Medical image segmentation using U-Net 3+

Neti Praveen
Assistant Professor
Department of Computer Science and Design,
SRKR Engineering College,
Bhimavaram, Andhra Pradesh, India
neti.praveen@srkrec.edu.in

Gedela Abhinay
Department of Computer Science and Design,
SRKR Engineering College,
Bhimavaram, Andhra Pradesh, India
saiabhinaygedela@gmail.com

Kollabathula Shyam Babu
Department of Computer Science and Design,
SRKR Engineering College,
Bhimavaram, Andhra Pradesh, India
kollabathulashyambabu12@gmail.com

Pepeti Ganesh
Department of Computer Science and Design,
SRKR Engineering College,
Bhimavaram, Andhra Pradesh, India
ganeshpepeti814@gmail.com

Perabathula Someswara Rao
Department of Computer Science and Design,
SRKR Engineering College,
Bhimavaram, Andhra Pradesh, India
somuperabattula25@gmail.com

Abstraction: Mesectomy of medical images has remained an unavoidable part of the contemporary healthcare sector due to the fact that it facilitates accurate analysis and diagnosis of different medical conditions. One of the most popular known deep learning models regarding medical image segmentation since it seems to segment small objects in an image such as tumors, lesions, and organs is the U-Net model, which was proposed in 2015. The quest to attain improved performance has fueled the effort to transform the encoder-decoder architecture in the current years which incorporates the skip connections which are constituted of the U-Net which has enhanced the performance of segmentation of the medical imaging tasks despite target structures that appear in the form of miniature structures. U-Net is extensive in medical imaging and judging by the academic journals, it has been cited over 2,500 times and this makes it a very far-reaching intervention in the various fields of medical studies and the medical practice. The paper provides the overall summary of the U-Net based architectures to the medical image segmentation, data on how the original U-Net architecture was enhanced, and how alternative and better versions were developed taking into account some novelties. We examine the differences of U-Net, their architectural adjustments, developments, and enhancing the quality of the segmentation and the efficiency of the computing. Moreover, the functions of loss and evaluation metrics and modules commonly used in medical image segmentation process are illustrated in the paper and their weaknesses and strengths in various clinical contexts are discussed. Discussing such formulations, we are going to give an organization of such approaches which will inform future research on the field of medical image segmentation. The results provided in the current paper can become a worthy asset to a researcher and a practitioner as they provide clues about the current progress in this crucial field of medical technology.

Abstract words: U-Net, U-Net+, U-Net3+, Medical Image fragmentation, Deep learning, Variations of architecture, Loss functions, Metrics of evaluation.

I. INTRODUCTION

Medical image segmentation is a process that is critical in the treatment and management of all forms of medical conditions. This has particularly been worsened by the fact that with the increased imaging equipments, which are more complex some of which are performed using MRI, CT scans, and X-rays among others, there can be no automated and accurate segmentation of medical images. Image analysis approaches that are manual though are reliable are time consuming, prone to Human error and subjective in some cases. It has led to the introduction of computer-aided diagnostic systems particularly the ones based on deep learning algorithm that enable the medical professionals to perform an adequate and efficient segmentation of medical images.[1]

One of the most useful deep learning frameworks in medical image segmentation is the U-Net suggested by Ronneberger et al. in 2015. U-Net was quickly introduced to experts because of the specifics of its design which was in fact modified to the biomedical image segmentation issues. Its encoder-decider structure, storage of skip connections and allow it to effectively extract global and local features that make the architecture particularly effective when it comes to segmentation tasks of small and complex structures like tumors, lesions or organs. The ability of the U-Net to handle high-resolution medical images can be reduced to capabilities in order to learn hierarchical nature and be able to preserve spatial data, which is highly significant in such tasks.[2]

U-Net architecture has been used in the research community extensively in the past few years and has been continuously improved. It has its versatility because it has been, and still is, used in an extensive number of medical imaging applications including, but not limited to, brain tumor segmentation, cardiac imaging, organ segmentation, and retinal image analysis. Further, the design of the U-Net was modular, which made it easy to apply various developments which include multi-scale feature extraction, attention models, and hybrid networks that have contributed to the enhancement of the degree of the segmentation accuracy and efficiency.[3]

The need to have a more precise and effective system of medical image segmentation has led to different versions of U-Net. The various variants are supposed to address the challenges to do with segmentation of complicated structures, processing of asymmetrical information, and the increase of the effectiveness of the judges. U-Net architectures also are further enhanced to significant U-Net architectures, such as U-Net++, U-Net3+, attention U-Net and so on. The models all offer refinements to the original U-Net architecture, such as the capacity to better utilize contextual information, more effective skip connections and more effective loss functions, and enhance the ability of the model to segment the structures of medical images more effectively.[4]

Medical imaging has not been an easy task in terms of segmentation algorithms grounded on deep learning. One of the main issues is the necessity to train such models on big and labelled datasets. Medical image datasets are usually small and unbalanced or of different quality that can lead to negative outcomes of segmentation models. Additionally, the question of interpretability of deep learning models remains open particularly in the field of medicine. The clinicians must have some confidence in the result of such algorithms and to this end, they should be very accurate and in addition, they should be able to know how such decisions are arrived at. To address these challenges, the recent trends have focused on the increases in the robustness and generalization of U-Net- based models to make them useful on diverse datasets and even in clinical practice.[5]

Another subject of interest is evaluation of segmentation models. Segmentation performance can be gauged using standard measures of segmentation performance such as Dice coefficient, Intersection over Union (IoU) and pixel accuracy. Such measures are, however, not normally sufficient in terms of deciding on the utility of a model in practice applied. In a situation like that, the high segmentation accuracy is not always significant about the clinical benefits, as all of them may be modeled by the model demonstrating a real capability to perform the diagnosis, treatment planning, etc. Therefore, more attention is now being paid to the improvement of further strength assessment criterion, including clinical relevance and interpretability to allow the identification of the true value of segmentation models, in the clinical healthcare practice.[6]

In addition to these challenges, medical image segmentation models are to be also efficient in design. Medical imaging data is typically huge and the training of the deep learning models on the data might be computationally intensive and time consuming. This can serve as a barrier to the broad application of the deep learning-based segmentation algorithms especially in situations where resources are scarce. Thus, the new studies reviewed have devoted much consideration to simplification of U-Net-based models to attain a short time training and low computational power with still high segmentation accuracy. The potential methods to make medical image segmentation systems more efficient have been proven to be model pruning, quantization, and lightweight network architecture methodology.[7]

The value of U-Net and its variations to the medical image segmentation issue cannot be overestimated. The development of accurate and quality segment detection algorithm has a potential in changing the medical diagnostics in the context of making the medical image analysis more precise, fast and automatic. The great accuracy with which U-Net can isolate small structures is particularly important in oncology, neurology, and cardiology where, the localization of abnormality and early detection are significant variables that would result in better patient outcomes. Moreover, the deep learning-based segmentation can be also used to track the development of the disease, surgical planning, and reaction to the treatment, which can lead to the overall quality of the medical care improvement.[8]

II. RELATED WORK

R. Smith-Bindman, M. L. Kwan, E. C. Marlow et al., “Trends in use of medical imaging in US health care systems and in Ontario, Canada, 2000-2016,” JAMA, vol. 322, no. 9, pp. 843–856, 2019 Sep 3.

The computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound increased the usage in the United States between 2000 and 2006. There was a high utilization of medical imaging services in comparison with other services provided by physicians among the beneficiaries of the Medicare. Sharp rises in imaging may be explained by technical advancement, demand, and physician and patient, as well as positive financial incentives. Medica imaging can be used to provide proper identification and treatment of the disease, though medical imaging may also increase the costs and cause other harms to the patient such as incidental finding, overdiagnosis, anxiety and radiation due to the potential of cancer. It is approximated that of every 30 or so imaging tests, a third or so is not essential and costs the US approximately 30 billion annually. Furthermore, in a recent research, a comparison of the United States and ten other states in regards to various indicators resulted in the United States as having the highest or second writing of CT and MRI scans performed per 1000 individuals. It was a research on 2000-2016 rates of medical imaging in individuals in the various health care systems in the US and determined how this medical imaging utilization has changed over the years regarding the country, the health system, and patient demographic variables.

Summary: An approximation of 30 percent or even more of imaging tests is estimated as unnecessary which costs the United States about 30 billion annually. In addition, in one of the recent studies, comparing the United States with ten other countries in terms of various measures, the United States was ranked first or second in terms of the number of CT and MRI scans performed per 1000 individuals.

O. Ronneberger, P. Fischer, and T. Brox, “U-net: convolutional networks for biomedical image segmentation,” in Proceedings of the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds., Springer, Munich, Germany, October 2015.

It is widely known that deep network training requires thousands of annotated training images. The network and training strategy presented in this paper assume that we exhaustively enhance data to better leverage the available data in the form of annotations. The architecture is made up of a contracting path of capturing context and a symmetric expanding path of localisation which is fine. On the ISBI segmentation task on electron microscopic stacks of neuronal structures, we prove that this type of network can be trained end-to-end using a very small number of images and it is better than the previous one (a sliding-window convolutional network). This network which was trained on transmitted light microscopy images (phase contrast and DIC) also won the 2015 ISBI cell tracking challenge in both categories. More so, the network is fast. The division of a 512x512 image can be made in less than a second in a current GPU. The trained networks and the entire implementation (on the basis of Caffe).

Summary: This paper proposes a network and training scheme that requires a lot of data augmentation to exploit annotated samples that are available better. These structures are captured context, and the expanding localization path which is a contracting path and an equal path respectively.

X. X. Yin, B. W.-H. Ng, and Q. YangA. Pitman, K. Ramamohanarao, and D. Abbott, “Anatomical landmark localization in breast dynamic contrast-enhanced MR imaging,” Medical, & Biological Engineering & Computing, vol. 50, no. 1, pp. 91–101, 2012.

We present the novel method of localising anatomical structures of a dynamic contrast-enhanced MRI of the breast: costal cartilage of the breast by the use of level sets. Localization of the current diagnosis of the breast MRI uses magnetic-resonance compatible needles. However, when the breast cartilage can be used in the structuration of the breast, it will not only be helpful in avoiding the invasive processes, but it will also simplify the movement of the breast by the action of the cardiac and respiratory movements that it is followed. The paper describes a novel algorithm of confidently determining the structure of the costal cartilage, and deriving it, applicable to existing motion artefacts, and the potential variability of the shape of the structure with uptake of a contrast agent, and the potential breast registration. This paper describes an algorithm that will help extract volume information of post-contrast MR images at three time slices that will be utilised to analyse motion artefacts and we test the validity of the existing algorithm with anatomic structure. Level-set method is used to choose the size of the region of interest. The form of contours obtained according to a level-set-based segment image is diverse, which determines the region of interest of the feature that may be utilized as a guide in acquiring the first masks to be employed in the feature extraction. Later, the algorithm applies a K-means method of disaggregating feature areas of other types of tissue and morphological change, and the addition of an appropriate structuring element, in achieving reliable masks and feature generation. Machines can remove masks and eventually acquire feature segments which will be utilized to perform a motion analysis of the breast and perform registration.

Summary: The paper introduces a novel algorithm that can be incorporated to identify and isolate reliably costal cartilage features which can be used to analyse motion artefacts, and shape changes that can be produced by uptake of contrast agent in addition to breast registration. The paper will describe an algorithm that will estimate the volume features of micropost contrast MR at these three time slices at these time slices, and the algorithm is checked on anatomic structure.

X.-X. Yin, S. Hadjiloucas, J.-H. Chen, Y. Zhang, J.-L. Wu, and M.-Y. Su, “Correction: tensor based multichannel reconstruction for breast tumours identification from DCE-MRIs,” PLoS One, vol. 12, no. 4, p. e0176133, 2017.

The water imaging modality such as computer-generated tomography (CT), magnetic resonance imaging (MRI), and ultrasound also gives the relative imaging rates among adults [18-64 years] in the United States and Ontario. Adult CT, MRI, and ultrasound imaging rates (1000 people) have been gradually growing over the recent years but their growth slowed down. Large slice count biomedical images are often blocky in volume space. An image processing algorithm in 2D is usually utilised so as to analyse a 3D image. Individual sorting and training of the data, however, would cost more, both in terms of computational cost and little efficiency. The volume images are, therefore, not easy to work with in most cases. These are the problems that are being addressed by designing a 3D U-Net with the help of the 2D U-Net. The proposed UNet 3+ can be used to make computations scale faster because of reduced network parameters. To sum it up, the creation of new ideas in the deep learning network, grounded in the U-Net, is not the primary motive why the specified work should be conducted, and the primary goal is to effectively apply methods of the U-Net-related deep learning networks in the multidimensional data segmentation so that the said methods could be applied to the biomedical issues of the work.

Summary: 2D image processing algorithm is frequently used in the processing of a 3D image. Individual classification and training of the data, on the other hand, would result in higher costs of computation, and lower efficiency. Therefore, in a variety of situations, the work with the volume photographs could be difficult. In order to address these issues, 3D U-Net model that is derived after using 2D U-Net is being created. The suggested UNet 3+ would be in a position to improve the level of computational productivity since it reduces network parameters.

III. PROPOSED SYSTEM WORKFLOW

A. Data Collection

In this study, we have used a medical image partition dataset, which is available on Kaggle since the dataset is related to the images of the lung and the chest X-ray. The data is completely focused on the training and analysis of deep learning models and, in specific, on the work of meditative image segmentation, specifically to identify different lung diseases with lung tumor, lesions, and other cancerous lesions. The images, which are represented in this dataset are labeled by the pixels and that is why, it is much more appropriate, in relation to the training of the models that demand the fine segmentation. This data set contains numerous varied chest and lung X-ray pictures along with the labels of ground facts to this picture segmenting task. It is a rich dataset and contains images at various stages/classifications of lung ailments and the model is thus trained on a variety of cases. The pictures are X-ray scans of high quality whose structures, textures and the differences that exist are diverse as far as displaying the disease is concerned. The variations serve to check the validity of the model of segmentation in the spirit that it can possibly survive problems in the real world where there may be variation in quality of the images and manifestation of diseases.[9]

Overall, the dataset is presented by thousands of X-ray images, which makes it possible to train and test the models. This information is subdivided into three components which are training, validation, and test and offers effective training and testing pipeline models. The labels are made by the medical experts manually, which is a representation of areas of interest i.e. tumors, lesions and other abnormalities and the labels are correct and consistent. With such a well-marked and diverse information, the objective is to enhance the efficiency of the segmentation models in detection and demarcation of the abnormal regions within the lung and chest imaging which is critical in the process of making a correct diagnosis and the succession planning in the clinical practice.

This data is very useful in the training of the deep learning algorithms that detect lung diseases as well as in developing an understanding about the complexity of the medical image. This would make an ideal base to test out complex segmentation methods and have their value in the different lung disease and lead to the creation of more and more useful medical image analysis programs, which are partially or fully automated.

B. Data Preprocessing

The performance of deep learning models and especially in segmentation of medical images largely depends on the quality and consistency of the data that is fed into the model. In this study, I have performed various preprocessing functions on the images of the lung and the chest X-rays to ensure that the data is in the optimal format of the segmentation models. The pictures have initially been downsized to a uniform size so that there is uniformity in the dataset. The process enables the reduction of the degree of computational complexity and all the input images should be of the same shape, which it is mandatory before they are fed into the neural networks. Images were resized to a typical image (ex: 224x224 pixels) which is typically an order of image classification and segmentation actions.

The second action involved normalizing the images to fit the pixels into the same range (this is usually 0-1). This component of normalization helps to stimulate the acceleration of training because the model weights are updated in a more gradual manner. To scale the values accordingly, the values of pixel intonation were scaled to the highest possible number of an 8-bit image, which is 255. The normalization is also helpful in dealing with the various intensity of the images to ensure that the model is not focused on the extreme intensities of an image hence being biased.

It also had data augmentation measures of increasing the number of data points artificially to improve the model in generalization. Medical image data are often non-variable and small, thus augmentation is helpful in overfitting the model and makes it stronger. In this case, the pictures were randomly rotated, flipped and slightly transferred, and made to look in the real world (with the model). These modifications too help in imparting a degree of invariance to the model which can be especially handy in the medical imaging scenario wherein the orientation and location of the structures of interest might vary.

C. Model Development

The reason is that a deep learning-based segmentation model was developed with the U-Net architecture, which has proven to be impressive in the medical image segmentation issues. The encoder-decoder structure based on the skip connections is particularly better adapted in the algorithms working with high-resolution images, i.e., segmentation of lung and chest X-rays. The encoder is employed to acquire high level features of the input images and the decoder is employed to acquire the segmentation map through the high level features with the skip connection taken care of the spatial information that may be crucial when the process involves segmentation of small and complex structures.

That began with the implementation of the U-Net simple architecture that uses the deep learning framework such as TensorFlow/Keras or PyTorch. The model was applied to apply a sequence of convolutional layers that were followed by batch normalization and activation functions like ReLU that were used to help the model to learn intricate features in an effective way. The encoder half of the U-Net then consists of a sequence of downsampling stages, either using 3x3 convolutional filters, to learn more and more large and abstract features in the input image in an extractive manner. The max pooling layers that followed each convolutional layer helped to restructure the spatial arrangements of the feature maps that were generated by the encoder to replicate the spatial resolution of the original image. It does it through the application of transposed convolution layers that increase the spatial dimensions. These leaping links amongst the encoder and decoder planes allow the model to save delicate-spatial information, which is vital in subdividing tiny objects, e.g. tumours or lesions, in other medical images.

Certain adjustments were carried out in order to enhance the performance of the simplified U-Net model. This was done with the help of the U-Net+, which is a more advanced version of U-Net with skip pathways applied on itself that has the added functionality of doing feature reuse as well as being able to learn finer detail. This variant also increases the precision of segmentation since it allows the model to be more representative of the multi-scale information that is required when dealing with medical images whose multi-scale and resolution arrangements are relevant to the image.Additionally, an attention mechanism has been incorporated into the U-Net structure that has corrected the attention mechanism that makes the model to be more centered on the most interesting portions of the image when performing segmentation. This helps the model to give priority to the areas of interest i.e. abnormalities or lesions and does not bother itself with the unnecessary background data.

D. Model Evaluation

These performance and effectiveness of a deep learning model evaluation is a crucial aspect in the model evaluation task because accuracy and reliability are critical aspects to medical image segmentation. The effectiveness of the U-Net-based segmentation model used in this paper had been adequately outlined by using various traditional indicators, which are uniquely established to segmentation. The Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) as well as pixel-wise accuracy were the important measures of evaluation. One widely used measure of the similarity of the segmentation maps predicted to that of the ground truth medical images used to determine the truth is the Dice Similarity Coefficient (DSC) which directly quantifies the similarity between the predicated segmentation and the ground truth. This can be between 0 (absolutely no overlap) and 1 (perfect overlap), and the higher the number, the better is the work of the segmentation. The DSC is particularly relevant to the medical imaging one, in which they consider both false positives and false negatives; therefore the consideration of small and imbalanced target structures like tumors or lesions is necessary. Iou is taken as a measure of the overlap between the predicted segmentation mask and the ground truth mask and the intersection between the two divided by the union. Just like the Dice coefficient, the scale of the IoU is 0 to 1 with higher values representing a higher level of performance.

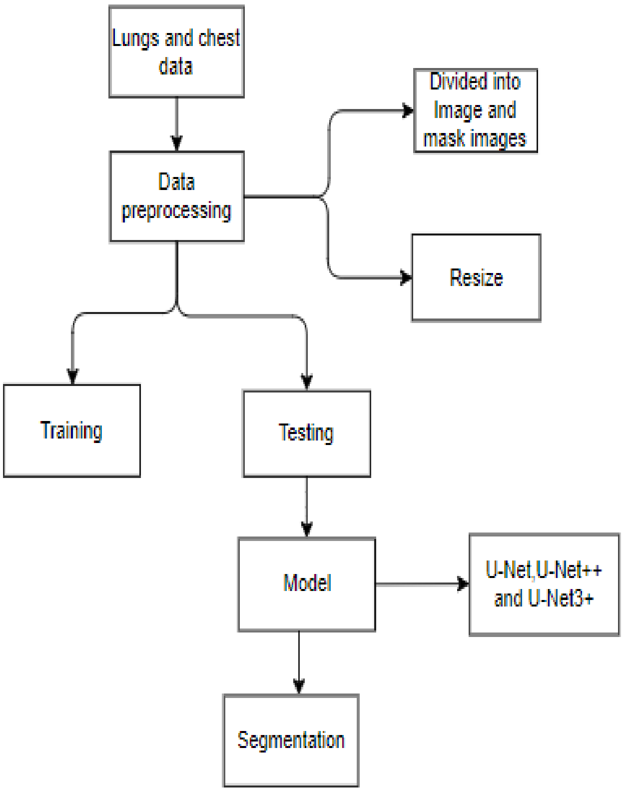


Figure 1 Project Work Flow

IV. METHODOLOGY

U-Net:

The most efficient and the most popular deep learning architecture that was used in the medical image segmentation is U-Net that was firstly introduced by Ronneberger et al. in 2015. It is directly meant to be used in biomedical image segmentation task and it has been successfully motioned since it can successfully do the right segmentation even in small or tricky shapes. No design that is more suitable to use in the application is the encoder-decoder design assumed to take the shape of the U-Net structure with skip connections, especially where a high level of emphasis is put in localising features of items well like in medical imaging where organs, tumour or lesions need to be accurately defined. These layers consecutively encode reduced spatial and increased feature map depths of the image so as to obtain high level features of the input image, e.g. edges, textures, forms and the such. The decoder component in turn takes the task of recombining the spatial resolution of the image by the successive upsampling of the feature maps to which it was injected directly, to its normal size. The skip connections are the characteristic that makes U-Net to be different among the rest of the mainstream convolutional neural networks (CNN).The u-net is performed through upsampling to preserve the spatial features and enhance the output of the segmentation process. These paths connect the corresponding layers on the encoder and decoder components in such a manner that it is possible both to directly transfer spatial data by the downsampling path and directly onto the upsampling one. This is especially essential in medical image segmentation wherein minute or tiny details of the image should be maintained in order to be able to segment small or minute abnormalities.U-Net has been shown to be most effective in a variety of medical imaging applications including brain tumor segmentation, organ segmentation, and detecting various diseases such as lung cancer and diabetic retinopathy. Its flexibility and capability to operate with images of high resolution with complicated structures have given it a preferred model in the medical industry whenever undertaking a segmentation task.

U-Net ++:

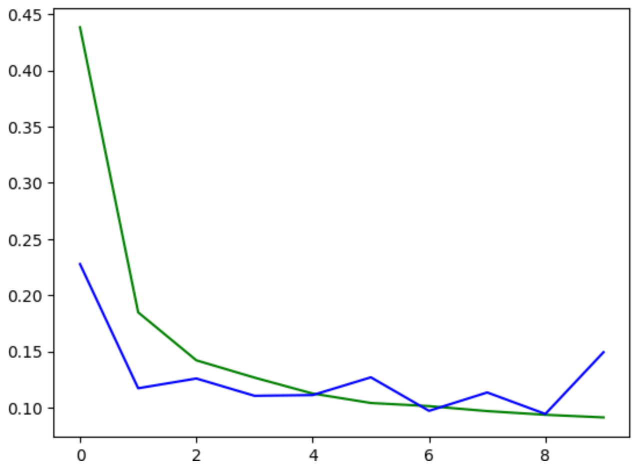
The U-Net++ is the enhanced version of the original U-Net architecture because it is developed on the basis of the performance improvement of the specific medical image segmentation task. U-Net++ is an improvement of the original U-Net, which was created by Zhou et al. in 2018, and which can thus make the following improvements to U-Net: giving it a more complex structure, which permits the reuse of the features as well as enabling the model to encode a much more detailed information across multiple scales. The dense skip pathway and deep supervision are the innovation of U-Net++ that significantly improve the precision of the performance of the segmentation, especially in the case of small and intricate structures that do not lend themselves to segmentation. These thick skip connections offer a closer relationship between feature maps between-levels in the network and enables greater reuse of features. The fact that the skip connections in the second are more profound, along with the fact that the network can capture more profound and context-sensitive features, in particular, when dealing with complicated segmentation, is another difference between the original U-Net and U-Net++. Moreover, U-Net++ uses the so-called deep supervision. The way is also that the auxiliary functions of loss are inserted in a number of intermediate network layers, and not the last output layer. It is possible to monitor the model on the intermediate results and thus learn multi-scale features separately on different levels of abstraction, which helps in: improving the quality of segmentation and increasing the speed of the train. It also allows more customization of the model to changes in image quality and disease presentation and therefore is more resistant to the real world complexities of medical images.

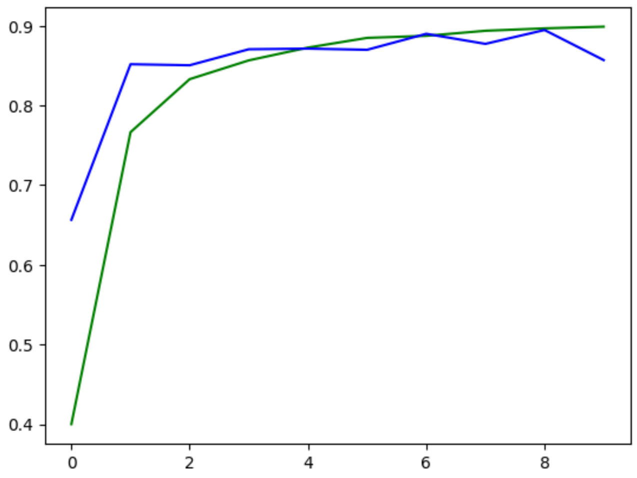
U-net 3+:

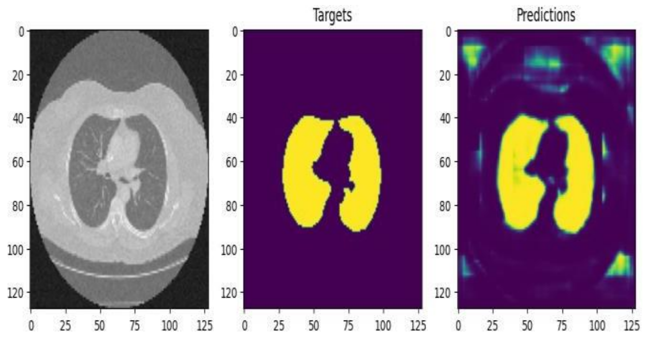
U-Net 3+ represents an enhanced version of the original U-Net architecture that was made by enhancing the performance of the segmentation of medical images through the refinement of multi-scale feature extraction and through the utilization of deep supervision in a manner that enables the attainment of even better results. U-Net 3+ introduced by Xie et al. is preoccupied with the fact that the models mentioned before have the limitation that these U-Net and U-Net++ are more effective in capturing low-level and high-level features. It particularly applies well to more difficult segmentation tasks, such as in medical imaging, where finer details of features are required and useful information about the context, the multi-scale fusions of features is used in the U-Net 3+ implementation to compute finer features at various scales and levels. The U-Net 3+ is more resistant to variations in the size, intensity, and resolution of the images because it has multiple encoders and decoders as its architecture. The network structure enables drawing of finer grained information whereby dense skip pathways are inserted between the encoder and decoder at an additional level and is similar to U-Net only that it uses multi-scale features to combine the features at the shallow and deeper layers of the network. The effect of this feature fusion aptitude is to render the model able to mix local features alongside global features that are essential in case of segmenting complicated medical images that comprise smaller (assuming the presence of lesions or tumors) and bigger features (assuming the presence of various organs). The multi-scale fusion, the model could use fine and coarse feature to produce an accurate prediction under a range of styles and conditions of an image and the final result; it also possesses a rich supervision system, which introduces auxiliary loss functions in several, intermediate layers. This inspires the network to focus on learning the multi scale representation of the same in various levels to facilitate the performance of the model and convergence. Deep supervision does not only help the model to learn more effectively, but also leads to a more effective regularization of the learning process, especially in those cases where the given dataset is not large enough, or imbalanced. The other advantage of U-Net 3 + is that it can withstand the imbalance of classes which normally appears in medical image segmentation. With the help of progressive methods as weighted loss functions or focal loss, U-Net 3+ can emphasize the less prevalent classes, whether small tumors or the initial lesions, and ensure the model does not focus on larger and more popular objects at the expense of smaller ones.

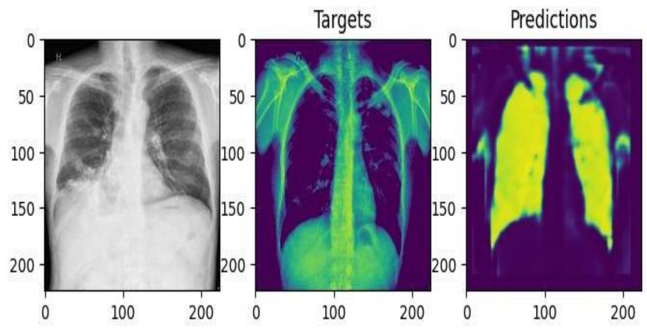
V. RESULTS

U-Net:



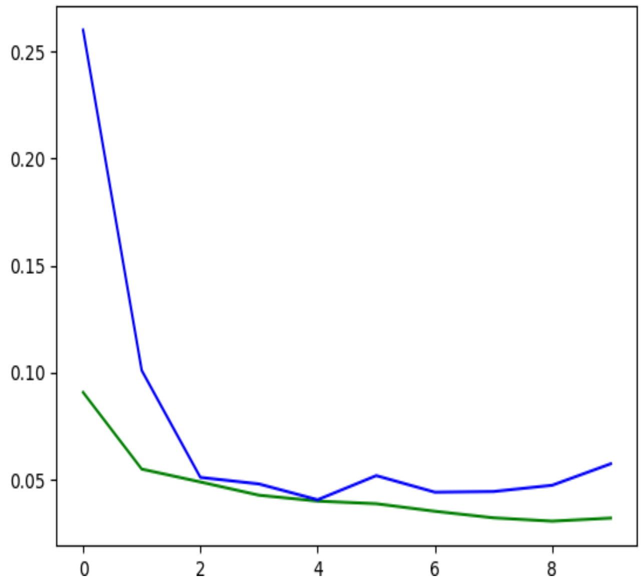


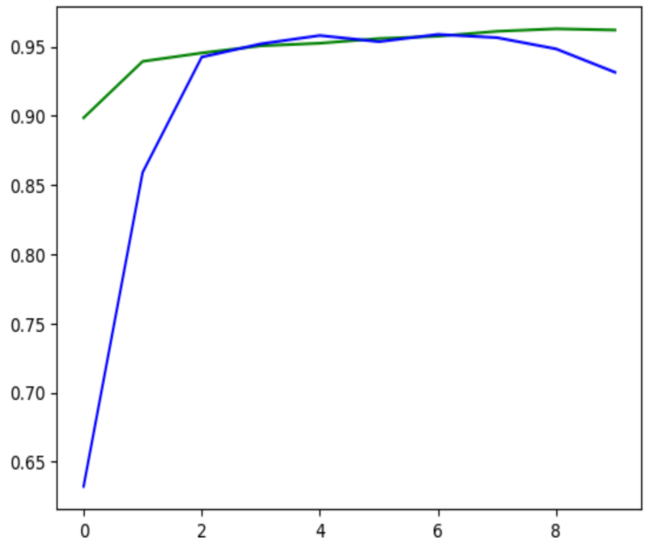


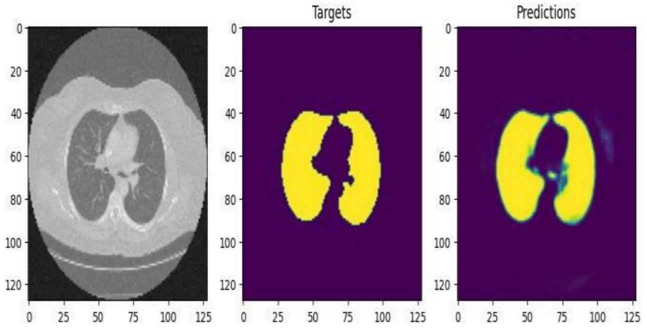


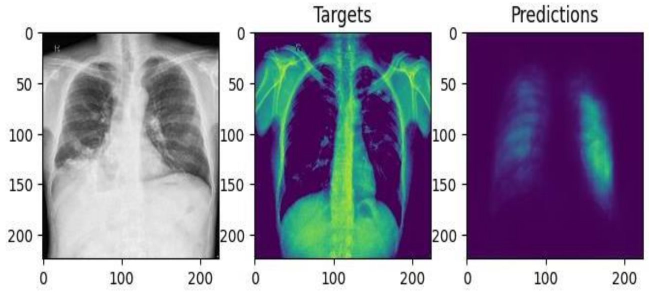
The model training, being trained with 10 epochs is shown to demonstrate gradual increases in the training and validation phase and brings out the ability of the model to learn and generalise over time. To start with, in the first epoch the model achieved a low training accuracy of 40.01 percent and a loss of 0.4383. However, during the second epoch, the accuracy is 76.65 with which the training loss is significantly reduced to 0.1847 of the learning process in the model. The first and the second epochs both improve the accuracy of validation (65.65 and 85.18 respectively) and this indicates that the model is getting to know how to generalize to unseen information.The performance of the model continues to enhance as the training accuracy and the validation accuracy rises with each subsequent epoch. The fifth epoch makes the model achieve an accuracy of 87.27 percent in training and validation of an accuracy of 87.14 percent. The training process gradually reduces the learning rate to achieve a perfect performance of the model. The learning rate will begin with 0.0010 in the first epoch and then decreases in the following epochs such that in the tenth epoch, the rate will be 0.00013422,

U-Net ++:



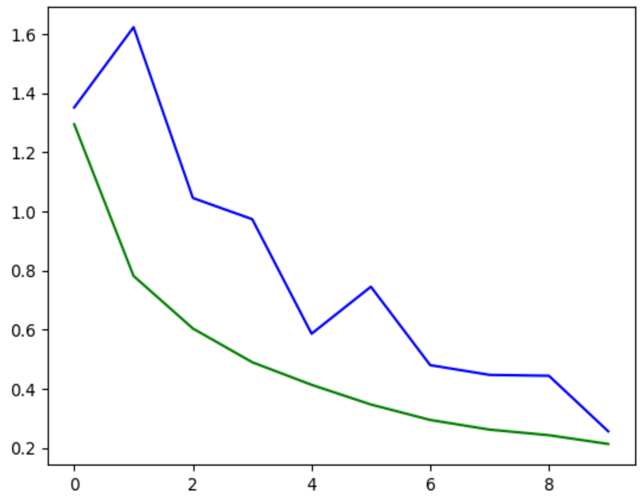


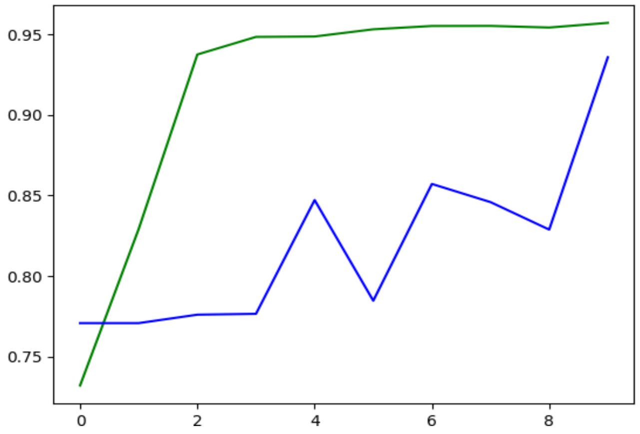


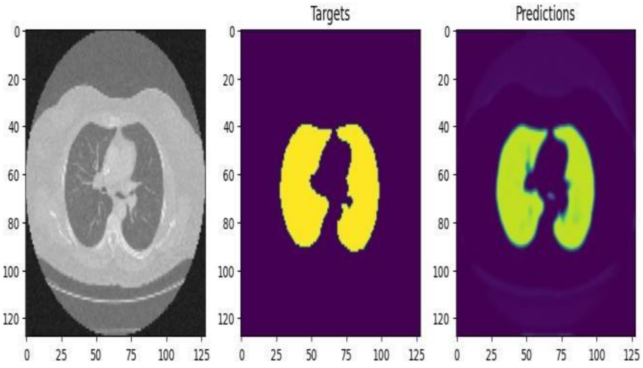


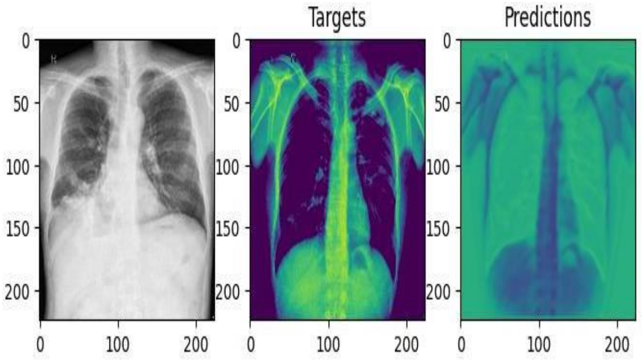
This is observed in the training of the model within the initial 10 epochs, the trend of which is comparable in terms of training and validation performance and oscillates about the validation loss. Initially, the loss (1.2942) and the accuracy (73.20) of the training set in the model is low and this is then followed by a loss (1.3510) and an accuracy (77.07) of the validation set. Accuracy in training the model reaches 82.91 (after the second epoch) and training loss also goes down to a great extent to 0.7812. However, the loss of validation is increasing to 1.6226 and the validation accuracy of the model is constant at 77.07, train accuracy is greatly increasing to 93.74, and the validation loss is also decreasing to a significant extent to 0.6034. Of this nature, though, there is a high loss of validation between 1.0447 to the slightest improvement in validation accuracy of 77.59 percent. The error of the training at the epoch 4 is high, 94.83, and the training loss is 0.4893. However, there is also a slight loss in validation value to0.9731, and there is an increase in validation accuracy to 77.65 i.e. the model is still having problems of generalization.

U-net 3+:









Among the remarkable facts of the model training process is that in the course of the 10 epochs, the model is able to learn and carry out generalization. The first epoch started with loss value of 1.2942 and 73.20 perfection in the training phase and the loss value of 1.3510 and the validation value of 77.07. This indicates that the model is new in the learning process and it can be enhanced especially in its ability to generalize to the validation data where the model training accuracy would be 82.91 and the model training loss be 0.7812 which is a big milestone in learning. The validation loss though rises to 1.6226 and the validation accuracy does not alter to 77.07. This implies that the model is equally likely to be overfitting the training data and unable to generalise to the validation data easily.The third epoch has made more improvement in training accuracy to 93.74 per cent with a training loss of 0.6034. Even though the performance is still not at the optimal level in training, the error of validation lost is quite high 1.0447, and the validation accuracy is increased not so much 77.59. The fourth epoch is identified with training accuracy of 94.83 and loss is even less with the value of 0.4893 but the validation performance is slightly stagnant with the values of the validation accuracy changing marginally by 0.7765 and the validation loss decreasing further to 0.9731.

VI. DISCUSSION

It can be seen that the medical image segmentation work of the deep learning model has been significantly improved during the 10 epochs based on the model training and evaluation results. The former is that the model possessed the typical overfitting traits, where the training performance was quite high, and the validation performance was declining slower, in particular at the first epochs. It may happen in the deep learning models, especially when working with multidimensional high-resolution medical data, where the model will initially learn certain features of the training data, but will not use them to the validation set. However, the loss of training progressively declined and the accuracy of the validation progressively rose signaling the challenges that are typically encountered during the training of deep learning models to identify medical image segments. This validation performance had certain minor deteriorations in the accuracy in certain epochs, and this might be attributed to a number of factors, including the optimization of the learning rate, the intricacy of the data, and the inherent noise of medical pictures. These waves were however neutralized by gradual improvement at the advanced training phases particularly during the final epoch when the model achieved a validation accuracy of 93.56 percent and a respective decrease in validation loss of 0.2555.

The interesting fact about the training process is that, the training loss and also the validation loss is reduced significantly at the end of the training. This reduction shows that the model had learned the intricate features of the data without overfitting as shown by the high validation accuracy in the final epoch. This ability of the model to generalize to unknown data is one of the major considerations, especially when the interest is the medical image segmentation where it is crucial to ensure that the model is commonly realistic on the diverse real world and unknown medical images.

VII. CONCLUSION

The model was a deep learning-based model, which was trained on the segmentation of medical images i.e., lung and chest abnormalities. The model had noteworthy variations in the training and validation performance in the course of the epochs, in which, loss is drastically lowered and accuracy is drastically high. The results show that the model had the ability to learn with the training data and generalize to other unknown validation data with a high validation accuracy of 93.56 at the final epoch. Despite the fact that they have some low variations of validation performance during the training process, the overall trend is that deep learning model has been trained to extract distinctive segments of complex medical images with high precision, which can greatly enhance the diagnostic process and automate it to detect abnormalities. However, the hardness of deep network training is also reflected on performance of the model, which have been demonstrated by overfitting and fine-tuning of deep network to achieve maximum generalization. The future work may be in the optimization of the model, with the help of more sophisticated data augmentation techniques, more sophisticated learning rate scheduling techniques, or more sophisticated regularization techniques to enhance further its stability and performance.

VIII. FUTURE ENHANCEMENT

Even though the current model has provided promising results in terms of the medical image segmentation, several areas can be enhanced in the future in order to enhance its functionality and adaptability. One of the key aspects that should be addressed is to address the issue of overfitting that had been experienced during some epochs. It may be mitigated in case of addition of some more-complicated regularization steps, such as dropout, weight decay, or batch normalization, which allow the model to better match the unknown data. Another possible improvement is more advanced architectures or hybrid models which can be fine-tuned and learning rate scheduling may be tried to maximize the training process and allow the model to converge faster or the desired model flipping around its optimum validation performance. These attention mechanisms or more advanced versions of U-Net, such as U-Net+ and U-Net3+, can possibly allow the model to pay more attention to what is relevant in the medical images, and specifically, smaller (or hard-to-find) abnormalities. The multi-scale feature extraction methods would also improve the accurateness of the segmentation particularly in cases of medical images that exhibit varying resolutions and complexity.

The other important step that is required to enhance the model is the data augmentation to widen the dataset. The model can also be trained to respond to more real-world scenarios by simulating different imaging conditions i.e. change in orientation, change in light or noise etc and hence augment the degree of its generalization. Furthermore, one can make the model more universal and general, apply more different data sets, images of different patients and imaging systems (e.g. CT scans, MRI) et cetera. In the medical sphere, healthcare specialists have to trust the result of the model. This could be aided by the explainability techniques such as Grad-CAM or saliency map to provide a visual representation of what the model is focusing on in the picture which would validate its predictions and make the clinical community more readily accept them.

IX REFERENCES

[1] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med Image Anal*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/J.MEDIA.2017.07.005.

[2] X. Liu, L. Song, S. Liu, and Y. Zhang, “A review of deep-learning-based medical image segmentation methods,” *Sustainability (Switzerland)*, vol. 13, no. 3, pp. 1–29, Feb. 2021, doi: 10.3390/SU13031224.

[3] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, “Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation,” Feb. 2018, Accessed: Nov. 18, 2025. [Online]. Available: http://arxiv.org/abs/1802.06955

[4] M. Khouy, Y. Jabrane, M. Ameur, and A. Hajjam El Hassani, “Medical Image Segmentation Using Automatic Optimized U-Net Architecture Based on Genetic Algorithm,” *Journal of Personalized Medicine 2023, Vol. 13, Page 1298*, vol. 13, no. 9, p. 1298, Aug. 2023, doi: 10.3390/JPM13091298.

[5] "Medical Image Segmentation Review: The Success of U-Net | IEEE Transactions on Pattern Analysis and Machine Intelligence." Accessed: Nov. 18, 2025. [Online]. Available: https://dl.acm.org/doi/abs/10.1109/TPAMI.2024.3435571

[6] R. Azad *et al.*, "Medical Image Segmentation Review: The Success of U-Net," *IEEE Trans Pattern Anal Mach Intell*, vol. 46, no. 12, pp. 10076–10095, 2024, doi: 10.1109/TPAMI.2024.3435571;SERIALTOPIC:TOPIC:ACM-PUBTYPE.

[7] A. Garbaz, Y. Oukdach, S. Charfi, M. El Ansari, L. Koutti, and M. Salihoun, "MLFA-UNet: A multi-level feature assembly UNet for medical image segmentation," *Methods*, vol. 232, pp. 52–64, Dec. 2024, doi: 10.1016/J.YMETH.2024.10.010.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[9] W. Ehab, L. Huang, Y. Li, W. Ehab, L. Huang, and Y. Li, "UNet and Variants for Medical Image Segmentation," *https://www.sciltp.com/journals/ijndi/articles/2504000040*, vol. 3, no. 2, p. 100009, Jun. 2024, doi: 10.53941/IJNDI.2024.100009.