3D CNN-Residual Neural Network Based Multimodal Medical Image Classification

B. SURYAKANTH, S. A. HARI PRASAD
Jain University, Bengaluru, Karnataka 560069, INDIA

Abstract:- Multimodal medical imaging has become incredibly common in the area of biomedical imaging. Medical image classification has been used to extract useful data from multimodality medical image data. Magnetic resonance imaging (MRI) and Computed tomography (CT) are some of the imaging methods. Different imaging technologies provide different imaging information for the same part. Traditional ways of illness classification are effective, but in today's environment, 3D images are used to identify diseases. In comparison to 1D and 2D images, 3D images have a very clear vision. The proposed method uses 3D Residual Convolutional Neural Network (CNN ResNet) for the 3D image classification. Various methods are available for classifying the disease, like cluster, KNN, and ANN. Traditional techniques are not trained to classify 3D images, so an advanced approach is introduced in the proposed method to predict the 3D images. Initially, the multimodal 2D medical image data is taken. This 2D input image is turned into 3D image data because 3D images give more information than the 2D image data. Then the 3D CT and MRI images are fused and using the Guided filtering, and the combined image is filtered for the further process. The fused image is then augmented. Finally, this fused image is fed to 3DCNN ResNet for classification purposes. The 3DCNN ResNet classifies the image data and produces the output as five different stages of the disease. The proposed method achieves 98% of accuracy. Thus the designed modal has predicted the stage of the disease in an effective manner.

Keywords:- Multimodal medical image; 3DCNN ResNet; Stereoscopic method; guided filtering.

Received: April 16, 2022. Revised: August 16, 2022. Accepted: September 13, 2022. Published: October 31, 2022.

1 Introduction

Multimodal Medical imaging technique has been increasingly used in the field of biomedical. In this technique, the usage of more than one modality on the same target has become a growing field. The multimodal image classification technique consists of simultaneous imaging of PET (positron emission tomography) and CT (computed tomography) or other imaging technique in the medical field that produce multimodal imaging and has become a standard clinical practice for a number of applications. Researchers have been working in the field of Medical informatics on data-driven ways to diagnose the illness automatically and detect the numerous dangerous diseases in the early years. Internal disease occurring is a unique problem which is difficult to identify in the early stages before impairment occurs, [1]. On the other hand, Medical imaging provides promise for earlier diagnosis of disease. The effects of the disease are identified based on the functioning and structure of the organs shown by computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET), [2]. MRI uses both strong magnet and radio waves to examine the body. The X-rays are used by CT to scan the entire body, whereas the X-rays are a type of ionizing radiation. The detection of disease probability is analyzed by images because each scan contains millions of pixels. Understanding such scans takes a long time for researchers and clinicians. Computer technology is used to diagnose disease probability, [3].

Machine learning techniques and other deep learning techniques have been developed recently to extract useful information in the medical field for data categorization, [4]. Usually, CNN (Convolutional Neural Network) is used for detecting conditions of the disease in medical images. The concept of CNN is based on deep learning, which has a more convoluted layer and hidden layer as well as more great image segmentation, [5]. Because of the data-driven nature, CNN can able to understand the minor difference between different classes than classic rule-based feature techniques like wavelets and principal component analysis (PCA), [6]. So, the prediction and classification of CNN will be effective on the medical dataset as benign or malignant. Likewise, many methods are available for classifying the disease like ANN, KNN, and cluster, but there are some constraints like overfitting, voxel imbalance and time-consuming during the loss function and training period, [7].

The 'overfitting' term refers to the inability of the model to generalize well. It has executed a tremendous job of learning the features of the training set, but when the same data is given for the second time, then the data will be significantly different to the same training data. So that it cannot generalize and reliably predict the outcome, [8]. Some of the existing methods use augmentation techniques to overcome the overfitting problem. Data augmentation is the method of increasing the dataset artificially by generating multiple visions of the unaffected dataset pieces in addition to the original, [9]. It is used to expand the amount of data utilized to train a model. Image, audio and Text data are all examples of data. However, the data augmentation method is suited better and has prominent growth for image data in the medical imaging field, [10].

- The multimodal medical image classification is executed with the input data as CT and MRI scans.
- Depth information is acquired by converting the 2D dataset (image) of CT and MRI to a 3D image.
- Single image filtering technique is used to combine both the 3D images of CT and MRI.
- The overfitting problem during the period of training is reduced by using data augmentation.
- A 3D CNN-ResNet classifier is used to analyze and predict multimodal image classifiers.

Multimodal data could reflect the biological mechanism of AD and MCI from different views and also could provide complementary information in classifications which is robust to noise and data heterogeneity.

The upcoming portion of the manuscript is organized as follows, and Section 2 illustrates certain research papers related to existing methods used for multimodal image classification. Section 3 describes briefly the description of the proposed methodology. Section 4 explains the results and performance metrics of the suggested framework. Section 5 concludes the entire research work.

2 Literature Review

Several algorithms have been introduced for better classification. The most commonly used classification techniques are CNN and FCN. Some of the existing techniques used in medical image classification are reviewed below.

The current state-of-the-art model [22] model achieves a 10-fold cross-validated accuracy of 86% on a train/validation/test split of 680/72/32 subjects, respectively. Their algorithm was aimed at distinguishing pMCI people who convert within 3 years; while also achieving a sensitivity of 87.5% and specificity of 85%. Their multimodal achieved these scores by using extensive preprocessing such as template registration.

Zhixian tang et al., [11], had designed a statistical shape model and three-dimensional thin-plate spline-based image augmentation strategy. This technique executes 3 procedures to detect the disease stage in MRI and CT datasets. At first, from the actual labelled images, the format information is designed with a statistical model. After that, a 3D thin-plate spline system is used to fill the generated shapes. At last, the disease is detected using a combination of generated and actual images. This technique achieves good accuracy. However, rebuilding a deep neural network is a tough undertaking with a high level of uncertainty about the outcome.

Chunyan Yu et al., [12], had designed a simple 2D-3D CNN-based novel HSIC framework implemented by collaboration among 2-D CNN and 3-D abstract levels. This technique achieves both spectral and spatial features immediately, and the strength of deep features has improved by using a convolution layer. The disadvantage of this technique is complexity and time cost, which occurred due to the enhancement in the number of 3-D kernels.

Roth, H.R et al., [13], had designed a cascaded 3D fully convolutional network-based medical image separation. In this model, there are two stages to detect the diseases, and the first stage is the conversion of 3D FCN to a roughly defined candidate region. In the second stage, FCN has to focus on a more detailed segmentation as well as classify approximately 10%. This approach achieves improved state-of-the-art outcomes; however, the loss function utilized for training is lowered because of the significant imbalance between high contrast voxels.

Horry et al., [14], had designed a COVID-19 Detection through Transfer Learning Using Multimodal Imaging Data. Detection of COVID-19 early can aid in detecting disease containment decisions and appropriate treatment plans. Through intelligent deep learning image categorization methods, this created model aims to give a second set of eyes. An appropriate Convolutional Neural Network (CNN) model is selected after a comparative examination of many prominent CNN models, followed by the selection of an optimized VGG19 model for the image modalities to demonstrate high rare and demanding COVID-19 datasets. The difficulties (including dataset size and quality) in using currently available COVID-19 datasets for developing usable deep learning models are discussed. An image preprocessing stage is utilized to produce a reliable image dataset for designing and testing the deep learning models.

Ahmadi, M et al., [15], had designed a convolutional neural network and robust PCA-based brain lesion location in MRI image detection. In this article, a convolutional neural network (CNN) is used to separate tumours in seven kinds of brain diseases, including Huntington's, Alzheimer's, Glioma, Meningioma, Alzheimer's plus, Pick and Sarcoma. Initially, the principal component analysis is employed as a feature reduction-based technique for robust finding tumour location and spot in a dataset. Then the CNN method is used to discover Brain tumours. Outcomes are illustrated based on the chance of tumour location in magnetic resonance images. Outcomes demonstrated that the allowed technique gives high accuracy (96%), sensitivity (99.9%), and dice index (91%) regarding other existing methods.

Rajalingam B. et al., [16], had designed Multimodal Medical Image Fusion based on Deep Learning Neural Network for Clinical Treatment Investigation. In this technique, a Siamese convolutional network is implemented to generate a weight map, which combines the motion of pixel data from two or even more multimodality medical images. The fusion process of medical image is passed via medical image pyramids to a multiscale manner for more reliability with the human visual sense. Furthermore, a comparison of the local-based scheme is applied for the decomposed coefficients to adaptively correct the fusion mode. An experimental outcome of the fusion technique gives the best fused multimodal medical images with the quickest processing time, leading to visualization and the highest quality in both objective assessment and visual quality criteria.

From the above-mentioned reviews, various methods are designed based on CNN, [11], 2D-3D CNN, [12], and FCN, [13], techniques. The accuracy, precision and error of the planned model are better compared to the present methods. Therefore, the selection of data augmentation with the CNN algorithm is designed in this proposed research for the effective classification of datasets.

3 Proposed Methodology

3D image classification is becoming more popular in today's globe because 3D is utilized in many sectors, such as medical and construction. In the medical profession, classification algorithms are frequently employed to give reliable prediction effects for identifying disease problems. The suggested approach is utilized to classify multimodal clinical databases that include both MRI and CT images. Normal 2D CT and MRI scans are not clear enough to provide precise information about a specific area of the body, but 3D imaging provides such information.

Initially, 2D multimodal medical images like MRI and CT scans are taken as the input dataset. These 2D images do not give deep information about the particular part, so it is needed to convert the 2D images into 3D images. A stereoscopic method is used in the conversion of 2D to 3D images. The 3D image is then taken fusion process, and both the 3D MRI and CT images are merged to get a single image. Using the Guided filtering technique, the combined image is filtered for further process. These fused images are then augmented. This augmentation is done using four different methods like brightness, contrast, saturation and hue. It alters the training dataset to generate an artificial dataset that is larger than the raw data. The primary goal of the data augmentation for fused image procedure is to reduce the overfitting issues during the training stage. 3D Residual Convolutional Neural Network (CNN-ResNet) classifier is used in the suggested approach. The fused data is given as an input for the 3D CNN-ResNet. This CNN has an arrangement of the combination of both convolutional layer and max-pooling layer. The dimension values of the dataset are divided by 16 because the 3D CNN-ResNet is used for classification purposes. Parameter Rectification Linear Unit (PReLU) is used for the activation function of this classifier. PReLU act as a threshold operator. During the training period, the value of the input is below zero, and the input is multiplied by a scalar value. 3D CNN-ResNet analyzes the given input dataset and produces an accurately predicted outcome. The below figure1 illustrates the architecture of the proposed approach.

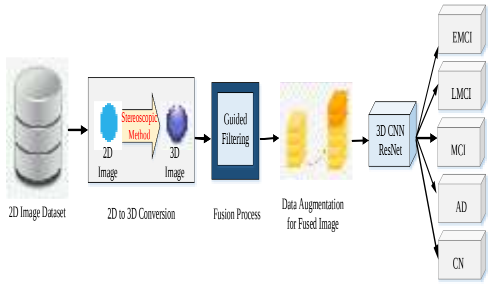


Fig. 1: Architecture of the proposed method

3.1 2D to 3D Conversion

The input dataset contains 2D images of a CT and MRI scan, these 2D images do not give detailed information about a particular part of the body, but 3D images give more information about the particular part when compared to 2D images. So it is required to convert 2D images into 3D images in the medical field. The proposed method uses the Stereoscopic method to convert 2D medical images into 3D medical images because 3D images give a clear vision of a particular part of the body.

(a)Stereoscopic Method

Many studies on the conversion of 2D images to 3D images have been undertaken throughout the world in the latest days. The proposed method uses the Stereoscopic method for 2D to 3D conversion. Stereoscopy is a method of improving or producing the illusion of three-dimensional depth from given two-dimensional images. This technology differs from 3D displays, which show an image in three dimensions, allowing the viewer to learn more about the three-dimensional objects. In the proposed method, 2D MRI and CT scan images of the human body are converted into 3D medical images by using this Stereoscopic method. This 3D image of the CT and MRI shows a clear view of the required part of the body.

Single view lenses were used to capture the 2D image. But to create a 3D image, two lenses set at a specified distance apart are used, [17]. The distance between the lenses is determined with the help of equation (1).

$$stereo = \frac{1}{30} \times distance\ of\ the\ object \quad (1)$$

Equation (1) is used to determine the distance between the lenses.

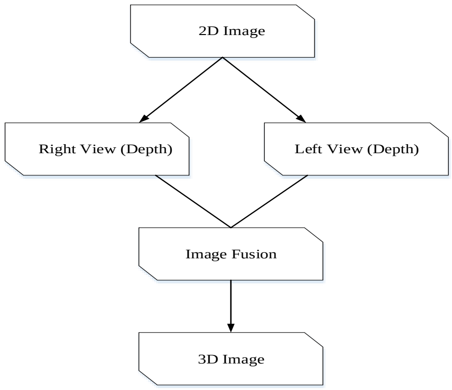


Fig. 2: 2D to 3D conversion

Figure 2 shows the process involved in 2D to 3D conversion. The 2D CT and MRI images are taken as input images. The depth value is used to create the right and left view images from the input 2D CT and MRI images. Then image fusion procedure is performed on the left and right views of the image. Finally, the depth of the 3D image is specified, and the 3D images of the CT and MRI images are produced. This 3D image gives a clear view than the 2D input image.

3.2 Fusion Process

Image fusion is the procedure of merging pairs of pictures into a series of images that combines the information from the individual photos. Medical image fusion methods combine the complementary characteristics of many medical images to create a single high-quality medical picture, decreasing lesion analysis uncertainty. Image fusion is a useful technique for a number of image processing and optical sensing applications, for instance, target detection and feature extraction. Image fusion allows you to combine many photos of the same scene into a single fused image. The combined image can give extra detailed data. The resulting image gives better information than any input image. It is utilized to produce fresh images that are better appropriate for human visual perception. The proposed method uses the Guided Filtering technique to combine the CT and MRI image.

(a)Guided Filtering

The guided filter is used initially in this proposed method for image fusion. 3D images of MRI and CT scans are combined using Guided Filtering, and the combined image is filtered for further process. This image filter can also remove noise or texture while keeping crisp edges.

$$Q_i = a_k I_i - b_k \forall_i \in w_k \quad (2)$$

$$a_k = \frac{1}{W} \frac{\sum i \in w_k - u_k E[t_k]}{\sigma_k^2 + \in} \tag{3}$$

$$b_k = E(t_k) - a_k \ u_k \tag{4}$$

The above equations (2), (3) and (4) are used to find the guided filtering of the image, [18].

Where, a_k and w_k are the unchanged factors. These factors are found using equations (3) and (4). The image data in the window on average w_k is returned by $E(t_k)$. u_k And σ_k^2 denote the mean and variance of the window featurew_k. The determined image filter has been widely utilized for image merging and has shown to be effective in merging multimodal clinical data. Guided filtering is an efficient and effective method in medical image application, including smoothing, image enhancement, image matting and joint up sampling.

3.3 Data Augmentation for Fused Image

Image data augmentation is a method of artificially boosting the size of a training dataset by creating different duplicates of the pictures in the dataset. This method alters the training dataset to generate an artificial dataset that gives large information than the original dataset. During the training period, the overfitting problem arises. To reduce this overfitting problem, a data augmentation process is used. The fused images are augmented in the proposed method using four methods like brightness, contrast, saturation and hue, [19]. Data augmentation of fused images is useful for better classification of the image.

Brightness: It works with the image's brightness to make the image darker or lighter in colour. The light in the image, or the lightning level, will be the variance between the novel image and the enhanced image. This bright image gives a clear view of the particular part of the body.

Contrast: The colour discrepancies between different areas of the image are dealt with using this approach. It gives better information about the particular part by differentiating each part image by different colours.

Saturation: Saturation is defined as the colour disparity between the image's multiple pixel hues. The depth or intensity of colour contained in a photograph is also referred to as saturation. Image duplication is generated by changing the pixel colours.

Hue: Hue is all about the picture's shadow visibility. A new image is generated by altering the image's colour hues.

By using these methods, create a huge set of images from the tiny set of input images. The huge data collection may be effectively used to produce image classification training data. Any medical application-related experimentation that requires augmentation is unavoidable since it improves the classification phases of the development process.

3.4 3D CNN ResNet Classifier

3D Residual Convolutional Neural Network is used in the proposed method for classification. 3D-RCNN can capture more information from the 3D spatial context for classification. 3D CNN is mostly used in 3D image data such as Computerized Tomography (CT) and Magnetic Resonance Imaging (MRI). The fused image data is given as an input for the 3D CNN ResNet. This 3D CNN ResNet contains the combination of both the convolutional layer and max-pooling layer. The dimension values of the dataset are divided by 16 because 3D CNN is used as a classifier. Parameter rectification linear unit (PReLU) is used for the activation function of this classifier. This 3DCNN is very much beneficial in the medical field. This classifier classifies the input data and determines the different conditions of the diseases. The below figure3 shows the architecture of the 3D CNN ResNet.

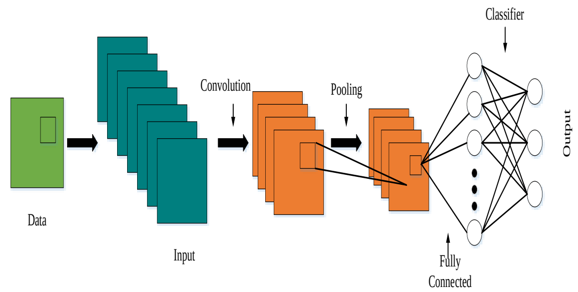


Fig. 3: Architecture of 3D CNN ResNet

3D CNN ResNet contains the input layer, convolution layer and max-pooling layer. It also contains the PReLU layer that is used for the activation function of the classifier. Fused image data given to the 3D CNN is present in the input layer. This input image is in matrix form.

Convolutional layer: It is the major block used in the 3D CNN ResNet. Features of the images are taken using the convolutional layer. The quantity and size of the kernels are specified in the convolutional layer. Mathematical processes are done between the input image and the filtered image. This mathematical operation performs using the convolutional layer. The size of the feature map is $m_2 \times m_3 \times m_4$. The output $Y_i^{(l)}$ of layer l consists of m_1^l feature maps of size $m_2^l \times m_3^l \times m_4^l$. The i^{th} feature map denoted $Y_i^{(l)}$ is computed as

$$Y_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{m_1^{(l-1)}} K_{i,j}^{(l)} * Y_J^{(L-1)} \quad (5)$$

$B_i^{(l)}$ is the bias matrix, $K_{i,j}^{(l)}$ is the filter size.

Max Pooling Layer: Pooling layers are used to decrease the size of the feature maps. The pooling layer expresses the features present in a region of the feature map produced by a convolutional layer. Max pooling layer selects the maximum component from the region of the feature map enclosed by the filter. Thus, the output after the max-pooling layer is the feature containing the most prominent features of the previous map. Figure4 shows the example of the max-pooling layer.

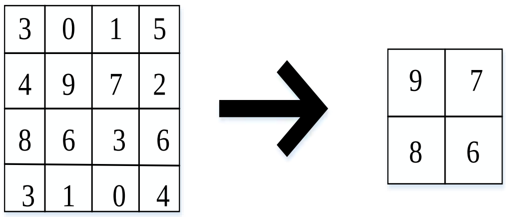


Fig. 4: Example of Max pooling

PReLU layer is an activation layer. It performs a threshold operation. For each channel, any input value is less than zero, and it is multiplied by the scalar value. It is represented in equation (6).

$$f(y_i) = a_i y_i \ \ if \ y_i \leq 0 \quad (6)$$

Fully connected layers are the layers where the entire input from a single layer is connected to every activation unit of the next layer. The input for the fully connected layer is the output from the pooling layer. Then the output layer shows the predicted output.

To extract combined spectral-spatial characteristics, use the 3D convolution technique. The spectral and spatial dimensions make up the input layer. The convolution kernel performs the convolution process in the input image. When the convolution process is done on the entire image, a new 3D feature map is acquired.

$$X_{i,j}^{x,y,z} = f\left(\sum_m \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-2} \sum_{n_3=0}^{N_3-3} W_{i,j,m}^{n_1,n_2,n_3} X_{(I-1),m}^{x+n_1,y+n_2,z+n_3} + b_{i,j}\right) \quad (7)$$

The 3D CNN ResNet equation is represented in equation (5). Where, $X_{i,j}^{x,y,z}$ denotes the output value of the jth feature graph at the location (x, y, z) of the ith layer, m represents the group of feature graphs connected to the present feature graph by the i-1 layer. The weight of the 3D convolutional kernel is represented as $W_{i,j,m}^{n_1,n_2,n_3}$ at the position (p, q, and r) in the mth feature graph. $b_{i,j}$ is represented as bias. The activation function for the Sigmoid is f. The length, breadth, and height of the convolution kernel are N1, N2, and N3, correspondingly, [20].

The proposed method using 3D CNN ResNet classifier classifies the input image dataset and determines the condition of the disease, and predicts the output as Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), Mild Cognitive Impairment (MCI), Alzheimer's disease (AD), Cognitively Normal (CN).

4 Result and Discussion

Our results suggest that the deep models outperform traditional shallow models for single modalities. The shallow models typically require handcrafted features by experts. This superior performance is due to its ability to extract relationships among features from different modalities.

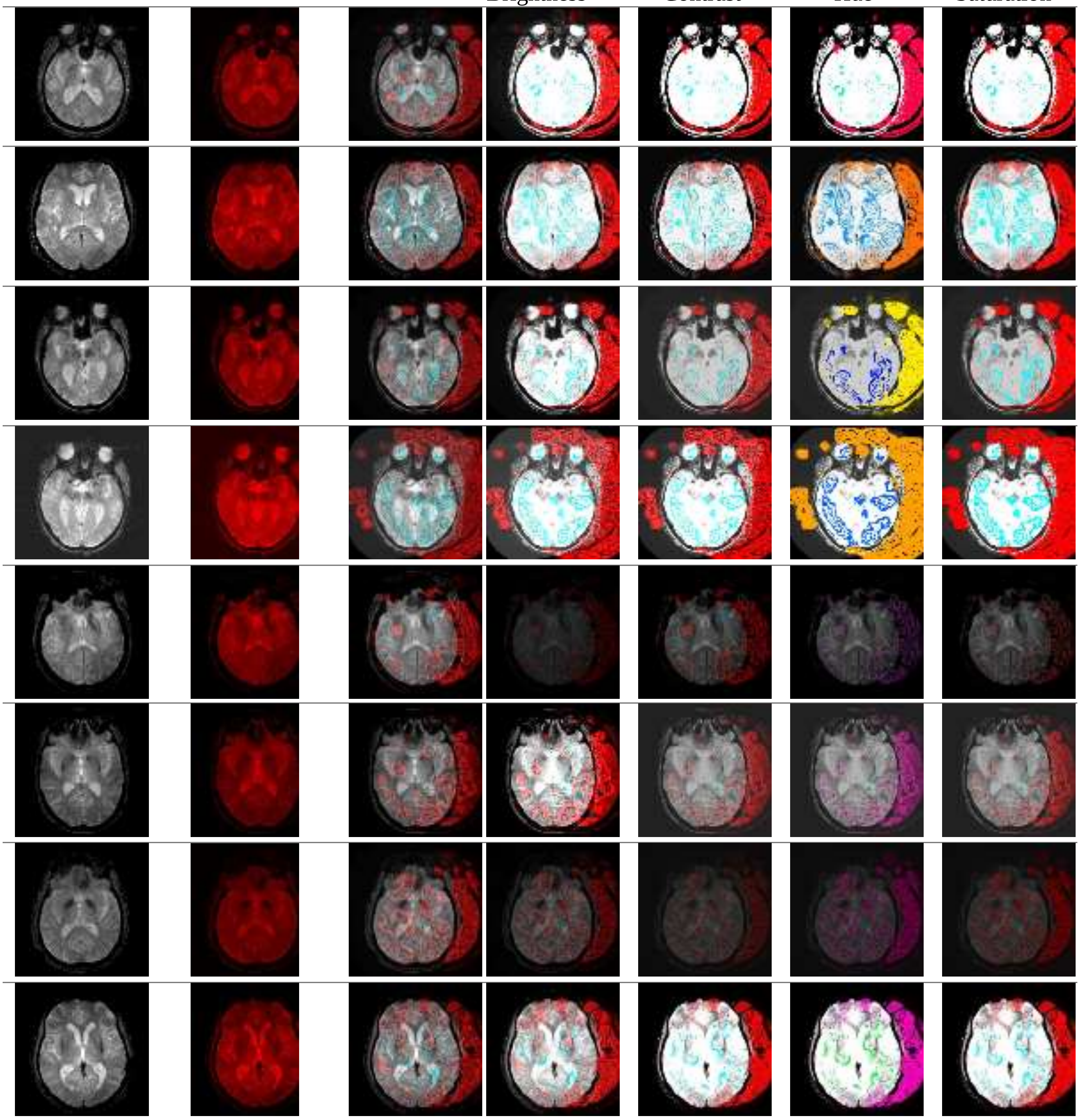
For years, medical informatics researchers have been developing data-driven methods to automate illness diagnosis procedures in order to detect a variety of severe diseases early. Inner disease is difficult to detect in the early stages before impairment emerges, which makes it a distinct challenge. The proposed method uses 3D-CNN Residual Neural Network for multimodal image classification. 2D to 3D conversion in the proposed method is performed with the help of MatlabR2021a, and the testing is performed with the help of Python 2021a with CPU: Intel Core i5, GPU: Nvidia GeForce GTX 1650 and 16GB RAM.

The data set used for the proposed approach is multimodal image data, [21]. The dataset consists of five different stages of the disease. Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), Mild Cognitive Impairment (MCI), Alzheimer's disease (AD), and Cognitively Normal (CN). Each stage of the disease is considered as each class. EMCI is represented as class1, LMCI as class2, MCI as class3, AD as class4 and CN as class5. Each class has a different set of images. Class0 has 171 images, class1 has 580 images, class2 contains 240 images, class3 contains 72 images, and in class4, 233 images are presented. These 2D multimodal image data are converted into 3D images using the Stereoscopic method. This 2D to 3D conversion is done using MatlabR2021a.

Then this image is fused and using Guided filtering, and the fused image is filtered for further process. This fused image is then augmented. This augmentation was done using four different methods like brightness, contrast, hue and saturation. This overall process is illustrated in table 1.

Table 1. Overall Process of the Proposed Method

Input image	Stereoscopic Image	Fused Image	Augmented Image			
			Brightness	Contrast	Hue	Saturation



Finally, the dataset was trained for classification purposes. The 3D CNN ResNet is used as a classifier in the proposed method. Figure5 shows the training and validation loss of the proposed method. In the figure red line indicates the training loss, and the blue line specifies the validation loss. When the number of epochs increases, the training and validation losses are decreased.

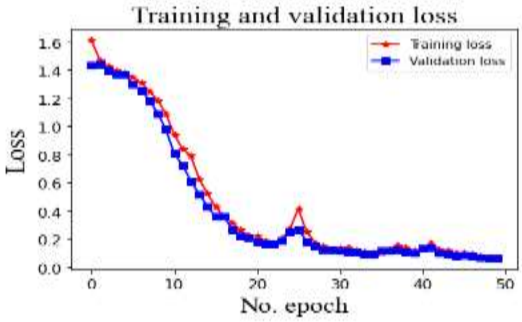


Fig. 5: Training and validation loss

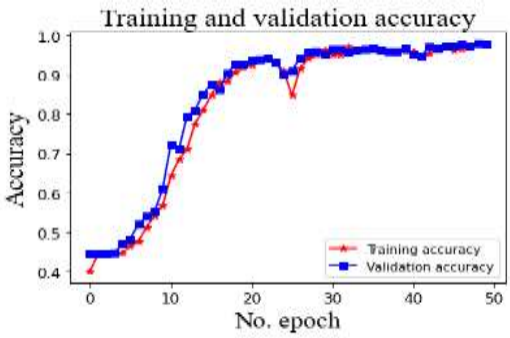


Fig. 6: Training and validation accuracy

Figure 6 illustrates the training and validation graph of the proposed method. The red and blue line in the graph indicates the training and validation accuracy. When the number of epochs increases, training and validation accuracy is increased.

Table 2. Accuracy Metrics of each class

Stages of diseases	Accuracy
Class 1	99.5%
Class 2	95%
Class 3	96%
Class 4	98.4%
Class 5	98.7%

Table 2 shows the accuracy metrics of the different stages of diseases. EMCI, LMCI, MCI, AD and CN are the five different stages of diseases. Each class has the accuracy rate of 99.5%, 95%, 96%, 98.4% and 98.7%. So the overall accuracy reached by the proposed method was 98%.

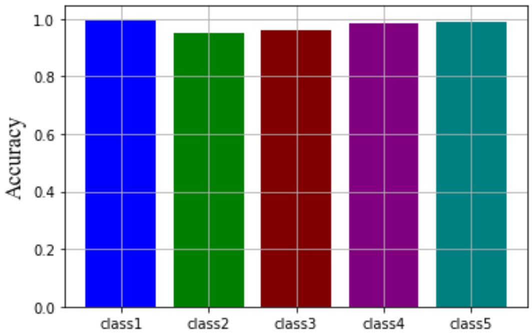


Fig. 7: Accuracy metrics of each class

The proposed method has five different classes Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), Mild Cognitive Impairment (MCI), Alzheimer's disease (AD), Cognitively Normal (CN). Figure7 shows the accuracy rate for five different classes. The accuracy rate achieved for EMCI class was 0.995, 0.95 for LMCI, 0.96 for MCI, 0.984 for AD and 0.987 for CN.

Table 3. Comparison investigation between the proposed and existing algorithms

Performance Metrics	3D-CNN	CNN-LSTM	VGG-NET	FCN	FC-LSTM
Accuracy	0.98	0.83	0.81	0.80	0.60
Precision	0.76	0.70	0.68	0.62	0.40
Recall	0.992	0.82	0.82	0.66	0.61
Error	0.02	0.15	0.17	0.20	0.40
Specificity	0.80	0.72	0.67	0.74	0.65

Performance Metrics	3D-CNN	CNN-LSTM	VGG-NET	FCN	FC-LSTM
F-1 score	0.90	0.74	0.80	0.70	0.66
Negative Predictive Value (NPV)	0.92	0.88	0.86	0.70	0.52
False Negative Rate (FNR)	0.10	0.20	0.25	0.28	0.30
False Positive Rate (FPR)	0.07	0.10	0.20	0.28	0.33
Mathew Correlation Coefficient	0.82	0.78	0.65	0.51	0.50

Table 3 shows the comparison investigation between the proposed and existing algorithms. The proposed method using 3D-CNN classifier achieves 0.88 accuracy rate, 0.76 precision, 0.92 recall, 0.09 error, 0.80 specificity, 0.90 F1_Score, 0.92 NPV, 0.10% FNR, 0.07 FPR and 0.82 MCC. Accuracy, precision, recall, specificity, F1_Score, NPV and MCC are high in the proposed method and error, FNR and FPR values are low in the proposed method compared to the existing method.

Table 4. Performance comparison of the proposed and existing methods

Methods	Accuracy	Error	Recall
Proposed approch	98%	2%	99.2%
Yasemin Turkan, et al. [2021]	94%	-	-
Chunyan Yu, et al. [2020]	97%	3%	-
Beheshti et al. [2017]	75%	-	-
Horry et al. [2019]	89%	11%	96%
Ahmadi, M et al. [2021]	96%	4%	-

Table 4 shows the performance judgement of the proposed approach and current techniques mentioned in the literature review part. When comparing the proposed approach with the existing methods mentioned in the literature review section, the accuracy and recall rate achieved in the proposed approach are high, and the error rate was low. The proposed method using a 3D CNN classifier is better than the existing methods used in the medical image classification. Some final remarks and interesting results exist in [22].

5 Conclusion

Multimodal medical images are very much important in numerous significant imaging applications. 3D image classification is becoming more popular in today's globe because 3D is utilized in many sectors, such as medical and construction. The proposed approach was utilized to classify multimodal clinical databases that include both MRI and CT images. The 2D multimodal medical image data was taken as an input in the proposed method. The 2D medical image dataset was converted into 3D images using the Stereoscopic method. 3D CT and MRT images were fused then, using guided filtering, the combined image was filtered for the classification purpose. This fused image was then augmented to reduce the overfitting problem. The 3D Residual Convolutional Neural Network (CNN ResNet) used in the proposed method classifies the augmented multimodal medical image data and classifies the different stages of the disease. The accuracy rate achieved by the proposed method using 3DCNN ResNet was high, and similarly, the false-positive rate (FPR) and error obtained were low. The result shows that the proposed multimodal medical image classification using the 3DCNN ResNet approach produced an ideal solution compared to the existing systems. So, the 3DCNN used in the proposed method was the best choice for the classification of the multimodal medical image. In future work, other advanced techniques or Artificial Intelligent will be used to detect the stages of diseases very accurately with low computational time.

Declarations

Funding: There is no funding provided to prepare the manuscript.

Conflict of Interest: The process of writing and the content of the article does not give grounds for raising the issue of a conflict of interest.

Ethical Approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Informal Consent: Informed consent was obtained from all individual participants included in the study.

Consent to participate: I have read and I understand the provided information.

Consent to Publish: This article does not contain any Image or video to get permission.

Data availability statement: If all data, models, and code generated or used during the study appear in the submitted article and no data needs to be specifically requested.

Code availability: No code is available for this manuscript.

References

[1] Dutta, K. Densely Connected Recurrent Residual (Dense R2UNet) Convolutional Neural Network for Segmentation of Lung CT Images. arXiv preprint arXiv:2102.00663, 2021.

[2] Kitanovski, I., Strezoski, G., Dimitrovski, I., Madjarov, G., & Loskovska, S. Multimodal medical image retrieval system. *Multimedia Tools and Applications*, 2017, vol.76, no.2, pp.2955-2978.

[3] Cheng, X., Zhang, L., & Zheng, Y. Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2018, vol.6, no.3, pp.248-252.

[4] Qayyum, A., Anwar, S. M., Awais, M., & Majid, M. Medical image retrieval using deep convolutional neural network. *Neurocomputing*, 2017, vol.266, pp.8-20.

[5] Hermessi, H., Mourali, O., & Zagrouba, E. Multimodal Medical Image Fusion Review: Theoretical Background and Recent Advances. Signal Processing, 2021, 108036.

[6] Tirupal, T., Mohan, B. C., & Kumar, S. S. Multimodal medical image fusion techniques–A review. *Current Signal Transduction Therapy*, 2021, vol.16, no.2, pp.142-163.

[7] Guo, Z., Li, X., Huang, H., Guo, N., & Li, Q., Medical image segmentation based on multimodal convolutional neural network: Study on image fusion schemes. *In 2018 IEEE 15th International Symposium on Biomedical Imaging* (ISBI 2018) 2018, April , pp. 903-907). IEEE.

[8] Singh, S., & Anand, R.S. (2019). Multimodal medical image fusion using hybrid layer decomposition with CNN-based feature mapping and structural clustering. *IEEE Transactions on Instrumentation and Measurement*, vol.69, no.6, pp.3855-3865.

[9] Manchanda, M., & Sharma, R. (2016). A novel method of multimodal medical image fusion using fuzzy transform. *Journal of Visual Communication and Image Representation*, vol.40, pp.197-217.

[10] Deeba, F., Dharejo, F. A., Zawish, M., Zhou, Y., Dev, K., Khowaja, S. A., & Qureshi, N. M. F. Multimodal-Boost: Multimodal Medical Image Super-Resolution using Multi-Attention Network with Wavelet Transform. arXiv preprint arXiv:2110.11684, 2021.

[11] Tang, Z., Chen, K., Pan, M., Wang, M. and Song, Z., An augmentation strategy for medical image processing based on statistical shape model and 3D thin plate spline for deep learning. *IEEE Access*, 2019, vol.7, pp.133111-133121.

[12] Yu, C., Han, R., Song, M., Liu, C. and Chang, C.I., A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial–spectral fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, vol.13, pp.2485-2501.

[13] Roth, H.R., Oda, H., Zhou, X., Shimizu, N., Yang, Y., Hayashi, Y., Oda, M., Fujiwara, M., Misawa, K. and Mori, K. An application of cascaded 3D fully convolutional networks for medical image segmentation. *Computerized Medical Imaging and Graphics*, 2018, vol.66, pp.90-99.

[14] Horry, M.J., Chakraborty, S., Paul, M., Ulhaq, A., Pradhan, B., Saha, M. and Shukla, N., COVID-19 detection through transfer learning using multimodal imaging data. *IEEE Access*, 2020, vol.8, pp.149808-149824.

[15] Ahmadi, M., Sharifi, A., Jafarian Fard, M., & Soleimani, N. Detection of brain lesion location in MRI images using convolutional neural network and robust PCA. *International Journal of Neuroscience*, 2021, pp.1-12.

[16] Rajalingam, B., & Priya, R. Multimodal medical image fusion based on deep learning neural network for clinical treatment analysis. *International Journal of ChemTech Research*, 2018, vol.11, no.06, pp.160-176.
[17] Chai, X., Shao, F., Jiang, Q. and Ho, Y.S., Roundness-preserving warping for aesthetic enhancement-based stereoscopic image editing. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, vol.31, no.4, pp.1463-1477.
[18] Rajalingam, B., Al-Turjman, F., Santhoshkumar, R. and Rajesh, M., 2020. Intelligent multimodal medical image fusion with deep guided filtering. Multimedia Systems, pp.1-15.
[19] Rani, V.V., Vasavi, G. and Kumar, K.K., A Detailed Review on Image Augmentation and Segmentation of Brain MRI Images Using Deep Learning.
[20] Zhao, J., Hu, L., Dong, Y., Huang, L., Weng, S. and Zhang, D., 2021. A combination method of stacked autoencoder and 3D deep residual network for hyperspectral image classification. *International Journal of Applied Earth Observation and Geoinformation*, vol.102, pp.102459.
[21]https://www.kaggle.com/madhucharan/alzheimersdisease5classdatasetadni.
[22] Yasemin Turkan,F.Boray Tek .Convolutional Attention Network for MRI-based Alzheimer's Disease Classification and its Interpretability Analysis. International Conference on Computer Science and Engineering.