Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020)
IEEE Xplore Part Number:CFP20K25-ART; ISBN:978-1-7281-4889-2

2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC) 978-1-7281-4889-2/20/$31.00 ©2020 IEEE 10.1109/ICCMC48092.2020.ICCMC-00079

Comparative Study of Deep Learning Models for Segmentation of Corpus Callosum

Siddharth Shrivastava
Dept. of Electronics & Communication
NIT, Raipur
Raipur, India
siddharthmanik16@gmail.com

Nitisha Singh
Dept. of Electronics & Communication
NIT, Raipur
Raipur, India
nitishasingh12390@gmail.com

Upasana Mishra
Dept. of Electronics & Communication
NIT, Raipur
Raipur, India
upmn24@gmail.com

Anjali Chandra
Dept. of Electronics & Communication
NIT, Raipur
Raipur, India
achandra.phd2017.etc@nitrr.ac.in

Shrish Verma
Dept. of Electronics & Communication
NIT, Raipur
Raipur, India
shrishverma@nitrr.ac.in

***Abstract*—A vital part of the Brain, which is responsible for the transmission of neural messages between the two hemispheres in the Brain is the Corpus Callosum. Many of the neurodegenerative diseases are related to the morphological properties of the Corpus Callosum. Therefore, its study and analysis become an essential part for the detection of such diseases. Examination of the Magnetic Resonance Images (MRI) through the mid-sagittal plane portrays their structure in the most distinguished manner. This paper carries out a comparative study of three deep learning models such as CE-Net, UNet++ & MultiResUNet for the segmentation of Corpus Callosum in the Brain MRI images using the dataset acquired from open source ABIDE platform. CE-Net gave the best dice similarity coefficient score of 0.9311, among all the three Deep Learning models. Thus, the CE-Net segmentation model can be further used for the classification of neurological disorders.**

***Index Terms*—Corpus Callosum, CE-Net, Deep Learning, Magnetic Resonance Imaging, MultiResUNet, UNet++.**

I. Introduction

One of the most important fields in Image Processing is Image segmentation [1] and is one of the fastest developing areas in terms of research. As a result, numerous image processing techniques and algorithms are there for image segmentation. Segmentation is the process in which an image gets divided into the foreground, i.e., the object and the background which are the constituent parts of the image.Mapping of each pixel of an image into their corresponding class labels is known as semantic segmentation [2]. Medical Imaging techniques use electromagnetic radiation for the diagnosis of abnormalities. No ionizing radiations of X-rays is use in this case. MRI is a technology based on imaging which is use to produce 3-D anatomical images. MRI images [3] examine internal body structures to diagnose brain functions, strokes, tumours and spinal cord injuries. The applied magnetic field is controlling an MRI scanner, the alignment of the proton (spin), but before reaching their original position, these protons get flip by applying an RF signal. It is use to detect abnormalities as the rate at which these protons return to their original position determines whether they are normal or abnormal. Different modalities of MRI images such as T1, T2 and flair are obtained by varying the sequence of RF pulse applied and collected. MRI scans have three different cross-sectional Views namely axial, sagittal and coronal. Analysis of superior and inferior portions are done through axial view refer to as a horizontal view. While the sagittal view use for visualizing left and right portions, the coronal view, also known as the vertical view is for anterior and posterior positions. Deep inside the brain , under the cortex, there is a bundle of nerve fibres acting as a connecting path between the two hemispheres, namely, left and right, which is Corpus Callosum. It is also the largest white matter [4] [5]. Each of these hemispheres is responsible for controlling the actions of the opposite half of the body. It mainly integrates the cognitive, motor and sensory functions between the cerebral cortex of both the halves. Diseases which affect the structure and biochemistry of the nervous system are called neurological disorders while the ones in which are caused due to the death of neurons leading to irremediable conditions are called neurodegenerative diseases. Alzheimer's and autism are one of the most common kinds of such diseases. Autism spectrum disorder (ASD) is a complex state of mind which leads to abnormalities in behaviour and communication. Alzheimer's is a condition of progressive memory loss. Segmentation of Corpus Callosum (CC) in the mid-sagittal plane (MSP) [6] plays a vital role in the diagnosis of various neurological diseases like Alzheimer's, dyslexia, autism as these diseases directly change the shape and size of CC. The MSP is use for analyzing the CC because the properties and the boundaries are well-defined. Tracing of CC in the MRI is a tedious and error-prone task as it requires an expert for their analysis, i.e. depends on the operator to a great extent. The task of segmentation of CC comes as a critical challenge in the field of medical image processing. The major issues include low contrast of MRI

scans and non-differentiating CC boundaries [7]. Hence, there is a need to come up with an automatic, robust and effective technique for segmentation. A technique based on Bayesian inference which used sparse representation as well as multi-atlas voting was proposed by Gilsoon park et al. [8]for the purpose of medical imaging. On comparing four different segmentation results, they concluded that Bayesian inference gave the best segmentation performance with mean dice index of 93.72% for OASIS dataset. Yue Li et al. demonstrated Adaptive main shift - automatic CC contour initialization - active geometric contour (AMS-ACI-GAC) model achieved accuracy of 0.95, sensitivity of 0.84, F1 score of 0.88 [9]. Another work propagated by Hamza Sharif et al. which used various machine learning (ML) algorithms such as linear discriminant analysis (LDA), support vector machine (SVM), random forest (RF), multilayer perceptron (MLP) and k nearest neighbour (KNN) and found LDA to be the best with an average accuracy of 55.93 [4]. A mechanized methodology for segmentation by localizing the anterior and posterior endpoints using point-matching was introduced by Wenpeng Gao et al., which achieved the endpoint localization error to be 0.85 mm [10]. Most of the proposed works till date have used ML algorithms for segmentation in the field of medical imaging. But the use of ML techniques in this field comes with the limitation of a high rate of manual dependency as the process of feature extraction is not automatic. This limitation is solved using Deep learning techniques [11] as they can recognize the new features automatically. Our proposed technique for semantically segmenting corpus callosum is shown in Fig.1

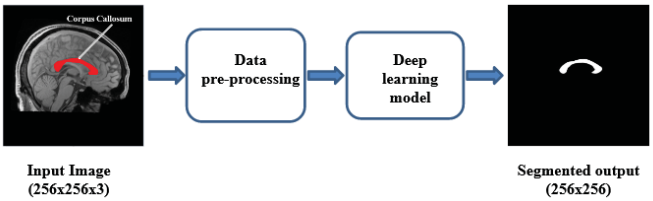

Fig. 1. Block diagram

II. Dataset

Generally, the collection and curation of the medical image dataset is one of the most strenuous and tedious jobs for the physicians as well as the researchers. The compilation of such a data which can be effectively used for the analysis comes as one of the major challenges due to various problems like the privacy issues of the patient and small number of input data for the rare diseases. However, such datasets have an upper hand over the genomics data or the pathology data as there are standard formats for such data.

The dataset that has been used is the Autism Brain Imaging Data Exchange (ABIDE). It is a standard open source dataset consisting of brain R-fMRI (resting state functional magnetic resonance imaging) data. It has samples of 1112 participants from 17 different international sites (for example: NYU, KKI and SDSU) which includes 539 autism and 573 healthy cases. The participants lie in the age range of 6.4 to 64 years [12].

III. Deep Learning Models

A. *UNet++*

UNet++ is another network from the Unet family which holds the Unet architecture as the backbone and is comparatively a more powerful architecture for the purpose of medical image segmentation. It is a deeply-supervised encoder-decoder network. A series of nested and dense skip pathways connect the encoder and the decoder sub-networks which help to reduce the semantic gap between their respective feature maps. This has been carried out with a belief of easier learning process by the optimizer when the feature maps from the two networks are semantically similar. The variation in the connectivity of the two sub-networks is brought by the re-designed skip pathways.

Unlike in U-Net where the feature maps of the encoder are directly received in the decoder [13], the UNet++ model as shown in Fig. 2 makes them undergo a nested convolution block. The pyramid level decides the number of convolution layers. The skip pathway between any two nodes will contain a dense convolution block. A concatenation layer will precede each of the convolution layer in it. This is used to fuse the output of the previous convolution layer of the same convolution block along with the corresponding output which has been up- sampled by the lower convolution block. Hence this brings down the semantic difference between the feature maps of encoder and corresponding decoder. Since a dense convolution block is being used along each skip pathway, therefore, all the previous feature maps accumulate at the current node [14].

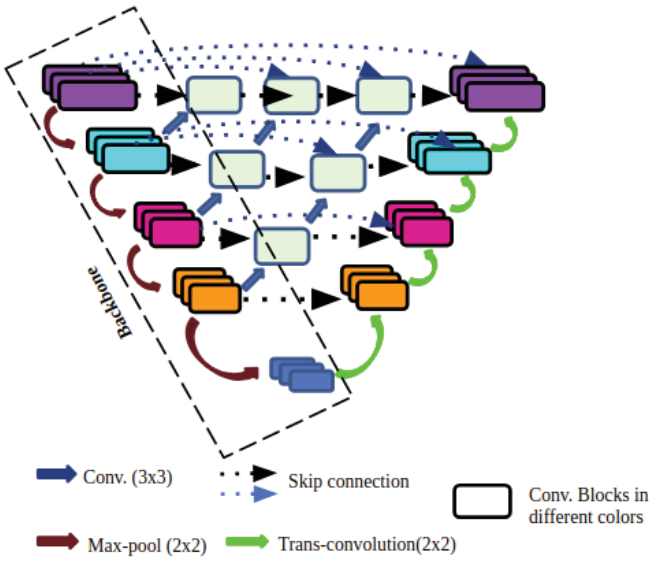

Fig. 2. Unet++ Architecture

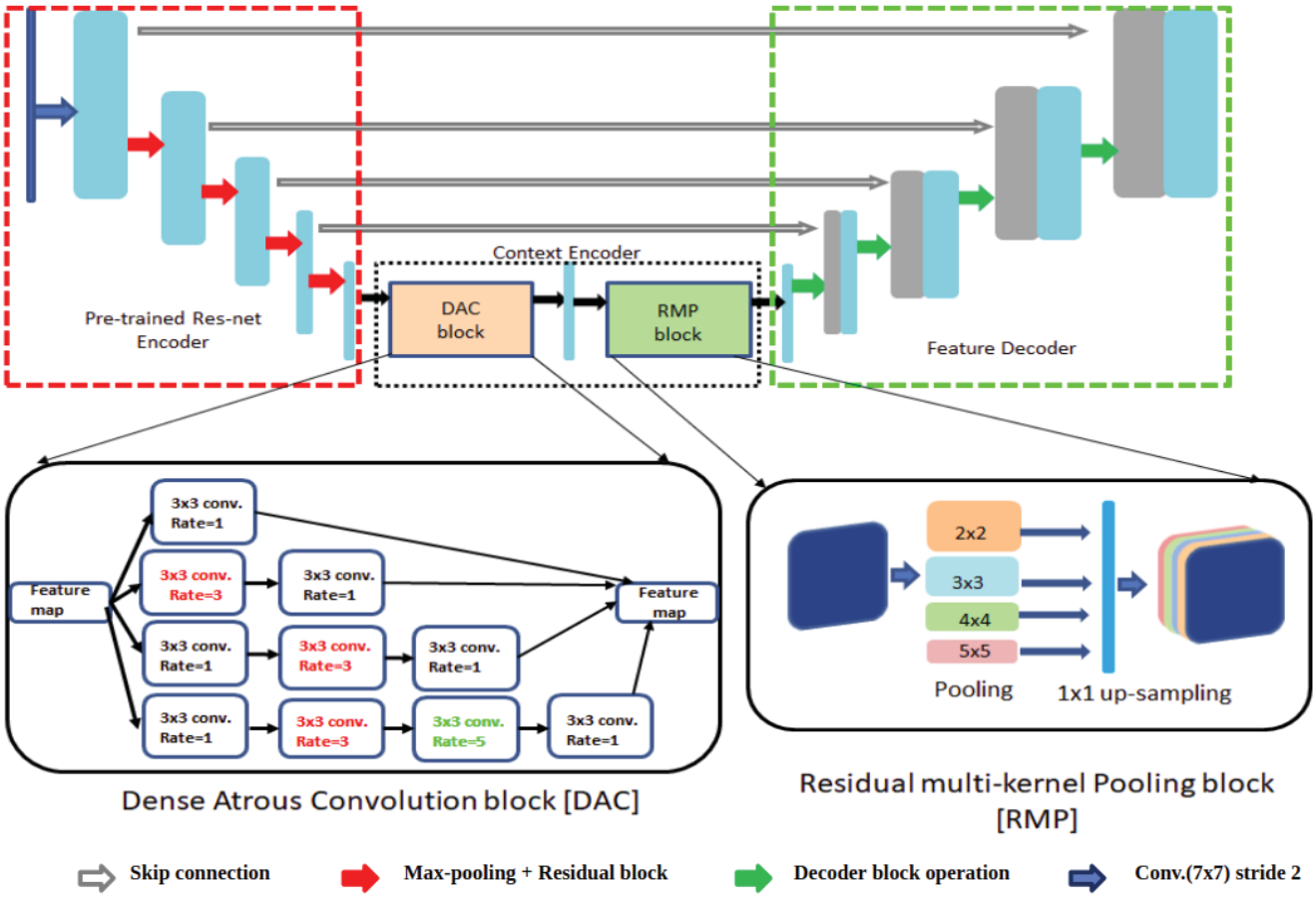

Fig. 3. CE-Net Architecture

B. CE-Net

In Deep Learning, information is generally lost due to the use of strided convolution or pooling operation. For segmenting images which are inherently a dense prediction task as every pixel in an image is being classified, there the usage of the above operations would significantly decrease the performance of the deep learning model. To overcome this challenge, CE-Net uses dilated convolutions along with residual pooling to preserve spatial information effectively. CE-Net comprises mainly of three parts as shown in Fig. 3, first being feature encoder then context extractor and decoder. Feature encoder is based on pre-trained Resnet34 architecture on imagenet dataset. It consists of one convolution block, batch normalization layer and a non-linearity followed by max-pooling operation. Then it uses starting four feature encoding layers of resnet without average pooling operations and densely connected networks. It solves the problem of vanishing gradient by introducing skip connections and which also helps in faster convergence rate. Context Extractor consists mainly of two blocks namely - Dense Atrous Convolution block (DAC) and Residual Multi-kernel Pooling block, which helps in extracting high-level semantic feature maps and preserves the spatial information among the pixels effectively. DAC block uses dilated convolution filters with a rate of 1, 3 and 5. This block is formed by adding five different layers with a receptive field of 1,3,7,9 and 19. The variability of the receptive field would help in generating more detailed feature map for large as well as small objects, as large receptive field convolution would take out more abstract feature knowledge of large objects in an image while small receptive fields would help detecting in small objects. Residual pooling block uses max pooling with a kernel size of 2, 3, 5 and 6 then utilizing 1x1 convolution to reduce the channel dimension followed by up-sampling layer through bilinear interpolation after every pooling layer, the feature maps so obtained are concatenated with the original feature map after dense atrous convolution. The decoder is based on vanilla Unet architecture using skip connections, also help in overcoming information loss. Up scaling is implemented by transposed convolution filter rather than interpolation operation since weights for transpose convolution can be learned and can be more task-dependent [15].

C. MultiResUnet

This model was proposed to overcome the shortcomings of Unet model i.e., the semantic gap between encoder and

decoder and undesirable results on dataset containing images of different scales. Each block of Unet model is replaced by MultiRes Block and the skip connections are incorporated with Res paths. The proposed model as shown in Fig. 4 contains simple Inception like block with different arrangements so as to fit in the spatial features from different context sizes. Instead of using 3×3, 5×5 and 7×7 convolutional filters in parallel, the bigger filters i.e. 5 x 5 and 7 x 7 which are also the more expensive filters, are factorized as a succession of 3 x 3 filters with the final feature map being a concatenation of the maps obtained from each of the filters in cascade form. These changes were done to aid in the reduction of the parameters and the memory requirement of the model. Each block also contains a residual connection followed by a 1 x 1 filter which conserves the dimensions.

The deviation in the semantic gap between the encoders and the corresponding decoder blocks is extenuated by passing the features of encoder through a series of convolutional layers along with residual connections making the learning easier. Except for the output layer, rests of the convolutional layers used are activated by the ReLU (Rectified Linear Unit) activation function and are batch normalized [16].

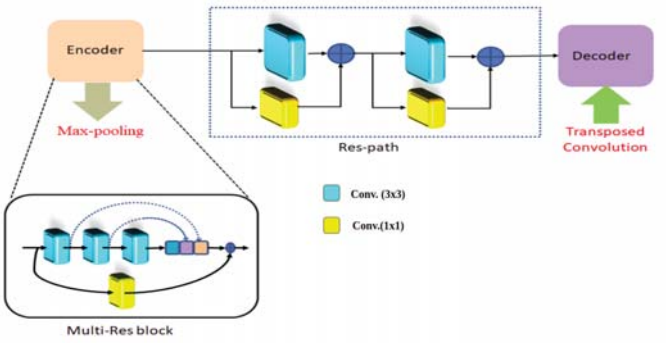

Fig. 4. MultiResUnet Architecture

IV. Experimental Results

A. Data Preprocessing

The original dataset consists of MRI scans comprising sagittal view slices. In order to obtain a generic model, data augmentation [17] was performed which included operations like random rotation by 90 degrees, flipping (both horizontal and vertical), random scaling by a factor sampled from uniform distribution between 0.9 to 1.1, random shifting and random Hue Saturation Value (HSV) [15]. These have been applied to both images as well as the masks. A total of 1112 data points were obtained after augmentation which was divided into 912 for training and 200 for validation.

B. Implementation Details

All the models have been trained on the ABIDE dataset using Pytorch as a Deep Learning framework for 30 epochs on Nvidia tesla p100 GPU with a capacity of 16 GB virtual RAM. The version of CUDA used is 10.0. For both validation and training, a batch size of 8 is taken Adam optimizer with a learning rate of 2 x 10^{-4} along with scheduler, which downscales the initial learning rate by a factor of 2 after every two patience. Input brain MRI size is considered as (256x256x3) while the mask size is (256 x 256). Binary Cross-Entropy (BCE) [18] loss is use as the objective function to optimize during the training of all the models.

C. Evaluation Criteria

The metrics which is use to evaluate ground truth and predicted image are as follows :

- Dice Similarity Coefficient(DSC) - It is a metric that is use to find similarity between two given binary images, as shown in equation (1).The dice similarity score is an evaluation metric which is same as the F1 score and is used to measure the similarity in terms of spatial overlap between two binary class images. It is preferred over Intersection over Union (IoU) for the classification of an unbalanced dataset.

$$DSC = \frac{2 * tp}{2 * tp + fp + fn} \tag{1}$$

- Accuracy (Acc) - It is the ratio of the total number of exact predictions to the number of overall predictions made, as shown in equation (2).

$$Acc = \frac{tp + tn}{tp + tn + fp + fn} \tag{2}$$

- Sensitivity (Sens) - It is the fraction of actual positive predictions to the total positive cases, as shown in equation (3).

$$Sens = \frac{tp}{tp + fn} \tag{3}$$

- Specificity (Spec) - It measures the proportion of actual negative predictions to the total negative instances, as shown in equation (4).

$$Spec = \frac{tn}{tn + fp} \tag{4}$$

Where tp,tn,fp,fn respectively refers to True Positive, True Negative, False Postive and False Negative.

- Binary Cross-Entropy (BCE) - It is a parameter used for analyzing the performance of binary classifiers. It is preferred over Dice loss as it overcomes the problem of exploding gradients, as shown in equation (5).

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \, log\left(p_i\right) + \left(1 - y_i\right) log\left(1 - p_i\right)\right) \tag{5}$$

where, N is batch size, y_i represents ground truth & p_i denotes predicted label.

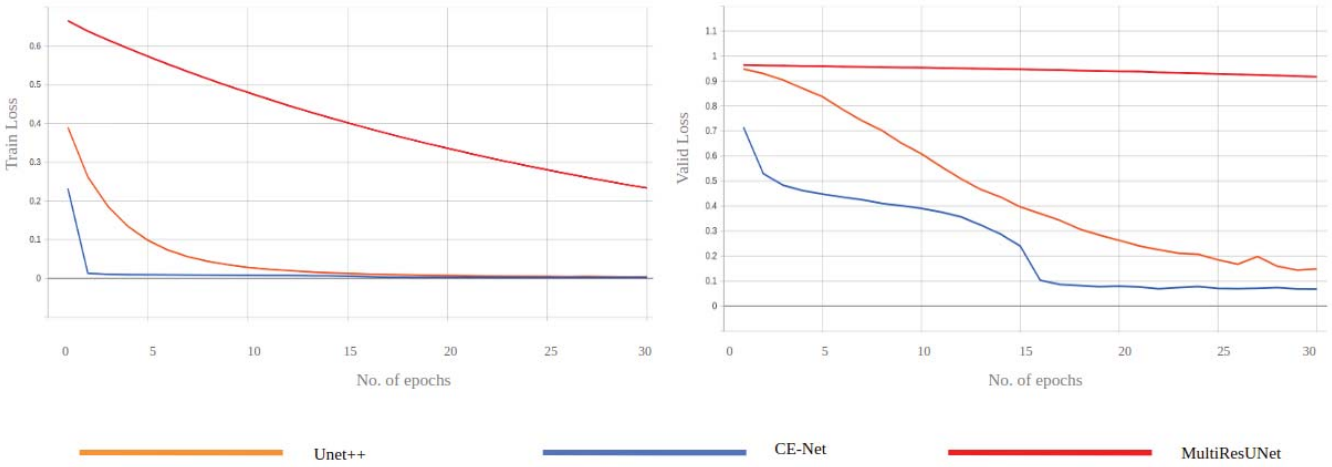

Fig. 5. loss curves

TABLE I
VALIDATION METRICS OF MODELS

Model	Dice Coefficient	Accuracy	Sensitivity	Specificity
UNet++	0.7168	0.9936	0.9178	0.9944
CE-Net	**0.9311**	**0.9989**	0.9346	**0.9994**
MultiResUNet	0.0802	0.7971	**0.9813**	0.7955

V. COMPARATIVE RESULTS

The outcomes reported in Table I infers us that the CE-Net model outperforms the rest of the Deep Learning models used for the task of semantic segmentation of corpus callosum. CE-Net makes use of transfer learning, atrous convolution as well as spatial pyramid polling layers which helped in faster and better convergence as compared to other models. Gradient vanishing can be one of the reasons for the underperformance of MultiResUnet. The validation result of all the model is shown in Fig. 6. The Training and validation loss curves are depicted in Fig. 5. Inferences on Training and Validation Curves are as follows:

- Training loss curves are steeper than validation loss curves for all the segmentation models and in due course, they represented good fit learning curves.
- CE-Net model outperforms other Deep Learning models in training and validation loss curves due to the inclusion of Pretrained ResNet34 weights in its encoder which are further improved via fine-tuning.
- Multi-ResUnet performance is found to be comparatively low, on account of multiple convolution filter weights in its residual path which lead to the problem of Vanishing Gradients
- Unet ++ performs fairly well on training and validation loss curves with nominal training parameters.

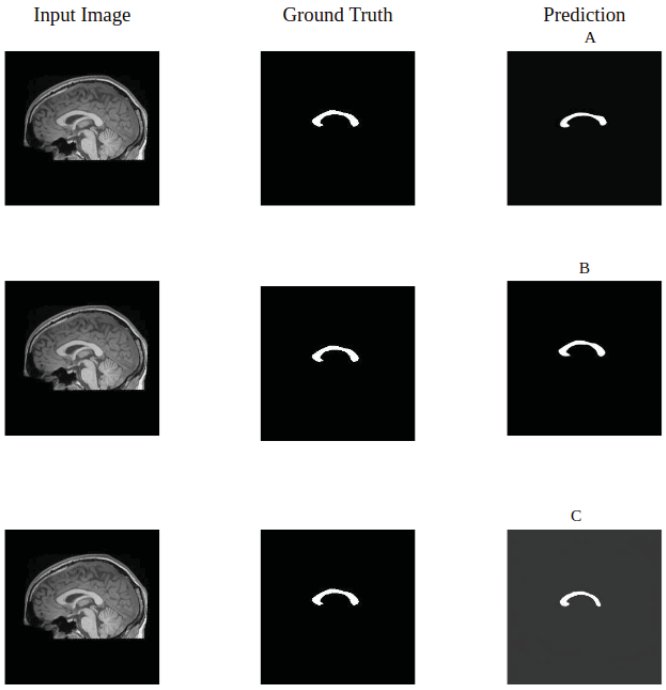

Fig. 6. A : UNet++, B : CE-Net, C : MultiResUnet

VI. CONCLUSION

This paper incorporates the implementation and evaluation of three different models for segmentation of Corpus Callosum, which is summarised in Table II. The parameters being used for analyzing the performence of these models are dice similarity coefficient, accuracy, sensitivity and specificity.
CE-Net was found to give the best results of all of them. It has a faster rate of convergence, has the highest accuracy, specificity and dice coefficient. It also comprises fewer number of trainable parameters as compare to other models.
Further, this model may be used for extracting regional

TABLE II
SUMMATION OF CE-NET, UNET++ & MULTIRESUNET

Architecture	Transfer Learning	Trainable Parameters	Salient Features
CE-Net	Fine-Tuning	38,969,176	Atrous Convolution & Spatial Pyramid Pooling
Multi-ResUnet	—	8,125,562	Revised Residual Path
UNet++	—	9,163,329	Series of Dense Nested Convolution

features which can be used for the diagnosis of various neurodegenerative diseases.

REFERENCES

[1] R. D. Deshmukh and C. Jadhav, "Study of different brain tumor mri image segmentation techniques," *International Journal of Science, Engineering and Computer Technology*, vol. 4, no. 4, p. 133, 2014.

[2] J. S. Sevak, A. D. Kapadia, J. B. Chavda, A. Shah, and M. Rahevar, "Survey on semantic image segmentation techniques," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE, 2017, pp. 306–313.

[3] S. J. He, X. Weng, Y. Yang, and W. Yan, "Mri brain images segmentation," in *IEEE APCCAS 2000. 2000 IEEE Asia-Pacific Conference on Circuits and Systems. Electronic Communication Systems.(Cat. No. 00EX394)*. IEEE, 2000, pp. 113–116.

[4] H. Sharif and R. A. Khan, "A novel framework for automatic detection of autism: A study on corpus callosum and intracranial brain volume," *arXiv preprint arXiv:1903.11323*, 2019.

[5] M. Pereira, G. Cover, S. Appenzeller, and L. Rittner, "Corpus callosum parcellation methods: a quantitative comparative study," in *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 10578. International Society for Optics and Photonics, 2018, p. 1057817.

[6] L. B. Hinkley, E. J. Marco, A. M. Findlay, S. Honma, R. J. Jeremy, Z. Strominger, P. Bukshpun, M. Wakahiro, W. S. Brown, L. K. Paul *et al.*, "The role of corpus callosum development in functional connectivity and cognitive processing," *PLoS One*, vol. 7, no. 8, p. e39804, 2012.

[7] A. Fitsiori, D. Nguyen, A. Karentzos, J. Delavelle, and M. Vargas, "The corpus callosum: white matter or terra incognita," *The British journal of radiology*, vol. 84, no. 997, pp. 5–18, 2011.

[8] G. Park, K. Kwak, S. W. Seo, and J.-M. Lee, "Automatic segmentation of corpus callosum in midsagittal based on bayesian inference consisting of sparse representation error and multi-atlas voting," *Frontiers in neuroscience*, vol. 12, p. 629, 2018.

[9] Y. Li, H. Wang, N. Ahmed, and M. Mandal, "Automated corpus callosum segmentation in midsagittal brain mr images." *ICTACT Journal on Image & Video Processing*, vol. 8, no. 1, 2017.

[10] W. Gao, X. Chen, Y. Fu, and M. Zhu, "Automatic extraction of the centerline of corpus callosum from segmented mid-sagittal mr images," *Computational and mathematical methods in medicine*, vol. 2018, 2018.

[11] M. Lai, "Deep learning for medical image segmentation," *arXiv preprint arXiv:1505.02000*, 2015.

[12] A. Lefebvre, A. Beggiato, T. Bourgeron, and R. Toro, "Neuroanatomical diversity of corpus callosum and brain volume in the autism brain imaging data exchange (abide) project," *bioRxiv*, p. 002691, 2014.

[13] A. Chandra, S. Verma, A. S. Raghuvanshi, N. D. Londhe, N. K. Bodhey, and K. Subham, "Corpus callosum segmentation from brain mri and its possible application in detection of diseases," in *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. IEEE, 2019, pp. 1–4.

[14] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 3–11.

[15] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, 2019.

[16] N. Ibtehaz and M. S. Rahman, "Multiresunet: rethinking the u-net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020.

[17] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.

[18] K. Janocha and W. M. Czarnecki, "On loss functions for deep neural networks in classification," *arXiv preprint arXiv:1702.05659*, 2017.