Biomedical Signal Processing and Control 96 (2024) 106550

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Ensemble-based multimodal medical imaging fusion for tumor segmentation

A. Karthik [a,*], Hatem S.A. Hamatta [b], Sridhar Patthi [c], C. Krubakaran [d], Abhaya Kumar Pradhan [e], Venubabu Rachapudi [f], Mohammed Shuaib [g], A. Rajaram [h]

[a] Institute of Aeronautical Engineering, Hyderabad, India
[b] Department of Applied Sciences, Aqaba University College, Al Balqa Applied University, Aqaba, Jordan
[c] Professor, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, India
[d] Dept. of AI&DS, St. Martin's Engineering College, Secunderabad, Telangana, India
[e] School of Technology, Woxsen University, Hyderabad, Telangana, India
[f] Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India
[g] Department of Computer Science, College of Engineering & Computer Science, Jazan University, Jazan, Saudi Arabia
[h] Department of Electronics and Communication Engineering, E.G.S Pillay Engineering College, 611002, Nagapattinam, India

ARTICLE INFO

Keywords:
Deep Learning
Multimodal Fusion
VGG-19
SqueezeNet
Dense Net
Segmentation
Ensemble Learning
Medical Imaging

ABSTRACT

The use of multimodal medical imaging is on the rise, both in academic and clinical settings. There was a meteoric growth in the use of multimodal imaging analysis (MIA) with the addition of ensemble learning techniques, which had particular advantages in the medical field. We provide an algorithmic framework that allows supervised MIA and Cross-Modality Fusing at the preprocessing phase algorithms for classification and decision-making levels, drawing inspiration from the current triumphs of deep learning approaches in medical imaging. We presented a method for picture segmentation that makes use of sophisticated convolutional neural networks to identify lesions produced by tumors in soft tissues. To do this, MRI tomography and PET scans are combined to provide multi-modal images. Networks trained with multimodal images outperform their single-modal counterparts. When it relates to tumor segmentation, fusing photos throughout the neural network (i.e., within the convolutional layer or totally connected layers) yields better results than photographs that merge the network's output. The proposed approach employs four pre-trained models, specifically VGG 19, ResNet 50, SqueezeNet, as well as DenseNet 121. Using a dataset of ISL images, the pre-trained models are fine-tuned. Subsequently, the ensemble learning technique is employed to combine the predictions generated by the three models. Here, ensemble is based on a weighted voting method. Impressive results were obtained with the proposed ensemble method: 98.1% accuracy, 97.5% F1 score, and 90.8% Kappa score. The ensemble method outperformed individual models and existing approaches for multimodal medical fusion and classification, with a Jaccard score of 93.8% and a recall of 98.2% demonstrate its effectiveness for multimodal medical fusion and classification.

1. Introduction

Most initial brain tumors are gliomas, and brain tumors are among the worst diseases ever. Brain tumor classification facilitates diagnosis by serving as a treatment guide as well as offering acquisition instruments for medical imaging that integrates different modalities for brain tumor categorization. Thus, prior research has used either 2D brain MRI image slicing or 3D brain images fused to accomplish this goal [1]. Decline in cognitive abilities and memory loss are symptoms of Alzheimer's disease, a neurological illness. Because they provide a broader view of the brain changes that define Alzheimer's disease, multimodal imaging techniques have been more popular for use in the diagnosis of the disease in recent years. This is because these approaches allow doctors to better monitor the progression of the illness over time.

* Corresponding author.
E-mail addresses: karthik011190@gmail.com (A. Karthik), Hatem@bau.edu.jo (H.S.A. Hamatta), sridharp35@gmail.com (S. Patthi), c.krubakaran@gmail.com (C. Krubakaran), abhaya08csc007@gmail.com (A. Kumar Pradhan), venubabu.r@gmail.com (V. Rachapudi), mohammed@gmail.com (M. Shuaib), drrajaram@egspec.org (A. Rajaram).

https://doi.org/10.1016/j.bspc.2024.106550
Received 2 April 2024; Received in revised form 18 May 2024; Accepted 7 June 2024

Because it integrates information from several types of imaging into a unified, more interpretable whole, medical image fusion is vital [2]. By supplementing medical pictures from many modalities, multi-modal image fusion methods help doctors make more accurate diagnoses. These methods improve the efficiency of medical condition analysis and result categorization [3]. With the advancement of medical image processing, picture fusion has become a practical option, automatically combining several pictures into one by extracting pertinent data. The identification and classification of brain tumors rely heavily on medical imaging methods like Magnetic Resonance Imaging, Computed Tomography, etc. Accurate disease diagnosis requires more than just one imaging approach [4]. Global Health Organization glioma classification system for 2021 states that glioma segmentation is a crucial foundation for genotype prediction and diagnosis. 3D multimodal magnetic resonance imaging of the brain is a useful diagnostic tool. Machine learning and deep learning in particular, have seen a surge in use for analyzing medical pictures within the last decade. Models pre-trained using large-scale datasets provide superior performance on a number of tasks [5]. This is all down to the creation of foundation models. In order to maximize yields and ensure healthy crop development, it is vital to identify illnesses in rice plants. The establishment of mitigation measures to provide large-scale food security and affordable rice crop protection may be aided by a real-time and precise plant disease detection system. Achieving site-specific use of agrochemicals might be made possible with a precise categorization of diseases of rice plants utilizing DL and computer vision. Using image research techniques efficiently allows for ongoing surveillance of plant health state and early identification of plant illnesses [6]. Recognizance, segmentation, as well as classification from RGB pictures are just a few of the computer vision tasks that typically use deep learning approaches. An assortment of sensors allows for the collection of industry-specific datasets, which in turn allow for the resolution of industry-specific problems. There is a wide range of modalities in the gathered datasets, suggesting that each picture has its own unique collection of channel number and pixel values. A complex technique is required to use deep learning algorithms on these multimodal data in order to get optimum results [7]. In order to better evaluate patients, guide treatment, treat them, or anticipate their results, multimodal medical image fusion efficiently integrates many imaging modalities. Because image fusion provides more crucial information, the accuracy of the combined picture from many medical imaging modalities greatly affects the prognosis for a disease. It is impossible to get comprehensive and accurate results from only one medical imaging modality [8]. When it comes to diagnosing medical issues, multimodal picture fusion using deep learning approaches has recently gained popularity. Unlike traditional approaches, deep learning techniques accomplish the fusion in the intermediate stages of deep neural networks. This allows for the implicit alignment of several visual modalities at the semantic level, bypassing the need for spatial alignment. As a result, many question the significance of spatial alignment during deep learning fusion [9]. The development of multimodal imaging methods for medicine has greatly aided advancements in clinical diagnosis and etiological analysis. The merging of multimodal medical pictures may provide a viable alternative to the inherent limitations of individual medical imaging modalities [10]. Improved accuracy as well as efficiency in clinical diagnoses can be achieved through multimodal medical image fusion, which entails combining medical images obtained by different sensors with the goal of improving image quality, reducing redundant information, and preserving specific features. Recent years have seen tremendous progress in picture fusion thanks to the advent of deep learning algorithms, which overcome the drawbacks of traditional approaches that need human intervention in the design of level of activity assessment as well as fusion rules [11]. For pixel-level medical picture fusion, a new sparse representation model called convolutional sparsity oriented morphological component evaluation is presented. Through the integration of multicomponent analysis as well as convolutional sparse representation, the CS-MCA model is able to accomplish both global and multicomponent SRs of the input pictures. Using pre-learned dictionaries, the CSMCA model in the current technique obtains the CSRs of the gradient and texture components [12]. Ultrasonography is a crucial imaging tool for evaluating breast lesions. The use of computer-aided diagnostic technology has greatly improved radiologists' ability to differentiate between benign and malignant tumors through automatically segmenting and detecting their features [13]. Numerous clinical contexts made extensive use of MMIF techniques. In order to aid in the development of diagnostic techniques, MMIF has the potential to provide a picture that contains anatomical and physiological information. Earlier, other models were suggested that were associated with MMIF. Prior approaches would need to have their functionality improved, albeit [14]. Because of its versatile nature, Multimodal sentiment evaluation is quickly becoming a popular tool. Effectively managing social media information with many modalities is challenging, as previous research has concentrated on SA of single methods, such texts or photographs. The majority of multimodal studies have failed to adequately address the complex interactions between the two modalities, leading to unsatisfactory results in sentiment classification [15].(See Tables 1 and 2).

The progressive and fatal brain disorder known as Alzheimer's Disease (AD) causes memory and cognition to deteriorate over time. However, deep-learning algorithms have shown potential in treating this neurodegenerative illness, which causes brain damage and mental degradation. The paper's main contribution is the creation of an algorithmic framework for multimodal medical image analysis (MIA) and classification that makes use of deep learning and ensemble learning methods. The inclusion of advanced convolutional neural networks, such as VGG 19, ResNet 50, SqueezeNet, and DenseNet 121, which have been optimized using ISL image data, is innovative and results in better performance. Applications of deep learning techniques include diagnostic decision assistance and pattern detection in medical images. Transfer learning, in which an established model is applied to a new task, could be very helpful when resources are few. The implementation of transferred learning has enabled the accurate diagnosis of AD. Data and MRI scans from Alzheimer's disease (AD) patients and healthy controls are processed using a mix of deep-transfer learning and bespoke models in this job. Clinical data characteristics are extracted using a two-layer fully connected network, whereas properties for input axial slices of magnetic resonance imaging (MRI) are obtained using SqueezeNet, ResNet 50, VGG 19, and DenseNet 121. During fivefold cross-validation, this results in a categorization accuracy of 99.65 % in both AD and NC.

The rest of this paper is organized as follows.In Section II, pertinent research using deep learning to segment brain tumors is reviewed. The research problem is outlined in Section III. Our suggested multimodal approaches and group Transfer Learning (TL) approach are described in Section IV. Extensive details, debate, and performance rating are provided in Section V. Section VI concludes with several recommendations and directions for future research.

2. Related work

This work explores the possibility of integrating MRI and PET data using Pareto optimal deep learning approaches. It does this by using pre-existing models from the Visual Geometry Group, including VGG11, VGG16, and VGG19 architectures [16]. When working with MRI and PET images, morphological operations may be carried out using Analyze 14.0. Next, the PET images are adjusted to the correct angle with the help of GIMP and then compared to the MRI scans. Improving the network's performance before image fusion is done through the addition of a transposed convolution component to the previously recovered feature maps. During that stage, features maps with the fusion weights that are necessary for fusion are constructed. The study's objective is to compare three VGG models' capacity to glean relevant features from PET and MRI images. Pareto optimization is used to optimize the model's hyperparameters. For this purpose, we use the Architectural Similarities Index

Table 1
Performance Analysis of Proposed vs. Existing Works.

	Ordinary gray image					Reconstructed gray scale image				
	Accuracy	F1 score	Kappa	Jaccard	Recall	Accuracy	F1 score	Kappa	Jaccard	Recall
SqueezeNet11	0.764	0.751	0.692	0.632	0.770	0.795	0.780	0.731	0.657	0.796
VGG19	0.752	0.738	0.683	0.618	0.759	0.878	0.885	0.845	0.792	0.891
ResNet 50	0.789	0.793	0.732	0.656	0.789	0.920	0.919	0.889	0.853	0.921
DenseNet 121	0.792	0.791	0.725	0.661	0.795	0.890	0.887	0.851	0.814	0.894
Ensemble	0.981	0.975	0.908	0.938	0.982	0.9880	0.964	0.931	0.97	0.98

Table 2
Augmentation Result.

Methods	**Augmentation**	**Accuracy**	**F1 score**	**Kappa**	**Jaccard**	**Recall**
SqueezeNet 1_1	No	0.816	0.817	0.738	0.687	0.817
	Yes	0.892	0.891	0.865	0.815	0.891
Vgg 19	No	0.741	0.731	0.618	0.596	0.739
	Yes	0.921	0.916	0.885	0.831	0.909
ResNet 50	No	0.759	0.759	0.667	0.651	0.768
	Yes	0.923	0.912	0.892	0.857	0.911
DenseNet 121	No	0.821	0.818	0.741	0.702	0.810
	Yes	0.911	0.929	0.891	0.851	0.923
Ensemble	No	0.921	0.918	0.947	0.902	0.91
	Yes	0.957	0.945	0.998	0.95	0.97

Technique in conjunction with E, SSIM, PSNR, MSE, as well as E to assess the models' operation on the ADNI dataset. The aforementioned fusion method makes use of an image fusion methodology that is based on a Siamese Neural Network and Entropy [17]. The fusion procedure is based on the entropy of a picture and the score of the SoftMax layer. In the end, the image is completed using Otsu Thresholding and Quick Fuzzy C Means Clusters Algorithms. Lastly, the segmented areas are used to extract a variety of attributes. A Logistic Regression classifier is used to do classification based on the characteristics that have been retrieved. The evaluation is conducted using a benchmark dataset that is accessible to the public. Results from experiments with several sets of medical pictures show that the suggested methods for fusing and classifying multi-modal images may hold their own against the current best practices described in the literature. Multimodal fusion with Principal Component Analysis is suggested in [18]. Once they get the fused output from multimodal fusion, they may utilize it to execute tumor classification and segmentation. That approach utilizes a CNN for classification with Otsu thresholding for tumor segmentation. All of the analysis for the brain tumor was done using the MATLAB App designer. Both numerically and qualitatively, the approach beats competing techniques, according to the experimental data. A new approach to automated brain tumor detection and classification utilizing multi-modal deep neural networks is described in [19]. At the outset, the suggested AMDL-BTDC model employs the bilateral filtering method to preprocess images. After that, authors use two pre-trained deep learning models, SqueezeNet and EfficientNet, to create feature vectors. To find the best hyperparameter values for the DL models, the Slime Mold Algorithms is used. Once the features have been fused, the last step in BT classification is to employ an auto encoder model. Thorough testing on a standard medical imaging datasets confirmed that the proposed model outperformed competing methods across a variety of metrics. For even more precise breast cancer categorization, see [20] for instructions on how to combine pathology pictures with structured data retrieved from clinical EMR. The authors of the research provide a new and improved fusion network for classifying benign and aggressive breast cancers using multimodal data. A more comprehensive multilevel feature description of the ill picture might be extracted from several convolutional layers, according to researchers' suggested technique, which would enhance its integration with unstructured EMR data. Instead of decreasing the dimensionality of the highly dimensional picture data to low-dimensional prior data fusion, they employ the denoising auto encoder to raise the size of the low-dimensional organized EMR data in order to decrease data loss for each modality. Furthermore, denoising autoencoder essentially broadens their method to accurately forecast structured EMR data that is partly missing. With an average rate of classification of 92.9 %, the suggested technique outperforms the state-of-the-art method, according to the testing results. With the goal of enhancing interpretability and guiding architecture selection, the authors of [21] provide a new technique for multimodal artificial neural networks that is based on gradients of feature significance. They build a validation system to set performance baselines; it mimics test scenarios and compares their feature importance approach's performance to ground truth; that allows us to show their technique. In the architecture's post-fusion phase, the gradient based approach for feature significance is used to get feature importance values for deep features. After that, we may calculate the significance for every multimodal input by summing the importances of every mode. The 58,830 medMNIST abdomen CT scans and generated clinical data are used to train their sample program. With that study, they have made a significant step toward making deep learning approaches more interpretable by estimating the relevance of features in multimodal machine learning models. In [22], the authors present a new CAD system that uses multimodal magnetic resonance imaging and a deep-learning architecture to identify thyroid nodules that may be malignant. In particular, their system is designed to fuse two MRI modalities—the apparent diffusion coefficient map and the diffusion weighted image—through the use of a multi-input CNN. Their system's primary contribution is multi-faceted. Three things stand out about that system: (1) it is the first of its kind to use CNN for classification purposes in thyroid DWI and ADC images; (2) it improves the likelihood of finding deep texture things in thyroid tumors by allowing for separate convolutional procedures for the DWI and ADC images; (3) it opens the door to integration via other imaging modalities as well as additional MRI scans by allowing for the addition of additional channels to each input. They compared their method to other fusion techniques and ML framework that use characteristics that are hand-crafted. Their algorithm was superior than the others, with a diagnostic success rate of 0.88, a precision of 0.82, with a recall of 0.82. Optimizing Sea Horses with Deep Learning-Based Improvements An technique reported in the literature [23] is acronymed as ESHODL-MFRPDC, which stands for Multimodal fusion for the Diagnosis and Categorization of Rice Plant Diseases. Utilizing a DL-based fusion technique with a hyperparameter tuning approach, the suggested

method enhances disease diagnosis in rice plants. During the pre-processing stage, the ESHODL-MFRPDC technique utilized Bilateral Filtering to eliminate noise and enhance contrast. Furthermore, the impacted regions in the leaf image were identified using Mayfly Optimization (MFO) segments using Multi-Level Thresholding. Three DL models—Xception, residual network (ResNet50), as well as NASNet—were used in the feature extraction procedure. The hyperparameters of the Quasi-Recurrent neural networks used to detect diseases in rice plants were established using the ESHO method. They checked the ESHODL-MFRPDC method's accuracy using the UCI database's dataset on rice leaf diseases. In a thorough comparison investigation, the proposed technique performed better than others. One workable solution to the problem mentioned in [24] is to use a data fusion method. Data fusion searches for the optimal solution by using and integrating all of the sensor data that is available. The research looks at three different forms of fusing in deep learning models—early, medium, and late—to categorize pictures of big rubbish. The model development and assessment methods both make use of multimodal datasets. The data set contains images of big garbage cans taken using thermal photography, terahertz, hyper spectra near infrared, and RGB cameras. The results show that compared to a single-sensor technique, multimodal sensor fusion enhances classification accuracy when applied to the provided dataset. The article [25] suggests a new way to fuse multimodal medical pictures using ELM and CNN. More and more people are turning to CNN in image processing since it is a classic example of deep learning. Nevertheless, CNN often encounters a number of limitations, including significant computational expenses and extensive human involvement. Therefore, by combining ELM with the conventional CNN model, the convolutional extreme machine learning model is built. To extract along with capture the characteristics of the source photos from multiple perspectives, CELM is an essential tool. It is possible to get the final fused image by integrating the key features. In addition to outperforming state-of-the-art methods on objective metrics and subjective visual performance, testing findings show that the suggested method enhances lesion identification and localization accuracy. Before reviewing the features and roles of various fusion modalities and outlining their interrelationships, the article [26] provides a comprehensive explanation of the multimodal medical picture fusion challenge. In order to provide a thorough overview of the latest advancements in medical image fusion from a deep learning standpoint, it then examines the theories and improvement methods linked to deep learning in that area. Among these advancements are unified models, methods for multimodal feature extraction using convolutional approaches, methods for signal processing based on convolutional sparse representation as well as stacked autoencoders, and methods based on adversarial learning. Finally, the article provides a concise overview of the strategies used to improve multimodal medical picture fusion, drawing attention to the serious problems and obstacles that deep learning approaches have in that field. Pyramid decomposition based on deep learning is used by the authors in [27]. As a technology, deep learning is currently somewhat demanding. Visual tasks such as object recognition, picture segmentation, and image restoration all make use of deep learning. That research presents a CNN based approach to medical picture fusion. Using a Siamese network, they directly map source pictures to a weight map that include the integrated pixel activities information. The main advantage of that method is that it may bypass the problem of artificial design by combining the assessment of activity levels with weight assignment via the application of network learning. Adaptive fusion selection modes and multi-scale processing are two well-known picture fusion technologies that provide aesthetically pleasing results. According to the findings of the experiments, the proposed approach has a high chance of producing satisfactory outcomes when it comes to both visual and subjective quality metrics. A multimodal fusion architecture is developed in [28] to distinguish between benign and malignant tumors using cropped B-mode and SE-mode ultrasound images of the lesion. The MFF is comprised of a decision-making network and an integrated feature analysis network. In contrast to previous recently published fusion strategies, the proposed MFF strategy has the ability to learn additional information from CNNs trained utilizing B-mode with SE-mode US pictures concurrently. Image classification is handled by DN with the use of CNN feature ensemble trained using the multimodal EmbraceNet model.. Radiologists' ability to correctly classify breast cancer in US pictures might be improved by using the suggested strategy. In [29], a new fusion model that uses deep learning and optimal thresholding is proposed as a potential answer. Using fusion principles similar to those of the shearlet transform, an improved monarch butterfly optimization finds the best threshold. The fusion rule is the primary determinant of the fusion process efficiency, and enhancing the fusion rule's performance is possible via optimization. Then, the deep learning method's extraction component was used to combine the subbands of high and low frequencies. A CNN was used to perform the fusion procedure. Research was conducted using MRI and computed tomography scans. After achieving the fusion results, it was shown that the suggested model provides effective performance with decreased error values and enhanced correlation values. A trio of DNNs is used by the model proposed in [30–32]. To identify the most meaningful parts of images and text from an emotional perspective, two separate neural networks are suggested. Additional discriminative characteristics are collected to ensure precise emotion categorization. Then, with the help of a self-attention strategy, they provide a multichannel combination fusion method that uses the inherent relationship between visual and textual features to gather emotionally rich data for joint sentiment categorization. Lastly, a decision fusion approach is used to combine the outputs of the three classifiers, making the suggested model more robust and generalizable. To build a strong and explainable visual-textual sentiment categorization model, they use the Local Interpretable System-agnostic Explaining system. Their MMF model beats both state-of-the-art methodology and single-model approaches, according to model evaluation criteria.

3. Problem statement

The research community has intensively researched multimodality data fusion using ML approaches. At different points, multiple modalities may be combined in a number of ways. You may use ML approaches at any step of the fusion process, depending on the analytic goals. Early fusion is the simplest method as it combines normalized data from several modalities into one classification pool. Modeling the interactions between the modalities using extracting complementary information has been done using early fusion approaches such stepwise logistic regression analysis, Gaussian processes, and SVM-based kernels. Nevertheless, when faced with heterogeneous data, early fusion falls short because it overemphasizes one modality with many characteristics and ignores the others. A distinct approach, intermediate fusion chooses characteristics separately for each modality. The research community has intensively researched multimodality data fusion using ML approaches. At different points, multiple modalities may be combined in a number of ways. You may use ML approaches at any step of the fusion process, depending on the analytic goals. Early fusion is the simplest method as it combines normalized data from several modalities into one classification pool. Modeling the interactions among the modalities and extracting complementary information has been done using early fusion approaches such stepwise logistic regression, Gaussian processes, and SVM-based kernels. Nevertheless, when faced with heterogeneous data, early fusion falls short because it overemphasizes one modality with many characteristics and ignores the others. An alternative approach, known as intermediate fusion, involves picking features separately for each modality, building a kernel-based matrix for every modality, and then merging them into a single fused matrix. Random Forest may also be used to create similarity matrices, which are then employed in the process of fusion. But these methods are sensitive to how various modalities are weighted, and if you don't give it some thought, you can end

up with subpar results.

4. Proposed work

Our innovative approach to fusing multimodal MRI as well as PET data takes into account key aspects that have been neglected in earlier research. To extract additional data for AD diagnosis while retaining the local structural information inherent to every modality, our suggested technique analyzes the inter with intra-modality relationships among PET and MRI. We also include a new relation called the Same-Subject Modalities-Interaction Matrices (SSIMI) to provide additional data that increases learning accuracy by enriching the training set. In order to make the most of the SSIMI communication, we suggest using an ensemble transfer learning framework to determine which features are best capable of capturing the highest amount of variance in a specific area using the data obtained from MRI and PET scans. At last, in order to reduce the computational expenses of the structure and identify useful biomarkers, we perform choosing features on the fused set. When using multimodality fusion to diagnose and track diseases, preprocessing images from imaging modalities like PET and MRI is essential. Segmentation of certain areas is made possible by registering PET pictures to their matching MRI scans. Disease categorization is made easier with the use of these segmented traits. Targeted localization is made possible by coordinating PET scans with MRI's anatomical data. The SSIMI interaction should be considered with intra- and inter-modality fusion. SSIMI delves at the ways in which two methods might interact with one other to capture the same area of a topic [33]. Examining this relationship may provide relevant and supplementary data that might enhance the precision of illness categorization. Our strategy for dealing with this issue comprises:

- To maintain the diverse structure of every information source, process as well as normalize each modality separately using Freesurfer.
- For the same area and topic, build a new set using the relationships between the various modalities.
- Merge all of the sets to boost learning efficiency.
- Get the most out of your classification work by training numerous classifiers.

In Fig. 1 we can see the structure of the approach that we have suggested. To isolate the most important aspects of PET and MRI data, we use attention-based transfer learning models. This is achieved by replacing the conventional FC layer with a set of networks that can extract spatial and semantic information at a high level: VGG-16, DenseNet, SqueezeNet, as well as ResNet. In the end, the characteristics that have been learnt are combined and then sent to the SoftMax classifier to diagnose diseases. Here we will provide you the rundown on the recommended technique [34].

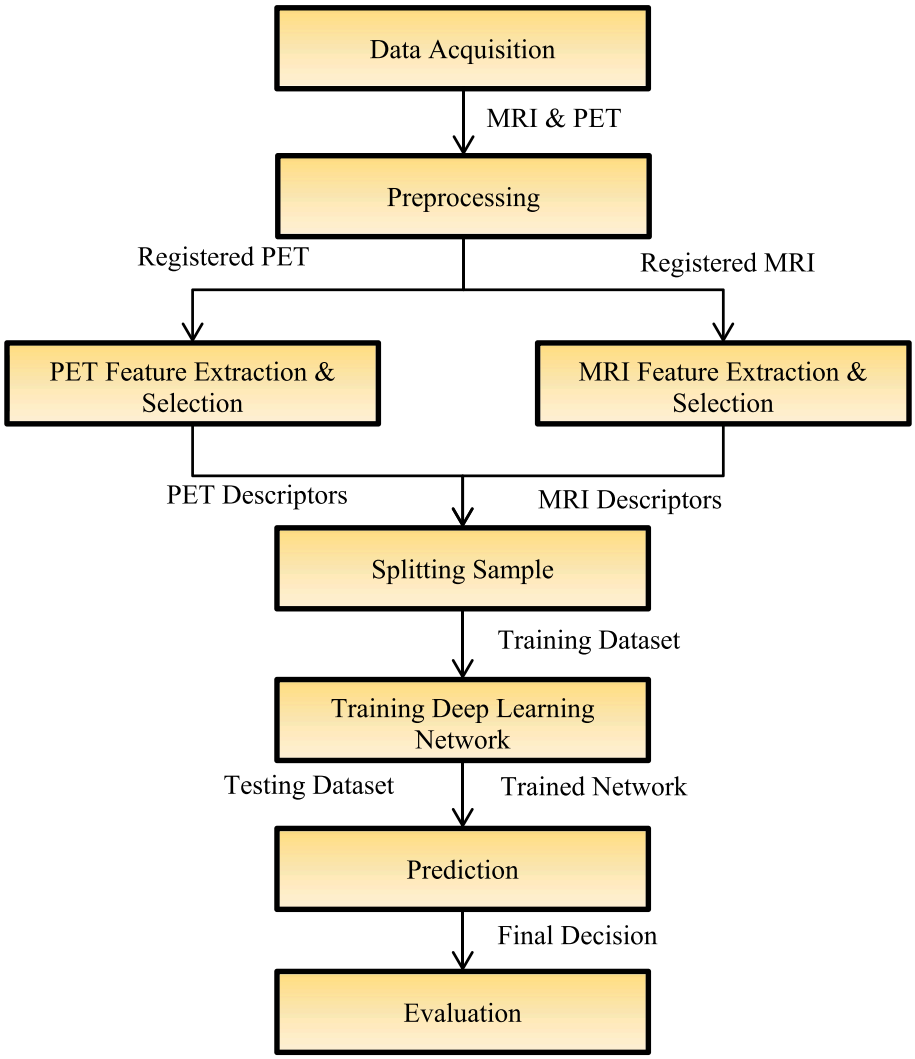

Fig. 1. Proposed System Architecture.

4.1. MRI data acquisition

The degree of brain shrinkage is a defining feature of the various AD phases. The use of volumetric MRI imaging methods allows for the detection of both global and regional changes in brain volume. In MRI, the proportion of bound to unbound molecules of water is used to magnify the variations in tissue matter. It is possible to quantify regional volumes because these ratios vary across various brain tissue types. T1 weighted MRI, which measures how long it takes for the vector of net magnetization to go back to its original shape after being spun by an RF pulse, was used in this endeavor. The T1 durations of tissues are shorter when the ratio of bound to unbound fluid is greater. Because the brain's tissue has a higher concentration of bound water than the cerebrospinal fluid (CSF) around it, a T1 picture will highlight this area. The primary raw MRI data set includes brain images acquired in successive slices that are aligned perpendicular to the transverse and coronal planes. Images were standardized to correspond to topology using an elastic warping technique, which preserves the morphological properties of an individual's brain [15]. Ventricular blood vessels, cerebrospinal fluid, and white and gray matter are separated from the original raw picture. After that, the image's visible brain regions may be identified using automated region interest (ROI) analysis, which enables the removal of skull-related areas and the computation of volumes for specified regions. Every side of the brain has fourteen of these regions identified, for a total of twenty-eight characteristics per picture. Prior to being used for classifier training, the volumes associated with these 28 regions were normalized relative to total intracranial volume [35].

The data used to train MudNet was sourced from the Alzheimer's disease, Neuroimaging Initiative (https://adni.loni. usc.edu/). The ADNI has collected a plethora of neuroimaging data from 1,821 subjects, including those with AD, moderate cognitive impairment (MCI), and controls with cognitively normal aging. This data includes results from MRI, PET, clinical assessments, and neurological tests. We used 559 individuals' baseline measurements as cross-sectional data to train MudNet. Clinical data, which comprised demographic information and outcomes from neurological assessments (ADAS-11, ADAS-13, ADASQ4, RAVLT, MMSE), in addition to structural MRI, were used. The data from all ADNI studies (namely ADNI1/GO/2/3) were amalgamated.

Prepare MRI scans: There is an enormous feature space that contains all characteristics that might be useful for forecasting pMCI conversion. Considering that MRI scans have dimensions that are comparable to 256 × 256 × 166, this means that each scan has 10, 878, 976 points of information. Hence, more data preparation is needed to simplify the data and enhance the extraction of pertinent visual characteristics of the brain as it becomes increasingly degraded in order to apply deep learning to the issue. 1) Signing up: Utilizing the MNI-152 templates space, all of the abovementioned methodologies record their MRI brain pictures, as illustrated in Fig. 1. The Neurological Centre of Montréal developed the T1-weighted MNI-152 space by combining the MNI-305 space with 152 normal MRI brain photos. The initial Talairach atlas has been superseded as the standard by the linearly documented MNI-152 template, according to the International Organization of Brain Mapping.

Research that is relevant registers the brain in the MNI-152 space using ADNI MRI data. A voxel depicting $1 \times 1 \times 1\ mm^3$ is provided by the registered pictures, which consist of 197 × 233 × 189 columns, rows, and slices. Because it enables the alignment of various brain regions, image registration is crucial for medical picture comparison. The pathological distinctions among sMCI and pMCI are better retained when paired non-linear techniques, such as affine with deformable transformations, are utilized. As a result, the spatial differences may be more precisely calculated using convolutional neural networks, which are capable of doing more exact spatial comparisons.

Brain-stripping: Magnetic resonance imaging (MRI) pictures depict a complex feature space. We may decrease the number of characteristics required to forecast pMCI conversion from sMCI by removing extraneous features from this region, such the eyes and skull. Skull-stripping, the surgical removal of the brain and skull, is a frequent pretreatment step in all of the approaches used to investigate current methodologies. This is accomplished by using grey matter extraction in both the convolutional and deep residual approaches. Therefore, the classification issue can be more simply partitioned via brain extraction, as only the most significant characteristics remain in this area. Optimizing weights and propagating errors may then zero down on the spatial variations within these important characteristics, shortening the training period and boosting the model's predictive power.

4.2. PET image acquisition

Different from magnetic resonance imaging (MRI), nuclear imaging methods like Positron Emission Tomography (PET) scan the body for gamma rays given off via a radioactive tracer which molecules with biological activity inject into it. Because fluorodeoxyglucose (FDG) is the chemical most often utilized for this usage, this imaging technique is also known as FDG-PET. In this investigation, FDG-PET pictures were taken around 30 min after an FDG injections and continued for another half an hour. The method from the Alzheimer's Neuroimaging Initiative (ADNI) was used to do the imaging. Stereotactic surfaces projection, a method that has been shown to be very useful in the identification of Alzheimer's disease, was used to evaluate these pictures. This research was carried out using Neurostat SSP, a computer program library, and the features used to train the classification algorithms were relative glucose rates for 43 designated regions of interest.

In the first step, known as Intensity Redistribution, the input pictures' pixel intensity values are normalized to a fixed value. In this stage, a normalized intensity is formed by aggregating the intensities of those neighbors that are most comparable. "Dynamic Intensity Specific Variance" is suggested as a means of standardizing intensities in this study.

It is discovered that the input image's pixel intensities are dispersed. Magnetic resonance imaging (MRI) scans of the brain show three distinct types of tissue at varying intensities: the cerebrospinal fluid, gray matter, and white matter.

$$In = \{In_1, In_2, In_3\} \tag{1}$$

The dissimilarity between various levels of intensity is expressed as,

$$dif_{1,2} = In_1 - In_2 \tag{2}$$

$$dif_{2,3} = In_2 - In_3 \tag{3}$$

$$dif_{1,3} = In_1 - In_3 \tag{4}$$

An expression that expresses the intensity-specific variability weights to a limited value is,

$$W_{ISV} = 2^{3\times s\left(dif_{1,2}\right)+s\left(dif_{2,3}\right)+s(dif_{1,3})} \tag{5}$$

$$s(z) = \begin{cases} 0, z > 0 \\ 1, z < 0 \\ 2, z = 0 \end{cases} \tag{6}$$

The weight proportional to changes in intensity difference may also be calculated in a similar way,

$$W_{IDV} = 2^{2-s\left(dif_{1,2}\right)-s\left(dif_{2,3}\right)-s(dif_{1,3})} \tag{7}$$

In order to calculate the ISV metric, one may use the following equations, which are expressed as,

$$ISV(In) = W_{ISV} \times W_{IDV} \tag{8}$$

This ISV measure is image-independent and adapts to new inputs. Next, we do Image Quantization to reduce the image's intensity values to a minimum. This is the part where we check the connection between the two pixels. It applies quantization based on the dissimilarity distance

between pixels. It is possible to express the function of relationships between the nodes as,

$$\mathscr{R} = \{\mathscr{R}_i | \forall i \in \mathscr{L}\} \tag{9}$$

Where, $\mathscr{L}$ indicated the grid where the pixels are located. An expression for the separation of the two pixels is,

$$D = \sqrt{(q_2(i,j) - q_1(i,j))^2 + (p_2(i,j) - p_1(i,j))^2} \tag{10}$$

Where $(p_1(i,j), q_1(i,j))$ and $(p_2(i,j), q_2(i,j))$ serve as the two pixels' coordinates, correspondingly. A new intensity is generated by substituting the center values for the intensity values. The intensities of pixel prior to and following normalization are shown in Fig. 2 (a) and (b).(See Figs. 3–7).

4.3. Multimodal techniques

In this case, the suggested architecture uses four convolutional neural networks (CNNs) to extract the main characteristics of the fusion image: VGG 19, ResNet 50,SqueezeNet, as well as DenseNet 121.

4.3.1. VGG

The structure of the visual geometry group is highly organized. The input image's size is reduced as the network is trained deeper; nevertheless, the explanation for this phenomenon is the continuous increase in the total number of convolution kernels. To improve the network's depth and breadth, a large number of 3 × 3 kernels with convolution are used to substitute the macrokernels. Therefore, the recognition capacity of the classification job and the richness of the recovered features are both enhanced by increasing the amount of activation functions.

4.3.2. SqueezeNet

SqueezeNet uses a large number of 1 × 1 kernels instead of a 3 × 3 convolution kernels to speed up CNN training and decrease computing cost, similar to AlexNet's findings on the ImageNet data set. The network's small size and great efficiency make it ideal for large-scale datasets.

4.3.3. ResNet

In contrast to VGG, ResNet addresses the deep network degradation issue by establishing connections between successive layers based on residuals of feature mapping. By addressing ill-posed challenges, researchers may train deeper neural networks to better represent tasks.

4.3.4. DenseNet

Based on ResNet's idea, DenseNet achieves dense skip connections by connecting one layer to all following levels by skipping connections. As the architecture is refined further, DenseNet's internal representation diverges greatly from ResNets.

Starting with the supplied feature map $X_{\text{in}} \in \mathbb{R}^{H\times W\times C}$ proceeds to a 1 × 1 convolutional layer that for the purpose of extracting local features and adjusting the dimensionality to align with the subsequent layer. This layer's output is $X_1 \in \mathbb{R}^{H\times W\times C'}$, where the original image's resolution (H, W), the total number of initial features (C), and the amount of convoluted dimensions (C^') are all variables.

Following that, we execute a patch embedding procedure that involves altering the picture and compressing the image patches. A series of N flattened 2D patches is created using the reshaped feature map X_1 X_p^i (Equation (11):

$$X_p^i = P \times P \times C, i \in \{1, 2, \cdots, N\} \tag{11}$$

N = H × W/P2 generates the total of all image patches, assuming that each picture patch has a resolution of (P,P). The space of embedding having D dimensions is created by Xp^i and an adaptive linear projections for the MLP layer, as seen in Eq. (12).

$$\mathrm{X}_2 = \left[\mathrm{X}_\mathrm{P}^1\mathrm{E}; \mathrm{X}_\mathrm{P}^2\mathrm{E}; \cdots; \mathrm{X}_\mathrm{P}^\mathrm{N}\mathrm{E}\right] + \mathrm{E}_{\mathrm{pos}}(2) \tag{12}$$

where X2 is the series of encoded images.

Equations (13), 14, the third step, involves transferring the processed information from sequence X2 into the MLP layer.

$$\mathrm{X2}' = \mathrm{Dropout}\left(\mathrm{Gelu}(\mathrm{FC}(\mathrm{X2}))\right) \tag{13}$$

$$\mathrm{X3} = \mathrm{Dropout}\left(\mathrm{FC}(\mathrm{X}')\right) \tag{14}$$

where the activation functions Gelu and Dropout are used to enhance training accuracy and avoid network overfitting. A fully connected layer, FC reduces the two-dimensional feature map's convolution output to a one-dimensional vector.

Once the MLP level is finished, the output is reorganized to match the original dimensions of the input picture $X_{\text{out}} \in \mathbb{R}^{H\times W\times C}$ (Eq. (15), as well as a classifier makes a prediction about the glaucoma group.

$$\mathrm{Xout} = \mathrm{rearrange}\ (X3, (hw)(p1p2c) \rightarrow c(hp1)(wp2)) \tag{15}$$

For automated classification, a decision-fusion strategy based on ensembles of classifiers was used. To create an ensemble-based system, it is necessary to combine a number of various classifiers. In most cases, the training parameters used to train each classifier are varied. If variety is enough, each classifier will make a unique mistake, which, when strategically combined, may lower the overall error. Many methods exist for achieving this variety, such as using various subset of the training information acquired by resampling, varying the settings of a classifier approach, employing many classifiers together, or utilizing distinct subsets of the characteristics included within the provided dataset.

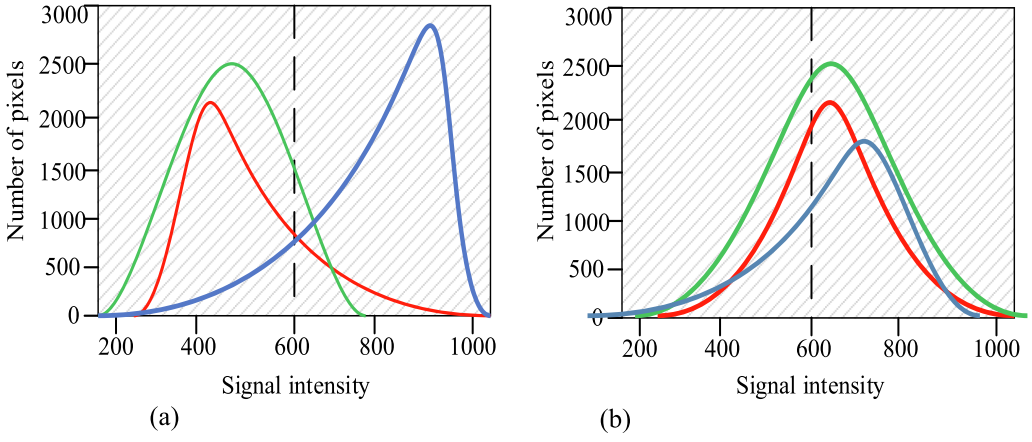

Fig. 2. (a) the intensity of the pixels before to normalizing, and the intensity of the pixels subsequent to normalization.

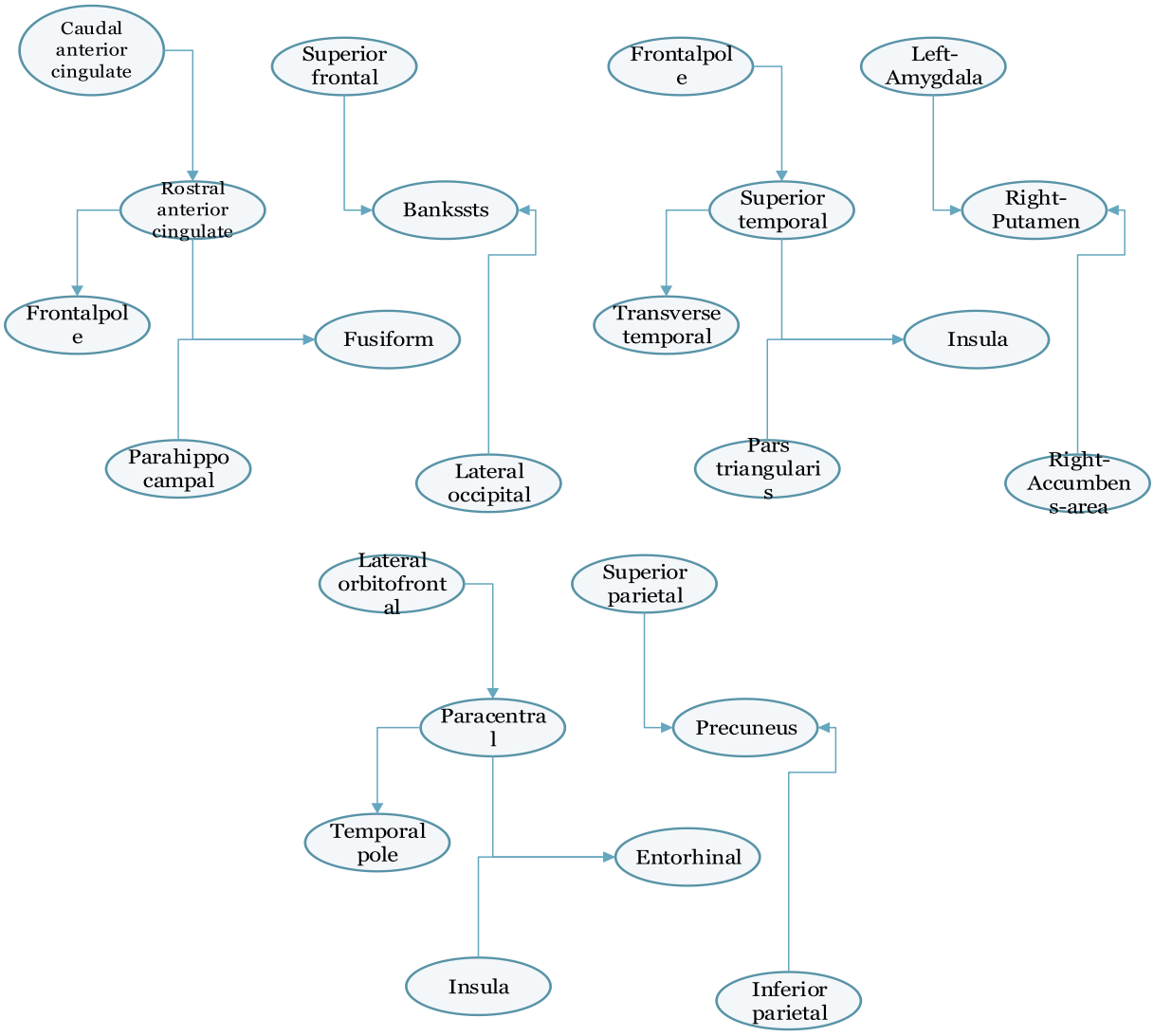

Fig. 3. Features Extraction using Ensemble Transfer Learning.

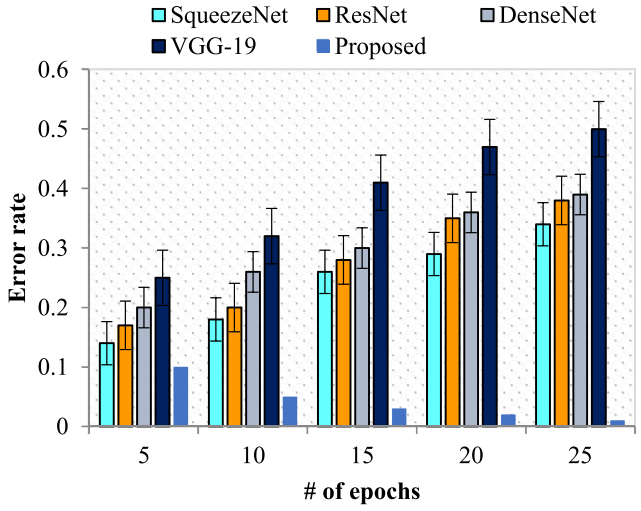

Fig. 4. Error Rate vs. No of Epochs.

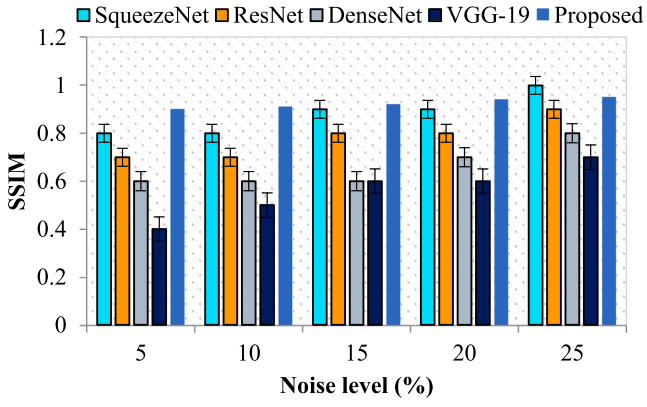

Fig. 5. SSIM vs. Noise Level.

Random subspace analysis is the name given to the second one. To get a final conclusion, an ensemble based approach combines the results from individual classifiers, which is similar to a decision fusion strategy. Compared to a system based on a single classifier, the objective is to achieve better generalization performance. On the other hand, data fusion applications, which mix data from several sources, are a perfect

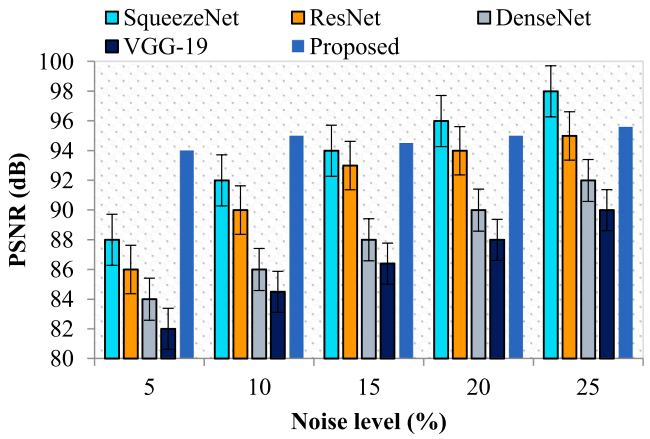

Fig. 6. PSNR vs. Noise Level.

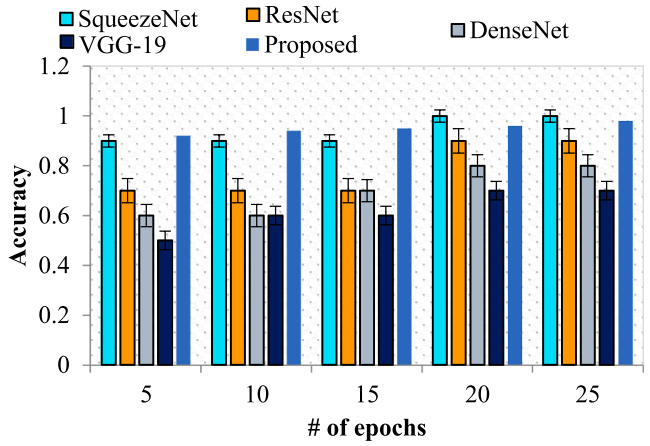

Fig. 7. Accuracy vs. Epochs.

fit for this approach's structure. When using data fusion applications, the aim is to make better decisions overall than when using only one source of data. The ensemble method is easy to implement in a data fusion setting: just train a different classifier on every set of data that originates from a different source, and then mix them using a suitable combination algorithm. For the sake of this research, we will be using an ensemble of classifiers for both data fusion and enhancing accuracy over a single classification algorithm, thereby combining the two applications of ensemble systems. Basically, we get three "experts" (one for Alzheimer's Disease data, one for MRI data, and one for PET data) by training a group of classifiers for every kind of data. Next, we merge these classifier ensembles to accomplish MRI, and PET data fusion by decision fusion. In order to better diagnose Alzheimer's disease, we want to determine whether these various methods provide further information. This overarching method is shown in Fig. 2.

Several combination rules are available for use once each of the classifiers have been produced. Among them, the majority of voting rule and the sum rule are perhaps the most utilized. We may say that the iˆth classifier's choice is $d_{i,j} \in \{0,1\} Ei = 1, \cdots, L$ and $j = 1, \cdots, c$, such that c = number of classes and L = number of classifiers. If i^{th} d_ij = 1 if class j is selected by the classifier, and zero otherwise. To implement majority voting, we first determine the general approval of S_j for class J by tallying up all of the votes, and then we choose the ensemble choice based on that.

$$S_j(\mathbf{x}) = \sum_{i=1}^{L} d_{ij}(\mathbf{x}), d_{ij}(\mathbf{x}) \in \{0,1\} \tag{16}$$

then declare victory to the class that garners the most votes. Another option is to utilize the sum rule to combine classifiers. This is effective when classifiers can offer continuous outputs for every class, signifying the support the class receives. After that, we total jointly every one of the class J outputs to determine the general backing for class J. Next, the class with the most total support will be chosen.

$$S_J(\mathbf{x}) = \sum_{i=1}^{L} d_{ij}(\mathbf{x}), d_{i,J}(\mathbf{x}) \in [0,1] \tag{17}$$

1) Rectified Linear Unit (RELU)

The goal of constructing convolutional neural networks out of the spatially structured perceptrons is to mimic brain function. Brain signals, which allow for thought and action, are the result of a complicated interaction between neuronal activations. Since the Rectified Linear Unit prevents the vanishing gradient issue that plagues many deep-layered network designs, it is the activation feature of choice for many CNN models. Activation functions' differentiability is crucial for back-propagation of error, which involves adjusting the weights to an optimum value. Because of its gradient, the sigmoid function is problematic when used; its partial derivatives of the error with respect to the weights determine the update to the current weights. As a result of layer-by-layer reductions, the model may stop updating its weights altogether. By maintaining a greater and more consistent gradient than the maximum sigmoid gradient, RELU, as a ramp function, mitigates the vanishing gradient issue. Quicker convergence is another feature of RELU (22). All of the proposed models in this technique (ELU) employ RELU or a variation thereof, which is not surprising.

2) Batch Normalisation

The majority of the aforementioned techniques use batch normalization on the output of the convolutional layers. The all convolutional technique, which had the lowest accuracy, didn't apply batch normalisation, which might explain why it wasn't as effective as other methods.

Many methods for ensemble learning have been created and studied within the domain of semisupervised learning. This approach aims to strengthen machine learning by merging many weak learners into one stronger one. Regarding instance-based ensemble transfer learning, however, very few proposals have been made for such a harmonic combination of several feature representations. Finding a way to successfully merge several weak student populations into a stronger one remains a significant issue for researchers in our field. To achieve this, we provide a new weighting method that optimizes the integration efficiency of the separate models on various feature representations. This scheme combines mutual information with weak learners to create a stronger learner. One of the most crucial aspects of ensemble learning for optimizing the model is learning how to weight the parameters. The mean achievement of every student in a conventional ensemble may be used to create the final ensemble learner. The weighting mechanism we suggest in our study may be expressed as (18) below, where y is the anticipated label vector from the ensemble learner.

$$y = \sum_{i=1}^{m} \exp\left(w_i^*\right) \phi_i[F_i(x)] \tag{18}$$

In the method above, w_i^* represents the weighted vector that has been normalized and which meets the conditions given by (19).

$$w_i^* = \frac{w_i}{\sum_{i=1}^{m} w_i} \tag{19}$$

Assuming a vector representing a predictability ^{y_i} of the case x_i, what is ultimately expected to be the instance's class label $Y_{T_i}^* = \arg\max_k \left(y_i^{(k)}\right)$. As there are n the assignment requires categorizing

various face expressions, the spectrum of k is $\{1,2,\cdots,n\}$. Theoretically, this ensemble learning method is comparable to Yang's weighted clustering ensemble approaches. Additionally, each MI value is bigger than 0 since all of them are computed in (1) w_i^* also exceeds zero. We next exhibit each weak leaner using an exponentially weighted strategy, as opposed to a linear weighting one. Then, the worth of exp $\left(w_i^*\right)$ not equal to 1. At some point, the projected outcomes taught by a stronger weak learner take on more significance.

5. Results & discussion

5.1. Experiment and discussion

Our study's experimental setup is described in this section. Afterwards, our architecture's efficacy is shown via four trials. In the end, the findings are presented and examined thoroughly. The proposed approach employs four pre-trained models, specifically VGG 19, ResNet 50,SqueezeNet, as well as DenseNet 121 are computed and compared with the Ensemble Transfer Learning architecture. Using a dataset of ISL images, the pre-trained models are fine-tuned. Subsequently, the ensemble learning technique is employed to combine the predictions generated by the three models. Here, ensemble is based on a weighted voting method. The performance of the proposed framework has been quantitatively analyzed on multiple evaluation parameters to demonstrate its effectiveness for multimodal medical fusion and classification.

5.2. Experimental setup

Our dedicated server, powered by a NVIDIA graphics card with a GeForce RTX 2060Ti GPU and equipped with 16 GB of RAM, runs every single test. The experimental framework is PyTorch, the programming language is Python, and the compiler is PyCharm.

The parameters of the suggested architecture are updated using the adaptive momentum estimating (Adam) optimizer in this study. The rate at which learning occurs is set to 0.0001, and the loss function is tuned to CrossEntropy Loss. With a value of 8, we've configured the batch size with epochs.

We use the accuracy, Kappa score, Jaccard score, recall, as well as F1 score to assess the efficacy of the given methodologies. The accuracy of a sample is defined as the percentage of the whole sample that is the accurate number. The percentage of positive samples that were actually anticipated to be positive is called recall. You may get the F1 score by dividing recall by accuracy. The Jaccard score is used to measure how similar or diverse a sample is. We use the Kappa score, which quantifies the degree to which actual and predicted classification results are consistent, to evaluate the performance of multi classification systems.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{20}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{21}$$

$$\text{Jaccard Score} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{22}$$

$$\text{F1 Score} = \frac{2 \bullet \text{precisionerecall}}{\text{precision} + \text{recall}} \tag{23}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{24}$$

$$P_e = \frac{(\text{TP} + \text{FN})(\text{TP} + \text{FP}) + (\text{TN} + \text{FN})(\text{TN} + \text{FP})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})^2} \tag{25}$$

$$\text{Kappa Score} = \frac{\text{Accuracy} - \text{P}_\text{e}}{1 - \text{P}_\text{e}} \tag{26}$$

the number of images that the categorization algorithm accurately identified, denoted as TP: FN stands for "false negative," which indicates how many pictures the algorithm mislabeled as belonging to certain categories; If TN is true negative, the classification technique correctly uses these categories to store non-category photographs; however, when FP is false-positive, it uses these categories incorrectly.

The suggested approach should not only be evaluated using traintest-split as well as cross-validation, but also on unseen data, especially data collected under diverse situations and with different procedures, to ensure it can generalize. We outperformed state-of-the-art approaches in terms of accuracy, precision, recall, along with F1-score, which bodes well for the future. Nevertheless, there are a few restrictions on our architecture. To start, we simply used two modalities—MRI and PET—to improve learning performance and classification accuracy. Additionally, PET segmentation alone, rather than relying on MRI segmentation, may improve classification accuracy. In addition, each data source should be treated independently since merging MRI with PET sets will affect the intra-relations. To get better classification results in the end, we used hyper-parameter tuning techniques during training and manually picked the parameters based on validation methods. It is also worth mentioning that the dataset utilized in this research was somewhat unbalanced and tiny in size. Despite using accuracy, specificity, recall, F1-score, as well as AUC to assess the learning models' efficacy, more effort is necessary to address the data's small size and imbalance. By using hyper-parameter tuning approaches, including the attention-based transfer learning technique, during training, we were able to enhance the final classification results, which were achieved by manual parameter selection for validation methods. It is also worth mentioning that the dataset utilized in this research was somewhat unbalanced and tiny in size. Despite using accuracy, specificity, recall, F1-score, as well as AUC to assess the learning models' efficacy, more effort is necessary to address the data's small size and imbalance.

6. Conclusion

In conclusion, our study demonstrates the efficacy of combining MRI and PET scans with attention-based transfer learning frameworks for more accurate classification of Alzheimer's disease (AD). By employing various evaluation metrics such as F1-score, AUC, recall, accuracy, precision, and specificity, along with rigorous validation techniques like train-test-split and cross-validation, we validated the robustness of our methodology. The results exhibited a significant improvement, particularly evident in the notable enhancement of classification accuracy. Notably, our research stands out due to its utilization of a larger sample size and a fusion framework that simultaneously considers all three modalities' relationships, setting a precedent for addressing similar classification challenges.

However, our study is not without limitations. We acknowledge the need for further exploration into incorporating demographic data and integrating gene expression data with MRI and PET scans to enhance learning performance. Additionally, we aim to refine intra-relationship modeling through non-linear graph modeling and augment the information extracted from PET measurements across various areas. Moreover, we recognize the potential of exploring advanced computer vision algorithms to leverage voxel characteristics and enhance diagnostic capabilities.

Moving forward, our future research endeavors will focus on addressing these limitations and expanding the scope of our methodology. Specifically, we intend to explore SSMI as a promising avenue for multi-class classification tasks in AD patient classification. Additionally, devising novel strategies for oversampling minority classes to address data imbalance in medical imaging datasets will be crucial. By prioritizing both accuracy and sensitivity in oversampling techniques, we aim to maintain the reliability of our classification model while ensuring its applicability in clinical settings. Overall, our findings pave the way for further advancements in AD classification and underscore the potential

of multimodal imaging and transfer learning in medical diagnosis.

7. Human and animal rights

No violation of Human and Animal Rights is involved.

CRediT authorship contribution statement

A. Karthik: Resources. **Hatem S.A. Hamatta:** Resources. **Sridhar Patthi:** Resources. **C. Krubakaran:** Resources. **Abhaya Kumar Pradhan:** Resources. **Venubabu Rachapudi:** Resources. **Mohammed Shuaib:** Resources. **A. Rajaram:** Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] I.K. Al-Tameemi, M. Feizi-Derakhshi, S. Pashazadeh, M. Asadpour, Interpretable multimodal sentiment classification using deep multi-view attentive network of image and text data, IEEE Access 11 (2023) 91060–91081.

[2] M. Safari, A. Fatemi, L. Archambault, MedFusionGAN: Multimodal medical image fusion using an unsupervised deep generative adversarial network, BMC Med. Imaging 23 (2023).

[3] S. Iqbal, A.N. Qureshi, M.A. Alhussein, I.A. Choudhry, K. Aurangzeb, T. Khan, Fusion of textural and visual information for medical image modality retrieval using deep learning-based feature engineering, IEEE Access 11 (2023) 93238–93253.

[4] Vasuki, A., & Malar, R.J. (2021). A Review on Multimodal Brain Image Fusion using Deep Learning for Alzheimer's disease.

[5] B. Liu, Y. Hao, H. Huang, S. Chen, Z. Li, E. Chen, X. Tian, M. Ren, TSCMDL: Multimodal deep learning framework for classifying tree species using fusion of 2-D and 3-D features, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–11.

[6] N. Tawfik, H.A. Elnemr, M. Fakhr, M.I. Dessouky, F.E. El-Samie, Multimodal medical image fusion using stacked auto-encoder in NSCT domain, J. Digit. Imaging 35 (2022) 1308–1325.

[7] M. Abdar, M.A. Fahami, L. Rundo, P. Radeva, A.F. Frangi, U.R. Acharya, A. Khosravi, H. Lam, A. Jung, S. Nahavandi, Hercules: Deep hierarchical attentive multilevel fusion model with uncertainty quantification for medical image classification, IEEE Trans. Ind. Inf. 19 (2023) 274–285.

[8] M.A. Khan, A. Khan, M. Alhaisoni, A. Alqahtani, S. Alsubai, M.S. Alharbi, N. A. Malik, R. Damaševičius, Multimodal brain tumor detection and classification using deep saliency map and improved dragonfly optimization algorithm, Int. J. Imaging Syst. Technol. 33 (2022) 572–587.

[9] R. Chanumolu, L. Alla, P. Chirala, N.C. Chennampalli, B.P. Kolla, Multimodal medical imaging using modern deep learning approaches, IEEE VLSI Device Circuit Syst. (VLSI DCS) 2022 (2022) 184–187.

[10] S.A. Adeshina, A.P. Adedigba, Bag of tricks for improving deep learning performance on multimodal image classification, Bioengineering 9 (2022).

[11] Ushaa, E., & Vishal, E. (2023). Unlocking clinical insights from medical images using deep learning. i-manager's Journal on Artificial Intelligence & Machine Learning.

[12] M.O. Odusami, R. Maskeliūnas, R. Damaševičius, S. Misra, Explainable deep-learning-based diagnosis of Alzheimer's disease using multimodal input fusion of PET and MRI images, J. Med. Biol. Eng. 43 (2023) 291–302.

[13] Hejazi, S.Z., Packianather, M.S., & Liu, Y. (2022). Novel Preprocessing of Multimodal Condition Monitoring Data for Classifying Induction Motor Faults Using Deep Learning Methods. 2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), 1-6.

[14] S. Liu, M. Wang, L. Yin, X. Sun, Y. Zhang, J. Zhao, Two-scale multimodal medical image fusion based on structure preservation, Front. Comput. Neurosci. 15 (2022).

[15] Yadav, A.K. (2021). FUSION OF MULTIMODAL BIOMETRICS OF FINGERPRINT, IRIS AND HAND WRITTEN SIGNATURES TRAITS USING DEEP LEARNING TECHNIQUE. Turkish Journal of Computer and Mathematics Education (TURCOMAT).

[16] M.O. Odusami, R. Maskeliūnas, R. Damaševičius, Pareto optimized adaptive learning with transposed convolution for image fusion Alzheimer's disease classification, Brain Sci. 13 (2023).

[17] Sangeetha Francelin Vinnarasi, F., Daniel, J., Anita Rose, J.T., & Pugalenthi, R. (2021). Deep learning supported disease detection with multi-modality image fusion. Journal of X-ray science and technology.

[18] Yuvasri (2021). Deep Learning based Automatic Brain Tumor Analysis using Multimodal Fusion.

[19] S.G. Sandhya, M. Senthil Kumar, Automated multimodal fusion based hyperparameter tuned deep learning model for brain tumor diagnosis, J. Med. Imaging Health Inform. (2022).

[20] R. Yan, F. Zhang, X. Rao, Z. Lv, J. Li, L. Zhang, S. Liang, Y. Li, F. Ren, C. Zheng, J. Liang, Richer fusion network for breast cancer classification based on multimodal data, BMC Med. Inf. Decis. Making 21 (2021).

[21] Azmat, M., & Alessio, A.M. (2022). Feature Importance Estimation Using Gradient Based Method for Multimodal Fused Neural Networks. 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 1-5.

[22] Naglah, A., Khalifa, F., Khaled, R., Razek, A.A., & El-Baz, A.S. (2021). Thyroid Cancer Computer-Aided Diagnosis System using MRI-Based Multi-Input CNN Model. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 1691-1694.

[23] Anandhi, D.F., & Sathiamoorthy, S. (2023). Enhanced Sea Horse Optimization with Deep Learning-based Multimodal Fusion Technique for Rice Plant Disease Segmentation and Classification. Engineering, Technology & Applied Science Research.

[24] M. Bihler, L. Roming, Y. Jiang, A.J. Afifi, J. Aderhold, D. Čibiraitė-Lukenskienė, S. Lorenz, R. Gloaguen, R. Gruna, M. Heizmann, Multi-sensor data fusion using deep learning for bulky waste image classification, Opt. Metrol. (2023).

[25] W. Kong, C. Li, Y. Lei, Multimodal medical image fusion using convolutional neural network and extreme learning machine, Front. Neurorob. 16 (2022).

[26] Wei, M., Xi, M., Li, Y., Liang, M., & Wang, G. (2023). Multimodal Medical Image Fusion: The Perspective of Deep Learning. Academic Journal of Science and Technology.

[27] Tanuja, N. (2022). Medical Image Fusion Using Deep Learning Mechanism.

[28] S. Misra, C. Yoon, K. Kim, R. Managuli, R.G. Barr, J. Baek, C. Kim, Deep learning-based multimodal fusion network for segmentation and classification of breast cancers using B-mode and elastography ultrasound images, Bioeng. Transl. Med. 8 (2022).

[29] Mergin, A., & Sebastin, G.P. (2023). Shearlet Transform-Based Novel Method for Multimodality Medical Image Fusion Using Deep Learning. Int. J. Comput. Intell. Appl., 22, 2341006:1-2341006:13.

[30] A. Rajaram, K. Padmavathi, S.K. Ch, A. Karthik, K. Sivasankari, Enhancing energy forecasting in combined cycle power plants using a hybrid ConvLSTM and FC neural network model, Int. J. Renew. Energy Res. (IJRER) 14 (1) (2024) 111–126.

[31] K. Salman Al-Tameemi, I., Feizi-Derakhshi, M., Pashazadeh, S., & Asadpour, M. (2023). Multi-Model Fusion Framework Using Deep Learning for Visual-Textual Sentiment Classification. Computers, Materials & Continua.

[32] Kalaivani, K., Kshirsagarr, P. R., Sirisha Devi, J., Bandela, S. R., Colak, I., Nageswara Rao, J., & Rajaram, A. (2023). Prediction of biomedical signals using deep learning techniques. *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1-14.

[33] P.A. Babu, A.K. Rai, J.V.N. Ramesh, A. Nithyasri, S. Sangeetha, P.R. Kshirsagar, S. Dilipkumar, An explainable deep learning approach for oral cancer detection, J. Electr. Eng. Technol. 19 (3) (2024) 1837–1848.

[34] Sucharitha, G., sankardass, V., Rani, R., Bhat, N., & Rajaram, A. (2024). Deep learning aided prostate cancer detection for early diagnosis & treatment using MR with TRUS images. *Journal of Intelligent & Fuzzy Systems*, *46*(2), 33 95-3409.

[35] J. Pradeep, S. Raja Ratna, P.K. Dhal, K.V. Daya Sagar, P.S. Ranjit, R. Rastogi, A. Rajaram, DeepFore: A deep reinforcement learning approach for power forecasting in renewable energy systems, Electr. Power Compon. Syst. (2024) 1–17.