CHAPTER 5

CONVOLUTIONAL NEURAL NETWORK BASED FEATURE EXTRACTION FOR CLASSIFYING STAGES OF ALZHEIMER DISEASE

5.1 Introduction

Classification of AD in medical diagnosis is vital due to the harshness of the disease, because it affects the overall wellbeing. Identifying the damaged neurons in the brain automatically using a screening system is necessary. An automated system classifies brain neuron images into categories with the aid of technical advancement. ML methods on MRIs aids physicians in AD diagnosis. Conventional ML use handcrafted feature extraction methods on MRIs which is complex and it requires an expert user. Consequently, the procedure could be streamlined by using DL approaches as an automatic feature extraction method. In this study, traditional ML based feature extraction techniques are applied to extract texture-based information, which includes GLCM, GLRLM, and GLDM. It is expected that it aids in the extracted spatial domain to produce results in a more effective classification-based union feature set. Also, a DL technique which is based on CNN is applied for feature extraction. It acts as an automatic extraction process for identifying AD using MRI. CNN model performance is evaluated using different metric measures.

A detailed description related to the traditional feature extraction techniques like GLCM, GLRLM, and GLDM are presented in the next subsection. The following sections provide a detailed description of the AlexNet model for multi-stage feature extraction. Performance metrics of the experimental findings are analyzed, and the outcomes of the outcome analysis and evaluation of the suggested method are provided.

5.2 Traditional Feature Extraction techniques

Traditional feature extraction techniques for images are needed in computer visualization and image processing. These methods benefit in altering raw pixel data into more significant and compact representations, allowing algorithms to analyze, classify, or recognize objects in images in a better manner. These findings develop the understanding of object shape and texture. It is easy to manage and extract features when the algorithm's input dataset size is reduced. Numerical data was created from the gathered pixels. This study uses some of the common classic feature extraction methods for images, including GLCM, GLRLM, and GLDM. The most utilized features are GLCM for statistical feature removal, GLRLM for run length feature extraction, and GLDM for grey level probability removal. Each technique's capabilities are given below:

5.2.1 GLCM Technique

A statistical method of texture observation that takes the spatial relationships between an image's pixels. It records the frequency with which pairs of pixels with exact values (gray levels) appear in a certain spatial relationship. It's used for texture analysis, pattern recognition, and image classification. In order to extract texture information and preserve a bond between pixels, GLCM calculates the grey level co-occurrence values. The primary GLCM theories are

i) Gray Level: Pixel intensity value, frequently signified on a scale from 0 (black) to 255 (white) for an 8-bit image.

ii) Co-occurrence: Existence of pixel values in pairs that have a stated spatial relationship.

iii) Spatial Relationship: Distance and route between pixel pairs. Common directions include 0°, 45°, 90°, and 135°.

The working of GLCM is given below:

i) Image Preprocessing: Converting the image into grayscale. For each pixel image, a spatial relationship with its neighbors (specified by a distance and route) is considered.

ii) Co-occurrence Matrix Construction: Matrix is built based on the intensity values of pixel pairs. The matrix size depends on the gray level count in the image. A 256-level image will have a 256×256 GLCM. The value of each pixel is compared to the value of its nearby pixels (depending on the provided distance and route). Each existence of pixel pair values is calculated and stored in the matrix.

iii) GLCM Normalization: By dividing each member by the total number of measured pixel pairings, the matrix is normalized. This makes the matrix sum equal to 1.

After GLCM is constructed, some statistical features are extracted to quantify the image texture. These features capture different characteristics of the image's texture, such as contrast, correlation, energy, and homogeneity. The common GLCM features are listed here (Rayen et al., 2021) Autocorrelation (AU), contrast, correlation, energy, Homogeneity1(H1), entropy, dissimilarity, Cluster Prominence (CP), Cluster Shade (CS), Maximum probability (MP), sum of squares (Variance), Sum Average (SA), Sum Variance (SV), Sum Entropy (SE), Difference Variance (DV), Difference Entropy (DE), Inverse Difference (ID), Invers Difference Normalized (IDN) and Inverse Difference Moment Normalized (IDMN). Based on these characteristics, the pixel intensity value and the intensity levels of adjacent voxels determines (H1).

5.2.2 GLRLM Technique

The length of consecutive pixel runs with similar strengths is another statistical method of texture analysis that may be used to represent the spatial correlations of pixel strengths in an image. It is effective for analyzing textures that exhibit repetitive patterns, like medical or industrial images, where textures incline to have long stretches of related pixel values.

The main theory of GLRLM are

i) Gray Level: strength or clarity of a pixel

ii) Run: sequence of successive pixels with the similar strength, beside a precise direction (horizontal, vertical, diagonal, or anti-diagonal).

iii) Run length: Successive pixel count in a run.

iv) Spatial relationship: Direction and distance of pixel pairs measured when analyzing runs (similar to GLCM).

The working of GLRLM is given below:

i) Image Preprocessing: Converting the image into grayscale because the GLRLM analysis emphases on the pixel intensity values.

ii) Run-Length Matrix Construction: Every pixel in the image is used to build the matrix, which is then used to measure the number of consecutive pixels (runs) with equal intensity values in a specified direction. The rows and columns relate to the gray levels and run lengths. For example, the entry at position (i, j) in the matrix represents the intensity count, 'i' appears successively for 'j' pixels.

iii) GLRLM Normalization: Similar to GLCM, by dividing all elements by the total number of pixel pairs measured. This makes the matrix sum equal to 1.

Once GLRLM is created, statistical features can be derived to describe the image texture. The image's runs' length, frequency, and distribution are captured by these features. The common GLRLM features are: a) The distribution of short runs in the image is measured using Short Run Emphasis (SRE). b) The distribution of long runs in the image is measured using Long Run Emphasis (LRE). A higher value specifies a higher frequency of long runs. c) Run Percentage (RP) measures the percentage of runs for a particular gray level relative to run count. d) Gray-Level Nonuniformity (GLN) measures the distribution of gray levels in the run-length matrix. High GLN specifies a greater variation in gray levels for a given run length. e) Run Length Nonuniformity (RLN) measures the non-uniformity of run length. A higher value specifies a greater variation in the lengths of the runs. f) Low Gray-Level Run Emphasis (LGRE) highlights the presence of shorter runs with lower intensity values. This feature is useful when the image contains large areas with low-intensity values that are uniformly distributed. g) High Gray-Level Run Emphasis (HGRE) highlights the presence of longer runs with higher intensity values. This is useful for analyzing images with regions of high-intensity textures. The stages of the GLRLM texture analysis are converting image to grayscale, selecting the spatial parameters, constructing the GLRLM, extracting features and finally using the features for classification.

5.2.3 GLDM Technique

An additional statistical method of texture analysis that measures the spatial correlation between an image's pixel intensities. Unlike techniques like GLCM or GLRLM, which focus on capturing run lengths or co-occurrence relationships, GLDM measures gray-level differences between neighboring pixels in an image, and provides valuable information about texture patterns.

The main theory of GLDM is

i) Gray Level: pixel intensity represents an integer value (0-255 for 8-bit images).

ii) Difference: is the change intensity between two neighboring pixels.

iii) Spatial Relationship: like texture analysis techniques, it denotes the direction and distance between pixel pairs.

iv) Co-occurrence Matrix: the focus is on intensity differences between pixels, and classifies the results in a matrix format.

The working of GLDM is given below:

i) Image Preprocessing: Converting the image into grayscale because the GLDM analysis emphases on the pixel intensity values.

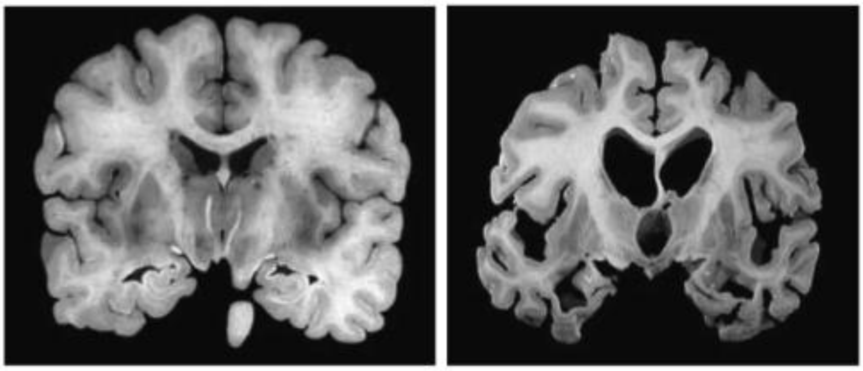
ii) Pixel Pair Analysis: Pixel pairs within distinct neighborhood or spatial relation is considered (horizontal, vertical or diagonal). The differences in intensity between each pair of neighboring pixels is computed. The extent of these differences is captured and stored.

iii) Matrix Construction: estimating the frequency of each pixel pair intensity difference. The pixel intensity value range determines the matrix dimensions. To guarantee that the total of all matrix values equals 1, the matrix is normalized.

After GLDM matrix construction, some texture features are derived. These features capture diverse characteristics of the image's texture, mainly focusing on how the differences in intensity vary spatially. The GLDM features are a) Contrast measures the intensity change between adjacent pixels which counts the image intensity variation. b) Correlation measures linear reliance among pixel pairs. A high link specifies consistent relationship in the pixel intensities. c) Energy reflects the consistency of the texture by measuring how much the intensity differences vary in the image. High energy specifies that the intensity differences are small and uniform. d) Homogeneity measures the nearness of distribution intensity differences. A high value specifies that the intensity differences are small and steady during the image. e) Entropy measures the randomness or complexity of the texture. High entropy specifies a more complex texture, with diverse intensity differences between pixels. f) Maximum Probability measures the frequency of the most regularly occurring intensity difference. A higher value specifies that the image has dominant pixel intensity differences. The stages of the GLDM texture analysis are converting image to grayscale, selecting the parameters, constructing the GLDM, extracting features and finally using the features for classification, segmentation, or pattern recognition.

5.2.4 Feature engineering technique

Figure 5.1 shows the outcome of feature extraction process of brain MRI scan Figure 5.1 (a) shows the AD patient and Figure 5.1(b) shows the healthy individual. At the same time that the ventricles widen, AD patients exhibit atrophy in the hippocampus, parahippocampal gyrus, and medial temporal lobes. Figure 5.2 shows the overall working process of feature extraction and classification using traditional feature extraction techniques. Table 5.1 shows the overall features employed to increase the vital gap between the class extents.



(a) (b)

Figure 5.1 Brain MRI of (a) an AD patient (b) a healthy individual after feature extraction

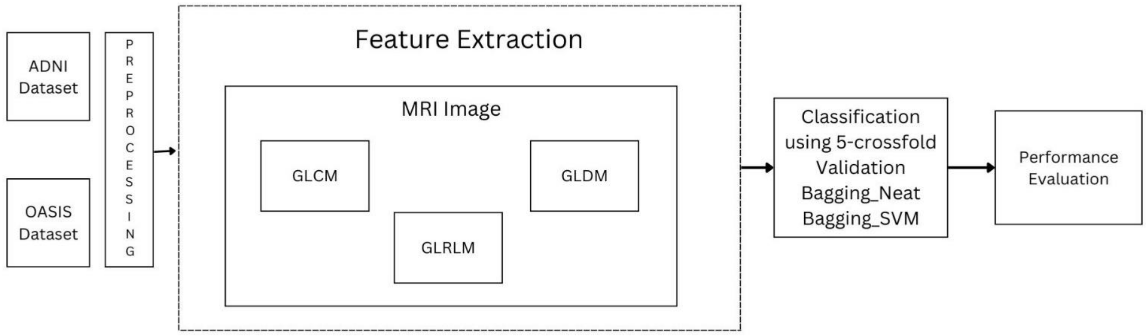


Figure 5.2 Overall working mechanism for traditional feature extraction techniques

Table 5.1 Overall Features using GLCM, GLRLM, and GLDM methods

Traditional Methods	Features
GLCM	Autocorrelation, contrast, correlation, cluster prominence, cluster shade dissimilarity, energy, entropy, homogeneity 1, homogeneity, maximum probability, sum of squares, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation 1 & 2, inverse difference, inverse difference normalized, inverse difference moment normalized,
GLRLM	SRE, LRE, RLE, GLN, RP, LGLRE, HGLRE
GLDM	Short Dependency Emphasis (SDE), Long Dependency Emphasis (LDE), Dependence Non-Uniformity (DNU), Dependence Non-Uniformity Normalized (DNUN)

The 33 distinct features are retrieved from the MRIs for AD classification. Feature extraction focuses on the feature strength of image pixels. The improved features offer information about image intensity, shape, texture, and location. GLRLM offers coarse image features, whereas GLCM offers spatial needs of the grey stages on different angles. .For higher level statistical data, GLRLM surges the strength. GLDM uses gray level mass feature identification. Finally, various ML methods are applied to categorize AD stages from MRIs.

5.3 Proposed pre-trained CNN model for feature extraction

A detailed presentation of the CNN model for feature extraction is given. CNN layers and model are elaborated. As explained in Chapter 4, the architecture model applies preprocessing techniques to the input in order to eliminate image noise. The MRI (pre-processed) are applied to AlexNet model to categorize AD stages. AlexNet model consists of a transfer learning model for feature extraction Components of the transfer learning model are also covered in the following subsections.

5.3.1 CNN Layers

The various CNN layers are discussed below:

i) Input layer: Image size is 227*227*3, it interprets image's height, weight & channel size. A trained network, at the start it is needed to shuffle the data and for every epoch beginning the shuffle happens automatically (Nawaz et al., 2021).

ii) Convolutional Layer (CL): It is the vital portion of CNN with learnable kernels. It's a key network layer with parameters and feature map count is defined. Padding is used to measure name pairs with padding as input.

iii) Batch normalization layer: Aids in network training to ease the optimization issues. Also, it aids to normalize incline and activate the network circulation. Similar to Rectified Linear Unit (ReLU) layers, which expedite network training, this layer is employed in between nonlinearities and CL.

iv) ReLU layer: An activation function used in the NN structures.

v) Max-pooling layer: Spatial size in feature map decline. The filter count in deep CL is increased without a surge in the per layer estimation due to sample decline. Maximum rectangular input count in a locality is given by this layer.

vi) Fully connected layer: Every neuron in this layer is fully connected to every other neuron in the preceding levels. A wide pattern is recognized and merges with formerly learned layers completely. It is the closing layer that joins all the features for classification. This layer's output is equal to the class count in the image dataset and with class count=4.

vii) Softmax layer: Fully connected layer's output is normalized. The output contains positive integers used by the classification layer.

viii) Classification layer: Final layer that uses softmax layer's chances to each input to allow restricted class and estimate the loss.

The training choices are given after the network architecture. Stochastic Gradient Descent with Momentum (SDGM) performs 10 epochs for training. Initially data is randomly started. The data correctness is estimated at regular intervals during training. Options for early AD detection include batch size, validation frequency, learning rate, and epoch. The residual data is fed into the network as testing data in order to estimate the performance of the training data. Testing data is assessed based on the classification of AD offset data (Nawaz et al., 2021). The CNN transfer learning model for AD classification is displayed in Figure 5.3.

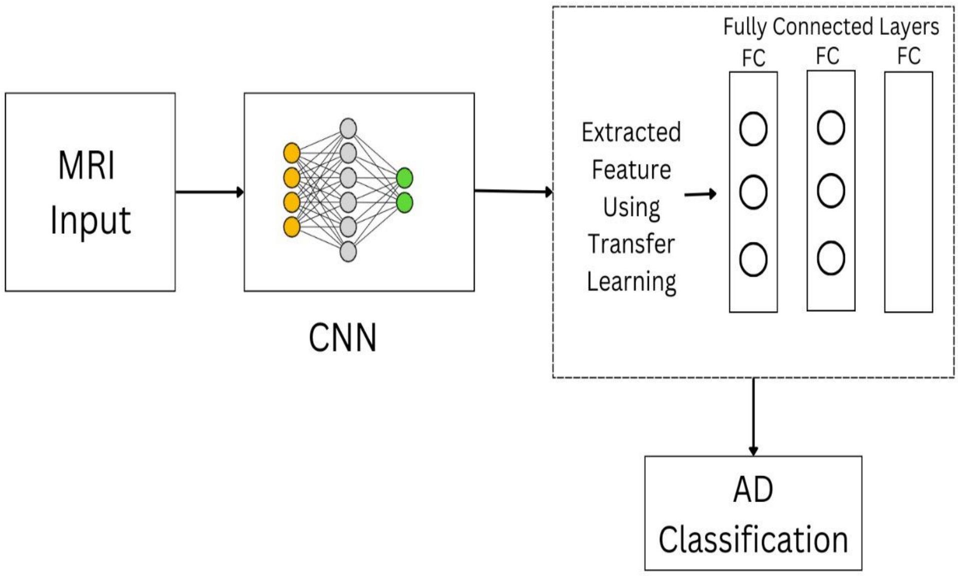


Figure 5.3 CNN transfer learning model

This model uses deep architecture, which consists of 5 CL and 3 FCL. The ReLU non-linearity operation is simple and quick. The output is produced only if the input is positive, otherwise, the output is 0. Overlap pooling is similar to regular max pooling layer, were the window moves through and overlays with the former window. The nearby pooling processes overlap with one another to reduce the network size and it's hard for the model to overfit. In this study, AlexNet is incorporated, which requires GPUs to enhance the computation period. Local Response Normalization (LRN) is performed to normalize the nearby channels in the network and inside a local neighborhood. To address the overfitting issues dropout layers are used to turn off the neurons and to regularize. This layer makes the model to learn robust features and end the model from shortening features in a single network area. This layer surges the training time yet, the model learns better. Figure 5.4 illustrates the architecture of the dropout layer.

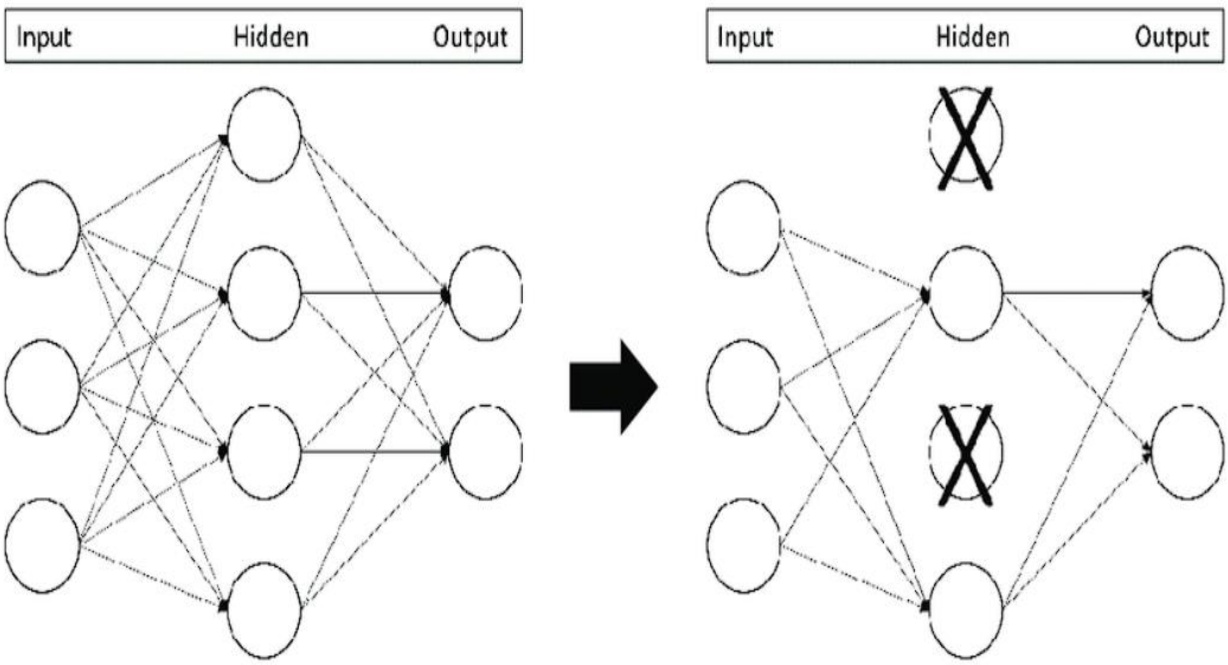


Figure 5.4 Dropout layers architecture

5.3.2 Architecture of AlexNet

It contains 8 layers, first 5 layers are CL, the convolutional filter sizes are 11×11, 5×5, 3×3, 3×3, and 3×3 for the particular CL. Certain CL are prior to max-pooling layers to minimize the spatial dimensions while holding vital features. The activation function in

ReLU, has greater performance when equated to sigmoid and tanh functions. After CL, there are 3 FCL and the network parameters are tuned with respect to the training performance. The vital components of the AlexNet architecture are listed below:

i) CNN is designed visual recognition tasks, which learn the features from image, mainly used for image classification.

ii) AlexNet has 8 layers, designed for extracting related patterns and features used for classification.

iii) ReLU activation function are used after every CL and fully connected layer. The network can handle complex data because of its non-linearity.

iv) Max pooling are applied after CL to minimize the spatial dimensions while holding the vital features. This process minimizes the computation and controls overfitting issues.

v) LRN is applied to enhance the normalization in the nearby channels

vi) Dropout layer to regularize the training duration. Applied to eliminate overfitting and enhance the model's performance

vii) Batch normalization is applied to normalize the outputs inside a small batch. Aims to increase the learning rate.

viii) Softmax activation is the last layer of AlexNet. Converts the raw output into possible classes for each input image.

ix) Training and optimization aims to adjust the learning rate and tune the parameters. Also aids to enhance the range of the training dataset.

5.3.3 CNN for feature extraction

Once the preprocessing is performed, features are removed from the image dataset using AlexNet. Its a transfer learning method used for training and testing networks. Figure 5.5 shows the different layers of AlexNet. Equation (5.1) shows a non-systematic and non-conformed system.

$$F(x) \text{ D } \max(x; 0) \qquad (5.1)$$

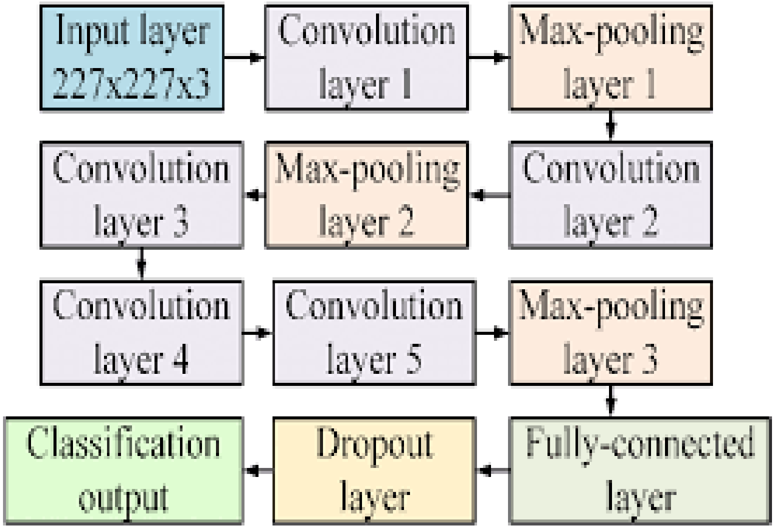


Figure 5.5 AlexNet Layers

This approach has hidden or visible neurons with 0 value to reduce the co-regulation effect. The steps involved in AlexNet architecture for image dataset are listed below:

i) Input layer: The preprocessed images of 227*227*3 size.

ii) CLl: It is the master layer used to collect productive feature based on CL size. It has a layout of 11*11 with 256 kernels. The CL functioning is given in equation (5.2).

$$x_k^l = f\left(\sum_{i \in Mk} x_i^{l-1} * Q_{ik}^{\ l} + b_k^{\ l}\right) \quad (5.2)$$

here, k denotes convolution feature map, *MK* denotes input map selection, Q_{ik} denotes the filter and b_k denotes the bias feature map.

iii) Max-pooling layer has 96 attributes with 27*27 size.

iv) CL2 has inputs with 5*5 size of convolutional pieces. The CL2 has 384 features of size 27*27. The attribute map can be achieved by combining all or few features and the leftover CL contains 384, 256, and 256 kernels.

v) Leftover max-pooling and CL has 13* 13 after convolution. With 40% precision when the primary layers are assessed.

vi) Fully connected layer has related neurons of the nearby layers. In AlexNet, 3 layers are FC, combined to have 4096 instances from 382 samples. The features are openly extracted from CL and is used for prediction and classification.

vii) Replace the last layer: With the preprocessed dataset, there is a need to substitute classification layer and shift initial 5 layers of AlexNet. The FCL framework includes weight learn factor, bias-learn, and output size. The FC output size and the number of output classes are equal. Weight components are used to measure the learning rate.

viii) A pre-trained network with three FCL, softmax, and classification layers make up the training process. The Alzheimer's data is used to train the new layer for categorization valuation. Only 77% of the images are required. Training options consists of batch volume, epochs count, learning volume, and validation frequency specified. Maximum of 10 epochs is used. The final layer learns the features of the data. The output is affected by changes to the learning parameters, learning rate, learning bias rate, and epoch count. Weights and bias rates range from 1e-4 and 10–100, whereas learning rates range from 1^{e-1} to 1^{e-10}. 10 is the MiniBatchSize (AlSaeed and Omar 2022).

ix) Network testing is performed on 23% dataset. Accuracy is measured based on the training network assessment. The testing result represents the training network’s Alzheimer’s classification (AlSaeed and Omar 2022).

x) The AlexNet altered sigmoid activation function to ReLU activation and parameters are initialized. Sigmoid function aids the model to train effectively. Table 5.2 shows the CNN parameters.

Table 5.2 CNN Parameters

S.No	Parameter	Value
1	Learning Rate	1e-1 to 1e-10
2	Momentum(SDGM)	10
3	Batch Size	10
4	Convolution Layer	5
5	Kernel Size	CL1: 11x11, CL2: 5x5, CL5: 3x3
6	Activation Function	ReLU
7	Filter Size	CL1: 256, CL2: 384, CL3: 384, CL4: 256, CL5: 256
8	Number of Epochs	10
9	Classifier	Softmax(Probabilities for each class)
10.	Optimizer	Sgd
11.	Dropout Rate	0.5
12	Depth	96
13	Loss function	Binary Cross Entropy
14.	No of Pooling Layer	3X3(2)

5.3.4 The overall structure of the proposed approach using CNN for feature extraction

Figure 5.6 shows the proposed structure of CNN model for feature extraction. A CNN model is built and validated for feature extraction and classification. The model is evaluated by examining the features extracted by CNN (AlexNet) from FCL. Five CL, three max-pooling layers, two normalized layers, two FCL, and one softmax layer make up AlexNet. Every convolution layer has a ReLU and a convolution filter. Because FCL occurs, the input size is fixed and the max-pooling function is carried out by the pooling layers. The input size 224x224x3, with padding 227x227x3 and has 60 million parameters. In this study, the ML classifiers are hybridized and applied (BAGGING_SVM, and BAGGING_NEAT) for each feature set and evaluated. AD diagnosis approaches are built as follows: i) MRI data collection stage, ii) image pre-processing is performed. Each MRI is resized for CNN model, iii) AlexNet are employed to extract MRI features and utilized for classification in the feature's extraction stage iv) two hybrid classifiers BAGGING_SVM and BAGGING_NEAT are proposed v) results are analyzed and evaluated for efficiency of each approach using few metrics vi) results are compared with recent studies in chapter 6 and chapter 7.

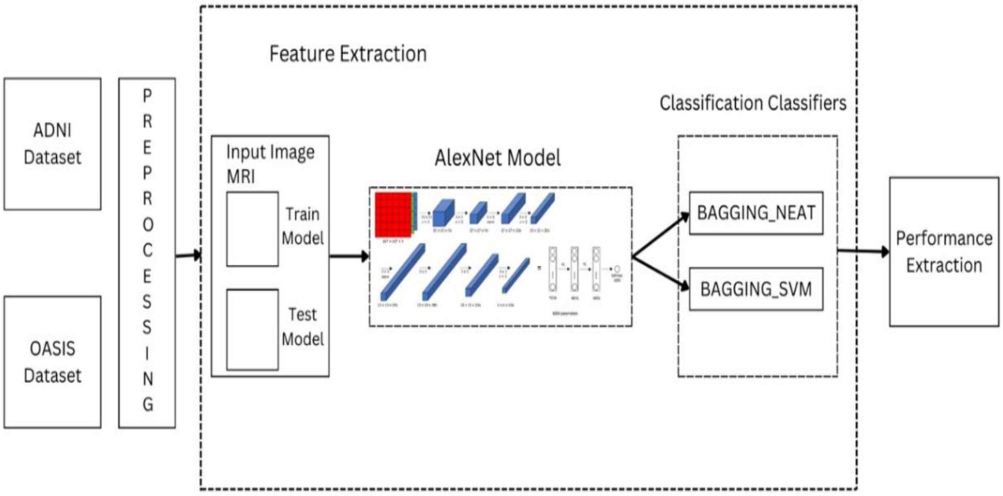


Figure 5.6 Overall structure of the proposed approach u sing CNN for feature extraction

5.4 Summary

This chapter discusses about an automated system for feature extraction using DL technique. The chapter begins with a detailed description about the traditional feature extraction techniques like GLCM, GLRLM, and GLDM. This study, proposed a real-time

DL approach for feature extraction which aids the classification approaches for an effective multiclass classification of AD. The CNN model aids in feature extraction. The procedure is assisted by an AlexNet model that has already been trained. A detailed discussion of the general framework of the suggested method for feature extraction using CNN is provided. The two different hybrid classifiers BAGGING_SVM, and BAGGING_NEAT are proposed in the subsequent chapters.