International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I10P236

Volume 70 Issue 10, 371-380, October 2022

Original Article

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

B. Suryakanth[1], S A. Hari Prasad[2]

[1,2]Jain University, Bengaluru, Karnataka, India.

[1]Corresponding Author : suryakanthbm@gmail.com

Received: 30 June 2022 Revised: 16 September 2022 Accepted: 20 October 2022 Published: 23 October 2022

Abstract *- Multimodal medical imaging has become incredibly common in biomedical imaging. From multimodality clinical visual information, meaningful information has been derived using clinical image classification. Computed tomography (CT) and Magnetic resonance imaging (MRI) are some imaging approaches. 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. Compared to 1D and 2D images, 3D images have a very clear vision. The suggested approach 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 2D image data. After employing guided filtering to integrate the 3D CT and MRI data, the resultant image is filtered for further processing. 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 model effectively predicted the disease's stage.*

Keywords *- Designed modal, Fused image, Guided filtering, Multimodal medical image.*

1. Introduction

Multimodal Medical imaging technique has been increasingly used in the field of biomedical. Using more than one modality on the same target has become a growing field in this technique. The multimodal image classification technique consists of simultaneous imaging of CT (computed tomography) and PET (positron emission tomography) or other imaging techniques in the medical field that produce multimodal imaging and has become a standard clinical practice for many applications. Researchers have been working in the field of Medical informatics on data-driven ways to diagnose the illness automatically and detect numerous dangerous diseases in the early years. The internal disease is a unique problem that is difficult to identify in the early stages before impairment occurs (Dutta 2021).

On the other hand, Medical imaging promises an 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), positron emission tomography (PET), and magnetic resonance imaging (MRI) (Kitanovski et al. 2017). MRI uses both strong magnets and radio waves to examine the body. CT uses X-rays to scan the entire body, whereas X-rays are a type of ionizing radiation. Images analyze the detection of disease probability 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 (Cheng et al. 2018).

Recently, methods for extracting pertinent data for data categorization in the medical profession have been created using machine learning and other deep learning methods (Qayyum et al. 2017). CNN (Convolutional Neural Network) is typically used to identify medical imaging illness problems. The concept of CNN is based on deep learning, which has a more convoluted layer and hidden layer, as well as greater image segmentation (Hermessi et al. 2021). Because of the data-driven nature, CNN can understand the minor difference between different classes than classic rule-based feature techniques like wavelets and principal component analysis (PCA) (Tirupal et al. 2021). 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 some constraints like overfitting, voxel imbalance, and time-consuming during the loss function and training period (Guo et al. 2018, April).

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. Still, when the same data is given for the second time, the data will be significantly different from the same training data. So that it cannot generalize and reliably predict the outcome (Singh et al. 2019). Some of the existing methods use augmentation techniques to overcome the overfitting problem. Data augmentation is increasing the dataset artificially by generating multiple visions of the unaffected dataset pieces in addition to the original Manchanda and Sharma (2016). 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 better suited and has prominent growth for image data in medical imaging (Deeba et al. 2021).

- 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.
- A 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.

The rest of the paper includes: portion 2 contains the paper related to multimodal image classification. Portion 3 represents the suggested approach. The result and discussion part of the proposed method is presented in portion 4. At last, portion 5 contains the conclusion part.

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.

Zhixian tang et al. (Tang et al. 2019) designed a statistical shape model and a 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 tough, with high uncertainty about the outcome.

Chunyan Yu et al. (Yu et al. 2020) designed a simple 2D-3D CNN-based novel HSIC framework implemented by collaboration among 2-D CNN and 3-D abstract levels. This technique immediately achieves both spectral and spatial characteristics, 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. (Roth et al. 2018) designed a cascaded 3D fully convolutional network-based medical image separation. The two stages of this model's illness detection process begin with the 3D FCN's transformation into a roughly defined candidate region. In the second stage, FCN must concentrate on a more precise segmentation and classify about 10% of the data. 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. (Horry et al. 2020) designed a COVID-19 identification via Transfer Learning utilizing Multimodal Imaging information. Detection of COVID-19 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. A suitable Convolutional Neural Network (CNN) model is chosen following a comparison of many well-known CNN models. A VGG19 model tuned for picture modalities is then chosen to demonstrate the high rarity and difficulty of COVID-19 datasets. The challenges of using the COVID-19 datasets that are currently available for building practical deep-learning models are explored, including dataset size and quality issues.

Ahmadi M et al. (Ahmadi et al. 2021) designed a convolutional neural network and robust PCA-based brain lesion location in MRI image detection. This study uses seven different types of brain diseases to differentiate tumours. Initially, the principal component analysis is employed as a feature reduction-based technique for robust finding tumour location and spot in a dataset. Brain tumours are then found using the CNN approach. The likelihood of a tumour's location in magnetic resonance scans is used to demonstrate outcomes.

Rajalingam and Priya (2018) designed a Multimodal clinical data Fusion utilizing Deep Learning Neural Network for Clinical Therapeutic Investigation. A Siamese convolutional network was utilized in this method to create a weight map that incorporates the motion of pixel data from two or even more multimodality clinical data. The fusion process of clinical data was passed via medical image pyramids to multiscale for more reliability with the human visual sense. Furthermore, the local-based scheme is compared for the decomposed coefficients to correct the fusion mode adaptively. 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.

Zhang J et al. (Zhang et al. 2019) designed a deep synergic learning (SDL) model by utilizing several deep convolutional neural networks concurrently and allowing them to learn from one another. As the input of a synergic network, which has a fully connected structure and predicts if the pair of input images belong to the same class, each pair of DCNNs' learned image representation is concatenated with its learned representation. In other words, if one DCNN correctly classifies, the second DCNN's error causes a synergic error, which adds further pressure to upgrade the system.

From the reviews mentioned earlier, various methods are designed based on CNN (Tang et al.2019), 2D-3D CNN (Yu et al. 2020), and FCN (Roth et al. 2018) techniques. The planned model's accuracy, precision, and error are better than the present methods. Therefore, the selection of data augmentation with the CNN algorithm was developed in this suggested research to classify datasets effectively.

3. Proposed Methodology

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

Initially, 2D multimodal clinical data 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. The combined image is filtered using the Guided filtering technique for further processing. These fused images are then augmented. This augmentation uses four methods: brightness, contrast, saturation, and hue. It alters the training dataset to generate an artificial dataset larger than the raw data. The primary goal of the data augmentation for the fused image procedure is to reduce the overfitting issues during the training stage.

A 3D Residual Convolutional Neural Network (CNN-ResNet) classifier is utilized in the suggested approach. The fused data is given as an input for the 3D CNN-ResNet. Convolutional also max-pooling layers are combined in this CNN's configuration. 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.

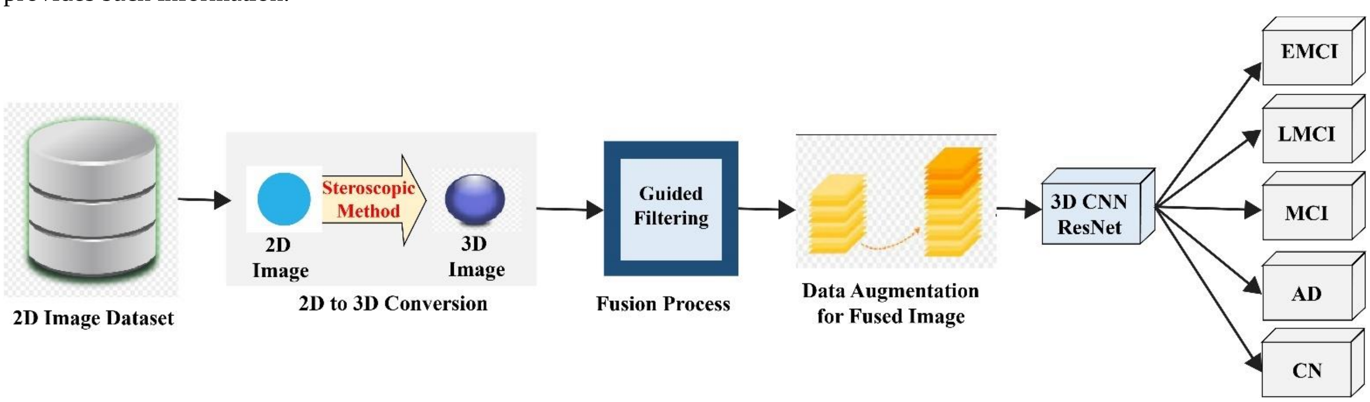

Fig. 1 Architecture of the proposed method

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. Therefore, converting 2D images to 3D images is necessary for the medical industry. 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 body part.

Many recent studies on converting 2D images to 3D images have been undertaken worldwide. 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 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 (Chai et al. 2020). The distance between the lenses is determined with the help of equation (1).

$$stereo = \frac{1}{30} \times distance\ of\ the\ object \tag{1}$$

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

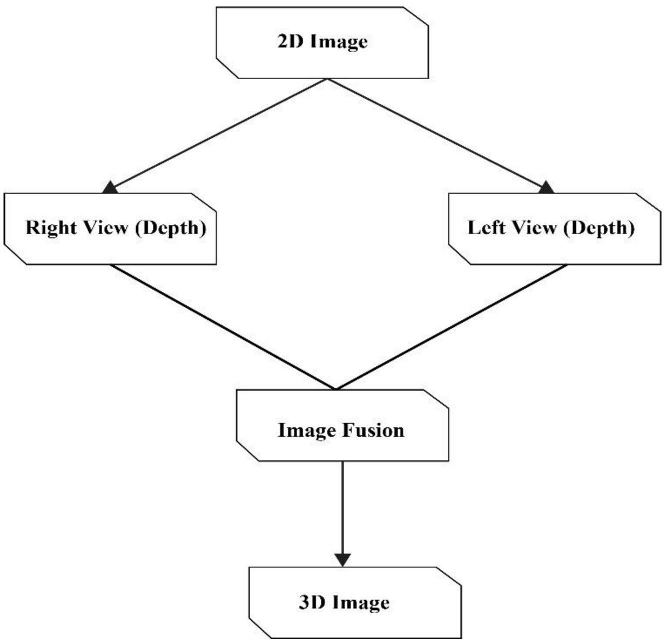

Fig. 2 2D to 3D conversion

Fig. 2 depicts the conversion process from 2D to 3D. The 2D CT and MRI images are taken as input images. The depth value is utilized to create the right and left view images from the input 2D CT and MRI images. The left and right perspectives of the image are then combined using an image fusion process. 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 combines two photos to create a series of images incorporating the data from separate shots. Clinical data fusion techniques integrate the beneficial aspects of numerous clinical photos to produce a single, excellent clinical image, decreasing lesion analysis uncertainty. Image fusion is a useful method for some photo processing and optical sensing applications, for instance, target detection and feature extraction. With the help of picture fusion, a single fused image is created by combining numerous photos of the same scene. 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 combines the CT and MRI images with the Guided Filtering technique.

3.2.1. 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 processing. This image filter can also remove noise or texture while keeping crisp edges.

$$Q_i = a_k I_i - b_k \forall_i \in w_k \tag{2}$$

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

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

The above equations (2), (3), and (4) are used to find the guided filtering of the image (Rajalingam et al. 2020).

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 average and dispersion of the window featurew_k. The determined image filter has been widely used for image merging and has been effective in merging multimodal clinical data. Guided filtering is an efficient and effective method in medical image applications, including smoothing, image enhancement, image matting, and joint upsampling.

3.3. Data Augmentation for Fused Image

It is a method of intentionally expanding the amount of a training dataset by producing various copies of the images 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: brightness, contrast, saturation, and hue (Rani, 2019). Data augmentation of fused images is useful for better classification of the image.

3.3.1. 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 and enhanced photos. This bright image gives a clear view of a particular part of the body.

3.3.2. Contrast

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

3.3.3. Saturation

Saturation is 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.

3.3.4. 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 effectively 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-RCNN can capture more information from the 3D spatial context for classification. The fused image data is input for the 3D CNN ResNet. This 3D CNN ResNet contains the combination of both the convolutional also max-pooling unit. 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.

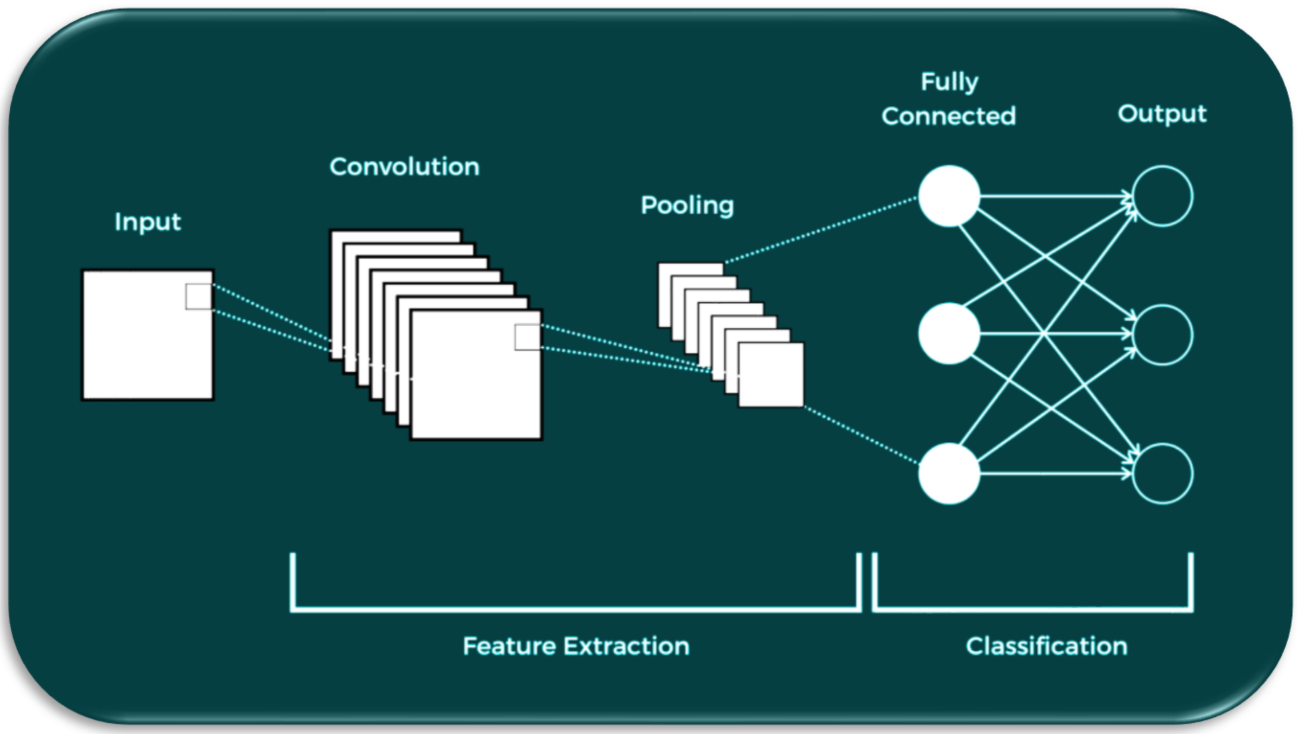

Fig. 3 Architecture of 3D CNN ResNet

3D CNN ResNet contains various input, convolution, and max-pooling units. It also contains the PReLU layer used for the classifier's activation function. Fused image data from 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 amount 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: The amount of feature maps is reduced by using pooling layers. The max pooling layer selects the largest element in the feature map area enclosing the filter. The feature that contains the most noticeable features from the preceding map is what is produced as a result of this unit. Figure 4 illustrates a max-pooling unit example.

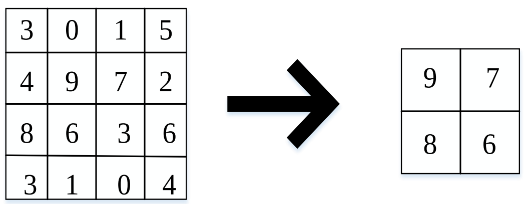

Fig. 4 Example of Max pooling

The PReLU layer is an activation layer. The threshold procedure is done in this unit. Any input value lower than zero for any channel is multiplied by the scalar value. It is represented in equation (6).

$$f(y_i) = a_i y_i \ \ if \ y_i \leq 0 \tag{6}$$

That unit is said to be fully connected when all input from one unit is linked to each activation unit in the layer above it. Input for the fully connected unit is the output from the pooling unit. Then the output layer shows the predicted output.

To extract combined spectral-spatial characteristics, use the 3D convolution technique. The input unit is composed of both spectral and spatial aspects. When the convolution process is done on the entire image, a 3D feature map is acquired.

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

The 3D CNN ResNet equation is represented in equation (5). Where, the symbol $X_{i,j}^{x,y,z}$ indicates the final result, and m indicates the collection of feature graphs. The weight of the 3D convolutional kernel is represented as $W_{i,j,m}^{n_1,n_2,n_3}$. $b_{i,j}$ represented as bias. Sigmoid's perceptron is denoted by f. N1, N2, and N3 denotes length, breadth, and height correspondingly (Zhao et al. 2021).

The proposed method using 3D CNN ResNet classifier classifies the input image dataset, determines the disease's condition, and predicts the output as five different classes.

4. Result and Discussion

According to our findings, deep models perform better for single modalities than conventional shallow models. Usually, bespoke features by professionals are needed for shallow models. Its capacity to extract correlations between features from many modalities accounts for its exceptional performance.

For years, medical informatics researchers have been developing data-driven methods to automate illness diagnosis procedures to detect various 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 GPU: Nvidia GeForce GTX 1650, CPU: Intel Core i5 and 16GB RAM.

The data set utilized for the suggested technique is multimodal image data (https://www.kaggle.com/madhucharan/alzheimersdisease5classdatasetadni). The dataset consists of five different stages of the disease. 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 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 using Guided filtering, and the fused image is filtered for further processing. This fused image is then augmented. This augmentation was done using four different methods: brightness, contrast, hue, and saturation. Table[1] depicts this overall process.

Finally, the dataset was trained for classification purposes. The 3D CNN ResNet is used as a classifier in the proposed method. Figure5 illustrates the loss of training and validation of the proposed method. The figure's red line indicates the training loss, and the blue line specifies the validation loss. Losses during training, as well as validation, are reduced as the value of epochs rises.

Figure 6 illustrates the training and validation graph of the suggested approach. The graph's red and blue line indicates the training and validation accuracy. When the epoch value increases, training also validation accuracy are increased.

Table 1. Overall Process of the Proposed Method

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

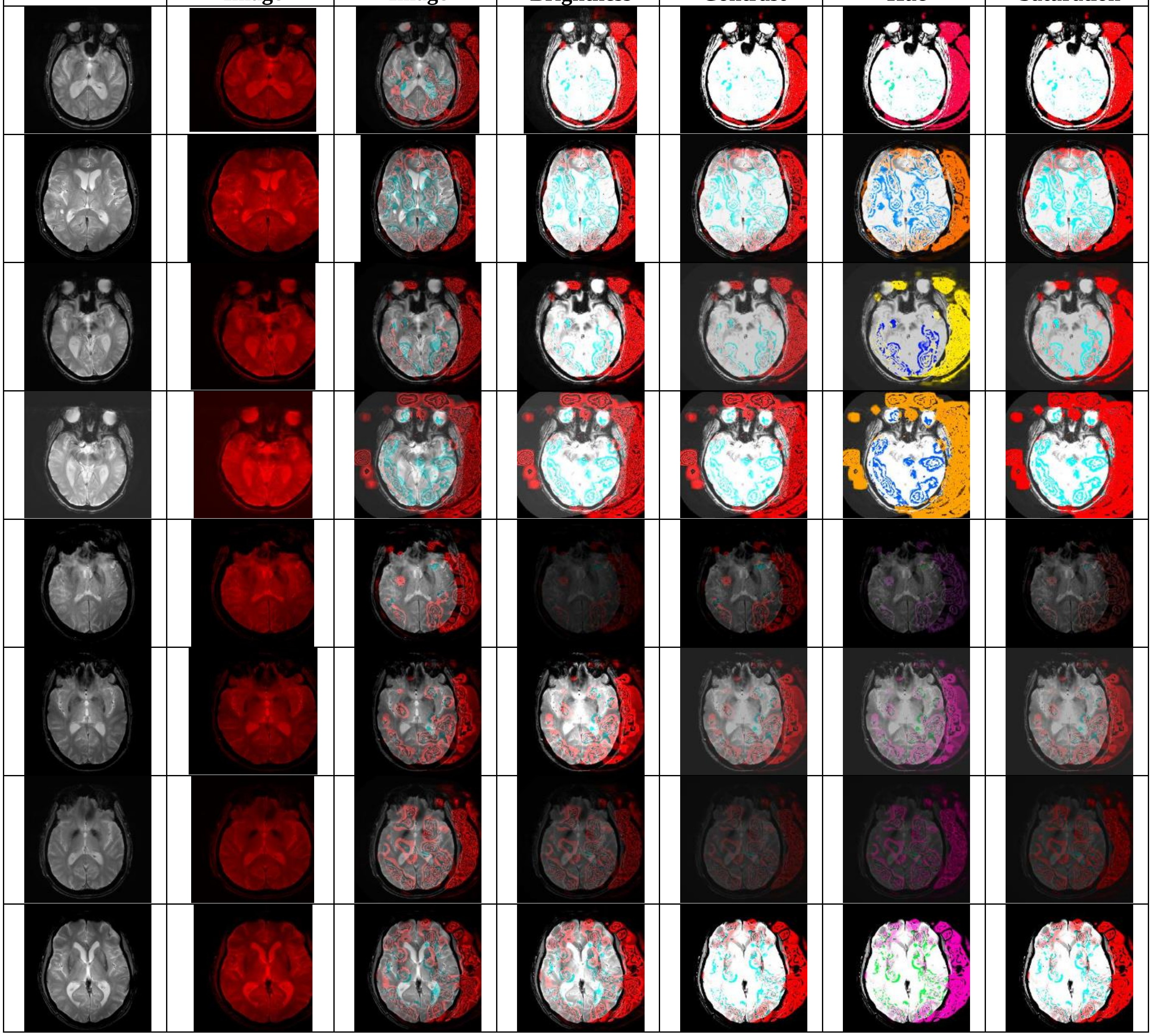

Table 2. Accuracy Metrics of each class

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 parameter for various 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%.

The proposed method has five different classes MCI, EMCI, CN, LMCI, and AD. 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 depicts the comparison investigation between the suggested 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, specificity, precision, F1_Score, NPV, recall, and

MCC are high in the proposed method, and error, FNR, and FPR values are low in the proposed method compared to the current algorithms.

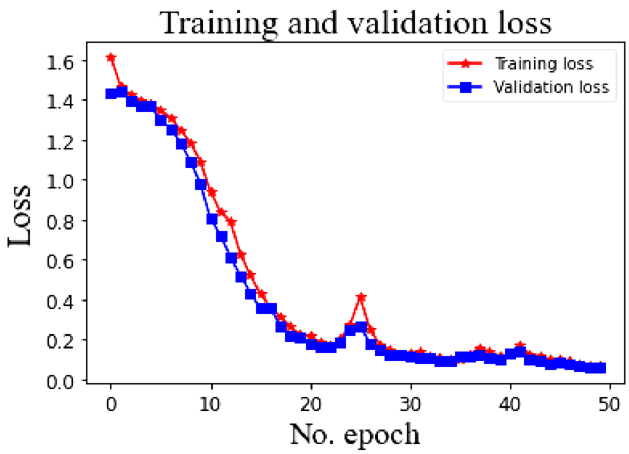

Fig. 5 Loss of validation and accuracy

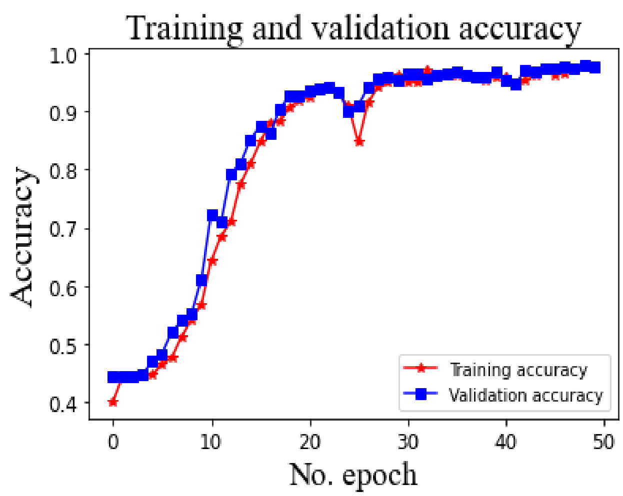

Fig. 6 Accuracy of validation and accuracy

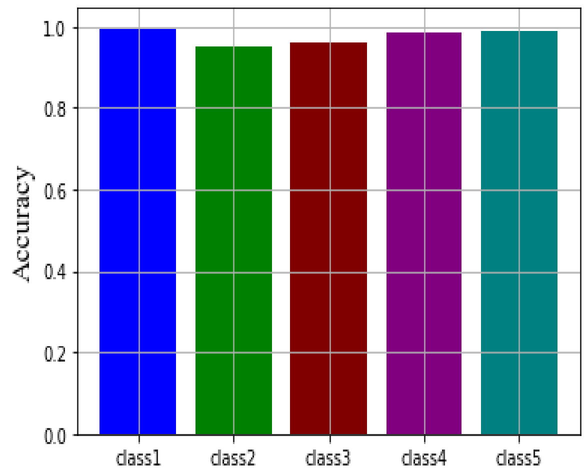

Fig. 7 Accuracy metrics of each class

Table 4 shows the performance judgement of the suggested algorithm and current methods mentioned in the literature review. When comparing the suggested technique with the current methods mentioned in the literature review section, the accuracy and recall rate achieved in the proposed approach are high, and the error rate is low. Thus the proposed method using a 3D CNN classifier is better than the existing methods used in medical image classification.

Table 4. Performance comparison of the suggested as well as Current techniques

Methods	Accuracy	Error	Recall
Proposed approach	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 3. Comparison investigation among the suggested and current algorithms

Performance Metrics	3D-CNN	CNN-LSTM	VGG-NET	FCN	FC-LSTM
Accuracy (%)	98	83	81	80	60
False Positive Rate (FPR) (%)	0.7	10	20	28	33
Precision (%)	76	70	68	62	40
Negative Predictive Value (NPV) (%)	92	88	86	70	52
Recall (%)	99.2	82	82	66	61
False Negative Rate (FNR) (%)	10	20	25	28	30
Error (%)	0.2	15	17	20	40
Mathew Correlation Coefficient (%)	82	78	65	51	50
Specificity (%)	80	72	67	74	65
F-1 score (%)	90	74	80	70	66

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 MRI and CT images. The 2D multimodal medical image data was taken as 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 classification purposes. This fused image was then augmented to reduce the overfitting problem. The 3D Residual Convolutional Neural Network (CNN ResNet) used in the suggested 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 classifying 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.

Acknowledgments

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of Interest

The authors declared that they have no conflicts of interest in this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Author Contributions

The corresponding author claims the paper's major contribution, including formulation, analysis, and editing. The co-author guides to verify the analysis result and manuscript editing.

Compliance with Ethical Standards

This article is a completely original work of its authors; it has not been published before and will not be sent to other publications until the journal's editorial board decides not to accept it for publication.

References

[1] K. Dutta, "Densely Connected Recurrent Residual (Dense R2UNet) Convolutional Neural Network for Segmentation of Lung CT Images," arXiv preprint arXiv: 2102.00663, 2021. https://doi.org/10.48550/arXiv.2102.00663.

[2] I. Kitanovski, G. Strezoski, I. Dimitrovski, G. Madjarov, and S. Loskovska, "Multimodal Medical Image Retrieval System," *Multimedia Tools and Applications,* vol. 76, no. 2, pp. 2955-2978, 2017. https://doi.org/10.1007/s11042-016-3261-1.

[3] X. Cheng, L. Zhang, and Y. Zheng, "Deep Similarity Learning for Multimodal Medical Images," C*omputer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization,* vol. 6, no. 3, pp. 248-252, 2018. https://doi.org/10.1080/21681163.2015.1135299.

[4] A. Qayyum, S. M. Anwar, M. Awais and M. Majid, "Medical Image Retrieval using Deep Convolutional Neural Network," *Neurocomputing,* vol. 266, pp. 8-20, 2017. https://doi.org/10.1016/j.neucom.2017.05.025.

[5] H. Hermessi, O. Mourali, and E. Zagrouba, "Multimodal Medical Image Fusion Review," *Theoretical Background and Recent Advances Signal Processing,* pp. 108036, 2021. https://doi.org/10.1016/j.sigpro.2021.108036.

[6] T. Tirupal, B. C. Mohan and S. S. Kumar, "Multimodal Medical Image Fusion Techniques–A Review," *Current Signal Transduction Therapy,* vol. 16, no. 2, pp. 142-163, 2021.https://doi.org/10.2174/1574362415666200226103116.

[7] S. Farjana Farvin, S. Krishna Mohan, "A Comparative Study on Lung Cancer Detection using Deep Learning Algorithms," *SSRG International Journal of Computer Science and Engineering,* vol. 9, no. 5, pp. 1-4, 2022. Crossref, https://doi.org/10.14445/23488387/IJCSE-V9I5P101.

[8] Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, "Medical Image Segmentation Based on Multimodal Convolutional Neural Network: Study on Image Fusion Schemes," *In 2018 IEEE 15th International Symposium on Biomedical Imaging,* IEEE, pp. 903-907, 2018. https://doi: 10.1109/ISBI.2018.8363717.

[9] S. Singh and R. S. Anand, "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, 2019. https:// doi: 10.1109/TIM.2019.2933341.

[10] M. Manchanda and R. Sharma, "A Novel Method of Multimodal Medical Image Fusion using Fuzzy Transform," *Journal of Visual Communication and Image Representation,* vol. 40, pp. 197-217, 2016. https://doi.org/10.1016/j.jvcir.2016.06.021.

[11] Lakshman Karthik Ramkumar, Sudharsana Premchand, Gokul Karthi Vijayakumar, "Sign Language Recognition using Depth Data and CNN," *SSRG International Journal of Computer Science and Engineering,* vol. 6, no. 1, pp. 9-14, 2019. Crossref, https://doi.org/10.14445/23488387/IJCSE-V6I1P102.

[12] F. Deeba, F. A. Dharejo, and M. Zawish, Y. Zhou, et al., "Multimodal-Boost: Multimodal Medical Image Super-Resolution using Multi-Attention Network with Wavelet Transform," arXiv preprint arXiv: 2110.11684, 2021. https://doi.org/10.48550/arXiv.2110.11684.

[13] Z. Tang, K. Chen, M. Pan, M. Wang, and Z. Song, "An Augmentation Strategy for Medical Image Processing Based on Statistical Shape Model and 3D Thin Plate Spline for Deep Learning," *IEEE Access,* vol. 7, pp. 133111-133121, 2019. https://doi: 10.1109/ACCESS.2019.2941154.

[14] C. Yu, R. Han, M. Song, C. Liu, and C. I. Chang, "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,* vol. 13, pp. 2485-2501, 2020. https://doi: 10.1109/JSTARS.2020.2983224.

[15] H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, and K. Mori, "An Application of Cascaded 3D Fully Convolutional Networks for Medical Image Segmentation," *Computerized Medical Imaging and Graphics,* vol .66, pp. 90-99, 2018. https://doi.org/10.1016/j.compmedimag.2018.03.001.

[16] Ms.S.Supraja, Dr.P.Ranjith Kumar, "An Intelligent Traffic Signal Detection System Using Deep Learning," *SSRG International Journal of VLSI & Signal Processing,* vol. 8, no. 1, pp. 5-9, 2021. Crossref, https://doi.org/10.14445/23942584/IJVSP-V8I1P102.

[17] M. J. Horry, S. Chakraborty, M. Paul, A. Ulhaq, B. Pradhan, M. Saha and N. Shukla, "COVID-19 Detection through Transfer Learning using Multimodal Imaging Data," *IEEE Access,* vol. 8, pp. 149808-149824, 2020. https://doi: 10.1109/ACCESS.2020.3016780.

[18] M. Ahmadi, A. Sharifi, M. J. Fard, and N. Soleimani, "Detection of Brain Lesion Location in MRI Images using Convolutional Neural Network and Robust PCA," *International Journal of Neuroscience,* pp.1-12, 2021. https://doi.org/10.1080/00207454.2021.1883602.

[19] B. Rajalingam and R. Priya, "Multimodal Medical Image Fusion Based on Deep Learning Neural Network for Clinical Treatment Analysis," *International Journal of ChemTech Research,* vol. 11, no. 6, pp. 160-176, 2018. http://dx.doi.org/10.20902/IJCTR.2018.110621.

[20] J. Zhang, Y. Xie, Q. Wu and Y. Xia, "Medical Image Classification using Synergic Deep Learning," *Medical Image Analysis,* vol. 54, pp. 10-19, 2019.

[21] X. Chai, F. Shao, Q. Jiang, and Y. S. Ho, "Roundness-Preserving Warping for Aesthetic Enhancement-Based Stereoscopic Image Editing," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 31, no. 4, pp. 1463-1477, 2020. https://doi: 10.1109/TCSVT.2020.3010181.

[22] B. Rajalingam, F. Al-Turjman, R. Santhoshkumar, and M. Rajesh, "Intelligent Multimodal Medical Image Fusion with Deep Guided Filtering," *Multimedia Systems,* pp.1-15, 2020. https://doi.org/10.1007/s00530-020-00706-0.

[23] V. V. Rani, G. Vasavi and K. K. Kumar, "A Detailed Review on Image Augmentation and Segmentation of Brain MRI Images using Deep Learning," *Paideuma Journal,* 2018. https://doi.org/10.1016/j.jocs.2018.12.003.

[24] J. Zhao, L. Hu, Y. Dong, L. Huang, S. Weng, and D. Zhang, "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, 2021. https://doi.org/10.1016/j.jag.2021.102459.

[25] Alzheimers Disease 5 Class Dataset ADNI, [Online]. Available: https://www.kaggle.com/madhucharan/alzheimersdisease5classdatasetadni.