International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

A HYBRID CNN-TRANSFORMER NETWORK FOR THE SEGMENTATION OF HIPPOCAMPUS FROM BRAIN MRI

Karthigai Selvi S [1], Krishan Veer Singh [2], Sanjiv Sharma [3], Karpaga Selvi S [4] A.R. Sivakumaran [5]

School of Computer Science and Engineering, Galgotias University

Greater Noida - 203201, India

email: karthigachandru@gmail.com [1], kv.jnu07@gmail.com [2], martin.mmmec@gmail.com [3], karpagaselvil@gmail.com [4], sivakumaran.ars@gmail.com [5]

Abstract

The physical features of Hippocampus (hc) in human brain indicate several neurological disorders such as Alzheimer. Magnetic Resonance Images (MRI) play a prominent role in radiotherapy planning. Hippocampus segmentation methods for radiotherapy planning make use of brain atlas, convolution neural network etc. We proposed a strategic CNN model with transformer layer which includes several human-engineered feature filters to segment hippocampus. We investigate the performance of the proposed features set in segmenting hippocampus from MRI without make much loss in detection. The feature set includes contrast of the MRI, wavelet implemented frequency components, energy level of grey values and regular statistical parameters. The features are provided as input to a deep learning model and ensured the opted activation function. The qualities are measured using dice, precision, recall, sensitivity, specificity and accuracy. The results ensured the proposed features pioneering the segmentation process.

Keywords: features, segmentation, hippocampus, active contour, binary pattern, weber's law, statistical features

1. Introduction

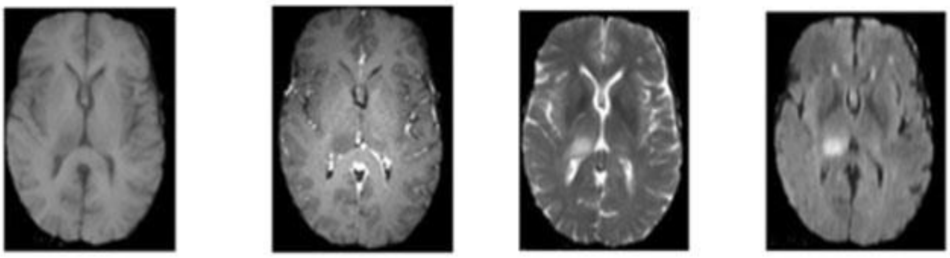
Hippocampus (hc) is a very complex structured small component in our brain that incorporates short term learning and memory process. Due to its sea-horse structure, it is called as Hippocampus [1][2]. The size of hc is indicating the presence of memory disorders such as Alzheimer and dementia [3] epilepsy and schizophrenia [4]. The tissue counting process of hc ensures the progressive of Alzheimer and guide to diagnose for the treatment plan. At present, Magnetic Resonance Imaging (MRI), which is non-invasive technology to mimic the morphology of human organs particularly brain [5]. The earlier detection of the disorder helps to get cure or avoid severity of the disorder. Radiotherapy is one of the disciplines to cure the neurological disorders but it requires segmented structure for therapeutic planning [1]. Accordingly, segmentation of the hc is going active and sensitive research topic.

Received: August 08, 2025

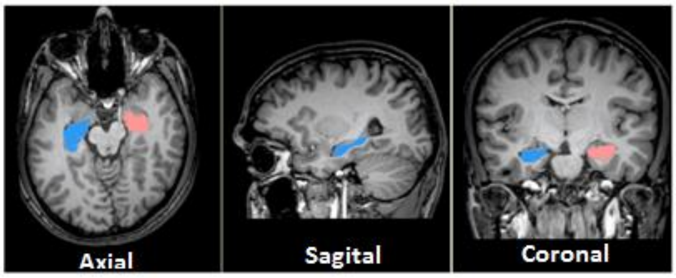
International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)



(a)



(b)

Figure 1. (a) Different contrast: T1-weighted, T1 contrast enhanced, T2-weighted and Flair respectively. (b) Different orientation: Axial, Sagital and Coronal respectively

MRI are good in contrast, produces different intensity varied images such as T1-Weighted (T1w), T2-Weighted (T2w) and Contrast enhanced (Tc) images. They are provided in Fig.1 (a), which shows different gray levels for same brain components. Additionally, MRI machine provides images from different orientations such as Axial, Coronal and Sagittal as given in Fig.1(b). The axial images are displays the brain from top to bottom, coronal images shows the brain structure from front to back and sagittal images shows slices from left to right. For diagnosis, different contrast and orientation images are used by the physicians. The colouring part of MRI in Fig.1.(b) represents hp. It shows the size of hc in different orientation images, sagittal orientation shows hc in large size. Usually, the hc segmentation process uses a sequential pipeline which encompasses pre-processing, registration and segmentation. Most of the algorithms exercise resampling as early processing procedure in which the collected image signals are resample at some millimetre voxel size that reduce the erroneous pixels. [1] resampled the images into 1.5mm voxel size and implemented augmentation techniques such as flipping, rotation etc.

Machine learning is an escalating technology which uses a set of features from a dataset to find the optimal parameters by trainning the network. Prior to machine learning, hc segmentation researches utilized registration and deformal methods. Wang et al. [6] followed multi atlas segmentation and fuse the similar labelled voxels. Along with the hc its subfields too extracted by using the multi-atlas registration method [7]. The atlas based methods provides higher result than manual segmentation but it consumes 8 hours for a single volume [8]. Currently Freesurfer is widely used by the medical image researchers. Though deep learning based methods yields

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

highest accuracy and consuming less time compared to other. The Random forest utilizes the local energy pattern feature set which were proposed in for the classification. The random forest method fabricated with 20 trees and the depth of each tree is empirically set as 25 [9]. The QuickNat network utilizes the convolution neural network for the segmentation process which results lower than the proposed method. The deep learning-based approach also achieved a DSC score of 0.85. When compared to other methods, the feature set and network deliver comparable performance with the smaller dataset and fewer features. The training process takes approximately 20 minutes, and testing requires around 10 minutes on a dual-core 2 processor machine [10].

1.1. Background

It is one of the subsets of machine learning (ML) algorithm which resembles the human brain. The neural network made up of several layers which include several neurons, the neurons transform the input into structured and unstructured huge amount of data. ML algorithms can automatically take decisions by the process of leaning the given data set. The learning process is similar to our human brain which takes the past data and learns logically to make significant decisions. Several ML methods have been implemented in the task of hc segmentation [11][12][13][14][15]. Deep learning one of the ML techniques and it is differ by its layers and neuron counts from ML technique and it takes decisions without the interference of the human. These features motivate the researchers to implement it in various industries, banking, e-commerce, health care and etc.

Deep learning techniques are generally divided into supervised and unsupervised learning approaches. Models such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, Recurrent Neural Networks (RNN), Radial Basis Function (RBF) networks, and Multilayer Perceptron (MLP) fall under supervised learning. On the other hand, Self-Organizing Maps (SOM), Deep Belief Networks (DBN), Restricted Boltzmann Machines (RBM), and Auto encoders are categorized as unsupervised learning methods.

Several deep learning methods were proposed by modifying the network structure and using different activation functions. The methods commonly use 2D or 3D images as inputs in the training phase. A CNN based method implemented a strategic input idea that cuts the volume in six more diagonal orientation and applied U-Net CNNs [16]. Xie ei al.[17] proposed a faster method than atlas based method. They converted the image as patches by tri-planar method. The patches from the central part of image where the hippocampus belongs to are provided as input during the training period that would minimized the overall segmentation time.

The structure of deep learning is modified as DeepNet which employed to obtain 25 structures from 3D image. The network is also called as 3D CNN architecture [18]. Roy, A.G., and et al. [10] introduced QuickNet which utilized multiple 2D U-Net architecture. Ataloglou, D. and et al. [4] utilizes multiple CNN outputs obtained from different orientation images. They uses U-Net, augmentation technique and error correction. This approach yielded a good result [4]. Thyreau, B. and et al.'s [19] model, named Hippodeep which uses CNNs, this model provided training in the region of interest (ROI). This model is starting with a planar analysis followed

International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

by layers of 3D convolutions and shortcut connections. The segmentation results of ML are depending on the training samples. Due to the irregular shape and lack in significant boundary detection in the segmentation process yields insignificant results [5].

Several recent studies have explored deep learning-based approaches for medical image segmentation, each presenting unique techniques and challenges. Dutta, P., and et al. [20] proposed an R-CNN combined with an epithelial network for prostate cancer prediction, achieving a notable accuracy of 99.07% and an AUC of 0.998. However, their approach lack in in-depth analysis for decision making and introducing a potential for bias. Aghili, M., and et al. [21] utilized a UNet architecture integrated with a transformer model to segment the hippocampus from brain MRI images. With 326 images, their method improved the mean segmentation value by 4.4% and achieved an 89% mean Jaccard value, though the poorly structured feature selection limited the model's learning potential. Another one study focused on 3D hippocampus segmentation using a combination of CNN and Vision Transformer models [22]. The results demonstrated high Dice coefficients across various datasets, including 90.6 for the Harp dataset, 90.3 for the Decathlon dataset (http://medicaldecathlon.com), and 87.8 for the Kulaga-Yoskovitz dataset (https://fcon_1000.projects.nitrc.org). Nevertheless, they acknowledged the need to test their approach on non-medical datasets for better generalization. Chen, J., and et al. [23] introduced TransUNet, a hybrid model combining Transformer and UNet structures for medical segmentation. Their experiments showed an improvement in Dice similarity from 77.86 to 83.46 when increasing image resolution from 224×224 to 512×512, with a notable 6.88% enhancement over existing models, although the model suffered from slower processing speeds. Cao, A., and et al. [24] proposed a three-dimensional CNN-MAMBA hybrid architecture for brain image segmentation. They reported a Dice similarity coefficient of 0.88383, volume similarity of 0.97076, and an average symmetric surface distance of 0.33604, validated using a Wilcoxon signed-rank test. Yet, the model's performance was constrained by its complexity and low parameter count. Recently, Zhu, H., and et al. [25] developed a Dual Branch Segmentation Network (DSNet) using a hybrid attention mechanism for hippocampus segmentation. Tested on the Kulaga-Yoskovitz dataset, their model outperformed others by 0.57% in Dice coefficient, although its effectiveness heavily depended on the quality and diversity of the dataset. The proposed model

This paper includes 4 sections, the section 2 discusses the data collection and evolution metrics, section 3 discusses the proposed method, section 4 explores the strength and weakness of the proposed method by demonstrating with experiment and the section 5 concludes the results.

2. Proposed Hybrid CNN Model (CNN-HM)

This methodology integrates both traditional feature extraction techniques such as edge, wavelet, statistical, and energy of each pixels with a deep learning based CNN architecture. The model aims to achieve robust and accurate hippocampus segmentation by combining handcrafted and learned features. CNN-HM includes 3 modules as given in Fig. 2, the first module collects the input and process with some filters. Nearly, twenty types of features used to represent the given images individually and pass the extracted feature values as a matrix to

International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

the next CNN module. Finally, the CNN outputs are fed into the UNET which able to understand the feature and train itself to detect the hc pixels. Finally, the net produces binary values; the values are reconstructed as images for the analysis.

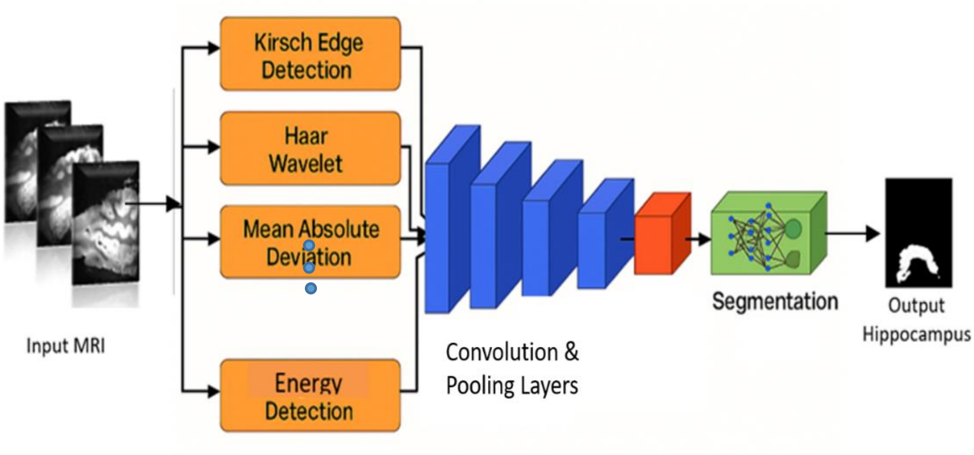


Figure 2. Architectural diagram of presented method

2.1. Feature Extraction

Features are the numerical or logical representation of image and it give the association with its neighbourhoods. The image intensity value and filtering values are taken out as numerical values and assigned to the net. This method utilizes nine feature extraction techniques that produces the twenty features for a pixel. The remaining section discusses the image feature extraction techniques. The visual results of the features are illustrated in Fig. 3

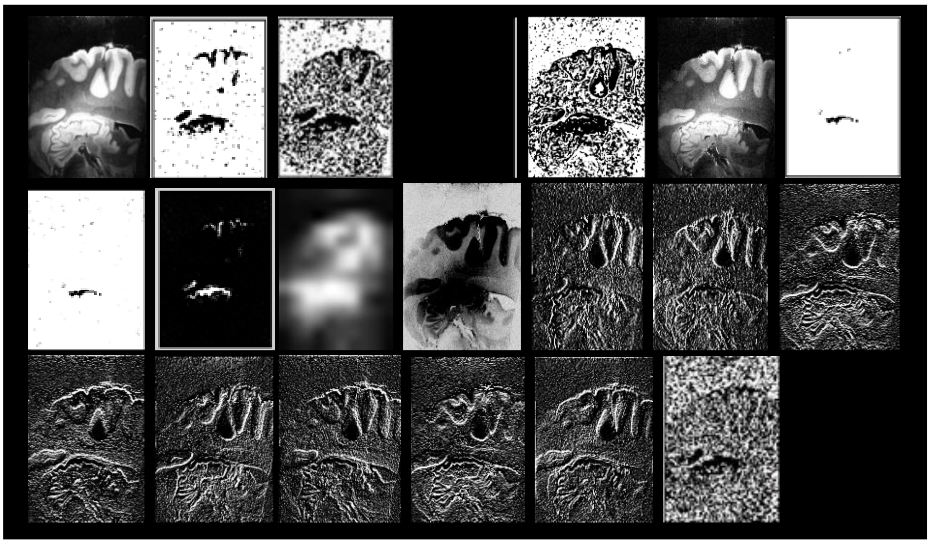


Figure 3: Features of brain MRI. Row 1: (from left to right) Original image, Skewness feature using mode, Kurtosis feature, Mean absolute percentage error, Pixel redundancy, Mean of 5×5 pixel neighbourhood, Variance among 5×5 pixel neighbourhood, Standard deviation among

Received: August 08, 2025

International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5×5 pixel neighbourhood, Mean difference and Haar wavelet feature respectively. Row 2: Energy level, Edge detection filtered images (North, South, East, West, North East, North West, South East, South West) and Standard variate respectively

Skewness: The measure of skewness verifies the symmetry in intensity distribution. The skewness is defined as,

$$f_1 = \frac{\sum_{i,j=1,b=1}^{m,n,k}\left(I_{i,j}-\bar{I}_b\right)^3/N}{S^3} \tag{1}$$

The skewness of a pixel is calculated using the 25 pixels (5×5) around each pixel (b). The centre pixel value is swapped with the computed value, denoted as f_1. The symbol $I_{i,j}$ refers to the gray value of the pixel in the array of the given image. $\bar{I}_b$ represents the mean intensity of the 5×5 neighbourhood. N represents the total count of pixels in the neighbourhood, array, and S denotes the standard deviation of the same neighbourhood. Instead of the mean ($\bar{I}_b$), the mode or median may use to calculate skewness mode and skewness median features. In this method, mode is used to extract the feature.

Kurtosis: It measures the frequency distribution of the grey values and it is defined as follows,

$$f_3 = \frac{\sum_{i,j=1,b=1}^{m,n,k}\left(I_{i,j}-\bar{I}_b\right)^4/N}{\left(\sum_{i,j=1,b=1}^{m,n,k}\left(I_{i,j}-\bar{I}_b\right)^2/N\right)^2} \tag{2}$$

The value of f_s is replacing the centre pixel value of the pixel array.

Mean Absolute Percentage Error (MAPE): Measures statistical dispersion by evaluating the average deviation from the mean. It is capturing texture irregularities and fine grained contrast features. In image processing, heterogeneity of a pixel among a block of pixels derived by using MAPE. The error to be measured by [26]

$$f_4 = \frac{1}{N}\sum_{i=j=1,b=1}^{m,n,k}\frac{\left(\overline{I_b}-I_{i,j}\right)}{I_b} \times 100 \tag{3}$$

The zero result ensures the homogeneity in pixel intensity distribution and high value represent the heterogeneity proportions with the mean of the block $\overline{I_b}$.

Pixel Redundancy: if a block of a pixels has more redundant grey value then it is ensuring that the pixel and its neighbours are belonging to a region. The redundancy count is taken as the grey value of its centre.

$$f_5 = |b_i| \tag{4}$$

Mean: *The mean value of the* 5×5 neighbourhood pixel represents the approximate value of the block. Each block centre updates with the mean value.

Variance: computing the variance among the 5×5 pixel array and replaces the centre pixel with variance.

Standard deviation: It is the usual calculation, computing by taking the square root of the variance and updating the value at the centre of the block.

International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Wavelet Feature: Usually, the discrete wavelet transform is using to separate the image into high frequency content and low frequency content. In this algorithm, the brain MRI image is decomposed into multiple levels to obtain the image size as 15 × 10. The low-pass image at this decomposition level highlights varying intensity values at each pixel, as noted by [26]. After this step, the image is resized back to its original dimensions through interpolation. As shown in Fig. 2, row 2, the pixels appear enlarged. Regions corresponding to the hippocampus and cerebrospinal fluid (CSF) show higher intensity the other areas appear with lower intensity, as described in [26].

Energy level: Identifies regions with high intensity changes to highlight structures of interest. Energy level highlights areas with significant anatomical changes such as the hippocampus as given in Fig. 3. The following mathematical framework helps to extract the energy level of each pixel as defined in [27],

$$f_{12} = \delta(\varphi)\left[\mu\, div\left(\frac{\nabla\varphi_{x,y}}{|\nabla\varphi_{x,y}|}\right) - (R(x,y) - C_1)^2 + (R(x,y) - C_2)^2\right] \quad (5)$$

where f represents the image, $\delta(.)$ known Heaviside function, the average intensity of the pixels more that 150 (>150) are denoted as C_1 and other (<150) are denoted as C_2 represent the average intensity of pixels above 150 (intensity >150) and below 150 (intensity <150) respectively. The contour is approximated by the Euler Lagrange's formulation. It is defined as,

$$div\left(\frac{\nabla\varphi}{|\nabla\varphi|}\right) = \frac{\varphi_{xx}\varphi_y^2 - 2\varphi_{xy}\varphi_x\varphi_y + \varphi_{yy}\varphi_x^2}{(\varphi_x + \varphi_y)^{3/2}} \quad (6)$$

Kirsch's Edge filters: Extracting edge features by detecting intensity gradients in multiple directions. This helps in highlighting hippocampal boundaries. The edge features are detecting intensity gradients in multiple eight directions. The edge filters are listed in Fig. 4 that are convolve with the image individually to obtain the edges from all directions individually [28].

Figure 4. Kirsch's Edge detection filters. North, South, West, East, South East, North East, South East and North West respectively from left to right.

$$\begin{bmatrix}-3 & -3 & 5\\ -3 & 0 & 5\\ -3 & -3 & 5\end{bmatrix}\begin{bmatrix}5 & -3 & -3\\ 5 & & -3\\ 5 & -3 & -3\end{bmatrix}\begin{bmatrix}5 & 5 & 5\\ -3 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}\begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & -3\\ 5 & 5 & 5\end{bmatrix}\begin{bmatrix}-3 & -3 & -3\\ 5 & 0 & -3\\ 5 & 5 & -3\end{bmatrix}$$

$$\begin{bmatrix}-3 & -3 & -3\\ -3 & 0 & 5\\ -3 & 5 & 5\end{bmatrix}\begin{bmatrix}5 & 5 & -3\\ 5 & 0 & -3\\ -3 & -3 & -3\end{bmatrix}\quad\begin{bmatrix}-3 & 5 & 5\\ -3 & 0 & 5\\ -3 & -3 & -3\end{bmatrix}$$

The convolution of image and each filter yields different fine edge information in various angles as provided in Fig.4. Each of these features is extracted in parallel and creating a multi-stream representation of the same input image.

Several features are obtained using various filters and mathematical tools. Hence, there is a chance of features dominations due to the scale difference. So normalizes each feature map to a consistent range [0, 1]. Stack the 2D maps along the channel axis to form a 3D tensor. This

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

tensor is now a multi-channel input as the feature counts. In this model, it builds 20 feature channels that is fed into convolution layers.

2.2. Convolution Layers

The CNN accepts multi-channel input and uses 1×1 convolution to reduce the dimensionality. Three convolution layers use to learn local spatial features from the enhanced inputs. The first layer includes 32 filters with 5 × 5 kernel size. The second and third layers use 64 and 128 filters respectively with the same kernel size. The filters extract patterns from the given multi-channel input relevant to the hippocampal region. Then the layer captures local dependencies, retains spatial hierarchy, and reduces the number of parameters compared to fully connected layers. The parameters used in the CNN model is summarized in Table I.

2.3. Pooling Layers:

Pooling identifies the significant features and reduces the spatial resolution and making the model computationally efficient and robust to spatial variations. The CNN-HM model uses the max pooling with the dimension of window is 2×2. Within the window, it extracts the maximum value and reduces the data size.

2.4. Segmentation Module

The feature maps are passed to a segmentation module, which likely includes fully connected layers. The model fabricates with 128 dense units which maps the features back to the spatial domain. This step predicts the pixel-wise labels, identifying whether each pixel belongs to the hippocampus region or not using the sigmoid activation function.

The final output is a binary mask where the hippocampus region is clearly segmented from the rest of the brain. This segmentation result can be visualized for further analysis in clinical applications such as Alzheimer’s diagnosis, hippocampal volume estimation, or brain disorder monitoring

3. Data Set and Validation Metrics

This work utilizes two types of datasets collected from clinics and the Alzheimer's disease Neuroimaging Initiative (ADNI) repository. The ADNI was established by Michael W. Weiner, MDon 2003. Each volume is converted into slices for training and testing purposes using the MRIcro tool. The dataset consists of axial data that is highlighting the hippocampus in its large structure. The images are not registered and are randomly selected for training.

Table I. Parameters used in the CNN

Component	**Parameter**	**Value / Setting**
Input Layer		
	Input Shape	(128, 128, 20) (stacked feature maps)
	Channels	20 (Kirsch, MAD, STD, Variance, Wavelet, Energy)

International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Conv Layer 1		
	Filters	32 (13 proposed filters and 19 default filters)
	Kernel Size	(5, 5)
	Activation Function	ReLU
	Batch Normalization	Yes
	Pooling	MaxPooling (2×2)
Conv Layer 2		
	Filters	64
	Kernel Size	(3, 3)
	Batch Norm	ReLU
	Activation	Yes
	Pooling	MaxPooling (2×2)
Conv Layer 3		
	Filters	128
	Kernel Size	(3, 3)
Fully Connected	Dense Units	128
	Dropout	0.5
Output Layer	Units	1 (binary output)
	Activation Function	Sigmoid (for segmentation mask/class)
Training	Loss Function	Binary Cross entropy
	Optimizer	Adam
	Learning Rate	0.001 (default)
	Evaluation Metrics	Accuracy

The efficiency of the model to be confirmed by quantitative measures. They are listed in the Table II. The column1 represents the name of the measure, column 2 represents the outcome of the measure, column 3 and column 4 explains the formula and its nature of the output value [29].

Table II. Evaluation metrics used for the result analysis

Evaluation Metrics	**Purpose**	**Formula**	**Result value**

International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Dice (DSC)	It measures the similarity between the segmented and ground truth [29]	$$DSC = 2.\frac{\lvert S \cap G \rvert}{\lvert S \rvert + \lvert G \rvert}$$	[0,1] '0' represents dissimilarity and '1' represents similarity
False Negative Dice (FND)	Identify the under prediction	$$FND = \frac{2\lvert \bar{S} \cap G \rvert}{\lvert S \rvert + \lvert G \rvert} \times 100$$	Less value is significant
Rand index (RI)	Measures similarity between pair of pixels in the predicted and the truth image	$$RI = \frac{x+y}{x+y+w+z}$$ Where x+y represents the agreement between the S and G. w+z represents the disagreement between the both sets.	Lower value represents the agreement between the both set. Higher value represents the disagreement between the both
Global consistency error (GCE)	Measures the extent to which segmentation can be viewed as a fine-tuning of the other	$$GCE = \frac{1}{n} min \left\{ \sum_i E(S,G,P_i) \, . \sum_i E(G,S,P_i) \right\}$$ E() refers the segmentation error	Ranges like [0,1]. '0' significant result.
Variation of information (VOI)	Computes the segmentation loss or gain	VOI(S, G) = H(S) + H(G) - 2I(S, G) H(s) entropy of segmented and H(G) entropy of ground truth	Lower value is significant
F1 Score	The F1 Score tells the association between the	$$\mathrm{F1} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$	[0,1], where 1 ensure the perfectness of the prediction

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

	precision and recall		and 0 indicates the worst prediction.

4. Results and Discussion

Initially, the experimental study performed over axial brain MRI data volumes which includes 50 patient images, and the human interpreted images (Ground truth). Out of the 50 volumes, 30 volumes are taken for training, 10 volumes taken for the validation and the remaining 10 volumes are taken for testing. The objective of the experiment is to resolve the queries about the finalization of the included features, confirm the network fitness, sure the perfectness of the proposed model, ensure best fitted activation function for hc segmentation and produce the alternative solution to achieve the aim of segmentation with best dice score.

4.1. Results of CNN-HM with Perceptron

The experiment commenced with the above stated feature images, convolution layer, pooling layer and perceptron. The intensity of hippocampus shows as the Cerebro-spinal fluid (CSF) in T2-weighted images.

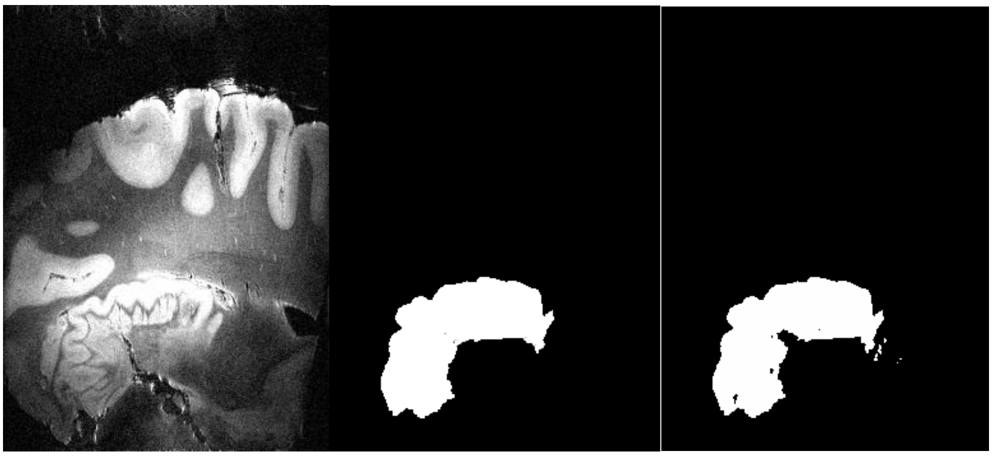


Figure 5. Original image, segmented and ground truth image respectively

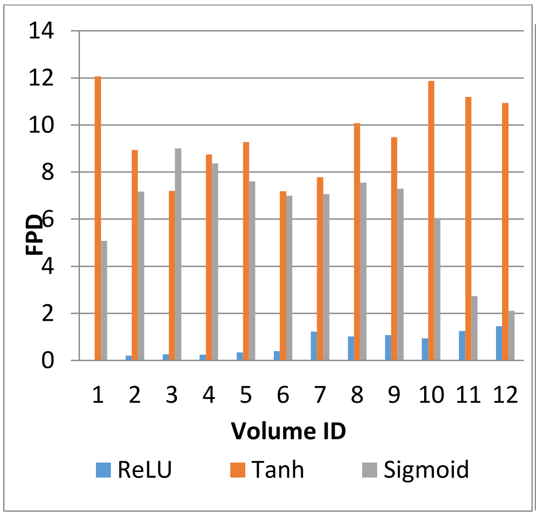
Next step to ensure the fitness of the net by its accuracy over the training set. Initially, five consecutive images took from each volume and provided at the moment of training which results 0.725 accuracy in 3 epochs. Strengthened the training set by selected several images from the volumes randomly for training. Subsequently, we increased the epochs up to 25. At this stage, the MSE was 0.0237 and the accuracy was 0.96. The results of the network are adjusted by increasing and deleting neurons. The similar experiment procedure is repeated for all activation function.

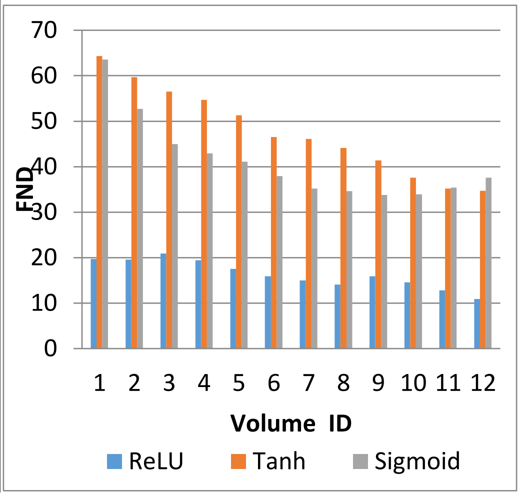
International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Usually, Dice similarity measure is popular to verify the segmentation process, this paper additionally measured the excess and underneath pixel extraction of the Sigmoid, Tanh and ReLU activation functions using FPD and FND. They are shown in Fig. 6 (a) and (b) respectively. In Fig. 6 (a) input volume ID is taken in x-axis and its FPD is taken in y-axis. It shows that the, Tanh function is always greater than 8 and also extracts more normal pixels along with hc pixels in all images compared to the sigmoid and ReLU functions and ReLU extracts very least number of other brain pixels. In Fig. 6(b), FND is taken in y-axis. It shows that the Tanh function under segments hc in all images compared to Tanh and Sigmoid activation functions; ReLU function shows very minimal and is also less than 20.





(a) (b)

Figure 6. (a) False Positive Dice (FPD), (b) False Negative Dice (FND)

The similarity of pairs of pixels in extracted region and the hc region in ground truth image was verified using RI. The RI was taken for the activation functions ReLU, Tanh and Sigmoid. They are compared using a line chart for better visual analysis and given in Fig. 7. In this figure, the image IDs are taken in x-axis and their RI are taken in y-axis. For all volumes, RI is closer to one and is also greater than 0.8 in ReLU function the other functions account greater than 0.6 and less than 0.8. It indicates that the ReLU function has yielded better results.

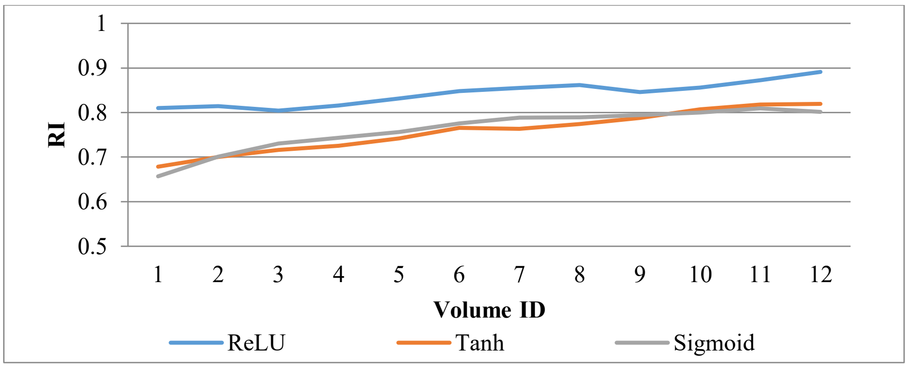


Figure 7. Similarity of pair of pixel (RI) detection of Activation functions ReLU, Tanh and Sigmoid in hp segmentation

Received: August 08, 2025

International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The segmentation error measure is necessary to analyse the performance of all methods. GCE is used to calculate the error measure the obtained results are shown in Fig. 8. In this figure, GCE is taken in y-axis and image ID is taken in x-axis. All activation functions provided low error measure and are less than 0.5, especially ReLU function confers less than 0.02 and is lowest GCE compared to the other activation functions. Tanh and Sigmoid functions provide almost equal CGE measure and ReLU function could not segment image 13 without error.

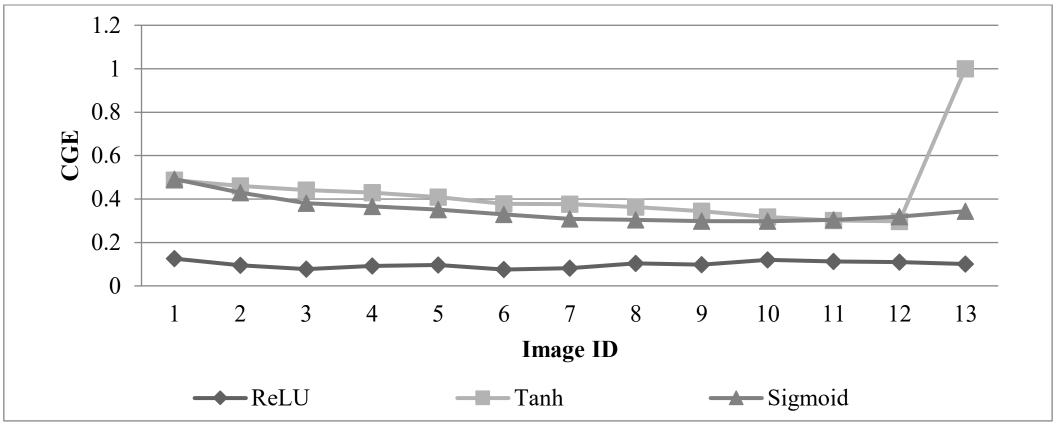


Figure 8. Segmentation error measure CGE for three activation functions in segmenting hc

Next, VOI between the extracted using various activation functions are demonstrated in Fig.9. In this figure, the obtained VOI is taken in y-axis and the image id is taken in x-axis. The VOI of ReLU for all images are less than 0.1 and in images 11 and 12 ReLU accounts zero error. The other two activation functions get VOI around 0.2. In the image 13 sigmoid function accounts nearly 1 but Tanh function gets less than 0.1 for the same image.

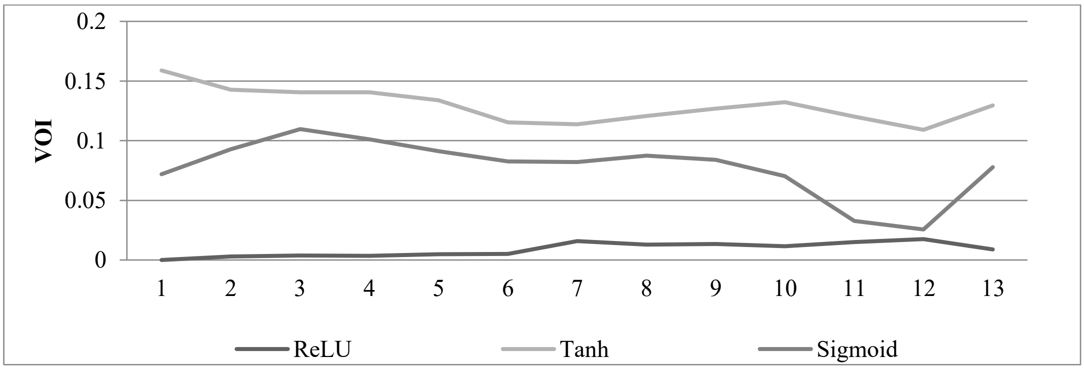


Figure 9. VOI of ReLU, Tanh and Sigmoid functions

The average values of all measures are tabulated in Table III. It details the averages performance measures of hc using Deep learning which uses several activation functions. Column1 shows the activation function names column 2 to column 6 show the mean value of the measures of Dice, FND, FPD, CGE, RI and VOI. The greater Dice and RI, and lower FPD, FND, GCE and VOI signify the better performance of segmentation. According to it, while comparing the activation function ReLU with the other functions, the ReLU function yields highest Dice and RI and lowest FPD, FND, GCE and VOI. Hence, we can conclude that the ReLU function works better than other functions in hc segmentation

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table III. Average values of Dice, FND, FPD, CGE, RI and VOI

	Dice	FND	FPD	CGE	RI	VOI
ReLU	**87.0394**	**0.702825**	**16.35623**	**0.098915**	**0.84225**	**0.00885**
Tanh	75.81428	9.56495	47.66859	0.431077	0.758133	0.129592
Sigmoid	76.22442	6.420142	41.13103	0.348069	0.762233	0.077667

The experimental results suggested that the proposed transformed learning strategy can boost the performances of hippocampus segmentation and ReLu activation function only reduces CGE error. Though, the dice score is not significant compared to the recent methodologies. Since, another set of experiments carried over the same set of input images. The results are described in the following section.

4.2. Results of CNN-HM with UNET

Next set of experiment includes the UNET instead of the simple perceptron in segmentation module which implemented with an EfficientNet-B0 encode. The training image features are taken as 3 batches and the model refines itself in 12 epochs to reduce the over-fitting effect. The entire processes are closely monitored with the training and validation volumes. The loss in each epoch is pictured in Fig. 10.

The losses are computed between the training and the testing which is given in the Fig. 11. The variation in training and validation loss over a series of epochs during the model training process. The chart pattern suggests that the model initially struggled with generalization but quickly learned from the data. The sudden drop and subsequent rise in both loss curves may also indicate early convergence and potential over-fitting.

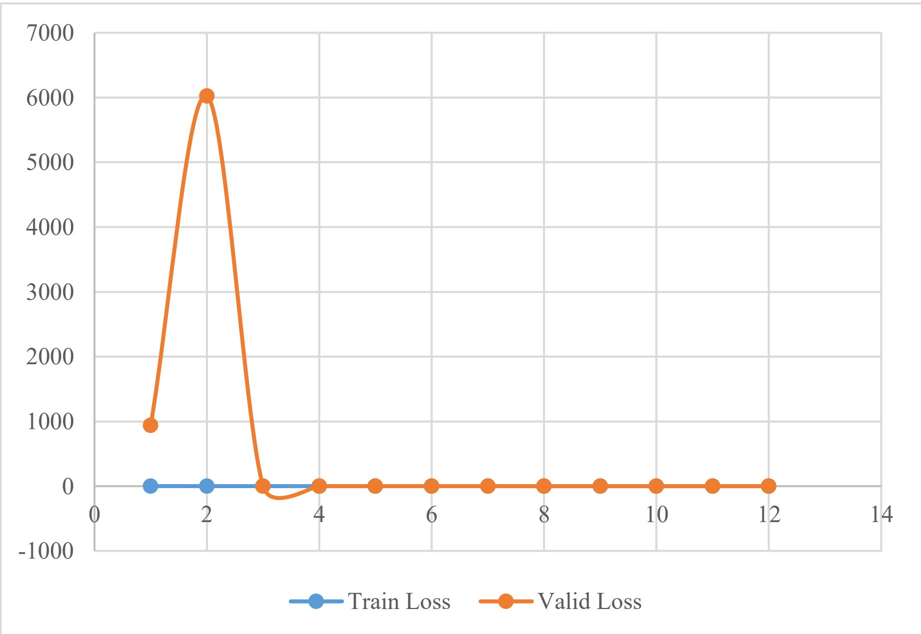
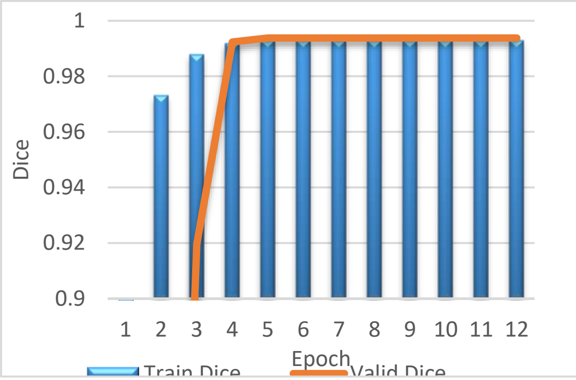
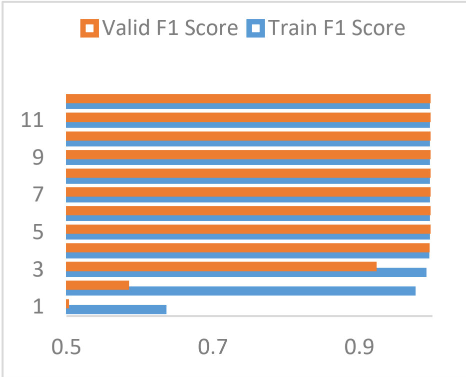


Figure 10. Training and testing loss

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)





(a)

(b)

Figure 11. Comparison of Dice and F1 score during training and validation. (a) Dice comparison, (b) F1 score comparison

The dice score at training time and testing time are depicted in Fig.11 (a). The Dice score during training starts slightly below 0.97 and increases steadily, reaching above 0.99 by epoch 4 and remains consistently high through to epoch 12. The Validation Dice shows a rapid increase from below 0.91 at epoch 2 to 0.99 by epoch 4, after it remains constant at a near-perfect score. The chart trend suggests that the model quickly achieves excellent segmentation performance on both training and validation data, with minimal over-fitting and strong generalization. The early plateau at a high Dice score may indicate that the model has reached its optimal performance early in training

Table IV. Comparison with similar methods

Method	**DSC Score**
QuickNat –Roy A.G., and et al. (2019)	0.84
Zhang J and et al. [9]	0.85
Hippodeep [18]	0.85
Roy et al., 2019	0.85
Random forest [12]	0.85
CNN-HM with Perceptron (Proposed Method)	0.87
Agili et al., 2023	0.89
Xiao, Z., [22]	0.97
CNN-HM with UNET (Proposed Method)	0.97
Dutta, P., and et al. [20]	0.99

The obtained F1 score is crafted in the Fig. 11 (b). The chart suggests that the model initially improved its classification ability rapidly over the first few epochs. The strong agreement

Received: August 08, 2025

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

between training and validation F1 scores in later epochs indicates good generalization and minimal over-fitting.

Additionally, the method verifies the present work with some existing works, the comparative results are illustrated in Table IV. The Table IV summarizes and compares the Dice Similarity Coefficient (DSC) scores achieved by various segmentation methods for medical image analysis, particularly focusing on hippocampal or brain structure segmentation. The table shows a clear progression in segmentation performance, with deep learning based and hybrid models. The CNN-HM with UNET significantly outperforming earlier traditional and shallow models. The proposed methods' contribution is significant in segmenting the hc. Specifically the result of proposed CNN-HM with UNET reflects the performance likely due to deeper or more refined network architectures.

5. Conclusion

This paper demonstrated twenty feature extraction techniques and transform the CNN to utilize the features. The proposed blended techniques evaluated with the two different deep learning models such as CNN-HM with Perceptron and CNN-HM with UNET. The models utilized handcrafted filters like edge detection and wavelets captures complementary features that may miss by CNN on small datasets. This helps the CNN focus on more informative regions, improving feature representation and learning efficiency. Wavelets and MAD bring out texture and frequency domain information since CNN extract hierarchical spatial features. Blending of the features allow better localization and boundary detection of complex and tiny structures like the hippocampus. These models use traditional filters that can be sensitive to noise in MRI scans. It will be the cause of false edges or distortions if not handled properly.

References:

[1] Annika Hansch, Jan H. Moltz, Benjamin Geisler, Christiane Engel, Jan Klein, Angelo Genghi, Jan Schreier, Tomasz Morgas, Benjamin Haas, Hippocampus Segmentation in CT using deep learning: impact of MR versus CT-based training contours. *Journal of Medical Imaging*, 7 , 6 (2020), 064001.

[2] W.B. Scoville, B. Milner, Loss of recent memory after bilateral hippocampal lesions, *The journal of neurophsychiatry and clinical neurosciences*, 12, 1 (2000), 103-a.

[3] A.Du, N. Schuff, D. Amend, M. Laakso, Y. Hsu, W. Jagust, K.Yaffer, J. Kramer, B. Reed, D. Norman, et al. Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, 71, 4 (2001), 441-47.

[4] D. Ataloglou, A. Dimou, D. Zarpalas, P. Daras, Fast and precise hippocampus segmentation through deep convolutional neural network ensembles and transfer learning, *Neuroinformatics*, (2019), 1–20.

[5] Manhua, Liu, Fan Li, Hao Yan, Kundong Wang, Yixin Ma, A multimodel deep convolutional neural network for automatic hippocampus segmentation and classification in alzheimer's disease, *Neuroimage*, 208 (2020), 116459.

Received: August 08, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[6] H. Wang, J.W. Suh, S.R. Das, J.B. Pluta, C. Craige, P.A. Yushkevich, (2013). Multi-atlas segmentation with joint label fusion, *IEEE transactions on pattern analysis and machine intelligence,* 35(3):611–623.

[7] J. Pipitone, M.T.M. Park, J. Winterburn, T.A. Lett, J.P. Lerch, J.C Pruessner, M. Lepage, A.N. Voineskos, M.M. Chakravarty, A.D.N. Initiative, et al., Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates, *Neuroimage,* 101 (2014), 494–512.

[8] Diedre Carmo, Bruna Silva, Alzheimer's disease neuro-imaging initiative, Clariss Yasuda, Leticia Rittner, Roberto Lotufo, Hippocampus segmentation on edpilepsy and alzheimer's disease studies with multiple convolutional neural networks, *Heiliyon,* 7 (2021), e06226.

[9] J. Zhang, J. Liang, H. Zhao, Local energy pattern for texture classification using self-adaptive quantization thresholds. *IEEE Trans Image Process,* 22, 1 (2013), 31–42.

[10] A.G. Roy, S. Conjeti, N. Navab, C. Wachinger, A.D.N. Initiative, et al., Quick-NAT: a fully convolutional network for quick and accurate segmentation of neu-roanatomy, *NeuroImage,* 186 (2019), 713–727.

[11] P. Cao, X. Liu, J. Yang, D. Zhao, M. Huang, J. Zhang, O. Zaiane, Nonlinearity -aware based dimensionality reduction and over-sampling for AD/MCI classification from MRI measures, *Comput. Biol. Med,* (2017), 91.

[12] L. Cao, L. Li, J. Zheng, X. Fan, Yin, F., Shen, H. & Zhang, J., (2018). Multi-task Neural Networks for Joint hippocampus Segmentation and Clinical Score Regression. *Multimedia Tools & Applications, , 1st77. Springer,* pp. 1(18).

[13] M. Liu, J. Zhang, D. Nie, P.T. Yap, D. Shen, A anatomical landmark based deep feature representation for MR images in brain disease diagnosis, *IEEE J. Biomed. Health Inform*, 22, 5 (2018), 1476–1485.

[14] J. Zhang, Y. Gao, Y. Gao, B. Munsell, D. Shen, Detecting anatomical landmarks for fast Alzheimer's disease diagnosis, *IEEE Trans. Med. Imaging,* 35 (2016), 2524–2533.

[15] J. Zhang, M. Liu, L. An, Y. Gao, D. Shen, ,. Alzheimer's disease diagnosis using landmark-based features from longitudinal structural MR images, *IEEE J. Biomed. Health Inform*, 21, 6 (2017), 1607–1616.

[16] Y. Chen, B. Shi, Z. Wang, P. Zhang, C.D. Smith, J. Liu, Hippocampus segmentation through multiviewensemble convnets, In: 2017 *IEEE 14th International Symposium on Biomedical* Imaging (ISBI 2017), IEEE (2017), 192–196.

[17] Z. Xie, D. Gillies, Near real-time hippocampus segmentation using patch-based canonical neural network, arXiv preprint arXiv:1807.05482 (2018).

[18] C. Wachinger, M. Reuter, T. Klein, Deepnat: Deep convolutional neural network for segmenting neuroanatomy, *NeuroImage,* 170 (2018), 434–445.

[19] B. Thyreau, Sato, H. Fukuda, Y. Taki, Segmentation of the hippocampus by trans-ferring algorithmic knowledge for large cohort processing, *Med. Image Anal.* 43 (2018), 214–228.

[20] P. Dutta, P. Upadhyay, M. De, R.G. Khalkar, Medical Image Analysis using Deep Convolutional Neural Networks: CNN Architectures and Transfer Learning, In: *2020*

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, (2020), 175-180, doi: 10.1109/ICICT48043.2020.9112469

[21] M. Aghili, et. al., Automated Hippocampus Segmentation and Volume Estimation Using a Transformer-based Deep Learning Architecture, (August, 2020), 1-24, DOI: https://doi.org/10.21203/rs.3.rs-3238001/v1

[22] Z. Xiao, et. al., Light3DHS: A lightweight 3D hippocampus segmentation method using multiscale convolution attention and vision transformer. *Neuroimage292*, 120608 (2024), 1-10. https://doi.org/10.1016/j.neuroimage.2024.120608

[23] J. Chen, et. al., Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306, (2021).

[24] A.Cao, Z. Li, J. Jomsky, A. F. Laine, J. Guo, MedSegMamba: 3D CNN-Mamba Hybrid Architecture for Brain Segmentation. arXiv preprint arXiv:2409.08307, (2024).

[25] H. Zhu, W. Cheng, K. Hu, et al., DSnet: a new dual-branch network for hippocampus subfield segmentation, Sci Rep, 14, 15 (2024), 317 https://doi.org/10.1038/s41598-024-66415-0.

[26] S. Karthigai Selvi, DWT based method to locate tumor from T2-w axial head scan, *international journal of engineering research and applications*, 9, 3 (2019), 63-70.

[27] T. Kalaiselvi, S. Karthigai Selvi, Energy Update Restricted Chan–Vese Model for Tumor Extraction from MRI of Human Head Scans, *International Journal of Computational Methods*, 15, 1 (2018), 1750081. Doi: 10.1142/S0219876217500815

[28] T. Kalaiselvi, S. Karthigai Selvi, Investigation of image processing techniques in mri based medical image analysis methods and validation metrics for brain tumor, Current Medical Imaging, 14, 4 (2017), 489-505.

[29] K. S. Sakthivelu, T. K. Raniprasanna, Modified Kirschs Method for Edge Based Image, *Recent Advances in Computer Science and Applications*, 1 (2015), 27-33.