Chapter 4

Computer Based Segmentation of Cancerous Tissues in Biomedical Images using Enhanced Deep Learning Model

Highlights of the Chapter

- *The proposed method presents efficient deep learning model for segmentation of medical images.*
- *Depth wise separable convolution based model.*

Contribution of the chapter

The emergence of deep learning-based techniques provided the opportunity to analyse the pathological regions in the biomedical images with great precision. For segmentation of pathological tissues infected with cancer, a deep learning based network is presented in this chapter. It is a challenging task to delineate the regions affected by cancer because of the irregular shape and boundaries. Moreover, the accurate boundaries of such regions should also be preserved for the diagnosis. The proposed method aims to figure out pathological cancerous tissues with good accuracy and precision in brain tumour and skin cancer affected images. For this purpose, an automatic end to end trainable segmentation network is proposed which incorporates depthwise separable convolution with bottleneck connections. The proposed network exhibits the qualitative images statistics which are close to the expected results as indicated by the evaluation parameters. The obtained feature maps are transferred directly to the successive layers for obtaining the information loss of features. In the resultant segmented images, the details of minute curvatures are also well preserved as compared with the resultant segmented images of other networks. A significant improvement of 3% and 4% was observed in the mIoU and BF score of the results obtained for brain tumour datasets without retraining

the network. The network showed an improvement of 3% and 5 % for skin cancer datasets without retraining. The network produced exceedingly good results when tested on real-time MR dataset without retraining. Further, for performance comparison, the external validation was performed with the manual segmentation generated by experienced radiologists.

4.1 Introduction

The computer-aided biomedical image processing is showing rapid advancements in the detection and segmentation of the cancerous regions for the diagnosis of cancer in patients. The identification of cancerous region with accurate boundary details is of prime importance as any minute error can lead to a blunder in the treatment of this disease. The cells grow in an uncontrolled manner in the regions affected by cancer which can lead to the miseries of the patient. The accurate delineation of irregular boundaries and shape of the cancerous regions is a challenging task in such type of segmentation. So, for distinguishing the pathological tissues from the healthy tissues, many methods have been proposed in the past. The methods used for segmentation include manual, semi-automatic and automatic segmentation. In manual segmentation, a radiologist explores his/her knowledge of anatomy and physiology gained through training and experience for segmenting such regions. The semi-automatic methods involve human arbitration for the initialization of the method as handcrafted features are required for these approaches. The final results can be reached by improvement through received about the results. The deep learning-based methods for medical image segmentation have evolved over a period of time as the training of shallow models based machine learning, and probability theory is becoming difficult because of different statistical properties of different modalities. The deep learning-based methods are capable of producing and learning the image features in the form of feature maps. Several deep learning approaches are proposed in recent times such as VGGNet[66], GoogleNet [72], ResNet [90], Alexnet [65], SegNet [94], DenseNet [108], FCN [109] and U-Net [30]. This paradigm shift in the performance of

segmentation is attributable to the computational efficiency provided by the GPUs in the recent past. The significant improvement in computational efficiency has enabled the deep learning-based models to extract features on their own without any console intervention. The automatic extraction of features is because of the massive amount of training data which is utilized by the deep learning networks.

Olaf Ronneberger proposed a more efficient network termed as U-Net by training the FCN end to end, which yielded better results even when trained on very few images. Seg-Net is another architecture developed for pixel-wise semantic segmentation. FCN based networks architectures were paid rising attention in segmentation of biomedical images. Deep contour network (DCAN) proposed by Hao Chen et al. [110] generates the predictions of both objects and contours for the segmentation process. It separates touching or overlapping objects by taking the boundary details into consideration [89]. A multi-scale CNN model capable of extracting size information in images and combining the pixel information of different scale was proposed by Zhao [111]. Kamnitsas developed a dual path CNN incorporated with two parallel paths with the similar size of the receptive field and the other path received patches from subsample representation [35]. The deep learning networks yields more reliable segmentation results, as indicated by the segmentation evaluation parameters. For accurate detection and diagnosis of lung tumour, Wang Yuanyuan proposed CNN based method which provided accurate quantitative analysis to compensate for grey-scale insensitive effects [112]. Peter Naylor et al. developed a method by replacing binary annotation with a distance map to train the network to predict pixel by pixel segmentation [113].

The proposed end to end trainable network incorporates depthwise separable convolution [114] with bottleneck connections for segmentation of medical images. The network takes the base structure of the SegNet network. This computer-aided method aims to produce accurate boundary and pixel details of the cancerous regions for efficient diagnosis of cancer. This

model once trained, produces remarkably good results on other datasets of the same type. The network delivered best in class performance in segmenting the pathological tissues contaminated with cancer in terms of the evaluation metrics.

4.2 Methodology:

This section gives detail information about the architecture and working of the proposed encoder-decoder based method. For exploring the boundary and region related information in the cancerous images, depth wise separable convolution is incorporated in the well-known segmentation network. The end to end trainable deep segmentation framework is modified by the use of depth wise separable convolution with bottleneck connections for the effective segment the biomedical images. The depth wise separable convolution reduces the cost of computation as the count of multiplications reduces in case of this convolution. Moreover, a lesser number of parameters are required to train in case of depth wise separable convolution as compared to regular convolution. Thus it reduces the overfitting effects in the network. The image is transformed once, so the computation complexity is reduced. The convolution blocks and activations of seg-Net are replaced by the depth wise separable convolution with the linear bottlenecks to achieve the higher efficiency in segmentation. The proposed architecture is shown in the figure 4.1.

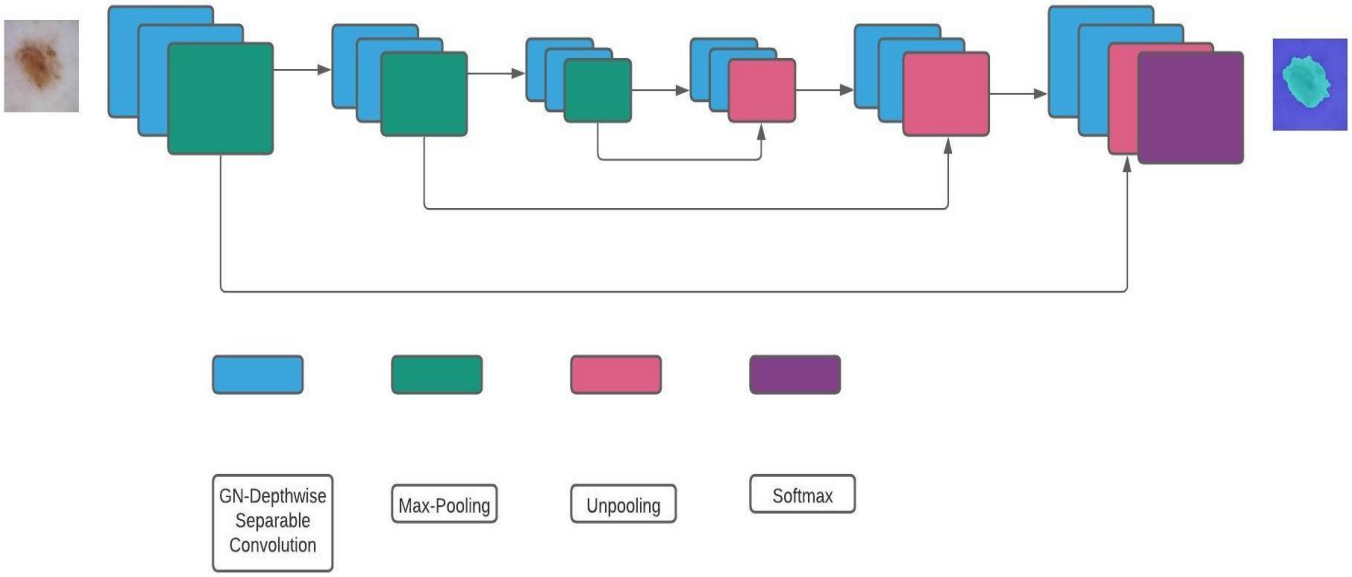

Figure 4.1: Architecture of the proposed network.

The basic building block for the proposed model is the depth wise separable convolution with the linear bottlenecks between the layers and the residual connections between the bottleneck blocks. The bottlenecks encode the in-between inputs and outputs. The transformation from lower-level units such as pixels to higher-level illustration such as image is performed by the middle layers. It consists of three convolution layers in the block. The layers are named as 1) Expansion Layer, 2) Depth wise convolutional Layer and, 3) Projection Layer. The group normalization technique is used instead of batch normalization. A single filter is applied to each input channel by the incorporation of depth wise convolution. For combining the outputs of depth wise convolution, a pointwise 1×1 convolution is applied. This combined process of filtering and combining is termed as depth wise separable convolution. This process is split into two parts, a different layer for filtering and a different layer for combining, as shown in the figure 4.2 when depth wise convolution is incorporated.

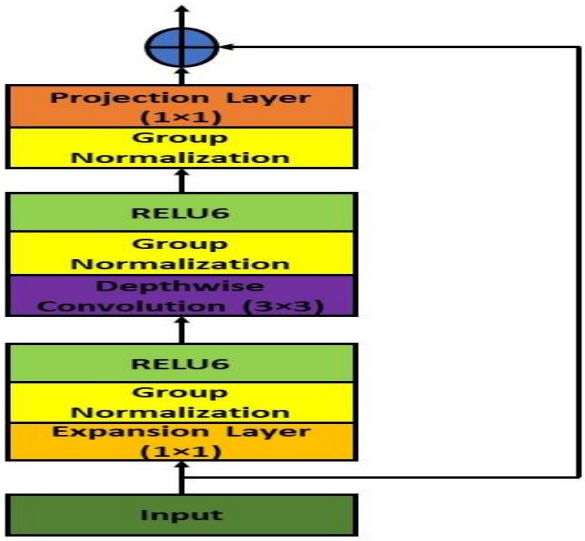

Figure 4.2: Depth wise Separable block with bottleneck connection.

Depth wise convolution with one filter per input channel (input depth) can be written as:

$$\hat{G}_{k.l,m} = \hat{K}_{i,j,m}.F_{k+i-1,l+j-1,m} \quad (4.1)$$

Where $\hat{K}$ is the depthwise convolutional kernel of size $D_K \times D_K \times M$ where the m^{th} filter in $\hat{K}$ is applied to the m^{th} channel in $\bar{F}$ to produce the mth channel of the filtered output feature map $\hat{G}$. The following equation gives the cost of convolution for depth wise separable convolution:

$$D_K.\ D_K.M.\ D_{F.}\ D_F + M.N.\ D_{F.}\ D_F \quad (4.2)$$

Equation 2 is the sum of depth wise convolution and 1×1 pointwise convolution. This 1×1 pointwise convolution layer causes the number of channels to be lower and is also known as Projection Layer and Bottleneck. The data with much higher dimensions is projected into a tensor with a lower count of dimensions using the projection layer. The initial layer in the figure 4.3 is the Expansion Layer. This layer is also a 1×1 convolutional layer. It expands the channel count in the data before it is passed to the depthwise convolution layer. Expansion Layer is just opposite to the projection layer as it has a higher number of output channels than the input channels, as shown in the figure 4.3. The expansion factor, which is a hyperparameter for depth wise convolution determines how much the data will be expanded. Figure 4.4 explains the whole process of cost calculation in depth wise separable convolution. In each layer, Group

Normalization [115] and activation function Relu6 [116] is incorporated. For low precision applications, RELU6 is more robust than normal RELU. It caps the upper ceiling of RELU to a value of 6. There is no activation function applied to the projection layer as it produces low-dimensional data, and using any non-linearity may destroy the useful information.

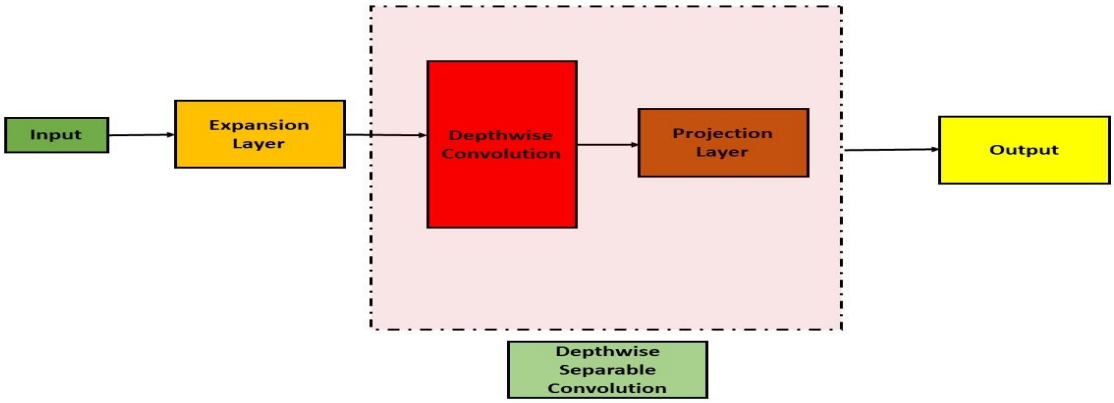

Figure 4.3: Functioning of layers in depthwise separable convolution block.

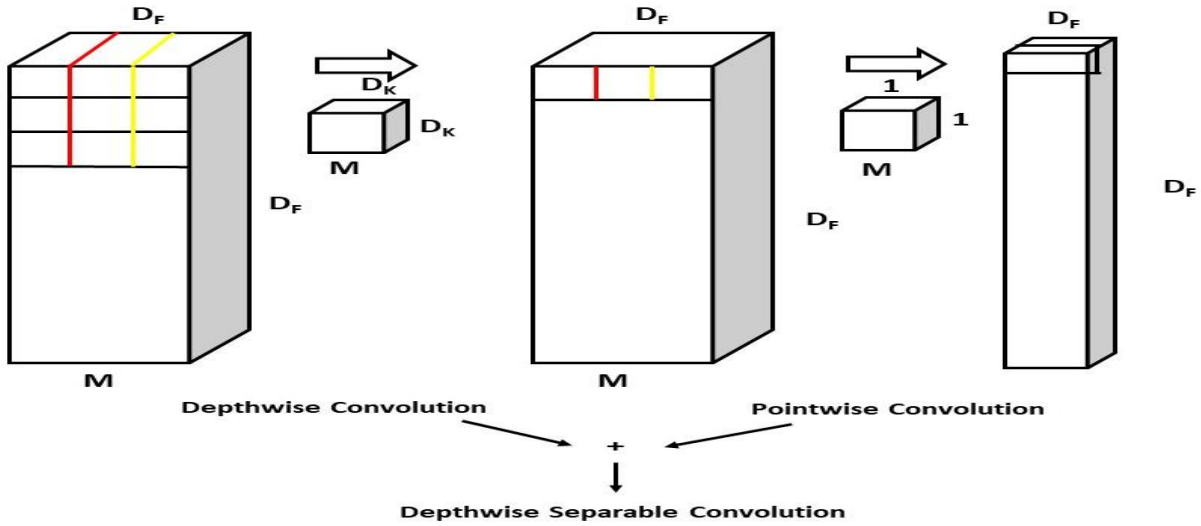

Figure 4.4: Depth wise Separable Convolution Functioning.

Group Normalization is for data normalization as it normalizes across the dimension of channels (Groups) instead of the batch dimension. The computation in group normalization is performed by dividing channels into groups. There is no dependence on batch sizes in case of group normalization, and it exhibits a stable accuracy over a wide batch size range [117]. The

process of group normalization requires the computation of mean and variance of the inputs over the specified groups of channels. The following equation describes group normalization:

$$\widehat{a_i} = \frac{a_i - \mu_g}{\sqrt{\sigma_g^2 + \varepsilon}} \quad (4.3)$$

Where a_i is the input, μ_g is the mean and is the variance. ε is the property to improve numerical stability for the small group variance. A max-pooling operator is used later with window size 2 × 2 and stride 2 with a non-overlapping window. The data is fed to the softmax layer after decoding. The softmax function is applied to the input data, which turns the raw score, i.e. logits into probability which sums to 1. It acts as a multiclass sigmoid function. The softmax function is used for the determination of probability computed once of multiclass [100]. The following equation represents the softmax function:

$$Y_r(x) = \frac{\exp(a_r(x))}{\sum_{j=1}^{k} \exp\left(a_j(x)\right)} \quad (4.4)$$

Where r signifies the class $0 \leq \mathrm{Yr} \leq 1$ and $\sum_{j=1}^{k} Y_j = 1$. The inputs and outputs of the layer are denoted by x and Y respectively. At last, each pixel is associated with a categorical label in the image by the pixel classification layer. For optimisation of network parameters, the cross-entropy loss function is used in the network [102]. The feature maps generated by the depthwise separable convolution have more information content regarding the object and boundary details. This is because in case of regular convolution an image is transformed m times where m denotes the number of channels while in case of depthwise separable convolution the image is transformed only once and elongated to m channels. In this process, the computational power is saved to a large extent. With each convolution, an extra multiplication is performed, and the boundary detail is lost during every such operation. Depth wise separable convolution eliminates the issue of blurring boundary by reducing the number of multiplications. The indices of feature maps with maximum values are stored and passed on to the decoder for

decoding purpose as every encoder corresponds to a decoder. For training all the networks in the study ADAM optimizer [118] [119] was used with the learning of 0.001. The cross-entropy loss function was used for training on the system having 8 GB GPU and 16 GB RAM.

4.2.2 External clinical validation of the segmented result:

The proposed model was trained on datasets of two types, i.e. brain tumour and skin cancer. For assessment of the effectiveness, the network trained once was tested on the corresponding datasets obtained from the distinct environment. The manual segmentation after the careful visual inspection of the dataset was performed by a senior radiologist. The region of segmentation included the lineament of the most solid components. The areas of necrosis cystic change and calcification were included within the segmented regions. The segmented results hence obtained were compared to the various automatically segmented images, including that obtained from our proposed network. The proposed method was compared with the two established deep learning methods for the performance evaluation.

4.2.3 Dataset acquisition

The segmentation work was carried out using five datasets. These datasets are as follows:

1. Dataset obtained from 'The Cancer Imaging Archive'. This dataset is sponsored by 'National Cancer Institute', and the images correspond to the TCGA lower-grade glioma collection with at least fluid-attenuated inversion recovery (FLAIR) sequence.
2. Dataset obtained from Figshare website. This dataset consists of slices of brain contrast-enhanced MRI (CE-MRI) with a large slice gap. The dataset of brain T1-weighted CE-MRI images consists of 3064 slices from 233 patients which is publicly available.
3. The dataset is obtained from the hospital of Institute of Medical Science, Banaras Hindu University. It contains 180 slices of brain tumour MR images. The corresponding masks were prepared in the School of Biomedical Engineering, IIT (BHU).

4. The third and fourth datasets are obtained from International Skin Imaging Collaboration (ISIC) challenge 2016 and 2018. This dataset consists of dermoscopic images of benign and malignant skin lesions.

4.3 Results:

The results obtained by the networks in the study are presented in this section. The networks were trained on one dataset and tested on some part of the same dataset as well as other datasets.

4.3.1 Results obtained for Brain Tumour Datasets:

The networks were trained on the Kaggle dataset and tested on Kaggle as well as on the Figshare dataset. Moreover, for the case of a brain tumour, the networks were tested on the real MRI data obtained from IMS (BHU) dataset. Figure 4.6 and Table 4-1 show the results obtained by testing the networks on Kaggle dataset while they were trained on the same dataset. The results attained when the networks were trained on Kaggle dataset and tested on figshare dataset are summarized in Figure 4.7 and Table 4-2.

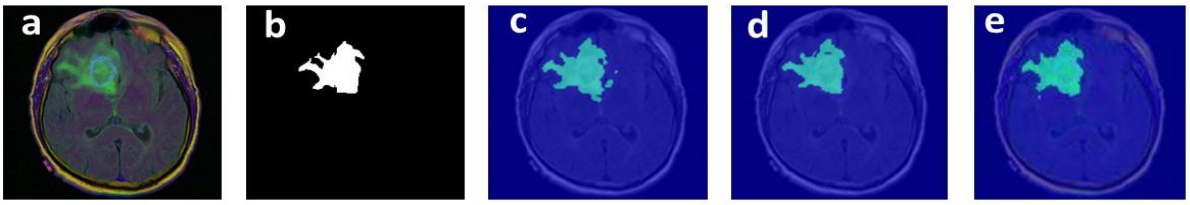

Figure 4.5: Results obtained when the network was trained and tested on kaggle dataset, (a) original image, (b) corresponding mask, (c) seg-net output, (d) u-net output, (e) proposed network output.

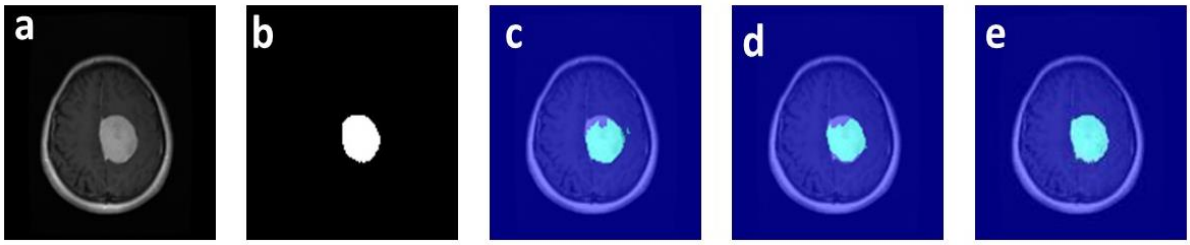

Figure 4.6: Results obtained when the network was trained on kaggle dataset and tested on Figshare dataset, (a) original image, (b) corresponding mask, (c) seg-net output, (d) u-net output, (e) proposed network output.

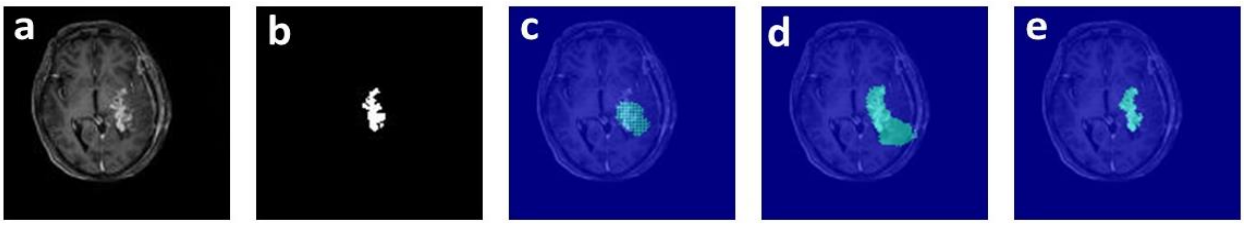

Figure 4.7: Results obtained when the network was trained on kaggle dataset and tested on IMS-BHU dataset, (a) original image, (b) corresponding mask, (c) seg-net output, (d) u-net output, (e) proposed network output.

Table 4-5: Comparison of evaluation metrics when the network was trained and tested on Kaggle dataset.

Networks	**Global Accuracy**	**Mean Accuracy**	**mIoU**	**Mean BF Score**
Seg-net	0.99772	0.9751	0.95387	0.917
U-Net	0.99818	0.97104	0.96222	0.92133
Proposed Network	0.99934	0.99438	0.98981	0.97616

Table 4-6: Comparison of evaluation metrics when the network was trained on Kaggle dataset and tested on Figshare dataset.

Networks	**Global Accuracy**	**Mean Accuracy**	**mIoU**	**Mean BF Score**
Seg-net	0.99664	0.96193	0.93318	0.88271
U-Net	0.99679	0.9373	0.94255	0.89674
Proposed Network	0.99727	0.98285	0.98857	0.97785

Table 4-7: Comparison of evaluation metrics when the network was trained on Kaggle dataset and tested on IMS-BHU dataset.

Networks	**Global Accuracy**	**Mean Accuracy**	**mIoU**	**Mean BF Score**
Seg-net	0.98654	0.65368	0.61026	0.71065
U-Net	0.99535	0.85571	0.82514	0.86229
Proposed Network	0.99612	0.92059	0.86056	0.90218

The real-time MRI data of brain tumour obtained from IMS (BHU) was also evaluated on the networks under study. Figure 4.8 and Table 4-3 give the results obtained by testing the networks on this dataset while networks were trained on Kaggle datasets.

4.3.2 Results obtained for Skin Cancer Datasets:

For segmentation of regions affected by skin cancer, the networks were trained on ISIC 2016 skin lesion dataset and tested on the same as well as on ISIC 2018 skin lesion dataset. Figure 4.9 and Table 4-4 show the results obtained when the networks were tested on ISIC 2016 skin lesion dataset. The results obtained when the networks were tested on ISIC 2018 skin lesion dataset are given in Figure 4.10 and Table 4-5. In both cases, the networks were trained on ISIC 2016 skin lesion dataset.

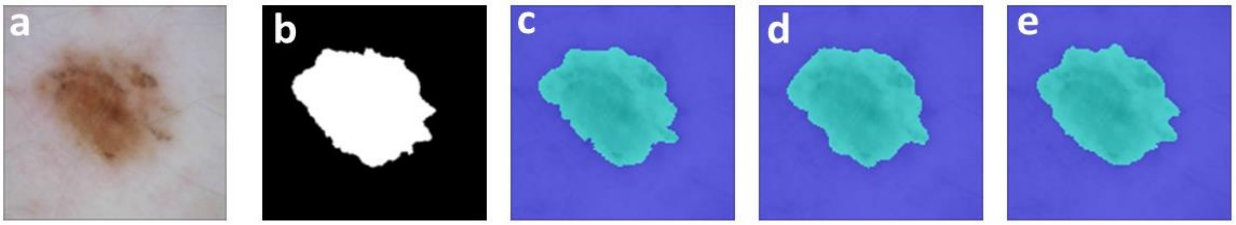

Figure 4.8: Results obtained when the network was trained and tested on ISIC 2016 dataset, (a) original image, (b) corresponding mask, (c) seg-net output, (d) u-net output, (e) proposed network output.

Table 4-4: Comparison of evaluation metrics when the network was trained and tested on ISIC 2016 dataset.

Networks	Global Accuracy	Mean Accuracy	mIoU	Mean BF Score
Seg-net	0.98237	0.97051	0.93557	0.87269
U-Net	0.99071	0.99098	0.95675	0.87503
Proposed Network	0.99361	0.99711	0.98995	0.91913

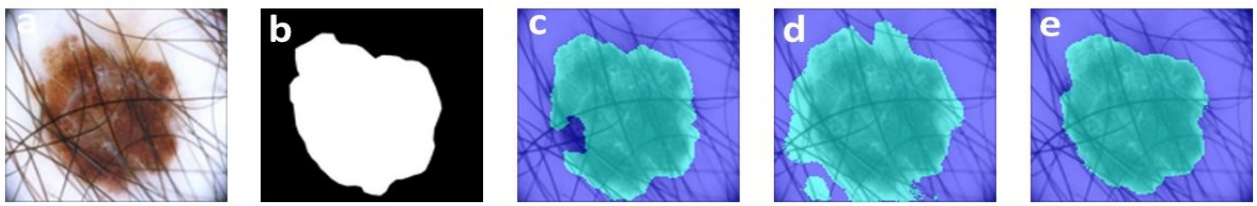

Figure 4.1: Results obtained when the network was trained on ISCI 2016 dataset and tested on ISCIC 018 dataset, (a) original image, (b) corresponding mask, (c) seg-net output, (d) u-net output, (e) proposed network output.

Table 4-5: Comparison of evaluation metrics when the network was trained on ISIC 2016 dataset and tested on ISIC 2018 dataset

Networks	**Global Accuracy**	**Mean Accuracy**	**mIoU**	**Mean BF Score**
Seg-net	0.97725	0.96426	0.91315	0.81088
U-Net	0.97551	0.9732	0.93393	0.86372
Proposed Network	0.98941	0.98772	0.96127	0.90877

4.4 Discussion

The accurate segmentation of regions affected by cancer in the biomedical images of various modalities is quite necessary. The detection of tumour or skin lesion with accurate boundary details can help the medical practitioner to efficiently diagnose cancer. So, for segmenting the regions affected by cancer, we presented a deep learning-based model which yields better performance as compared with other methods [120]. The model was tested on the datasets of brain tumour and skin lesion. The results obtained were quite remarkable as without retraining the network, it gave the best performance. Moreover, for brain tumour, the dataset was tested on real-time MRI data and produced exceedingly good results.

4.4.1 Quantitate Analysis:

The network proposed in this study not only efficiently delineates the cancerous regions in brain tumour and skin cancer images but also preserves the boundary details. The network yields improvement of above 3% in mIoU score for Kaggle dataset and figshare dataset respectively when the network was trained on Kaggle dataset and tested on figshare dataset. For skin lesion datasets the network produces results with improved mIoU score of 3.5% and 3% respectively when the network was tested on ISIC 2016 and ISIC 2018 dataset (trained on ISIC 2016 dataset). The improved mIoU score depicts how well the segmented results overlap with the ground-truth data, i.e. how close the obtained result is to the expected output [121]. Thus, it can be inferred from the boosted mIoU score that the results obtained by our proposed

model are closer to the expected results and they are more suitable for medical applications as shown in figure 4.9 and table 4-4. The BF score matches the class boundary with the expected ground-truth images. It analyses precision and recall values using distance error tolerance for deciding the matching of the boundary point. The proposed network exhibits the improvement of about 6% and 4.5% in BF score for Kaggle and figshare datasets respectively. Moreover, it shows an improvement in BF score of about 5% for ISIC skin datasets when the network was trained on ISIC 2016 dataset and tested on 20% part of ISIC 2016 dataset as well as on ISIC 2018 dataset. For brain tumour segmentation, we used an additional dataset obtained from Institute of Medical Sciences, BHU on which all the networks in the study were tested without retraining of network. A remarkable improvement of about 4.2% was obtained in mIoU score while BF score improved by 4.6% as shown in table 5-3. This good improvement in the results obtained by the network on real-time data shows the remarkable versatility of our proposed model.

4.4.2 Performance Analysis:

The prominent figure of merit for the proposed network is the ability to produce exceedingly good results without retraining the network. Training the network, again and again, is a time-consuming and tedious task, so if the network is capable of giving good results without retraining, it can be seen as an added advantage. For analysis of segmentation results, the combination of mIoU and mean BF score provides a better judgment for the output. Both these metrics obtains the highest score for all the datasets, which depicts better segmentation capability of the proposed network [122]. The proposed network performs best on the public as well as real-time datasets [122]. The figure 4.10 can be analysed for the results obtained for such dataset. In the case of images analysing the cancerous regions, the minute details are of great importance as a small mistake can lead to the blunder in the process of treatment. The number of correctly classified pixels is quite larger in case of our proposed network as depicted

by the higher values of global and mean accuracies [123] (shown in Tables 4-1 to 4-5). The network is incorporated using depthwise separable convolution and clipped RELU (ceiling fixed at 6). The clipped RELU caps the upper ceiling of the RELU and prevents the over-saturation of the network [124]. When the regular convolution and max-pooling is used in the network, the network focuses on learning the detailed segmentation characteristics. In this combined operation the network tends to lose the spatial resolution of feature maps. The spatial resolution is important to locate the contour of the object to be segmented. The depthwise separable convolution reduces the number of convolution multiplications by a large margin. By the reduction in multiplications, the computation complexity is decreased as well as the spatial resolution is preserved because every convolution operation tends to learn more about the object but losses the curvature details [74]. The bottleneck connections of the depthwise separable blocks enrich the network with the capabilities of Resnet as the network updates better in the presence of such connections. The layers near the centre update better, and the problem of vanishing/exploding gradient is minimised by a large margin [96] [125].

4.4.3 Qualitative Analysis:

The qualitative assessment can be made by visual inspection of the results obtained by the networks in the study. The visual inspection of the figures 4.5 - figure 4.10 shows the better-segmented boundary and more similarity of the results with the ground-truth in the output obtained by the proposed model [126]. The delineation of the boundary contour is much finer in case of the proposed network, which is quite a necessary improvement achieved by the proposed network. Moreover, the similarity of the results with the contours and curvatures of the ground-truth indicates the suitability of the results for the specified purpose. The results obtained by the proposed network were clinically more pertinent as the results accurately included pathological tissues and excluded the normal brain tissues projecting within the areas of complex tumoral anatomy [127]. Hence, the results obtained by the proposed model have

more information content as compared with other networks in the study, which is desirable for biomedical applications.

4.5 Conclusion:

The network proposed in this chapter performed the segmentation task exceedingly well in different domains of medical imaging. The network eliminated the need for retraining by yielding the best segmentation results when tested on images of other datasets once trained on the different dataset. For biomedical images, the region of interest is of utmost importance, especially in case of cancer tissues and the proposed network accurately delineated the region of interest for medical diagnosis. The proposed network effectively segmented the brain tumour as well as the skin cancer regions. Moreover, the network performed notably well when the real-time brain tumour dataset was tested on the trained network. The class accuracy of brain tumour segmentation was best when tested on real-time MRI. The minute curvature and boundary details are well preserved in the results obtained by the proposed model. This computer-based automatic segmentation method can be helpful for saving the precious time of the doctors and the burden on the medical system.