/83	88.8/93/85	89.7/90/95	84.4/83/86	87/87/88	90.6/88/94
Batch 2		84/83/87	Imaginary part does not exist	91.1/87/95	85.3/86/85	Imaginary part does not exist	91.5/93/90
Batch 3		85/81/89		**95.6/96/94**	87/85.5/90		**94.8/95/93**
Batch 4		81/87/77		86.3/86/87	83.5/89/78		89.7/93/87
Batch 5		78.1/77/79		83.5/83/84	80.8/89/72		85.3/77/94
Batch 6		87.1/88/86		90.2/91/89	87.9/88/88		91.1/88/94

Tables 6.3 and 6.4 show that Batch 3 is the best performer on both datasets using DD-DTCWT. Tables 6.3 and 6.4 include the experimental results of KNN classifier. The performance of other classifiers like CSVM, LSVM, k-NN, DT, TSVM, NB, FNN, ELM, and LDA has also been analysed. These results are not included due to their low performance in comparison to KNN classifier. The excellent classification outcomes of Batch 3 validate the effectiveness of the proposed feature. This feature captures the grey matter density and structural atrophies and retains the ability to identify AD in its earliest stage i.e., MCI stage. Figs. 6.6 and 6.7 show the ROC plots for classifying AD with MCI and MCI with NC using KNN classifier on the OASIS and ADNI datasets respectively.



(a) (b)

Fig. 6.6 KNN classifier performance on OASIS dataset (a) AD/MCI (b) MCI/NC



(a) (b)

Fig. 6.7 KNN classifier performance on ADNI dataset (a) AD/MCI (b) MCI/NC

In Figs. 6.6 and 6.7, Batch 3 is the best performer compared to other batches. The ROC curve of batch 3 climbs sharply towards the upper-left section of the plot. This reflects the model's effective classification capability, with high TPR and minimal FPR.

6.3.2. Proposed Method against Existing Methods

The proposed method has been benchmarked against fourteen existing methods. All methods belong to the different categories of feature extraction methods. All the existing methods have been implemented on MATLAB 19 using the same datasets, i.e. OASIS and

ADNI. All the algorithms have been executed ten times, and their average results are noted in Table 6.5.

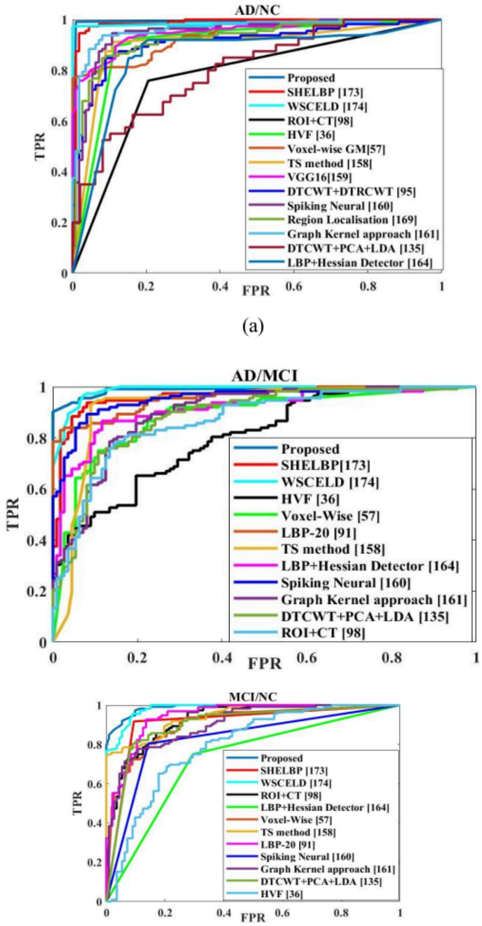
Table 6.5 Proposed vs Existing Methods

		OASIS			ADNI		
Method	**Classifier**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**	**Acc/Sen/Spe**
Proposed	**KNN**	**99.1/98.2/100**	**95.8/96.0/93.0**	**95.6/96.0/94**	**99.5/98.9/100**	**95.1/94/96**	**94.8/95/93**
Shifted ELBP [173]	KNN	95.1/95/96	92/90/94	91.5/93/90	94.8/94/95	91.9/90/92	91.1/92/91
WSCELD [174]	DT	97.3±1.6/97.1±1.2/97.2±1.1	94.6±1.1/96.1±1.2/93.1±1.1	93.8±1.4/92.4±1.5/96.2±1.2	97.1±1.6/96.7±1.2/98.2±1.1	93.6±1.1/94.1±1.2/91.1±1.1	92.8±1.4/91.4±1.5/93.2±1.2
DTCWT+PCA+LDA [135] [2018]	ELM	88.2/86.5/84.4	80.2/78.4/76.3	82.1/76.3/85	88.4/89.5/87.6	78.4/75.2/79.8	81.4/75.4/84.2
LBP+ Hessian detector [164] [2021]	CNN	87.5/89.0/86.0	86.2/83.0/89.0	71.9/72.0/71	86.2/87/87	85.9/85/86	72.8/70/76
ROI +CT [98] [2022]	SVM	81.1/83.3/79.2	78.1/83.0/73.0	80.4/82.0/79	92.9/94.1/91.2	82.7/84.6/80	83.1/87.8/76.6
HVF method [36] [2015]	SVM	85.3/77.0/94.0	74.0/75.0/73.0	78.1/77.0/79	85.7/77.6/91.2	71.5/75.9/71.2	76.2/72.3/81.5
Voxel-wise GM [57] [2012]	SRC ensemble	89.3/87.0/92.0	81.7/86.0/78.0	83.5/88.0/79	90.8/86.3/94.8	83.5/89/78	81.7/87/77
TS method [158] [2015]	SVM	88.5/89.0/86.0	86.2/87.0/86.0	85.3/86.0/85	89.9/92.3/86.9	81.5/83.7/80.1	82.4/84.1/80.5
VGG16 [159] [2022]	NN	86.7/85.2/87.3	-	-	85.5/86/84	-	-
DTCWT+DTRCWT [95] [2022]	FNN	90.7/91.2/92.4	-	-	90.2/91.7/88.3	-	-
Spiking Neural [160] [2021]	-	90.5/92.0/89.0	89.7/89.0/90.0	86.2/85.0/88	90.1/94/87	87.1/89/85	83.9/80/86
Region localisation-based method [169] [2023]	-	88.4/85.0/92.0	-	-	87.5/85/90	-	-
Graph-kernel Approach [161] [2023]	SVM	92.0/90.0/94.0	82.6/88.0/77.0	81.3/82.0/80	90.8/94/87	83.5/85/82	82.1/78/87
3D-DWT+LBP-20 [91] [2020]	SVM	-	91.5/92.0/91.0	89.3/84.0/95	-	92/91/93	90.3/90.2/90.2

As demonstrated in Table 6.5, the proposed method attains remarkable performance over existing algorithms. The proposed method is comparatively simpler and more accurate than existing algorithms. It does not require any segmentation process, which is required with existing techniques [36,57]. The brain's structure is complicated, which makes it difficult to perform accurate segmentation. The proposed method used only 160 features for AD classification, while existing techniques [173,174] use 768 features [173] and 4096 features [174] for AD classification. The existing techniques use PCA to reduce feature dimensionality. With only 160 features involved in classification, the proposed approach eliminates the need for dimensionality reduction while retaining spatial details.

The significant contribution of the proposed work lies in the introduction of a new feature batch i.e. mean energy of real and imaginary components taken independently for AD classification. The proposed method gives outstanding outcomes for early AD detection at the

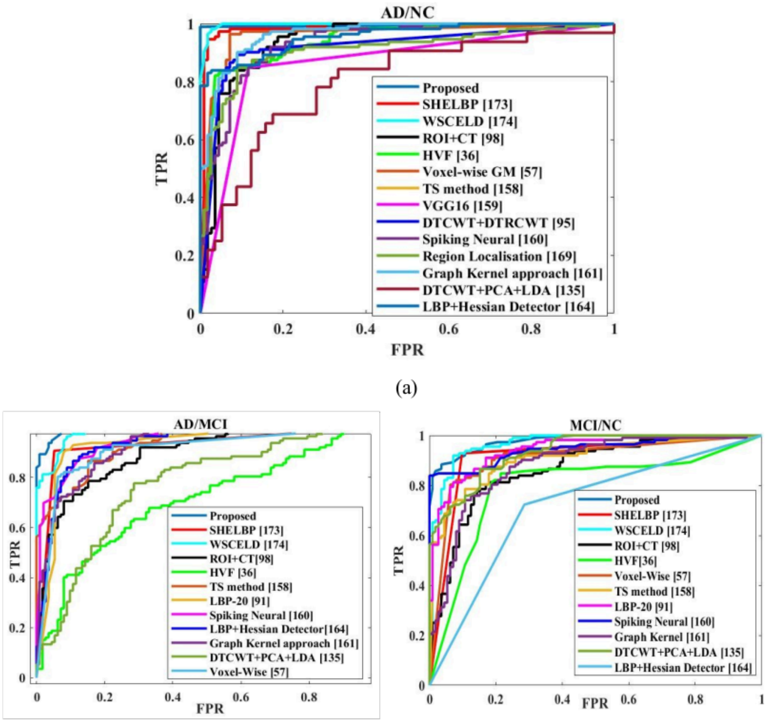
MCI stage as the technique captures the information about grey matter and structural changes with the advancement of the disease. Fig. 6.8 and 6.9 display the ROC curves comparing the proposed method with existing approaches on both datasets.



(a)

(b) (c)

Fig. 6.8 Evaluation of Existing and proposed methods through ROC curves (a) AD/NC (b) AD/MCI (c) MCI/NC on OASIS dataset



(a)

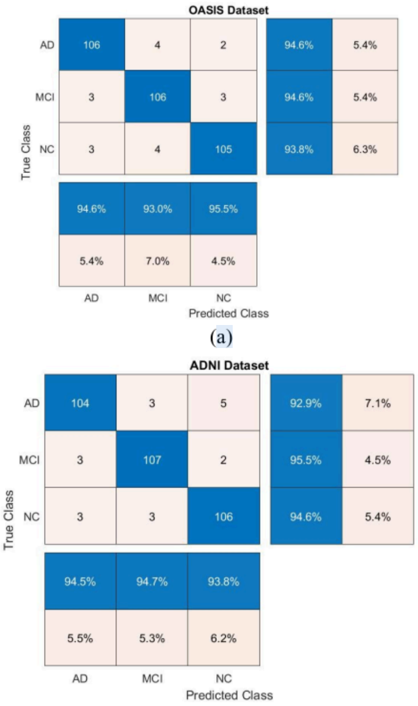
(b) (c)

Fig. 6.9 Evaluation of Existing and proposed methods through ROC curves (a) AD/NC (b) AD/MCI (c) MCI/NC on ADNI dataset

The ROC plots in Figs. 6.8 and 6.9 for the proposed method rise steeply towards the upper-left quadrant. This highlights that the proposed method holds high TPR value and low FPR value which is responsible for high sensitivity and specificity.

6.3.3. Multi-class Classification Results

The proposed method has been further assessed for multi-class classification, and the corresponding confusion matrices for both datasets are demonstrated in Fig. 6.10.



(a)

(b)

Fig. 6.10 Confusion Matrices (a) OASIS Dataset (b) ADNI Dataset

The proposed technique provides 96.2% accuracy,94.3% sensitivity and 97.1% specificity for multiclass classification on OASIS dataset. With ADNI dataset, these figures are 96.2% accuracy,94.3% sensitivity and 96.8% specificity

6.4. Summary

This chapter introduces mean energy-based feature extraction technique for early AD detection. The proposed technique introduces a novel feature, i.e., mean energy of real and imaginary components taken independently for early AD detection. This feature efficiently captures the grey matter density and structural abnormalities in the brain and retains the ability to detect AD in its earliest stage, i.e., the MCI stage. The proposed technique provides exceptional results for binary as well as multiclass classifications on both datasets, thus minimizing the problem of scalability. This technique is suitable for early AD detection due to its outstanding performance in both AD/MCI and MCI/NC classification. This technique is simple and utilizes only 160 features for AD classification. It does not require any feature dimensionality reduction and feature selection method, thus preventing the spatial information loss resulting from feature dimensionality reduction methods.

In proposed techniques, DD-DTCWT has been applied on 2-D MR images to get sixteen sub-band images and then mean energy features have been obtained from these sixteen sub-band images. The proposed technique investigates the performance of different types of features like mean energy of real coefficients, mean energy of imaginary coefficients, mean energy of real and imaginary coefficients used independently, mean energy of complex coefficients, variance, and entropy of complex coefficients. The results with mean energy of real and imaginary coefficients taken independently are outstanding. The proposed method obtained an accuracy of 99.1%, sensitivity of 98.2%, and specificity of 100% for AD/NC classification with OASIS dataset. For AD/MCI classification, the corresponding values are 95.8%, 96%, and 93%, while for MCI/NC classification, they are 95.6%, 96%, and 94%, respectively. With ADNI dataset, these figures are 99.5%, 98.9% and 100% for AD/NC classification ,95.1%,94.0% and 96.0% for AD/MCI classification ,94.8%,95.0% and 93.0% for MCI/NC classification respectively. With multiclass classification, respective results are 96.2%, 94.3%, and 97.1% on OASIS dataset and 96.2 %, 94.3%, and 96.8% on ADNI dataset.

The publication related to this chapter

- Neha Garg and M.S. Chaudhary, "Implementation of Dual Tree Complex Wavelet Transform with Mean Energy Features to detect Alzheimer's Disease," 2021 7th International Conference on Signal Processing and Communication (ICSC), 2021, pp. 188-193, doi: 10.1109/ICSC53193.2021.9673484
- Neha Garg, M. S Choudhry, & R. M. Bodade, "Detection of Alzheimer's Disease using Double Density Dual-Tree Complex Wavelet Transform with Mean Energy Features." **Expert Systems with Applications**, SCIE Impact factor-7.5 (**Communicated)**

Chapter 7

Conclusion, Future Scope and Social Impact

7.1. Introduction

This research work proposes and effectively applies seven feature extraction methods that contribute to the early diagnosis of AD. Early AD detection is a cornerstone of effective Alzheimer’s disease management. It speeds up research, lessens societal and financial responsibilities, empowers people and families, and helps healthcare professionals to provide focused treatment. By prioritizing early detection, the door to a future is opened where Alzheimer's disease person is better understood, treated, and maybe prevented. Early AD detection faces several challenges. These challenges may result from shortcomings in social awareness, ethical issues, healthcare accessibility, and diagnostic techniques. This is an important problem in the research field.

This research work presents different feature extraction techniques for early AD detection. These techniques extract features from the whole brain. In contrast to ROI-based techniques, these techniques avoid extracting any affected areas of the brain. The proposed techniques in the research work use complex wavelets, i.e., DD-DTCWT, for extracting features from 2-D MR images. The complex wavelet provides imaginary coefficients that extract structural details in the image and help in capturing structural atrophies in the brain due to AD progression. The proposed techniques are applied to high frequency sub-bands, which helps in providing the high frequency variation in grey matter density due to the progression of AD. This is the significance of complex wavelets for early AD detection. The proposed techniques provide both biomarkers, viz. structural atrophies and grey matter density variation for AD detection. On the other hand, existing techniques focus on a single biomarker; for example, VBM focuses on grey matter density maps, and ROI-based methods focus on structural atrophies in the brain. The entire work is organized into a sequence of seven chapters.

The introductory content is covered in Chapter 1, with Chapter 2 presenting the foundational literature review. Chapter 3 outlines the first methodology for early AD detection. In this technique, detail coefficients at fifth level are used for AD classification.

This technique provides good accuracy for AD classification with MCI and NC. However, it suffers from high dimensionality and processing time. Chapter 4 presents histogram-based feature extraction techniques. These techniques operate with lower dimensionality, making them more efficient compared to the direct coefficient-based feature extraction technique. These techniques provide directional local features that are more robust to noise in comparison to direct coefficient-based feature extraction techniques. Chapter 5 presents the statistical feature extraction techniques. Statistical features provide a compact representation (e.g., mean, variance, skewness, entropy) that significantly reduces dimensionality while preserving key information. Statistical features are less sensitive to noise as they focus on the overall distribution of data rather than individual bins. Chapter 6 presents Mean-Energy based feature extraction technique for early AD detection. This technique is quite simple in comparison to previous techniques. This technique does not suffer from a dimensionality problem and thus does not require any dimensionality reduction or feature selection technique. The results with this technique are outstanding for binary as well as multi-class classification.

7.2. Conclusion

The overall findings of this research lead to the following conclusions.

- An Intensive Literature Review has been done, which indicates the importance of wavelet transform-based methods over existing methods for extracting features in AD classification.
- The features extracted through Complex wavelets provide better classification results in comparison to real wavelets. A complex wavelet provides real and imaginary coefficients, while a real wavelet provides only real coefficients. Imaginary coefficients capture the edge and structural details nicely.
- The details sub-bands have been used for feature extraction, which provides directional and localised features compared to approximate sub-bands. The detail sub-bands also capture the high frequency variation in grey matter density, which is an important biomarker for detecting AD.
- Among all classifiers like SVM, NB, DT, and RF, the best performance is shown by the KNN classifier in most of the proposed methods. This is so because KNN is non-parametric in nature, i.e., KNN does not require assumptions regarding the datasets & it

successfully adjusts based on neighbour's parameters. It tends to constantly evolve & immediately adapt to new training data during real-time applications.

- Among all feature selection methods like PSO, ACO, and GA, the best results are shown by the Genetic Algorithm in both proposed methods [Direct-coefficients and GLCM-based] using feature selection process. This is so because GA has a tendency not to trap in local optima while ACO and PSO are more prone to trap in local optima. Local optima refer to a situation where the algorithm converges to a solution that is better than neighbouring solutions but is not the best possible solution.
- The classification results with statistical features obtained from WSELBP, WSCELD, and GLCM are better than the case when EWLBP and WSCELD histograms are used as features. Statistical features summarize the histograms into meaningful values, thus providing more compact, interpretable, robust, and lower dimensional features.
- Energy features have shown better performance than the other statistical features like entropy and variance. This is verified from two proposed techniques (Energy version of WSCELD with histogram features and Mean Energy of real and imaginary coefficients). This is so because energy is proportional to the squared magnitude of coefficients and, thus, less sensitive to noise. LBP-based and GLCM-based techniques are more sensitive to noise in comparison to energy-based methods. LBP is sensitive to small local changes due to noise, and thus, texture patterns can be easily disturbed, which leads to errors in classification. In the case of GLCM, the spatial relationship between the pixels gets disturbed due to noise, which lowers the classification results.

7.3. Future Scope

Future research can be targeted to deep learning approaches to extract features and fuse these features with wavelet-based features. The proposed techniques can be evaluated with multimodalities. In future, research can be focussed on the advancements in imaging, biomarkers, AI-driven analytics, and multi-modal data integration. Advancements in technology, data science, and medicine open several opportunities to enhance AD detection, progression tracking, and personalized treatment. Key directions for the future are outlined below:

7.3.1. Improved Imaging and Data Fusion

Multimodalities can be used for feature extraction. They can reflect structural as well functional details of the image.

- **Multimodal Analysis:** Experiments may be conducted on multimodalities like PET and f-MRI to see the performance of the proposed techniques.
- **Integration with Genomic Data:**
  Imaging data can be fused with genetic information (e.g., APOE ε4 allele) for a better risk stratification model.
- **Higher-resolution Imaging:**
  Ultra-high-field MRI scanners (7T and beyond) can be used for detecting subtle structural and functional changes.

7.3.2. Integration of CSF and Blood Biomarker

CSF Biomarkers such as tau, phosphorylated tau, and amyloid-β can be combined with blood biomarkers like Plasma Aβ42 and Aβ40 to enhance early AD diagnosis. It can help to track the progression from preclinical to symptomatic stages.

7.3.3. Longitudinal and Predictive Modelling

Longitudinal studies can track disease progression. Virtual models can be designed to predict the disease trajectory.

- **Disease Progression Models:**
  The models may be designed to use baseline and follow-up scans to predict the disease trajectory.
- **Digital Twins:**
  The personalized virtual models of a patient's brain may be created to simulate disease progression and response to interventions.

7.3.4. Enhanced Wavelet Transform Techniques

Different transform techniques can be adopted to obtain more informative and directional features for AD detection. Enhanced wavelet transforms offer greater noise reduction, enhanced feature extraction, and higher diagnostic accuracy for Alzheimer's disease diagnosis.

- **3-D Wavelets:**
  Advanced methods like 3D wavelet transforms can be explored to capture spatial and temporal features from volumetric MRIs.
- **Fusion of features at different scales:**
  Features at multiple scales can be combined to capture both local and global patterns efficiently.

7.3.5. Deep Learning-Driven Feature Extraction

Deep Learning provides automated, accurate, and early-stage diagnosis, which is crucial for timely intervention. It reduces human bias by avoiding the needs of hand-crafted features.

- **Pretrained Models:**
  Fine-tune pre-trained models (e.g., 3D-CNNs, transformers) can be used on medical imaging datasets to extract high-level features.
- **Autoencoders and Generative Models:**
  Deep autoencoders or variational autoencoders (VAEs) can be used to derive compact and informative latent features.

7.4. Social Impact of Early AD Detection

Early AD detection has significant social implications, affecting individuals, families, healthcare systems, and society. These effects are extensive and significant, enabling better outcomes for individuals impacted and more effective use of resources by society. Early AD detection has a revolutionary effect on society. Apart from improving the standard of living for individuals and their families, it also promotes a society that is healthier, more compassionate, and uses resources more effectively. Soon, Alzheimer's disease will be proactively controlled by making investments in early detection. This will result in reducing its negative effects on society and enhancing overall well-being. Early AD detection can impact society directly or indirectly in the following ways:

7.4.1. Direct Impact

An immediate effect on individuals and society due to early AD detection comes under the direct impact. These can be summarized as:

7.4.1.1. Improved Quality of Life and Better Planning

Early AD detection can significantly enhance patients' quality of life. Patients can make plans for their family and future and can live life with more comforts.

- **Early Interventions**: Early detection allows for timely interventions, which helps in slowing down the disease progression, maintains independence, and improves life quality for patients. Cognitive therapies, brain exercises, and lifestyle modifications (diet, exercise) help to maintain mental function longer.
- **Planning and Support**: People can organize their financial and legal affairs, plan, and make well-informed judgments while still cognitively capable.
- **Mental Health Benefits**: Families become emotionally stable by early AD detection. Timely psychological support can be taken, which can lower depression and isolation.

7.4.1.2. Reduced Caregiver Burden

Caregivers for AD patients generally suffer from emotional, physical, and financial stress. Early AD detection helps in reducing their stress by allowing them for better preparation, support, and resource allocation.

- **Preparation:** Families might better prepare emotionally and practically for caring tasks .
- **Support Systems:** Early AD detection gives caregivers early access to resources, support systems, and education.
- **Economic Relief:** Long-term care planning is made possible by early AD detection, which reduces unanticipated financial strain.

7.4.1.3. Advancing Research and Treatment Development

Early AD detection bridges the gap between research and real-world application, ensuring that treatments are developed faster and are more effective.

- **Clinical Trials:** Early AD detection speeds up research and drug development by expanding the pool of eligible clinical trial participants.
- **Biomarker Studies:** Early AD detection helps researchers in their understanding of the disease's early stages, which could result in better preventative measures or treatments.

7.4.1.4. Cost Savings for Healthcare Systems

Early AD detection reduces the cost burden of healthcare systems by avoiding emergencies and providing home care to patients, which is more affordable and sustainable.

- **Reduced Emergency Costs**: Early management can decrease emergency visits and hospitalizations due to late-stage complications.
- **Efficient Resource Allocation**: Early detection enables healthcare organizations to spend money on preventive and early care instead of expensive late-stage therapies.
- **Delaying Institutionalization**: Keeping patients in home settings longer reduces societal costs associated with nursing homes and assisted living facilities.

7.4.2. Indirect Impact

The long-term and broader impact on society due to early AD detection is called an Indirect Impact. These can be summarized as:

7.4.2.1. Ethical and Policy Implications

The legal and insurance systems must adapt to address ethical, financial, and privacy concerns with the advancement in diagnostic capabilities like biomarker tests, AI-powered imaging, and genetic screenings.

- **Healthcare Policies**: Early AD detection highlights the necessity of regulations facilitating resource access and cognitive health screening.
- **Insurance and Legal Frameworks**: Early AD detection encourages reforms to ensure that insurance covers early diagnosis and intervention strategies. Regulations like Health Insurance Portability and Accountability Act and the General Data Protection Regulation govern how medical data, including genetic tests and cognitive screenings, are stored and shared

7.4.2.2. Economic Benefits to Society

Early detection can reduce the rate of severe AD cases, which in turn decrease public healthcare costs. More individuals remain in the workforce longer, reducing economic strain.

7.4.2.3. Social Awareness and Stigma Reduction

Misconceptions regarding AD can be reduced by spreading proper awareness of AD symptoms among patients. Early detection helps AD patients to get emotionally supported by society.

- **Education:** Campaigns can educate people about the signs and symptoms of AD so that they can differentiate between normal aging and AD. Spreading information through media, health care institutions, and community support can lessen stigma and misconception.
- **Inclusion**: Inclusion can reduce stigma by allowing individuals with early-stage AD to remain engaged in workplaces, social activities, and decision-making.
- **Empathy**: Early AD Detection fosters a more sympathetic view of people with neurodegenerative diseases in society.

Early AD detection is vital for effective treatment and management, offering a critical opportunity to slow its progression. This thesis explores multiple methods for feature extraction to enhance the reliability and performance of Alzheimer's disease detection systems through medical imaging and related data. The research emphasizes the significant importance of advanced feature extraction methods in identifying subtle patterns linked to the early progression of AD.

Wavelet-based techniques, such as DD-DTCWT, have demonstrated their effectiveness in capturing multi-scale features that represent fine structural details in brain images. Texture-based approaches, including LBP and its variants such as EWLBP, shifted ELBP, and WSCELD have proven useful for detecting textural changes in brain regions commonly affected by AD. Additionally, histograms and energy-based features have contributed to the understanding of the distribution and intensity patterns of brain tissues, complementing other feature sets.

The incorporation of these techniques with machine learning models has demonstrated potential in enhancing classification accuracy. Moreover, combining imaging features with biomarkers and leveraging multimodal data has further enriched the diagnostic potential, opening the door for more robust and clinically useful systems.

This work underscores the prominence of continuous innovation in feature extraction techniques to address the challenges of early AD detection. Future directions may include the development of hybrid approaches that integrate multiple feature extraction methods, utilizing deep learning-based features, and the inclusion of longitudinal data to track disease progression. By refining these techniques and ensuring their applicability across diverse populations and imaging protocols, early AD detection systems can become more reliable, accessible, and impactful in clinical practice.

The methodologies and insights presented in this thesis lay a foundation for advancing diagnostic tools, bringing the field closer to realizing the goal of timely and accurate AD diagnosis.

Thesis Final -modified - plag.pdf

ORIGINALITY REPORT

18%	9%	17%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	iieta.org Internet Source	4%
2	Neha Garg, Mahipal Singh Choudhry, Rajesh M Bodade. "Alzheimer’s disease detection through wavelet-based shifted elliptical local binary pattern", Biomedical Signal Processing and Control, 2025 Publication	4%
3	Neha Garg, Mahipal Singh Choudhry, Rajesh M Bodade. "A review on Alzheimer’s disease classification from normal controls and mild cognitive impairment using structural MR images", Journal of Neuroscience Methods, 2023 Publication	3%
4	Linda Delali Fiasam, Yunbo Rao, Collins Sey, Esther Stacy E. B. Aggrey et al. "Domain Contrastive Learning for Multi-Site Alzheimer's Disease Classification", 2023 20th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2023 Publication	<1%
5	Neha Garg, Mahipal Singh Choudhry, Rajesh M Bodade. "Alzheimer’s Disease Classification Using Wavelet-Based Image Features", Traitement du Signal, 2024	<1%

	Publication	
6	www.researchgate.net Internet Source	<1 %
7	ijritcc.org Internet Source	<1 %
8	Neha Garg, Mahipal Singh Choudhry, Rajesh M Bodade. "A review on Alzheimer’s Disease Classification from Normal Controls and Mild Cognitive Impairment using structural MR Images", Journal of Neuroscience Methods, 2022 Publication	<1 %
9	link.springer.com Internet Source	<1 %
10	Xinxing Zhao, Candice Ke En Ang, U. Rajendra Acharya, Kang Hao Cheong. "Application of Artificial Intelligence techniques for the detection of Alzheimer’s disease using structural MRI images", Biocybernetics and Biomedical Engineering, 2021 Publication	<1 %
11	www.mdpi.com Internet Source	<1 %
12	Neha Garg, Mahipal Singh Choudhry, Rajesh Mungasaji Bodade. "Alzheimer’s Disease Detection using Particle Swarm optimization", 2022 8th International Conference on Signal Processing and Communication (ICSC), 2022 Publication	<1 %
13	coek.info Internet Source	<1 %

14	Jianfu Li, Weiguo Gong, Weihong Li, Xiaoying Liu. "Robust pedestrian detection in thermal infrared imagery using the wavelet transform", Infrared Physics & Technology, 2010 Publication	<1%
15	Neha Garg, Mahipal Chaudhary. "Implementation of Dual Tree Complex Wavelet Transform with Mean Energy Features to detect Alzheimer’s Disease", 2021 7th International Conference on Signal Processing and Communication (ICSC), 2021 Publication	<1%
16	kluedo.ub.uni-kl.de Internet Source	<1%
17	escholarship.org Internet Source	<1%
18	Sreelakshmi Shaji, Rohini Palanisamy, Ramakrishnan Swaminathan. "Structural irregularities in MR corpus callosal images and their association with cerebrospinal fluid biomarkers in Mild Cognitive Impairments", Neuroscience Letters, 2023 Publication	<1%
19	"Pattern Recognition", Springer Science and Business Media LLC, 2020 Publication	<1%
20	jmss.mui.ac.ir Internet Source	<1%
21	Computer Vision for X-Ray Testing, 2015. Publication	<1%

22	"Proceedings of Second Doctoral Symposium on Computational Intelligence", Springer Science and Business Media LLC, 2022 Publication	<1%
23	Martins, Laura Afonso Lourenço. "Alzheimer's Disease-Related Gray Matter Atrophy Using Quantitative MRI: A Statistical and Machine Learning Study.", Universidade de Lisboa (Portugal), 2024 Publication	<1%
24	Submitted to University of Central Florida Student Paper	<1%
25	repositorium.uminho.pt Internet Source	<1%
26	www.brightfocus.org Internet Source	<1%
27	Rohollah Hedayati, Mohammad Khedmati, Mehran Taghipour-Gorjikolaie. "Deep feature extraction method based on ensemble of convolutional auto encoders: Application to Alzheimer’s disease diagnosis", Biomedical Signal Processing and Control, 2021 Publication	<1%
28	Nidhi Pandey, Oshin Sharma. "A stacked ensemble deep learning framework for Alzheimer’s severity ranking and classification using MRI scans", Neural Computing and Applications, 2025 Publication	<1%
29	"Biomedical Engineering Science and Technology", Springer Science and Business Media LLC, 2024 Publication	<1%

30	www.frontiersin.org Internet Source	<1 %
31	Advances in Intelligent Systems and Computing, 2016. Publication	<1 %
32	Motulsky, Harvey. "Intuitive Biostatistics", Oxford University Press Publication	<1 %
33	Yanjie Zhang, Fan Li, Lingguo Bu, Su Han, Yuanyuan Bu. "VR-based approach for MCI assessment system using fNIRS and graph convolutional network", Biomedical Signal Processing and Control, 2026 Publication	<1 %
34	Submitted to Nanyang Technological University, Singapore Student Paper	<1 %
35	vdoc.pub Internet Source	<1 %
36	Bing Xue, Mengjie Zhang, Will N. Browne. "A Comprehensive Comparison on Evolutionary Feature Selection Approaches to Classification", International Journal of Computational Intelligence and Applications, 2015 Publication	<1 %
37	Nunes, Ana Teresa Casimiro. "Exploring Texture Biomarkers of Neurodegeneration in the Retina", Universidade de Coimbra (Portugal) Publication	<1 %
38	enggjournals.com Internet Source	<1 %

39	www.coursehero.com Internet Source	<1 %
40	Submitted to University of Pittsburgh Student Paper	<1 %
41	ecsjournal.org Internet Source	<1 %
42	"Biological, Diagnostic and Therapeutic Advances in Alzheimer's Disease", Springer Science and Business Media LLC, 2019 Publication	<1 %
43	Hesheng Huang, Witold Pedrycz, Kaoru Hirota, Fei Yan. "A multiview-slice feature fusion network for early diagnosis of Alzheimer’s disease with structural MRI images", Information Fusion, 2025 Publication	<1 %
44	Xue Hua, Derrek P. Hibar, Christopher R.K. Ching, Christina P. Boyle et al. "Unbiased tensor-based morphometry: Improved robustness and sample size estimates for Alzheimer's disease clinical trials", NeuroImage, 2013 Publication	<1 %
45	airccse.org Internet Source	<1 %
46	www.ncbi.nlm.nih.gov Internet Source	<1 %
47	Submitted to European University of Cyprus Student Paper	<1 %
48	Mingchang Xu, Zhenbing Liu, Zimin Wang, Long Sun, Zhibin Liang. "The Diagnosis of	<1 %

	Alzheimer's Disease Based on Enhanced Residual Neutral Network", 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2019 Publication	
49	oro.open.ac.uk Internet Source	<1 %
50	"Smart Data Intelligence", Springer Science and Business Media LLC, 2024 Publication	<1 %
51	K. Srinivasan, Jude Hemanth, V. Rukkumani. "Transforming Healthcare - Artificial Intelligence, Machine Learning, and 5G Innovations for Enhanced Patient Care", CRC Press, 2025 Publication	<1 %
52	assets-eu.researchsquare.com Internet Source	<1 %
53	ia800701.us.archive.org Internet Source	<1 %
54	"Hybrid Artificial Intelligent Systems", Springer Science and Business Media LLC, 2012 Publication	<1 %
55	Begg, R.. "A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data", Journal of Biomechanics, 200503 Publication	<1 %
56	Roy, Dipon. "Imaging of Amyloid-β in Neurodegenerative Diseases and Review of the Relationship with Cancer, Neurodegenerative Diseases and Type II	<1 %

	Diabetes", The George Washington University, 2022 Publication	
57	shropsys.com Internet Source	<1%
58	Kim-Han Thung, Chong-Yaw Wee, Pew-Thian Yap, Dinggang Shen. "Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion", NeuroImage, 2014 Publication	<1%
59	Lecture Notes in Computer Science, 2015. Publication	<1%
60	ghettoyouths.com Internet Source	<1%
61	www.nature.com Internet Source	<1%
62	Boo-Kyeong Choi, Nuwan Madusanka, Heung-Kook Choi, Jae-Hong So et al. "Convolutional Neural Network-based MR Image Analysis for Alzheimer's Disease Classification", Current Medical Imaging Formerly Current Medical Imaging Reviews, 2020 Publication	<1%
63	Submitted to Glyndwr University Student Paper	<1%
64	Jasmine Ion Titapiccolo, Manuela Ferrario, Sergio Cerutti, Carlo Barbieri et al. "Artificial intelligence models to stratify cardiovascular risk in incident hemodialysis patients", Expert Systems with Applications, 2013 Publication	<1%

65	Revati M. Wahul, Sarita Ambadekar, Deepesh M. Dhanvijay, Mrinai M. Dhanvijay et al. "Multimodal approaches and AI-driven innovations in dementia diagnosis: a systematic review", Discover Artificial Intelligence, 2025 Publication	<1%
66	Yanjiao Ban, Xuejun Zhang, Huan Lao. "Diagnosis of Alzheimer's Disease using Structure Highlighting Key Slice Stacking and Transfer Learning", Medical Physics, 2022 Publication	<1%
67	ebin.pub Internet Source	<1%
68	theses.hal.science Internet Source	<1%
69	www.tdx.cat Internet Source	<1%
70	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security", CRC Press, 2023 Publication	<1%
71	Submitted to UT, Dallas Student Paper	<1%
72	Ying Tan. "Anti-Spam Techniques Based on Artificial Immune System", CRC Press, 2019 Publication	<1%
73	pergamos.lib.uoa.gr Internet Source	<1%
74	Chengwei Wu, Weiran Yao, Wei Pan, Guanghui Sun, Jianxing Liu, Ligang Wu.	<1%

	"Secure Control for Cyber-Physical Systems Under Malicious Attacks", IEEE Transactions on Control of Network Systems, 2022 Publication	
75	Frederick Kungu, Aaron Awere-Duodu, Eric S. Donkor. "Prevalence and Antimicrobial Resistance of Typhoid Fever in Ghana: A Systematic Review and Meta-Analysis", Diseases, 2025 Publication	<1%
76	He, Xiaolan, Yili Wu, and Yiwei Wu. "Texture Feature Extraction Method Combining Nonsubsampled Contour Transformation with Gray Level Co-occurrence Matrix", Journal of Multimedia, 2013. Publication	<1%
77	Mohammed Abdul Matheen, Zainulabedin Hasan, Amairullah Khan Lodhi, Shaikh Abdul Waheed, Altaf C.. "chapter 26 Ethical and Privacy Considerations in AI-Driven AD Research", IGI Global, 2024 Publication	<1%
78	Sandeep Kumar, Anand Nayyar, Anand Paul. "Swarm Intelligence and Evolutionary Algorithms in Healthcare and Drug Development", Routledge, 2019 Publication	<1%
79	ai.jmir.org Internet Source	<1%
80	instapedia.com Internet Source	<1%
81	5dok.org Internet Source	<1%

82	Anish Bhandarkar, Pratham Naik, Kavita Vakkund, Srasthi Junjappanavar, Savita Bakare, Santosh Pattar. "Deep learning based computer aided diagnosis of Alzheimer’s disease: a snapshot of last 5 years, gaps, and future directions", Artificial Intelligence Review, 2024 Publication	<1%
83	Submitted to Carroll Community College Student Paper	<1%
84	Submitted to Miami University of Ohio Student Paper	<1%
85	Tuan Vo, Ali K. Ibrahim, Hanqi Zhuang, Chiron Bang. "Extraction and interpretation of EEG features for diagnosis and severity prediction of Alzheimer’s Disease and Frontotemporal dementia using deep learning", Biomedical Signal Processing and Control, 2026 Publication	<1%
86	moldstud.com Internet Source	<1%
87	"Machine Learning Algorithms", Springer Science and Business Media LLC, 2025 Publication	<1%
88	"Machine Learning, Optimization, and Big Data", Springer Science and Business Media LLC, 2015 Publication	<1%
89	A. Diana Andrushia, K. Martin Sagayam, Hien Dang, Marc Pomplun, Lien Quach. "Visual-Saliency-Based Abnormality Detection for MRI	<1%

	Brain Images—Alzheimer's Disease Analysis", Applied Sciences, 2021 Publication	
90	Submitted to Birla Institute of Technology Student Paper	<1 %
91	Chen Luo, Wei Pang, Zhe Wang, Chenghua Lin. "Hete-CF: Social-Based Collaborative Filtering Recommendation Using Heterogeneous Relations", 'Institute of Electrical and Electronics Engineers (IEEE)', 2015 Internet Source	<1 %
92	Submitted to Dublin City University Student Paper	<1 %
93	Hari Mohan Rai, Joon Yoo, Syed Atif Moqurrab, Serhii Dashkevych. "Advancements in traditional machine learning techniques for detection and diagnosis of fatal cancer types: Comprehensive review of biomedical imaging datasets", Measurement, 2023 Publication	<1 %
94	Mohammad Osmani, Freshteh Osmani, Mohammad Amin Pourhoseingholi. "Comparison of Decision Tree and Logistic Regression for Prediction of Functional Dyspepsia and Gastroesophageal Reflux Disease in Tehran Province Using Rome III", Modern Care Journal, 2019 Publication	<1 %
95	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent	<1 %

	Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	
96	Submitted to Universiteit van Amsterdam Student Paper	<1%
97	core.ac.uk Internet Source	<1%
98	digital.library.adelaide.edu.au Internet Source	<1%
99	libweb.kpfu.ru Internet Source	<1%
100	scholar.ppu.edu Internet Source	<1%
101	Submitted to CECOS College London Student Paper	<1%
102	Mingkan Shen, Peng Wen, Bo Song, Yan Li. "Muti-band Morlet mutual information functional connectivity for classifying Alzheimer's disease and frontotemporal dementia with a deep learning technique", Computers in Biology and Medicine, 2025 Publication	<1%
103	Mingxing Zhang, Yang Yang, Hanwang Zhang, Fumin Shen, Dongxiang Zhang. " and sample constraint based feature selection and classification for AD diagnosis ", Neurocomputing, 2016 Publication	<1%
104	S. R. Reeja, Bore Gowda, Y. S. Rammohan, Ganesan Prabu Sankar, G. Jayalatha. "Engineering Science and Technology: Innovations for the Future", CRC Press, 2025	<1%

Publication

105	Zhang, Mingxing, Yang Yang, Hanwang Zhang, Fumin Shen, and Dongxiang Zhang. "L2,p-norm and sample constraint based feature selection and classification for AD diagnosis", Neurocomputing, 2016. Publication	<1%
106	pmc.ncbi.nlm.nih.gov Internet Source	<1%
107	shura.shu.ac.uk Internet Source	<1%
108	"CNC Milling Tool Head Imbalance Prediction Using Computational Intelligence Methods", Lecture Notes in Computer Science, 2015. Publication	<1%
109	Reed, James Christopher. "Environmentally Adaptive Iterative Learning for Performance Optimization of Tethered Energy Systems", North Carolina State University, 2023 Publication	<1%
110	Rohit M. Thanki, Vedvyas J. Dwivedi, Komal R. Borisagar. "Multibiometric Watermarking with Compressive Sensing Theory", Springer Science and Business Media LLC, 2018 Publication	<1%
111	Sabrine Abassi, Hela Mahersia, Youssef Zied Elhechmi. "Dyskalemia detection based on wavelet scattering features and deep learning models: A comparative analysis", Biomedical Signal Processing and Control, 2026 Publication	<1%
112	Submitted to Universidade do Porto Student Paper	<1%

113	dspace.bits-pilani.ac.in:8080 Internet Source	<1%
114	macsphere.mcmaster.ca Internet Source	<1%
115	uacyti.uagro.mx Internet Source	<1%
116	"Automation of Complex Systems: Bio-Medicine, Energy and Manufacturing", Springer Science and Business Media LLC, 2026 Publication	<1%
117	"Machine Learning in Healthcare Informatics", Springer Science and Business Media LLC, 2014 Publication	<1%
118	123dok.com Internet Source	<1%
119	Alfardus, Asma. "Evaluating Machine Learning for Intrusion Detection in CAN Bus for In-Vehicle Security", Howard University Publication	<1%
120	Hejie Li, Ying Tan, Jiaqing Miao, Ping Liang, Jinnan Gong, Hui He, Yuhong Jiao, Fan Zhang, Yaolin Xing, Donghan Wu. "Attention-based and micro designed EfficientNetB2 for diagnosis of Alzheimer’s disease", Biomedical Signal Processing and Control, 2023 Publication	<1%
121	Jupalli Pushpakumari, Janapareddy Uttara Alekhya, Jakkapu Nagalakshmidevi, Mangu Maanasa et al. "A Novel Self-Supervised Swin Transformer with Wavelet Feature Extraction	<1%

	for Early Alzheimer’s Disease Recognition", International Journal of Basic and Applied Sciences, 2025 Publication	
122	K.R. Kruthika, Rajeswari, H.D. Maheshappa. "CBIR System Using Capsule Networks and 3D CNN for Alzheimer’s disease Diagnosis", Informatics in Medicine Unlocked, 2018 Publication	<1 %
123	Kirpalani, Ashwin L. "Comparison of speckle noise reduction techniques using near shift invariant wavelet transforms with application to SAR imagery", Proquest, 20111004 Publication	<1 %
124	Susmita Bandyopadhyay. "Business Analytics - Theory and Practice", CRC Press, 2025 Publication	<1 %
125	Wei Feng, Nicholas Van Halm-Lutterodt, Hao Tang, Andrew Mecum et al. "Automated MRI-Based Deep Learning Model for Detection of Alzheimer’s Disease Process", International Journal of Neural Systems, 2020 Publication	<1 %
126	hrcak.srce.hr Internet Source	<1 %
127	ijece.iaescore.com Internet Source	<1 %
128	iris.unica.it Internet Source	<1 %
129	mobt3ath.com Internet Source	<1 %
130	publishoa.com Internet Source	

<1%

131	scholar.uwindsor.ca Internet Source	<1%
132	tud.qucosa.de Internet Source	<1%
133	worldwidescience.org Internet Source	<1%
134	www.allproducts.com.tw Internet Source	<1%
135	www.docme.ru Internet Source	<1%
136	www.joca.cn Internet Source	<1%
137	www.jsr.org Internet Source	<1%
138	www.science.gov Internet Source	<1%
139	www.science.org Internet Source	<1%
140	"Artificial Intelligence in Decision Support Systems for Diagnosis in Medical Imaging", Springer Science and Business Media LLC, 2018 Publication	<1%
141	"Smart Trends in Computing and Communications", Springer Science and Business Media LLC, 2022 Publication	<1%

142	Acharya, U., S. V. Sree, M. Mookiah, L. Saba, H. Gao, G. Mallarini, and J. S. Suri. "Computed tomography carotid wall plaque characterization using a combination of discrete wavelet transform and texture features: A pilot study", Proceedings of the Institution of Mechanical Engineers Part H Journal of Engineering in Medicine, 2013. Publication	<1 %
143	Ameh, Emmanuel Esem. "Multimodal Deep Learning Algorithms for Predictive Modeling of Cardiovascular Diseases Onset and Progression", Capitol Technology University, 2025 Publication	<1 %
144	Anjan Bandyopadhyay, Tanvir Habib Sardar, Saurav Mallik, Ruhul Amin Hazarika, Mahendra Kumar Gourisaria. "AI and Data Engineering for Healthcare - Real-World Applications and Case Studies", CRC Press, 2025 Publication	<1 %
145	Aryaz Baradarani, Q.M. Jonathan Wu, Majid Ahmadi. "An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTWT filtering", Pattern Recognition, 2013 Publication	<1 %
146	Biao Jie, Daoqiang Zhang, Bo Cheng, Dinggang Shen. "Manifold regularized multitask feature learning for multimodality disease classification", Human Brain Mapping, 2015 Publication	<1 %

147 Chao Zhang, Wentao Fan, Bo Wang, Chunlin Chen, Huaxiong Li. "Self-paced semi-supervised feature selection with application to multi-modal Alzheimer’s disease classification", Information Fusion, 2024 <1%

Publication

---

148 D. B. Hann, A. M. S. Smith, A. K. Powell. "Classification of off-diagonal points in a co-occurrence matrix", International Journal of Remote Sensing, 2003 <1%

Publication

---

149 Dattatray G. Takale, Parikshit N. Mahalle, Sachin S. Bere, Piyush P. Gawali. "Applied Machine Learning in Healthcare - Case-Based Approach", CRC Press, 2025 <1%

Publication

---

150 Dipankar De. "Introduction to Differential Geometry with Tensor Applications", Wiley, 2022 <1%

Publication

---

151 Dode, Pragati. "Exploring Classification Methods for Motor Imagery and Execution EEG Signal Fluctuations.", University of Washington <1%

Publication

---

152 Gleitsmann, Anke. "Exploiting the Spatial Information in High Resolution Satellite Data and Utilising Multi-Source Data for Tropical Mountain Forest and Land Cover Mapping", Georg-August-Universitaet Goettingen (Germany) <1%

Publication

153	Hung-Hsiang Huang, Pai-Yu Cheng, Chung-You Tsai. "Exploring artificial intelligence in functional urology: A comprehensive review", Urological Science, 2025 Publication	<1 %
154	Inna Berezovska, Hichem Fettaka, Thierry Salmon, Dominique Toye, Peter Lodewyckx. "Redistribution of a mixture of organic vapours inside an activated carbon filter", Chemical Engineering Journal, 2015 Publication	<1 %
155	K. Emily Esther Rani, S. Baulkani. "Alzheimer disease classification using optimal clustering based pre-trained SqueezeNet model", Biomedical Signal Processing and Control, 2025 Publication	<1 %
156	M. K. Flaks. "Short Cognitive Performance Test: Diagnostic Accuracy and Education Bias in Older Brazilian Adults", Archives of Clinical Neuropsychology, 07/29/2009 Publication	<1 %
157	Prashant Upadhyay, Pradeep Tomar, Satya Prakash Yadav. "Comprehensive Systematic Computation on Alzheimer's Disease Classification", Archives of Computational Methods in Engineering, 2024 Publication	<1 %
158	Sinclair, Eleanor K.. "Evaluating Sebum as a Biofluid for Parkinson's Disease Diagnostics Using Mass Spectrometry-Based Metabolomics", The University of Manchester (United Kingdom), 2025 Publication	<1 %

159	Sukhpreet Kaur, Amanpreet Kaur, Manish Kumar. "Recent Advances in Computational Methods in Science and Technology - Volume 2", CRC Press, 2026 Publication	<1%
160	Udara Yedukondalu, V Vijayasri Bolisetty. "Advancing Innovation through AI and Machine Learning Algorithms - Computational Intelligence for Virtual System Optimization. A proceeding of ICMVRCET – 2025", CRC Press, 2025 Publication	<1%
161	Submitted to University of Bath Student Paper	<1%
162	Submitted to University of East London Student Paper	<1%
163	Wellington Pinheiro dos Santos, Juliana Carneiro Gomes, Valter Augusto de Freitas Barbosa. "Swarm Intelligence Trends and Applications", CRC Press, 2022 Publication	<1%
164	Xiong, Lisa. "From Midlife Dementia Risk Factors to Late-Life Functionality: An Examination of Targets to Support Healthy Aging", University of Toronto (Canada) Publication	<1%
165	Zheng Alan Zhao, Huan Liu. "Spectral Feature Selection for Data Mining", Chapman and Hall/CRC, 2019 Publication	<1%
166	arxiv.org Internet Source	<1%

167	brieflands.com Internet Source	<1%
168	dokumen.pub Internet Source	<1%
169	inass.org Internet Source	<1%
170	ir.amu.ac.in Internet Source	<1%
171	mdpi-res.com Internet Source	<1%
172	open.library.ubc.ca Internet Source	<1%
173	pubmed.ncbi.nlm.nih.gov Internet Source	<1%
174	researchonline.gcu.ac.uk Internet Source	<1%
175	www.asterhospitals.in Internet Source	<1%
176	www.scilit.net Internet Source	<1%
177	"Proceedings of the 12th International Conference on Soft Computing for Problem Solving", Springer Science and Business Media LLC, 2024 Publication	<1%
178	"SK298 topic 8 dementia WEB153413", Open University Publication	<1%
179	Chu Yu, Sao-Jie Chen. "VLSI implementation of 2-D discrete wavelet transform for real-time	<1%

video signal processing", IEEE Transactions on Consumer Electronics, 1997
Publication

---

180 Devendra Kumar, Shraddha Manish Gupta, Nagendra Singh Chauhan, Kamal Shah. "Alzheimer's Disease - Diagnosis, Molecular Targets, and Therapeutic Modalities", CRC Press, 2025
Publication
<1%

---

181 Hermanus Vermaak, Philibert Nsengiyumva, Nicolaas Luwes. "Using the Dual-Tree Complex Wavelet Transform for Improved Fabric Defect Detection", Journal of Sensors, 2016
Publication
<1%

---

182 Mingyang Li, Wanzhong Chen, Tao Zhang. "Automatic epilepsy detection using wavelet-based nonlinear analysis and optimized SVM", Biocybernetics and Biomedical Engineering, 2016
Publication
<1%

---

183 Peiqi Luo, Guixia Kang, Xin Xu. "A Novel Feature Selection and Classification Method of Alzheimer's Disease based on Multi-features in MRI", Proceedings of the 2020 10th International Conference on Bioscience, Biochemistry and Bioinformatics, 2020
Publication
<1%

---

184 Prats-Montalban, J.M.. "Prediction of skin quality properties by different Multivariate Image Analysis methodologies", Chemometrics and Intelligent Laboratory Systems, 20090315
Publication
<1%

185	Sabrina Bonarota, Giulia Caruso, Carlotta Di Domenico, Sofia Sperati et al. "Integration of Automatic MRI Segmentation Techniques with Neuropsychological Assessments for Early Diagnosis and Prognosis of Alzheimer's Disease. A systematic review", NeuroImage, 2025 Publication	<1 %
186	Shi, Bibo. "Diversification and Generalization for Metric Learning with Applications in Neuroimaging.", Ohio University, 2017 Publication	<1 %
187	Sujata Dash, Subhendu Kumar Pani, Joel J. P. C. Rodrigues, Babita Majhi. "Deep Learning, Machine Learning and IoT in Biomedical and Health Informatics - Techniques and Applications", CRC Press, 2022 Publication	<1 %
188	Xinyu Sun, Xin-Na Geng, Tao Liu. "Due-window assignment scheduling in the proportionate flow shop setting", Annals of Operations Research, 2020 Publication	<1 %
189	Zhou Xu. "Chapter 4 Class Imbalanced Learning-Based Defect Prediction", Springer Science and Business Media LLC, 2026 Publication	<1 %
190	orbi.uliege.be Internet Source	<1 %
191	rgu-repository.worktribe.com Internet Source	<1 %
192	"Advances in Signal Processing and Intelligent Recognition Systems", Springer Science and	<1 %

	Business Media LLC, 2018 Publication	
193	"Inventive Systems and Control", Springer Science and Business Media LLC, 2022 Publication	<1 %
194	A. Balasundaram, Sruthi Srinivasan, A. Prasad, Jahan Malik, Ayush Kumar. "Hippocampus Segmentation-Based Alzheimer’s Disease Diagnosis and Classification of MRI Images", Arabian Journal for Science and Engineering, 2023 Publication	<1 %
195	Abdul Wahid, Praveen Kumar Donta. "Advanced Techniques for Anomaly Detection - Beyond the Basics", CRC Press, 2025 Publication	<1 %
196	Huan Lao, Xuejun Zhang. "Regression and Classification of Alzheimers Disease Diagnosis using NMF-TDNet Features from 3D Brain MR Image", IEEE Journal of Biomedical and Health Informatics, 2021 Publication	<1 %
197	J.R. Sveinsson, J.A. Benediktsson. "Double density wavelet transformation for speckle reduction of SAR images", IEEE International Geoscience and Remote Sensing Symposium, 2002 Publication	<1 %
198	Kerang Cao, Zhongqing Lu, Kaidi Zhang, Yu Liu, Ye Ma, Jun Wang, Hoekyung Jung. "Local–global feature fusion for enhanced Alzheimer’s detection: a multi-scale attention-driven MRI classification model", Complex & Intelligent Systems, 2025	<1 %

	Publication	
199	Kim, DongKyu. "Automatic Artifact Annotator for EEG Waves Using Recurrent and Convolutional Neural Networks.", The Cooper Union for the Advancement of Science and Art, 2020 Publication	<1 %
200	Neeraj Kumar, Aaisha Makkar. "Machine Learning in Cognitive IoT", CRC Press, 2020 Publication	<1 %
201	Palak Goyal, Rinkle Rani, Karamjeet Singh. "An efficient ranking-based ensembled multiclassifier for neurodegenerative diseases classification using deep learning", Journal of Neural Transmission, 2024 Publication	<1 %
202	Rashmi Gupta, Shashi Shekhar, Jeetendra Kumar, Gaurav Dhiman. "Ambient Assisted Living - Technological Advances", CRC Press, 2025 Publication	<1 %
203	Ruilin Wang, Wei Zhang, Wenbo Nie, Yao Yu. "Gastric Polyps Detection by Improved Faster R-CNN", Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition, 2019 Publication	<1 %
204	Yini Chen, Yiwei Qi, Yiying Hu, Xinhui Qiu, Tao Qiu, Song Li, Meichen Liu, Qiqi Jia, Bo Sun, Cong Liu, Tianbai Li, Weidong Le. "Integrated cerebellar radiomic-network model for predicting mild cognitive impairment in Alzheimer's disease", Alzheimer's & Dementia, 2024	<1 %

Publication

Exclude quotes	On	Exclude matches	< 4 words
Exclude bibliography	Off		

Thesis Final -modified - plag.pdf

GRADEMARK REPORT

FINAL GRADE

/100

GENERAL COMMENTS

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51

PAGE 52

PAGE 53

PAGE 54

PAGE 55

PAGE 56

PAGE 57

PAGE 58

PAGE 59

PAGE 60

PAGE 61

PAGE 62

PAGE 63

PAGE 64

PAGE 65

PAGE 66

PAGE 67

PAGE 68

PAGE 69

PAGE 70

PAGE 71

PAGE 72

PAGE 73

PAGE 74

PAGE 75

PAGE 76

PAGE 77

PAGE 78

PAGE 79

PAGE 80

PAGE 81

PAGE 82

PAGE 83

PAGE 84

PAGE 85

PAGE 86

PAGE 87

PAGE 88

PAGE 89

PAGE 90

PAGE 91

PAGE 92

PAGE 93

PAGE 94

PAGE 95

PAGE 96

PAGE 97

PAGE 98

PAGE 99

PAGE 100

PAGE 101

PAGE 102

PAGE 103

PAGE 104

PAGE 105

PAGE 106

PAGE 107

PAGE 108

PAGE 109

PAGE 110

PAGE 111

PAGE 112

PAGE 113

PAGE 114

PAGE 115

PAGE 116

PAGE 117

PAGE 118

PAGE 119

PAGE 120

PAGE 121

PAGE 122

PAGE 123

PAGE 124

PAGE 125

PAGE 126

PAGE 127

PAGE 128

PAGE 129

PAGE 130

PAGE 131

PAGE 132

PAGE 133

PAGE 134

PAGE 135

PAGE 136

PAGE 137

PAGE 138

PAGE 139

PAGE 140

PAGE 141

PAGE 142

PAGE 143

PAGE 144

PAGE 145

PAGE 146

PAGE 147

PAGE 148

PAGE 149

PAGE 150

PAGE 151