Integrating Transfer Learning and Deep Neural Networks for Accurate Medical Disease Diagnosis from Multi-Modal Data

Dr.Chamandeep Kaur[1], Dr.Abdul Rahman Mohammed Al-Ansari[2], Dr. Taviti Naidu Gongada[3], Dr. K. Aanandha Saravanan[4], Divvela Srinivasa Rao[5], Ricardo Fernando Cosio Borda[6], R. Manikandan[7]

Lecturer, Department of CS & IT, Jazan University, Saudi Arabia[1]
Department of Surgery, Salmanyia Hospital, Bahrian[2]
Assistant Professor, Dept. of Operations-GITAM School of Business, Gitam University, Visakhapatnam[3]
Department of ECE-Vel Tech Rangarajan, Dr. Sagunthala R&D Institute of Science and Technology[4]
Sr. Assistant Professor, Department of AI & DS, Lakireddy Bali Reddy College of Engineering, Mylavaram[5]
Universidad Autónoma de Ica, Peru., Peru[6]
Vel Tech Rangarajan, Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India-600062[7]

***Abstract*—Effective patient treatment and care depend heavily on accurate disease diagnosis. The availability of multi-modal medical data in recent years, such as genetic profiles, clinical reports, and imaging scans, has created new possibilities for increasing diagnostic precision. However, because of their inherent complexity and variability, analyzing and integrating these varied data types present significant challenges. In order to overcome the difficulties of precise medical disease diagnosis using multi-modal data, this research suggests a novel approach that combines Transfer Learning (TL) and Deep Neural Networks (DNN). An image dataset that included images from various stages of Alzheimer's disease (AD) was collected from kaggle repository. In order to improve the quality of the signals or images for further analysis, a Gaussian filter is applied during the preprocessing stage to smooth out and reduce noise in the input data. The features are then extracted using Gray-Level Co-occurrence Matrix (GLCM). TL makes it possible for the model to use the information gained from previously trained models in other domains, requiring less training time and data. The trained model used in this approach is AlexNet. The classification of the disease is done using DNN. This integrated approach improves diagnostic precision particularly in scenarios with limited data availability. The study assesses the effectiveness of the suggested method for diagnosing AD, focusing on evaluation metrics such as accuracy, precision, miss rate, recall, F1-score, and the Area under the Receiver Operating Characteristic Curve (AUC-ROC). The approach is a promising tool for medical professionals to make more accurate and timely diagnoses, which will ultimately improve patient outcomes and healthcare practices. The results show significant improvements in accuracy (99.32%).**

Keywords—Transfer learning; deep neural network; disease diagnosis; multi-modal data; Alexnet; GLCM; DNN; pre-trained model

I. INTRODUCTION

In the fields of machine learning and artificial intelligence, TL is a potent and well-liked technique that aims to use information learned from one task or domain to enhance the performance of another task or domain that is related to it [1]. It is predicated on the notion that knowledge obtained while resolving one problem can be applied and transferred to resolve another problem that is unrelated but nonetheless similar more effectively [2]. Models are created from scratch for each distinct task using traditional machine learning techniques and lots of labelled data. However, this procedure can be time-consuming, costly in terms of computation, and it may call for a sizable amount of labelled data, which isn't always available. Transfer learning overcomes these constraints by applying previously learned features or representations from a pre-trained model, referred to as the "source task," to a new target task with a smaller dataset. Pre-training and fine-tuning are typically the two essential steps in the TL process. A DNN is trained on a sizable dataset from the source task, such as image recognition on a sizable image dataset, during the pre-training phase [3]. The pre-trained model gains knowledge of broader features and patterns that can be used for a variety of tasks. Numerous industries, including healthcare, speech recognition, computer vision, and NLP, have found extensive use for TL. It has made it possible to create complex models that perform better even when there is a dearth of labelled data. TL encourages knowledge sharing and transfer across tasks by reusing learned representations, resulting in ML models that are more effective and efficient [4].

Alzheimer's disease is a progressive and irreversible neurological condition that primarily affects memory, thought, and behavior in the brain's cognitive regions [5]. It is the most typical cause of dementia, which is a group of brain disorders marked by a decline in memory, communication, and reasoning skills and the inability to perform daily tasks. Usually, a disease takes time to develop and gradually gets worse. People may experience mild memory loss in the early stages, as well as trouble finding words or organizing their thoughts. People with advanced Alzheimer's may experience confusion, mood swings, difficulty solving problems, and a loss of recognition of familiar faces and environments [6]. A vital component of healthcare is accurate disease diagnosis, which enables patients to receive timely and efficient care. In recent years, new opportunities for enhancing patient care and improving diagnostic accuracy have emerged due to the accessibility of a variety of medical data from various modalities. Data like genetic profiles, clinical reports, medical images, and textual information are included in this [7]. However, because of their complexity, heterogeneity, and the requirement to capture complex relationships between various modalities, analyzing and integrating these multi-modal data sources present significant challenges [8].

This research suggests a novel method that combines TL and DNN for precise medical disease diagnosis from multi-modal data to address these issues [9]. With the ability to transfer knowledge from a source domain to a target domain, transfer learning has become a potent machine learning technique [10]. TL enables the adaptation of learned representations and weights to new datasets, even when the data distributions differ significantly, by utilizing pre-trained models on large-scale datasets [11]. This knowledge transfer improves the ability of models to generalize and makes it easier to make an accurate diagnosis in the target domain [12]. In a number of industries, including computer vision, natural language processing, and healthcare, deep neural networks have achieved remarkable success [13]. These networks provide a strong framework for analyzing multi-modal medical data due to their capacity to learn intricate patterns and relationships from high-dimensional data. Recurrent neural networks (RNNs) capture temporal dependencies in sequential data, while convolutional neural networks (CNNs) are excellent at extracting features from images [14]. Furthermore, attention mechanisms allow the network to concentrate on pertinent data within the multi-modal data, increasing diagnostic precision [15].

There are several benefits to combining TL and DNN when diagnosing medical diseases using multimodal data [16]. First of all, it permits the use of prior knowledge and learned representations from comparable tasks or domains, which can greatly improve the performance of models on small medical datasets. Second, it combines the advantages of various network architectures to enable thorough analysis of multi-modal data by capturing both spatial and temporal dependencies [17]. Last but not least, it offers the possibility of individualized and accurate diagnosis, resulting in enhanced patient outcomes and improved treatment strategies [18]. A branch of ML and AI called "deep learning" focuses on teaching artificial neural networks how to carry out challenging tasks [19]. DNN, which are made up of multiple layers of interconnected nodes (neurons) that process and transform data, are used in this process [20]. These networks can learn and recognize patterns and features from enormous amounts of data because they are built to mimic the structure and operation of the human brain.

The goal of this study is to combine DNN and TL to create a reliable and accurate framework for diagnosing medical diseases [21]. The superiority of the approach over conventional methods and single-modal analyses through extensive tests and evaluations. Additionally, ablation studies are carried out to investigate the influence of various network architectures and TL strategies on diagnostic precision [22].

The key contributions of the paper is,

- The suggested model was trained using an image dataset that was taken from the Kaggle repository and contained images from different stages of AD.
- Before further analysis or feature extraction, the preprocessing stage of the data is where the Gaussian filter is used to smooth the input data and reduce noise.
- The study uses a pre-trained model for transfer learning called AlexNet, which is a well-known DNN architecture.
- The GLCM is used to extract features from the input data, capturing the spatial relationships and occurrence patterns of various pixel intensities in the image, which provides useful texture information for subsequent analysis or classification tasks.
- In order to accurately and efficiently classify diseases, Artificial Neural Networks (ANNs), a type of DNN, are used in the classification task to learn complex patterns and relationships from the input data.
- The research employs a number of evaluation metrics to assess the performance of the proposed approach for AD diagnosis. Accuracy, precision, recall, miss rate, F1-score, and the Area under the Receiver Operating Characteristic Curve (AUC-ROC) are some of these metrics.

The rest of this article is structured as follows: An overview of related research is given in Section II. The problem statement is presented in Section III. The methodology and architecture of our suggested approach are described in Section IV. Section V discusses the findings and subsequent discussion, and Section VI discusses the conclusion.

II. Related Works

Alzheimer's disease (AD) has become a growing problem among older people [23]. It is crucial for AD treatment to accurately identify mild cognitive impairment in the initial phases of the symptoms. Nevertheless aren't many samples of brain images and they come in a variety of methods, making it very challenging for machines to correctly categorize images of the brain. Through re-transfer training and multi-modal learning, the present research suggests an extremely fine brain image identification method to identifying AD. Diffusion tensor images (DTI) are initially finely classified into four different groups using a throughout its entirety DNN classification system called CNN4AD. Additionally, the re-transfer technique for learning is suggested on the basis of multipurpose theory of learning and is in accordance with the features of the multi-modal brain image data collection. The suggested strategy achieves greater precision with fewer labelled samples for training, according to the experiment's findings. This might aid in a quicker and more precise diagnosis of AD by medical professionals.

AD is a severe and unchangeable dementia of the central nervous system that impairs memory and ability to think [24]. DL algorithms have been proven successful in medical applications in treating this neuro-degenerative illness that results in neurological impairment and mental decline. DL techniques have been discovered to be efficient for duties like recognizing trends in imaging data and helping with diagnosis. When there is a lack of information, which applies an established model to a novel assignment, might prove especially helpful. TL has been used by investigators to successfully identify AD. AD is a severe and unchangeable dementia of the central nervous system that impairs memory and ability to think. DL algorithms have been proven successful in medical applications in treating this neuro-degenerative illness that results in neurological impairment and mental decline. DL techniques have been discovered to be efficient for duties like recognizing trends in imaging data and helping with diagnosis. When there is a lack of information, which applies an established model to a novel assignment, might prove especially helpful. TL has been used by investigators to effectively detect AD.

In modern healthcare environment, diagnostic testing has taken on an important function [25]. Brain cancer, among a particularly deadly disease and the main cause of death worldwide, is a significant area of study in the discipline of healthcare imaging. A rapid and reliable diagnosis made using MRI can enhance the study and outlook of brain tumors. Healthcare visuals needs to be recognized, divided, and categorized in order for automated diagnosis techniques to help physicians in determining the presence of brain tumors. It is crucial for establishing a computerized approach because radiologists find it tedious and prone to errors to manually identify malignancies in the brain. As an outcome, an accurate strategy for identifying and classifying brain tumors is given. The suggested process entails five phases. The borders in the original image are found using a linear contrast enhancement in the initial step. The following step involves creating a unique, 17 layered DNN architecture for segmenting brain tumors. The next step involves training the altered MobileNetV2 design employing for extracting features. The most desirable characteristics were chosen in the following step using an entropy-based regulated technique and a M-SVM. On the information sets from BraTS 2018 and Figshare, the approach that was suggested was tested. An investigation demonstrates that the suggested approach for classifying and detecting brain tumors surpasses existing techniques both qualitatively and in quantitative terms, with accuracy rates of 97.47% and 98.92%, accordingly The XAI technique is then used to clarify the outcome. The suggested approach performed better than existing approaches for identifying and classifying brain tumors. These results show that the suggested method performed better in the context of increased quantitative assessments with better accuracy as well as visual appeal.

After surgical procedures, tumors in breasts patients frequently experience recurring and metastases [26]. For the creation of accuracy therapy, forecasting a person's likelihood of metastatic growth and recurrence is crucial. In the present investigation, histopathological images that were stained with H&E, medical records, and information concerning gene expression to offer an innovative multi-modal DL forecasting model. To be more precise, DNN to record every image inhibit into a 1D incorporate vector after segmenting tumor spots in H&E into image segments. The probable likelihood of recurrence as well as metastasis for every participant was then predicted by the attention-getting component, which scored every region of the H&E-stained images and paired visual characteristics with a medical and gene transcription information gathered. All 196 cancerous breast specimens with concurrently accessible clinical, expression of genes, and H&E data from the cancer genome database to test the hypothesis. The geographic distributions of the collected data were subsequently maintained among the two databases by centralized collection as the specimens were split into the training and testing sets in a ratio of 7:3. On the evaluation set, the multi-modal model outperformed those that relied merely on H&E image, arranging the information, and medical information, each achieving an AOC value of 0.75. This research could potentially be useful in the clinical setting to determine individuals with breast cancer who are at high risk for responding well to following surgery adjuvant therapy.

A neurodegenerative condition known as Alzheimer's disease (AD), it affects numerous individuals all over the world [27]. Although AD remains one of the most prevalent brain disorders, it can be challenging to identify, and in order to distinguish comparable trends, it needs a classification depiction of its features. Neural networks are commonly used in research to address increasingly difficult issues, including AD detection. Researchers and scientists without specialized expertise in artificial intelligence see those methods as well-understood and even sufficient. Therefore, it is crucial to find a detection technique that is both fully automated and simple for non-AI specialists to utilize. To quickly streamline the creation of neural networks and consequently democratize artificial intelligence, the approach should determine effective settings for the modeling variables. Multi-modal medical image fusion also provides deeper multimodal characteristics and a better capacity for information representation. For more precise diagnostics and more effective therapy, a fusion image is created by combining pertinent and related information from many input images. In order to diagnose Alzheimer's disease, the present research introduces a MultiAz-Net, a novel optimized ensemble-based deep neural network learning model that incorporates heterogeneous data gathered from PET and MRI scans. The study provides an automated method for anticipating the early start of AD according to characteristics identified from the fused data. The suggested structure involves three steps: picture fusion, feature extraction, and classification. A multi-objective optimization technique called the Multi-Objective Grasshopper Optimization technique (MOGOA) is provided to optimize the MultiAz-Net's layers. To do this, the required functions of objectives are enforced and the appropriate values for the proposed variables are looked for. Using the openly accessible Alzheimer neuroimaging dataset, the suggested deep ensemble model was empirically evaluated to complete four tasks for grouping Alzheimer's illness, three binary categorizations, along with a multi-class categorization task.

III. Problem Statement

Limited diagnostic accuracy results from traditional machine learning methods' inability to fully grasp the subtleties found in multimodal medical data. As each modality offers distinct insights into a patient's status, it has become clear that it is necessary to combine different sources of information, such as pictures and clinical data. The accuracy of diagnosing various medical disorders might be greatly increased by creating an integrated framework that smoothly integrates multiple modalities. Additionally, DNNs' performance in image analysis tasks points to their promise in this situation. On the other hand, over fitting and unsatisfactory outcomes frequently arise from training DNNs from scratch on scant medical data. Therefore, there is a need for a technique that efficiently uses transfer learning to adapt pre-trained models to medical diagnosis tasks while maximizing the use of multi-modal data. The development of appropriate data fusion strategies that capture the complimentary information provided by each modality is necessary for the integration of multi-modal data.

To guarantee an effective translation of pre-trained models to the medical domain, where the data distribution may differ greatly from general datasets, transfer learning algorithms must be developed. Choosing the right neural network topologies and optimization techniques to handle the complicated, high-dimensional medical data is also essential. The suggested framework should also take into account the ethical ramifications of using AI in medical diagnosis, including transparency and interpretability. This study's main goal is to provide a novel method for precise medical condition detection utilizing multimodal data that blends transfer learning and deep neural networks. The suggested framework seeks to intelligently leverage the pre-trained knowledge from other domains and efficiently incorporate various medical data sources in order to get beyond the constraints of existing machine learning approaches and increase diagnosis accuracy. The research's ultimate goal is to advance the area of medical diagnostics by enabling more accurate and early illness identification, which can enhance patient outcomes and make healthcare systems more effective [28].

IV. Proposed TL-DNN Framework

The methodology involved using an image dataset from Kaggle that contained images from various stages of AD to train the suggested model. Data preprocessing used the Gaussian filter to smooth out and reduce noise before analysis. The pre-trained AlexNet DNN architecture was used with transfer learning. The GLCM was used for feature extraction in order to record spatial relationships and pixel intensity patterns for texture data. ANN were used to extract complex patterns from the input data and classify diseases. Different metrics were also used in performance evaluation. Fig. 1 shows the proposed methodology.

A. Data Collection

The effectiveness of the study would be enhanced by having access to a comprehensive dataset that contains a wide range of disorders, several imaging modalities (such as MRI, CT, and X-ray), and a variety of patient demographics [29]. Such a dataset would enable the examination of the suggested approach's performance across various medical diseases, imaging modalities, and patient groups, providing a more thorough evaluation of the method. It would be possible to evaluate how well the technique generalizes and adjusts to various illness presentations and demographic characteristics using this extensive dataset, which would eventually increase its potential usefulness and relevance in real-world clinical settings. The intended database originated from a publicly accessible Kaggle repository [30]. These MRI images of the brains of individuals who are Very Mildly Demented (VMD), Moderately Demented (MOD), Non-Demented (ND), and Mildly Demented make up this dataset. An image dataset that included images from various stages of AD was used to train the suggested model. Table I shows the overall amount of image samples used as input after enhancement, broken down by class.

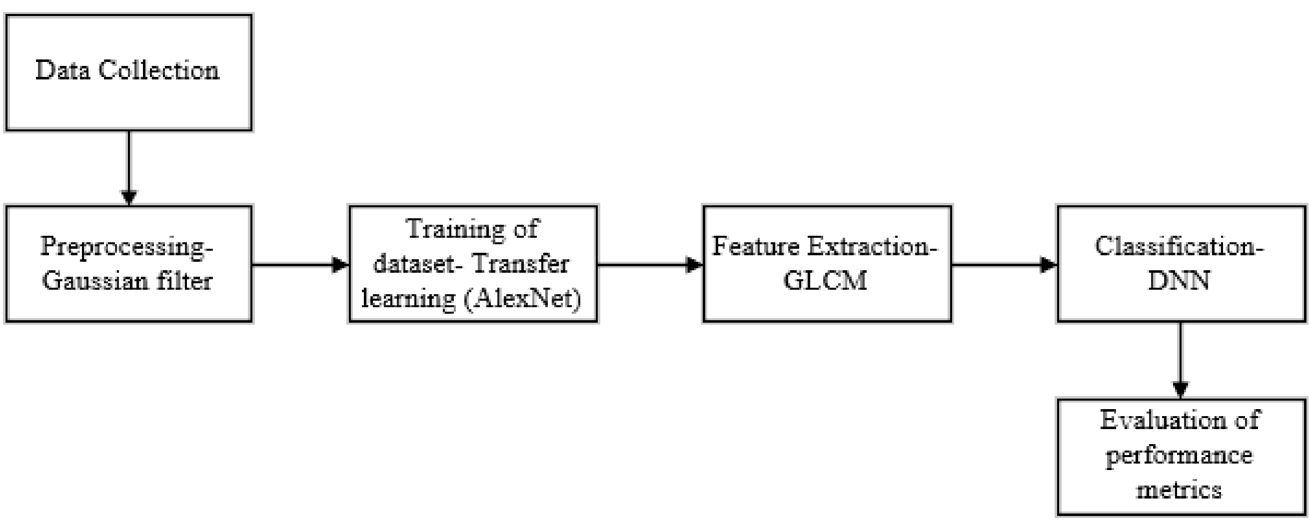

Fig. 1. Proposed methodology.

TABLE I. DATASET PARAMETERS

Mental State	No. of image samples
VMD	1792
MOD	1024
ND	2560
MD	1017

B. Multi-Modal Data Fusion using Autoencoders

AI-driven multi-modal data fusion approaches have evolved as a way to tap into the deep insights contained in these diverse sources. Utilizing autoencoders, a kind of neural networks created expressly for unsupervised feature learning and data reduction, is one well-known strategy in this field. Autoencoders provide a strong foundation for multi-modal data fusion since they are typically used for dimensionality reduction and data reconstruction tasks. They serve as modality-specific encoders in this situation, extracting unique patterns and characteristics from each data source. The models may learn representations that capture modality-specific information while abstracting away noise and unnecessary features by training distinct autoencoders for each modality. In order to merge the encoded representations from various modalities, a shared latent space must be created, and this is where the multi-modal fusion's key challenge resides. The model may learn shared characteristics and relationships that might not be obvious in individual modalities alone by using this joint space as a bridge to enable the integration of modalities. The constraints presented by multi-modal medical data, where complicated interrelationships frequently influence diagnostic findings, are well matched with the flexibility of autoencoders in capturing nuanced connections.

The possibility for higher data quality is one of the intrinsic benefits of utilizing autoencoders for multi-modal fusion. Autoencoders naturally filter out unimportant fluctuations by condensing noisy and high-dimensional input into a latent space, improving the signal-to-noise ratio. This can result in conclusions that are more reliable and generalizable, especially when working with noisy medical data. The clinical relevance and interpretability of multi-modal data fusion employing autoencoders ultimately determines its effectiveness. The combined representations should increase diagnostic precision while also giving doctors useful new information. Making sure that the AI-driven fusion effectively integrates with medical expertise and decision-making involves translating these learnt qualities back into clinical words that can be understood by patients.

C. Preprocessing

Applying a Gaussian filter to an image as part of preprocessing it with a Gaussian function helps to smooth it out and reduce noise. The Gaussian function, a mathematical function that has a bell-shaped curve, serves as the filtering operation's kernel or mask. A Gaussian kernel is used in the filter's operation, and each pixel is given a weight based on how close to its neighbors it is. The values of the Gaussian function at each point along the kernel are used to calculate the weights. The degree of image smoothing can be altered by modifying the Gaussian filter's parameters, such as the kernel size and standard deviation. Smaller kernel sizes and lower standard deviations preserve finer details while larger kernel sizes and higher standard deviations result in more extensive smoothing. A preprocessed image with less noise, fewer high-frequency details, and a smoother overall appearance is the end result. A common preprocessing method in image processing, Gaussian filtering is especially beneficial for tasks like demising, feature extraction, and improving image quality. Below is the equation for the Gaussian function.

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{z^2}{2\pi\sigma^2}} \quad (1)$$

Where the standard deviation of the distribution is denoted as σ. The distribution is assumed to have a mean of 0.

D. Employing Transfer Learning for training the Dataset

1) Pre trained model – AlexNet: A convolutional network framework that has already been trained is called AlexNet. Transfer learning is the process for employing an approach that has been trained, and it is currently widely employed in applications that employ DL. In the suggested technique, the study employed an updated form of this AlexNet framework. AlexNet is an eight-layer structure with accessible variables, five of which comprise layers of convolution that combine maximum pooling with three layers that are completely interconnected. ReLU is a nonlinear function of activation that is present in every layer. Images obtained from the Pre-processed layers are retrieved by the network inputs layer. Pre-processing, which may be performed in a variety of methods, such as by enhancing specific image characteristics or decreasing the image, is an essential phase in order to produce appropriate datasets. Image scaling is a necessary procedure since images come in a variety of sizes. As a result, images were reduced in dimension to 227 * 227 * 3, where 227 * 227 denotes the input images' height and breadth and three indicates the total amount of channels.

The approach incorporates the previously trained CNN network as well as AlexNet, which has a significant influence on contemporary deep learning techniques. After being modified to accommodate the needs, this CNN network was used to feed the preprocessed images to the suggested AlexNet transfer learning system. The resultant of the categorization layer, completely interconnected layer, and softmax layer constitute three substantially adjusted layers in the design that correspond to the issue specification. Fig. 2 depicts the altered network utilized for transfer learning.

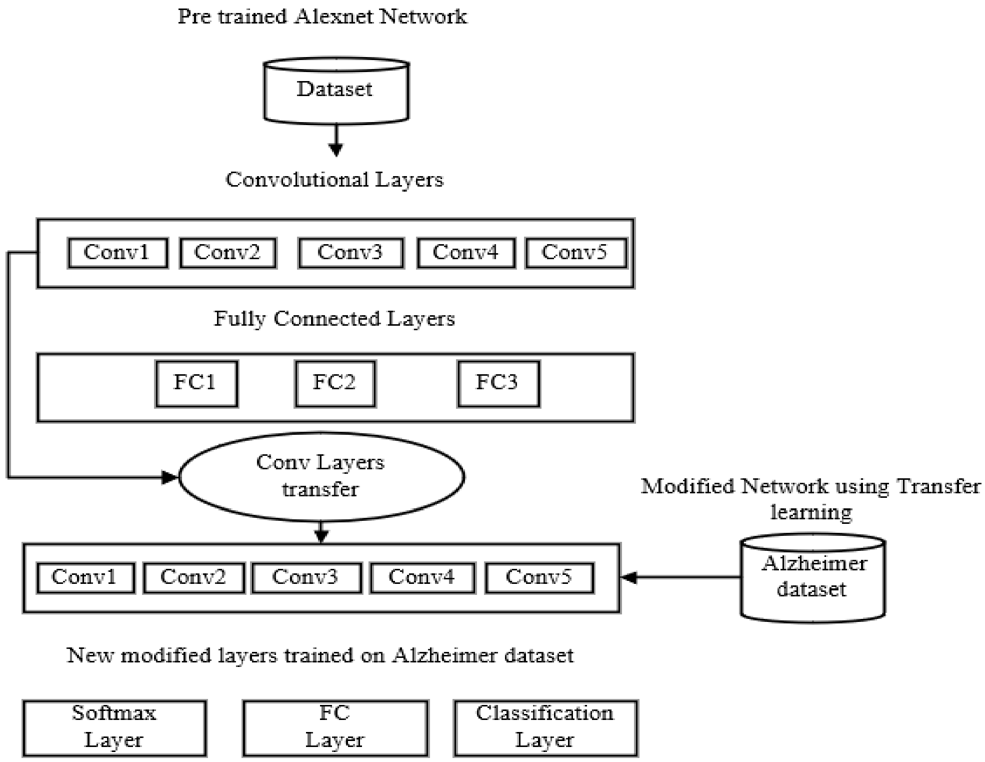

Fig. 2. Pre trained transfer learning model.

Although the subsequent modification layers are learned with Alzheimer datasets, the first five segments of the network, which had been trained employing AlexNet, become unchanged. The remaining three sections are set up to categorize the images into their specific category labels based on the resultant labels assigned to each class. The dimensions of the results, which is made up of several categories, is one of the input variables for completely interconnected layers. A completely linked layer's generated size is equal to the entire quantity of labelled classes. Softmax functions are applied to the data provided using softmax layers. A layer that is completely connected is used for acquiring the class-specific characteristics needed to distinguish across categories, whereas a layer made up of convolutions represents generalized visual characteristics like edge recognition throughout training. As a result, the class-specific characteristics are adjusted for fully linked layers. To categorize images into several classes, the suggested model is trained employing multi-class labelling of Alzheimer's disease.

Numerous variables can be utilized for preparing the system for training, or there are other training choices that can be provided. The remaining three layers of AlexNet—fully interconnected layers, outputs categorization layers and SoftMax layers—are not included in the extraction process since they are not necessary for transfer learning.

The following variables can be utilized as training alternatives: learning rates, the amount of iterations, the rate of validation, and the total amount of epochs.

The different components of CNN are in responsibility for extracting the universal properties from the images, which are referred to as the domain of origin. The resulting learning characteristics can then be applied to determine and categorize a variety of additional operations, such as the identification of Alzheimer's disease. Customized generated models that may be utilized for validation are positioned on clouds. The approach that has been trained evaluates each image and classifies them according to their corresponding categories, which are MD, MOD, VMD and ND. The trained system has precisely the characteristics for image processing that it obtained throughout the training phase. Thus, it was shown that the effectiveness of Alzheimer's disease recognition is influenced by the identification of Alzheimer's disease phases in individuals and the transfer of information from big datasets.

E. Feature Extraction using GLCM

By computing various statistical measures from the GLCM matrix, different facets of the image texture can be captured during feature extraction. These measurements, also referred to as GLCM features or texture features, offer numerical data about the spatial relationships between gray-level values in an image.

1) Energy: By adding up the squared values of each component in the GLCM matrix, the energy also known as the angular second moment or uniformity is determined. It gauges the texture of the image's overall intensity or contrast. While a lower energy value reflects a more complex or heterogeneous texture with variations in gray-level values, a higher energy value denotes a more homogeneous texture where the gray-level values are evenly distributed.

$$E = \sum_{p} \sum_{q} \{N(p,q)\}^2 \tag{2}$$

Where the images are denoted as N, and the image's squares with grey levels are labeled as (p, q).

2) Contrast: In terms of the variations in their grey levels, contrast quantifies the intensity contrast between pixel pairs. It displays how much local variation or abrupt texture transitions there are. Low contrast values imply a more uniform or smooth texture, while high contrast values suggest significant differences between adjacent pixel pairs.

$$C = \sum_{y=0}^{I_q} y^2 \left\{ \sum_{p=1}^{I_q} \sum_{q=1}^{I_q} N(p,q) \right\} \tag{3}$$

I stand for the grayscale of the images, N for the images, and (p, q) for the square of an image's grayscale.

3) Correlation: Correlation measures how closely the linear associations between the gray-level values of adjacent pixels are related to one another. It shows how the values of the grey levels vary consistently or predictably between adjacent pixels. A stronger or more nonlinear relationship between the gray-level values is indicated by a higher correlation value than by a lower correlation value.

$$C_o = \frac{\sum_{p} \sum_{q} (p,q) N(p,q) - \mu_u \mu_v}{\sigma_u \sigma_v} \tag{4}$$

In the images, the values of mean, as well as standard deviation, are μ_u, μ_v, $\sigma_{u,}$ *and* σ_v are characterized as row and column.

4) Entropy: The amount of information or uncertainty related to the texture is measured by entropy. It displays the distribution of gray-level values among adjacent pixel pairs. A texture with a higher entropy value is more varied or heterogeneous, with dispersed and unpredictable gray-level values. A lower entropy value, on the other hand, denotes a more regular or uniform texture, where the values of the grey levels are concentrated or predictable.

$$En = -\sum_{p} \sum_{q} N(p,q) \log\big(N(p,q)\big) \tag{5}$$

F. Classification using DNN

DNN have an intricate network model and adhere to the same structure as standard ANN. It aids in the development of models and the clear definition of complex structure. It has 'n' layers that are hidden that analyze information obtained from the layer preceding it, which is referred to as the initial layer. Following every moment, the rate of errors of the input information is going to be gradually lowered by changing the weights for every node, back promoting the network's structure and continuing until it achieves improved outcomes. In the input layer, any amount of the inputs can be designated as input nodes. In order to intensify the conditioning analyze, DNN typically has multiple nodes than the data it receives from the layer. As distinct nodes for output in the layer that produces the results, any amount of generates can be stated. The total quantity of points in the data for input and output, bias, developing rate, starting weights for modification, the amount of hidden layers, the total number of nodes in each hidden layer, and prevent the requirements for stopping the operation of the times are considered to be the variables that are utilized by the DNN. In order to prevent network leads to from being invalidated discrimination value is typically set to 1 in any neural network design. Additionally, the rate of learning is set to 0.15 by standard and is later arbitrarily impacted through trial and error to produce different results compared to the equation. The system can determine the beginning weights of the nodes at arbitrary, modify it throughout replication by determining its error rate, and modify it frequently after every phase. The amount of inputs and the dimension of the information determine the amount of secret layers as well as node locations in each hidden layer. Both the system reaches the ideal amount of periods or the anticipated outcome from the approach to learning is realized is referred to as the network's removal state. It requires a greater amount of resources to train the computational framework if there are more sections and node locations in the framework.

1) Artificial neural network: ANN can be used to classify cases of AD. An example of a ML model is an ANN, which takes inspiration from the design and operation of neural networks in biology. They are made up of interrelated "neurons," or nodes, organized in levels. Every neuron takes in information, uses a stimulus to create an output, and subsequently sends the result to the neural layer below. Fig. 3 shows the architecture of ANN.

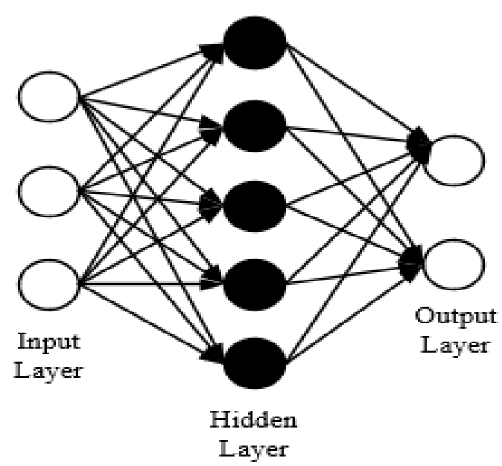

Fig. 3. Architecture of ANN.

To avoid the biased impact caused by the input data due to the different ranges, each input distinctive, $b_i; 1 \leq i \leq 5$, had been normalized inside the same range [y, a]. Let's say an additional range is defined as [y, x]. Let b_{max} and b_{min} represent the upper and lower bounds of the concept's natural range. The distinctive values b_i within the range of values [y, x] could be normalized using the formula that follows symbolized as b_i.

$$b_i' = \frac{(y-x)(b_i-b_{min})}{b_{max}-b_{min}} + y \tag{6}$$

A symmetrical function has x and y parameters of 0.1 and 0.9, respectively, whereas a tangent hyperbola function has x and y coordinates of 0.9 and 0.9 in both cases. The symmetrical function is shown below.

$$f(b) = [1 + \exp(-b)]^{-1} \tag{7}$$

The highest and lowest values of the hyperbola's tangent activating function were limited to [-1, 1], while the highest and lowest values of the symmetrical function had been limited to [0, 1]. The tangential hyperbolic function of activation is shown in Eq. (3).

$$f(b) = [\exp b - \exp(-b)]/[\exp(b) + \exp(b)] \tag{8}$$

This limits the network's expected output for symmetrical and hyperbolic angular functions of activation, respectively, to [0, 1] and [1, 1]. On the contrary hand, the system's outputs don't accurately reflect the data's true worth. The output value has to be changed to its actual value, b_d using the following,

$$b'_d = b_{min} + [\frac{b_d-y}{x-b}] \tag{9}$$

Equations 10 and 11 calculate the generalization error using the summation squared errors (SSE) and regression analysis error (R2):

$$summation\ squared\ error = \sum(b'_d - b_t)^2 \tag{10}$$

$$Regression\ analysis = 1 - [\sum(b'_d - b_t)^2 / \sum(b'_d - b)^2] \tag{11}$$

Where b is the measurement value of the evaluation sequence t and b_t is the average of the data collected.

V. Results and Discussion

The methodology involved training the suggested model using an image dataset from Kaggle that included images from various stages of AD. Before analysis, noise in the data was smoothed out and reduced using the Gaussian filter. Transfer learning was used in conjunction with the pre-trained AlexNet DNN architecture. In order to capture spatial relationships and pixel intensity patterns for texture data, the GLCM was used for feature extraction. ANN were used to classify diseases and extract intricate patterns from the input data. Additionally, various metrics were applied when assessing performance.

A. Accuracy

The system model's overall performance is assessed using accuracy. Essentially, it is the notion that each encounter will be accurately predicted. Accuracy is provided in equation (12),

$$Accuracy = \frac{T_{Pos} + T_{Neg}}{T_{Pos} + T_{Neg} + F_{Pos} + F_{Neg}} \tag{12}$$

TABLE II. Comparison of Accuracy

Method	Accuracy (%)
Inception V4 [31]	73.75
Landmark based feature extraction [31]	79.02
ADDTLA [31]	91.7
TL-DNN	99.32

The accuracy of the suggested TL-DNN and the existing methods is shown in Table II. Due to its accuracy value of 99.32%, the suggested method is determined to be more effective than the others. The accuracy is shown in Fig. 4.

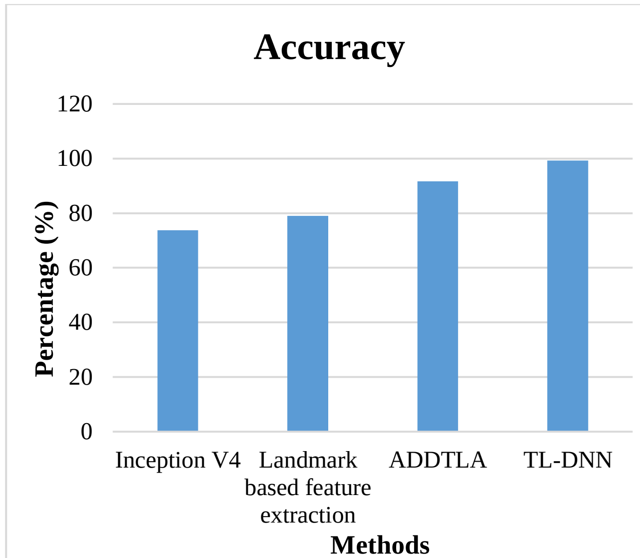

Fig. 4. Comparison of accuracy.

B. Precision

Besides to being correct, precision also refers to how closely two or more calculations resemble one another. The relationship between accuracy and precision demonstrates how repeatedly a finding can be made. Equation (13) can be used to calculate precision.

$$P = \frac{T_{Pos}}{T_{Pos} + F_{Pos}} \tag{13}$$

C. Recall

Recall is the proportion of all pertinent results that the methods were effectively sorted. The ratio among the true positive and false negative values is used to calculate the appropriate positive for those numbers. It is referred to in equation (14).

$$R = \frac{T_{Pos}}{T_{Pos} + F_{Neg}} \tag{14}$$

D. F1-Score

The F1-Score formula combines recall and accuracy. Precision and recall are used to calculate the F1-Score that is given in equation (15).

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (15)$$

TABLE III. COMPARISON OF PERFORMANCE METRICS

Method	Precision (%)	Recall (%)	F1-score (%)
DNN [32]	97	97	97
SMO [32]	94.1	93.9	96.3
LDA [32]	95.7	95.5	95.5
KNN [32]	89.2	86.4	86.6
TL-DNN	98.12	98	98.5

Table III compares the precision, recall, and F1-score of current methods with the suggested TL-DNN. The precision of the recommended TL-DNN is higher than that of the other approaches. In Fig. 5, it is shown.

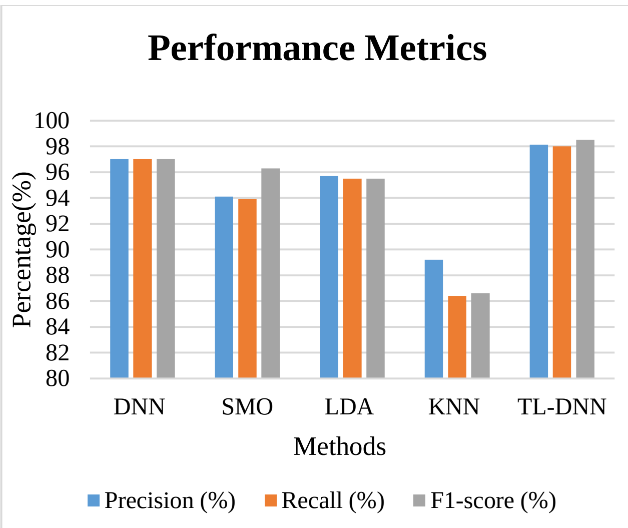

Fig. 5. Performance Metrics.

E. Miss Rate

A binary classification model's effectiveness is measured by the miss rate, also referred to as the false negative rate or Type II error rate. It calculates the percentage of positive instances that the classifier actually classifies as negative.

$$MissRate = \frac{F_{Neg}}{T_{Pos} + F_{Neg}} \quad (16)$$

The miss rate of the proposed TL-DNN is compared with the miss rate of other methods, which is given in Table IV. The miss rate of TL-DNN is 5.1% which lower than other methods. It is represented in Fig. 6.

F. Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

The area under the ROC curve is known as the AUC. It evaluates the classifier's overall performance in separating the two classes while taking into account all potential classification thresholds. A classifier's performance is graphically represented by the ROC curve. Fig. 7 depicts the AUC-ROC.

TABLE IV. COMPARISON OF MISS RATE

Methods	Miss Rate (%)
Inception V4 [31]	26.25
Landmark based feature extraction [31]	20.98
ADDTLA [31]	8.3
TL-DNN	5.1

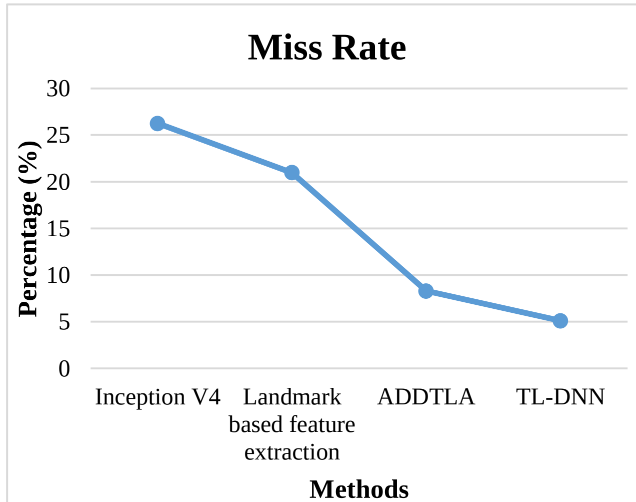

Fig. 6. Comparison of miss rate.

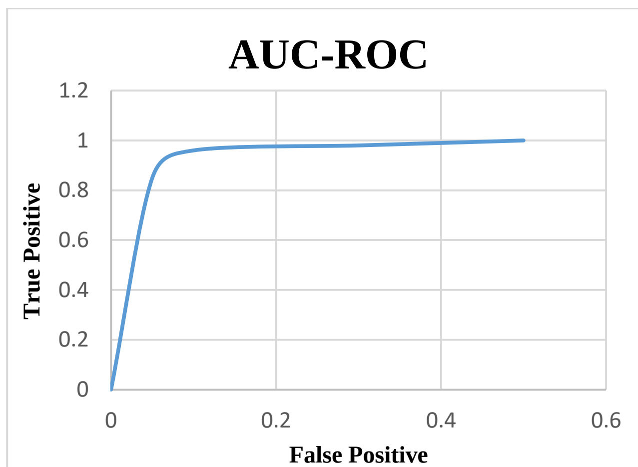

Fig. 7. Area under the receiver operating characteristic curve.

G. Discussion

The end result discusses how various metrics can be used to evaluate a system model's effectiveness when performing binary classification tasks. Evaluation of the accuracy, precision, miss rate, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) are the main considerations. Accuracy is a binary classification metric that shows how frequently the model predicts correctly. The accuracy of various methods is compared in the table and figure, with TL-DNN having the highest accuracy (99.32%). Less false positives are implied by higher precision. Comparing precision, Table III shows that TL-DNN has the highest precision (98.12%). The miss rate for the TL-DNN is the lowest (5.1%). The F1-Score and recall values achieved by TL-DNN are 98% and 98.5%, respectively. The outcome offers a thorough analysis of the model's performance using a variety of metrics, enabling a thorough evaluation of its efficiency in the binary classification task. The outcomes demonstrate that the TL-DNN method performs remarkably across a variety of evaluation metrics.

VI. Conclusion

For accurate medical disease diagnosis using multimodal data, the combination of TL and DNN shows to be a highly effective and promising approach. This study effectively illustrates the potential to improve diagnostic precision by utilizing information from pre-trained models and imaging data sources. Even with little training data, the model can generalize to different medical specialties and diseases by using transfer learning. Due to its robustness and dependability in diagnosing medical conditions, this integrated approach has a lot of potential for use in the real world. The model's achieved high generalization and accuracy highlight its practical value in aiding medical professionals in making prompt and accurate diagnoses. This strategy can result in better treatment choices, better patient outcomes, and possibly lower healthcare costs by accurately diagnosing diseases. Despite these noteworthy developments, there are still some issues that need to be resolved. Maintaining the model's performance in real-world medical settings requires careful consideration of data quality and mitigating potential biases. This study offers insightful information into the field of medical disease diagnosis and demonstrates the enormous potential of fusing DNN and TL. The deployment of extremely precise and trustworthy diagnostic tools that will revolutionize medical procedures and have a big impact on patient care by overcoming obstacles and further improving the strategy. Innovative and ethical solutions to benefit patients and the healthcare industry as a whole will be developed as the field of artificial intelligence in healthcare develops as a result of ongoing research and collaboration between AI experts and healthcare professionals. The legal implications of utilizing AI in medical diagnosis, including transparency, interpretability, and possible biases, are not adequately covered in the research. The investigation of other designs that can perhaps produce superior results may be constrained by the employment of a specific neural network architecture (AlexNet).

References

[1] P. Xi, C. Shu, and R. Goubran, “A Unified Deep Learning Framework for Multi-Modal Multi-Dimensional Data,” in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, IEEE, 2019, pp. 1–6.

[2] F. Zhang, Z. Li, B. Zhang, H. Du, B. Wang, and X. Zhang, “Multi-modal deep learning model for auxiliary diagnosis of Alzheimer’s disease,” *Neurocomputing*, vol. 361, pp. 185–195, 2019.

[3] G. Lee, K. Nho, B. Kang, K.-A. Sohn, and D. Kim, “Predicting Alzheimer’s disease progression using multi-modal deep learning approach,” *Scientific reports*, vol. 9, no. 1, p. 1952, 2019.

[4] G. Sreeja and O. Saraniya, “Image fusion through deep convolutional neural network,” in *Deep learning and parallel computing environment for bioengineering systems*, Elsevier, 2019, pp. 37–52.

[5] Z. Xiang *et al.*, “Self-supervised multi-modal fusion network for multi-modal thyroid ultrasound image diagnosis,” *Computers in Biology and Medicine*, vol. 150, p. 106164, Nov. 2022, doi: 10.1016/j.compbiomed.2022.106164.

[6] Y. Gulzar, “Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique,” *Sustainability*, vol. 15, no. 3, Art. no. 3, Jan. 2023, doi: 10.3390/su15031906.

[7] L. Xu, M. Bennamoun, F. Boussaid, S. An, and F. Sohel, “Coral classification using densenet and cross-modality transfer learning,” in *2019 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2019, pp. 1–8.

[8] U. Evci, V. Dumoulin, H. Larochelle, and M. C. Mozer, “Head2Toe: Utilizing Intermediate Representations for Better Transfer Learning,” in *Proceedings of the 39th International Conference on Machine Learning*, PMLR, Jun. 2022, pp. 6009–6033. Accessed: Aug. 22, 2023. [Online]. Available: https://proceedings.mlr.press/v162/evci22a.html

[9] “Preparing for the next pandemic via transfer learning from existing diseases with hierarchical multi-modal BERT: a study on COVID-19 outcome prediction | Scientific Reports.” https://www.nature.com/articles/s41598-022-13072-w (accessed Jul. 13, 2023).

[10] Q. Zhu, N. Yuan, J. Huang, X. Hao, and D. Zhang, “Multi-modal AD classification via self-paced latent correlation analysis,” *Neurocomputing*, vol. 355, pp. 143–154, 2019.

[11] J. He, C. Zhou, X. Ma, T. Berg-Kirkpatrick, and G. Neubig, “Towards a Unified View of Parameter-Efficient Transfer Learning.” arXiv, Feb. 02, 2022. doi: 10.48550/arXiv.2110.04366.

[12] N. E. Benzebouchi, N. Azizi, A. S. Ashour, N. Dey, and R. S. Sherratt, “Multi-modal classifier fusion with feature cooperation for glaucoma diagnosis,” *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 31, no. 6, pp. 841–874, 2019.

[13] A. S. K. Reddy *et al.*, “Multi-modal fusion of deep transfer learning based COVID-19 diagnosis and classification using chest x-ray images,” *Multimed Tools Appl*, vol. 82, no. 8, pp. 12653–12677, Mar. 2023, doi: 10.1007/s11042-022-13739-6.

[14] Y. Pan, M. Liu, C. Lian, Y. Xia, and D. Shen, “Disease-image specific generative adversarial network for brain disease diagnosis with incomplete multi-modal neuroimages,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*, Springer, 2019, pp. 137–145.

[15] J. Hu *et al.*, “Towards accurate and robust multi-modal medical image registration using contrastive metric learning,” *IEEE Access*, vol. 7, pp. 132816–132827, 2019.

[16] J. Kalajdjieski *et al.*, “Air Pollution Prediction with Multi-Modal Data and Deep Neural Networks,” *Remote Sensing*, vol. 12, no. 24, Art. no. 24, Jan. 2020, doi: 10.3390/rs12244142.

[17] W. Liu *et al.*, “Research on medical data feature extraction and intelligent recognition technology based on convolutional neural network,” *IEEE Access*, vol. 7, pp. 150157–150167, 2019.

[18] “Sensors | Free Full-Text | Learning for a Robot: Deep Reinforcement Learning, Imitation Learning, Transfer Learning.” https://www.mdpi.com/1424-8220/21/4/1278 (accessed Aug. 22, 2023).

[19] K. Oh, Y.-C. Chung, K. W. Kim, W.-S. Kim, and I.-S. Oh, “Classification and visualization of Alzheimer’s disease using volumetric convolutional neural network and transfer learning,” *Scientific Reports*, vol. 9, no. 1, p. 18150, 2019.

[20] “RadImageNet: An Open Radiologic Deep Learning Research Dataset for Effective Transfer Learning | Radiology: Artificial Intelligence.” https://pubs.rsna.org/doi/full/10.1148/ryai.210315 (accessed Aug. 22, 2023).

[21] O. Pelka, F. Nensa, and C. M. Friedrich, “Branding-fusion of meta data and musculoskeletal radiographs for multi-modal diagnostic recognition,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[22] S. Ismail, B. Ismail, I. Siddiqi, and U. Akram, "PCG classification through spectrogram using transfer learning," *Biomedical Signal Processing and Control*, vol. 79, p. 104075, Jan. 2023, doi: 10.1016/j.bspc.2022.104075.

[23] M. Fang, Z. Jin, F. Qin, Y. Peng, C. Jiang, and Z. Pan, "Re-transfer learning and multi-modal learning assisted early diagnosis of Alzheimer's disease," *Multimed Tools Appl*, vol. 81, no. 20, pp. 29159–29175, Aug. 2022, doi: 10.1007/s11042-022-11911-6.

[24] S. K. G. U., A. P. R., and S. K. G., "Detecting Alzheimer's Disease using Multi-Modal Data: An Approach Combining Transfer Learning and Ensemble Learning," in *2023 International Conference on Control, Communication and Computing (ICCC)*, May 2023, pp. 1–6. doi: 10.1109/ICCC57789.2023.10165454.

[25] S. Maqsood, R. Damaševičius, and R. Maskeliūnas, "Multi-Modal Brain Tumor Detection Using Deep Neural Network and Multiclass SVM," *Medicina*, vol. 58, no. 8, Art. no. 8, Aug. 2022, doi: 10.3390/medicina58081090.

[26] Y. Yao *et al.*, "ICSDA: a multi-modal deep learning model to predict breast cancer recurrence and metastasis risk by integrating pathological, clinical and gene expression data," *Briefings in Bioinformatics*, vol. 23, no. 6, p. bbac448, Nov. 2022, doi: 10.1093/bib/bbac448.

[27] W. N. Ismail, F. R. P. P., and M. A. S. Ali, "A Meta-Heuristic Multi-Objective Optimization Method for Alzheimer's Disease Detection Based on Multi-Modal Data," *Mathematics*, vol. 11, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/math11040957.

[28] C.-Y. Wee *et al.*, "Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations," *NeuroImage: Clinical*, vol. 23, p. 101929, 2019.

[29] A. El-Said *et al.*, "Assessing the Impact of Demographic Factors on Presenting Conditions or Complaints Among Internal Medicine Patients in an Underserved Population in Central Florida," *Cureus*, vol. 14, no. 8, p. e27811, doi: 10.7759/cureus.27811.

[30] "Alzheimer's Dataset (4 class of Images)." https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images (accessed Jul. 14, 2023).

[31] T. M. Ghazal *et al.*, "Alzheimer Disease Detection Empowered with Transfer Learning," *Computers, Materials & Continua*, vol. 70, no. 3, pp. 5005–5019, 2022, doi: 10.32604/cmc.2022.020866.

[32] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68–71, Jun. 2018, doi: 10.1016/j.fcij.2017.12.001.