U-net Based Autonomous Fetal Segmentation From 2D and 3D Ultrasound Images

by

Maliha Tabassum Mithila
18101459
Tasnim Ahsan Prome
18101420
Elmi Tabassum
18101222
Sameha Kamrul
18101523
Nausheen samiha
18101108

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

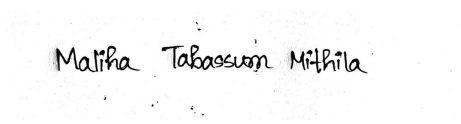
Department of Computer Science and Engineering
BRAC University
May 2022

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

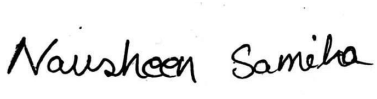


Maliha Tabassum Mithila
18101459

Tasnim Ahsan Prome
18101420

Elmi Tabassum
18101222

Sameha Kamrul
18101523



Nausheen Samiha
18101108

Approval

The thesis/project titled "U-net Based Autonomous Fetal Segmentation From 2D and 3D Ultrasound Images" submitted by

1. Maliha Tabassum Mithila (18101459)
2. Tasnim Ahsan Prome (18101420)
3. Elmi Tabassum (18101222)
4. Sameha Kamrul (18101523)
5. Nausheen Samiha (18101108)

Of Spring, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on May 24, 2022.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Golam Rabiul Alam
Associate Professor
CSE
BRAC University

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Associate Professor
CSE
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Ethics Statement

Our goal is to estimate 3 fetal biometric parameters which are head circumference, abdominal circumference and femur length in order to monitor fetal development and compute gestational age of the fetus based on the estimated fetal length. We have worked with extensive fetal data from first and third trimester in order to segment the fetal head, abdomen and femur area accurately by processing ground truth or binary mask images. We have assured complete transparency of our evaluation process in addition to providing visual interpretation of the output generated by the model.

Abstract

There are various biometric parameters of the fetus that need to be evaluated to monitor prenatal diagnosis during pregnancy. Biometric parameters such as head circumference, abdominal circumference, cortical volume, the volume of the brain, crown-rump length, femur length, etc. play a very important part in the characterization and detection of the development of the fetus. Gestational age is one of the most effective parameters for monitoring fetal growth and development, as well as diagnosing any abnormalities, among several quantitative indices. To estimate the gestational age, birth size, weight, and to monitor prenatal abnormalities, many biometric parameters such as head circumference (HC), abdomen circumference (AC), and femur length (FL) must be measured. We can extract these parameters from the segmentation of an MRI scan. However, performing full manual segmentation is exhaustive and time-consuming. Ultrasound imaging has been shown to be more efficient than MRI for measuring such biometric characteristics.. Also, in this case, manual segmentation requires experts' experience and skills, clinical experience of the staff which is time-consuming. As a result, we propose a fully autonomous segmentation method based on U-Net architecture for fetal biometric parameters such as head circumference (HC), abdomen circumference (AC), and femur length (FL), which eliminates the need for manual intervention, reduces computational complexity, and greatly speeds up the segmentation process. U-Net is a convolutional neural network that was created for performing segmentation on biomedical images. Our goal is to train the network such that it can create high-resolution 2D and 3D ultrasound images of each segmented fetal area.

Keywords: Biometric Parameters, Gestational Age, U-net, Semantic Segmentation, Head Circumference, Abdominal Circumference, Femur Length, Deep Neural Network, Convolution, Autonomous

Dedication

We would like to dedicate this thesis work to our loving parents. As well as, all the amazing faculties we came across and learnt from in the course of pursuing our Bachelors degree. It's been a worthwhile experience...

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our supervisor Mr. Golam Rabiul Alam sir for his kind support and advice in our work. He helped us whenever we needed help.
And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

Declaration i

Approval ii

Ethics Statement iii

Abstract iv

Dedication v

Acknowledgment vi

Table of Contents vii

List of Figures ix

List of Tables x

Nomenclature xii

1 Introduction 1
- 1.1 Fetal Development and Segmentation Process . . . 1
- 1.2 Problem Statement . . . 2
- 1.3 Research Objective . . . 4
- 1.4 Challenges . . . 5
- 1.5 Thesis Outline . . . 5

2 Literature Review 7
- 2.1 Existing Works . . . 7
- 2.2 Background Study . . . 11
  - 2.2.1 U-Net Architecture . . . 11
  - 2.2.2 Why Ultrasound Image? . . . 14
  - 2.2.3 Biometric Parameters . . . 14
  - 2.2.4 Why Autonomous Segmentation? . . . 14

3 Research Methodology 16
- 3.1 System Model . . . 16
- 3.2 Data Collection . . . 19
- 3.3 Data Pre-processing . . . 19

3.3.1 Annotation and Masking . . . 19
3.3.2 Converting to Grayscale . . . 20
3.3.3 Resizing . . . 20
3.3.4 Normalization . . . 21
3.4 Train-Test Split . . . 21

4 Implementation and Performance Evaluation 23
4.1 Model Implementation . . . 23
4.2 Performance Metrics and Loss Function . . . 24
4.2.1 Dice-coefficient (F1 Score) . . . 24
4.2.2 IoU (Intersection-Over-Union) . . . 25
4.2.3 Dice Loss . . . 26
4.3 Model Fitting & Training . . . 26
4.4 Evaluating the Model . . . 26
4.5 Circumference & Length Measurement . . . 31
4.5.1 Head & Abdominal Circumference Measurement . . . 31
4.5.2 Femur Length Measurement . . . 33
4.6 Gestational Age Measurement . . . 34

5 Results & Discussion 36
5.1 Training Result Analysis . . . 36
5.2 Segmented Image Analysis . . . 37
5.3 Circumference, Length & Gestational Age Results . . . 40

6 Epilogue 41
6.1 Why U-Net Architecture & Modified Version? . . . 41
6.2 Limitations . . . 41
6.3 Future Work . . . 42
6.4 Conclusion . . . 42

Bibliography 45

List of Figures

2.1 U-Net Architecture 13

3.1 The Work Flow of Proposed U-Net Based Fetal Segmentation Model 18
3.2 Annotated Image vs. Generated Mask Image 20
3.3 Raw Image vs. Resized Grayscale Image 21

4.1 Implemented Model Architecture 24
4.2 Training and Validation Accuracy 27
4.3 Training and Validation Dice Score 28
4.4 Training and Validation IoU 29
4.5 Training and Validation Loss 30

5.1 Bar Chart of Training Performance Metrics 36
5.2 Bar Chart of Validation Performance Metrics 37
5.3 Segmented Image of Head Circumference 38
5.4 Segmented Image of Abdominal Circumference 39
5.5 Segmented Image of Femur Length 39

List of Tables

4.1 Estimated Head Circumference Result 32
4.2 Estimated Abdominal Circumference Result 33
4.3 Estimated Femur Length Result 34
4.4 Estimated Gestational Age Result 35

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AC Abdominal Circumference

AI Artificial Intelligence

ASD Average Symmetric Contour Distance

BPD Biparietal Diameter

CAD Computer-Aided Detection

CNN Convolutional Neural Network

CRL Crown-Rump Length

CT Computed Tomography

D Dice Coefficient

DP Dynamic Programming

DSC Dice Similarity Coefficient

EAC Estimated Abdominal Circumference

EFL Estimated Femur Length

EGA Estimated Gestational Age

EHC Estimated Head Circumference

FCN Fully Convolutional Network

FL Femur Length

GA Gestational Age

HC Head Circumference

HD Hausedorff Distance

HL Humerus Length

HT Hough Transform

IoU Intersection Over Union

J Jaccard Index

$JSON$ JavaScript Object Notation

$MFP-U-NET$ Multi-feature Pyramid U-Net

MRI Magnetic Resonance Imaging

MSD Maximum Symmetric Contour Distance

MSI Microsoft Installer

OFD Occipitofrontal diameter

$ReLu$ Rectified Linear Activation Function

$RMSD$ Root Mean Square Symmetric Contour Distance

ROI Region of Interest

SE Squeeze and Excitation

SGD Stochastic Gradient Descent

US Ultrasound

Chapter 1

Introduction

1.1 Fetal Development and Segmentation Process

Fetal development is one of the most crucial factors for monitoring fetal health. Anomalies in different stages of fetal development can be detected by evaluating various biosignatures such as fetal head circumference (HC), abdominal circumference (AC), femur length (FL), biparietal diameter (BPD), etc. These biometric parameters can be used to determine gestational age, as well as to facilitate the process of prenatal diagnosis. Fetal segmentation is the most efficient process for determining these biometric parameters. By fitting an ellipse, line or rectangle on the segmented areas, these parameters can be easily measured which can be further used for monitoring fetal growth. Ultrasound (US) imaging is the best fit to be used for segmentation procedures as it is affordable, swift in manner and non-invasive, hence, painless and far better than other imagining technologies like magnetic resonance imaging (MRI) or computed tomography (CT) in terms of detecting anomalies. Despite the benefits of US, there exist some limitations that make it difficult for interpretation and require the intervention of experts to be overcome. To reduce such intervention and to ensure faster operation, an autonomous segmentation method is needed to be implemented.

Haar-like characteristics, also known as digital picture features, are utilized in object recognition to detect edge, rectangle, line, and center-surround features. In previous studies, for locating the skull of the baby, a random forest classifier was trained by using those characteristics. In general, randomized Hough transform, semi-supervised patch-based graphs, multilevel thresholding, texton-based features, circular shortest paths, boundary fragment models, Haar-Like features, active contouring or compound methods, and intensity-based features are some of the methods for extracting skull features [8]. Later, Hough transform (HT), dynamic programming (DP), and elliptic skull fitting were utilized for additional HC measures [15].

Convolution neural networks (CNNs) have lately acquired prominence due to their better performance in image processing tasks such as classification, object identification, registration, computer vision challenges, and semantic segmentation. A study used fully convolutional networks (FCN) to segregate fetal tissue in 3D ultrasound images. They created a two-stage CNN to extract structural information

[7]. Another study employed multi-task deep neural networks based on Link-Net to segment fetal ultrasound images [15]. FCN was used in another study [9] to segment the fetal head from second-trimester ultrasound images.

Finally, due to its skip connection layers, which may be utilized to detect symmetric structures in an image, U-Net has become useful for biological image segmentation. It decreases computational complexity by detecting the skull, abdominal boundary, and femur target for fetal HC, AC, and FL measurements. The encoders and decoders in the U-Net architecture are used to extract features from the image and then cascade those encoded feature maps to extract spatial features [3]. As a result, it performs well while reducing warping, rand, and pixel mistakes.

1.2 Problem Statement

It is necessary to check the baby's development and to check for any abnormalities at the time of pregnancy. Ultrasound examination is a test that employs high-frequency sound waves to visualize whether or not things are running as they ought to and it is a commonly used method to monitor pregnant women throughout the Western world. Parents get a chance to have a glance at the infant by this inspection. Along with this, another significant diagnostic feature of ultrasound imaging incorporates crucial information in regards to the baby's health. [23].

In low-resource countries, access to care during pregnancy and childbirth is a key barrier to improving maternal and perinatal health outcomes. There is a huge disparity between the developed and developing worlds, with the least developed countries experiencing higher maternal mortality and pregnancy complications. Among the illnesses for which obstetric ultrasound is the principal diagnostic modality are multiple gestation, congenital deformities, fetal development limitation, and abnormalities of placental implantation. An ultrasound can offer critical information regarding pregnancy verification, gestational age, multiple gestation, detection of inborn defects, placental issues, and monitoring fetal position, growth, and amniotic fluid disturbances etc. which can be used for diagnostic purposes.

It is vital to check on the fetus's health on a regular basis during pregnancy. There is a close link between the fetal brain's volume and its stage of development. An MRI scan is performed when a fetus' abnormality is suspected, and the volume and structural properties of the brain can be retrieved from the MRI scan's segmentation. 3D images of the inner parts of the body can be produced by MRIs and magnets are used for doing so. Any modification or transformation in the tissues caused by injury can be identified by an MRI scan. Performing segmentation on medical images using MRI is tough due to ambiguous or nonexistent or overlapping anatomical structure, artifacts, and noise. [13]. On the contrary, in case of ultrasound imaging, images of internal organs and structural shapes are produced using high-frequency sound waves. For examining fetal anatomies, ultrasound exams are preferable as it is economical, harmless, and capable of producing immediate responses. As a result, it is extremely important in obstetrics. The fetal volume can now be used as a biometric parameter to track its development thanks to the emergence of three-dimensional ultrasonography.

Segmentation is the process of isolating or distinguishing a region of interest from its surroundings. It is frequently one of the earliest and most significant phases in image processing, and it is a common and important phase in the quantitative and qualitative analysis of medical pictures. Segmentation, which includes designating the areas in a digital image that relate to the object of interest, is required for the estimation of fetal volume. However, there is no standard method for determining fetal volume. To determine fetal measurements, fetal diameters or outlines are extracted manually using the images found from the current ultrasound scanners [2]. Generally, specialized or expert radiologists or clinicians execute the process of manual segmentation. Apart from performing this using a slice-by-slice technique, it is also achievable for 3D images by bounding the region of interest via annotation [21]. For example, in the case of femur length, radiologists use an interactive marker device to show the femur ends; nonetheless, these measures are subjective and unreliable, and most studies depend on the segmentation which is performed manually. The signal-to-noise ratio of ultrasound images is low because of a property of ultrasonic imaging process, making observation and interpretation difficult. Furthermore, defects such as speckle noise and acoustic shadows, as well as attenuation and low contrast across regions of interest, make ultrasonic image segmentation extremely challenging and frequently result in weak or missing edges, as well as the appearance of artificial edges. As a result, manual fetal component measurements are frequently imprecise and inconsistent. According to several studies, substantial random mistakes in measurements done via manual process are a key reason for estimating fetal weight erroneously, necessitating the reduction of measurement error. It is thus desirable to have an autonomous process to obtain fetal biometrics that is resistant to the position and aspect fluctuations seen in the fetal images [2]. Analysis of ultrasound images provides unique problems, notably for successful segmentation of discrete objects, due to the poor contrast, speckle noise, and aberrations inherently associated with ultrasound images.

Utilizing the proficiency of experts is one of the mentionable benefits of manual segmentation procedure whereas being tedious and susceptible to inconsistency regarding intra and inter observations are some of its downsides [21]. Performing full manual segmentation takes 70-80 minutes and is exhaustive, also boring. Furthermore, both intra- and inter-observer ultrasound image quality differ substantially between institutes and manufacturers. It also relies a lot on how many operators or diagnosticians there are and how much experience they have. All of these factors make hand segmentation more variable, affecting quantitative and geometric assessments of ultrasound pictures. Ultrasound segmentation issues have been popular research subjects in recent years, and they have grown quickly. Autonomous segmentation techniques based on deep learning have lately gained popularity, demonstrating considerable improvements in picture classification and recognition tests.

In short, manual segmentation places far too much reliance on the staff's professional expertise and clinical experience. As a result, the solution is to develop an automated fetal biometric segmentation approach. Automated fetal segmentation is a critical first step toward developing a comprehensive neuroimaging analytic methodology for fetal development research. A reliable segmentation method would

allow for precise measurement of morphological structures, which might be used for prenatal development monitoring and characterization, as well as earlier diagnosis and intervention.

There were numerous flaws in all of the previously published publications. Some studies focused on femur length rather than density. Some studies used procedures that did not produce an accurate result while being near, and they also failed to demonstrate adequate skull shape. Some authors have employed a relatively small dataset to get more accurate results, but this does not guarantee that precise results will be obtained on a big dataset. Only two-dimensional images were utilized in some articles. Furthermore, no other work has revealed the segmentation of three separate locations in order to estimate femur length, stomach circumference, and head circumference. As a result, we've chosen to present a method based on U-net that will take both 2D and 3D images and perform autonomous segmentation, providing us with accurate measurements of head circumference, belly circumference, and femur length. Additionally, we are interested in measuring the circumference of the abdomen and the circumference of the skull, developing an autonomous fetal segmentation method to measure the length of the femur by defining the targeted femur area via improvement in shape detection, as well as implementing a faster method to calculate the length of the femur quickly.

To sum up, based on U-Net architecture, we offer a unique and efficient autonomous technique to perform segmentation on 2D and 3D ultrasound images of fetal head, abdomen, and femur to delimitate fetal skull, abdomen, and femur target. Following segmentation, our method applies elliptical fitting to the extracted skull and abdomen boundaries, which is then utilized to calculate the fetal head circumference and abdominal circumference. In addition, a line fitting technique is used on the femur to quickly estimate the femur length [3], [10], [2]. Our proposed method may segment these three fetal areas simultaneously, which is novel because no other research has done so before.

1.3 Research Objective

The goal of this research is to develop an autonomous method to segment head, abdomen, and femur of fetus which can further be used for head circumference, abdominal circumference and femur length estimation and eventually computing gestational age, newborn's size, weight, due date and monitoring fetal growth. This autonomous segmentation can be achieved using U-Net architecture due to its simplicity and being widely used for biomedical image segmentation. The objectives of this research are:

1. To perform fetal head, abdomen, and femur segmentation.

2. To estimate the head circumference, abdominal circumference, femur length.

3. To compute gestational age, monitor fetal growth.

4. To deeply understand the architecture, advantages of U-Net and procedure of

accomplishing autonomous segmentation.

5. To evaluate the model.

6. To make future recommendations to improve the model.

1.4 Challenges

It is quite evident that conducting the entire proposed methodology was very difficult for us. As the world is presently confronting the most provoking time because of the outbreak of Covid-19 epidemic, it was quite challenging for us to analyze the research experiments without any hitch. Our first focus was to gather ultrasound images of pregnant patients from different hospitals. Due to the Covid-19 circumstances, it was difficult to be physically present in the hospitals to collect patients' data as we had to undergo numerous lockdown periods. Many hospitals are reluctant to share data due to privacy concerns and other legitimate reasons. However, we attempted to gather datasets from hospitals through virtual correspondence, but they were not eager to cooperate because of the limitations as well as the Coronavirus circumstances. Hence, we had to visit the hospitals physically to persuade them. Finally, we managed to acquire the data of various patients from different hospitals. Additionally, as it was difficult to assemble physically, we, the team members, had to communicate virtually to facilitate the collaboration in regards of conducting the research.

1.5 Thesis Outline

The focus of this research was to perform autonomous segmentation on 2D and 3D ultrasound images using U-Net architecture. After training the model with the training dataset containing ultrasound images of two types: 2D and 3D, we aimed to test the process of training using the validation dataset and later using the dataset kept for testing. Evaluating the performance of the segmentation process was another goal of this research.

In the first chapter, denoted as Chapter 1, an overview regarding fetal development and segmentation has been included. Next, the problem statement addresses the concurrent issues or problems of existing methods and the research objective has been added for clarifying the goals and proposals of this research. Lastly, some challenges have been mentioned that are faced while conducting the research.

The next chapter (Chapter 2) highlights the previous works that have been done regarding the autonomous segmentation of fetal and other biomedical images. Major findings of such research works have been summarized in this section, and the shortcomings or inadequacies of previous different approaches have also been discussed. The factors affecting fetal development, different biometric parameters that are vital in determining fetal growth, the significance of using ultrasound images over MRI and a brief overview regarding the necessity of autonomous segmentation

have also been highlighted.

In the following chapter (Chapter 3), the fundamental assessment of U-Net architecture and different biometric parameters have been explored. A generalized workflow diagram has been added to facilitate the research pipeline. Next, the research methodology has been introduced in a detailed manner. The research procedure is explained thoroughly, starting from data collection, then data preprocessing to train-test split.

In the next chapter (Chapter 4), the implementation and model evaluation procedure has been discussed in detail. A comparative analysis of different performance metrics and loss functions to evaluate the implementation has also been added. Next, a detailed explanation has been provided regarding model fitting, training and evaluating the model using the performance metrics that were discussed earlier. Using the segmented images, the procedure of circumference, length and gestational age measurements has also been highlighted.

The next part (Chapter 5) provides a detailed analysis of the results found. Actual values and values found from our experiment regarding fetal head circumference, abdomen circumference, femur length and gestational age have been subjected to a comparative analysis using tables and charts.

Finally, in Chapter 6, the reasons for choosing our proposed model and its limitations have been described. The paper has been concluded highlighting the plans and scopes for future research.

Chapter 2

Literature Review

2.1 Existing Works

This research effort [2] introduces an effective and instinctive approach for fetus's femur segmentation from ultrasound pictures and an efficient system for swiftly determining the segmented femur's length, as femur measurement is one of the most significant elements of fetal segmentation. [2] presents the automatic morphology-based femur detection process first, and then the model is used for detecting the length of femur secondly. The major entropy-based segmentation approach is as follows: firstly, a median filter is applied to smoothen the contents of the image and for decreasing the unnecessary noises and after that, it uses entropy based segmentation for detecting possible pixels from the images and later it looks out for narrow and tall objects and lastly using the morphology and object layout, it chooses the best possible outcome from this technique. Then there's the alternate edge-based segmentation method: this is a segmentation method that only works when the first method fails to identify anything. It marks down the horizontal boundaries and in the preprocessing step, these edges get stretched, similar to the main approach but without the smoothing filter. The object density and the balance ratio of the images are used to construct the thin and lengthy object selection. The procedure used here produced possible result alongside comparing to the ground truth images, and the assessment was followed through distance-based metrics where the measurements is calculated in mm for the Maximum Symmetric Contour Distance (MSD), Average Symmetric Contour Distance (ASD), and Root Mean Square Symmetric Contour Distance (RMSD). The authors have implemented Philips HD9 mid-range ultrasound equipment to capture the images. The results of this paper on automatic femur segmentation and length measuring are are very well-planned, compatible and sufficient. This highly methodical and detailed model is written in C, although C++ or Java might be used instead. It also primarily displays femur length rather than femur density.

A research report [4] illustrates the application of the texton-based supervised method for the accurate segmentation and measurement of ultrasonography fetal head (BPD, OFD, HC) and femur to assure time efficiency and cost-effectiveness over previously done research (FL). It is used to determine the fetus' gestational age (GA) by estimating its total weight and proportion and to detect aberrant fetal growth trends. This topic will increase the quality of biometric measurements

of fetuses while also reducing time and expense. There are various steps to this operation. The first is a nonlinear diffusion approach, which is used to minimize speckle noise. After that, the authors made a presumption that cross-sectional intensity profiles that are made for femur length and fetus skull will be continuously registered to Gaussian-like curves, and To bring out texton properties specific to ultrasonography fetal bone structure, they design a multi-scale and multi-orientation filter bank. Lastly, a closed head contour for the fetal head is constructed using a minimal square elliptical fitting method, while a closed femur contour for the fetal femur is formed by joining the specified femur borders. They gain through extracting features in the texton generation stage which follows, as opposed to the previous method. This improves feature extraction in the subsequent texton generation stage and decreases the false-positive ratio of border identification substantially. In the second stage, these tests provide 32 textons, which are sufficient to build femur and head primitives. To create comparable primitives, the texton map is a grayscale image. After the final stage, it can be seen that the total precision of the two experts for fetal head segmentation are above 97 percent, with a standard deviation of less than 1%, and all other metrics (accuracy, sensitivity, and specificity) values are around 99 percent. In the fetal femur, expert 1 exceeds expert 2 in terms of precision, accuracy, specificity, ASD, RMSD, and their associated standard deviations. Even if it is close, this procedure does not produce an accurate result. This study used photographs from the United States, which are notoriously difficult to deal with. This method aims to generate a whole skull structure, but it only shows parts of the structure at a time, resulting in an incomplete structure. Discriminations between bone and non-bone features are also difficult to identify in the femur. These issues can lead to clinical application disagreements.

Another investigation [8] used computer-aided detection (CAD) to determine the Fetal head circumference in 2D ultrasound pictures. They obtained the HC's ultrasound photos from the Radboud University Medical Center's Department of Obstetrics database in Nijmegen, the Netherlands. The CAD method has two steps: first, to detect the embryonic skull, Haar-like features were obtained from ultrasound pictures and they are used to train a random forest classifier. Second, the Hough transform, dynamic programming, and an elliptical fit were used to find the HC. The CAD system was trained on 999 photographs and validated on 335 images from all trimesters in a separate test set, indicating that it is not valid for additional images. The average difference between the reference GA and the GA estimated by the experienced sonographer in the first place, second, and third trimesters examination was 0.8 2.6, 0.0 4.6, and 1.9 11.0 days, respectively. The average time difference between the reference GA and the GA of the medical researcher was 1.6 2.7, 2.0 4.8, and 3.9 13.7 days, respectively. The GA difference between the reference and CAD systems on average was 0.6 4.3, 0.4 4.7, and 2.5 12.4 days, respectively. To have a clear image of the fetus's progress, researchers identified 1334 2D images from 551 pregnant women. This study's accuracy was significantly greater than in previous studies.

Following this study, CNN was utilized in research work [13] to acquire a better understanding of the image of brain segmentation, and U-net was also employed to get a better understanding of it. From the middle of the mask's brain region (512

x 512), all images were automatically cropped to 256 x 256. This ensures that the brain area appears in every fresh photograph. Following that, each cropped image is filtered using a small 3 x 3 median filter window. This study used a deep learning approach to cover 200 fatal brains with over 11,000 2D photos, which was implemented by U-original net's architecture. Here, Deep U-net was used to automate 2D fetal brain segmentation of MR data, resulting in a clear view of images with fuzzy areas. Large datasets were not previously employed, but this study places a greater emphasis on them. In addition to this, a U-Net design based CNN network is included in this paper. As compared to the original U-Net and its upgrades, in Automated 2D Fetal Brain Segmentation of MR Images that used a deep U-Net 13, the proposed approach produced significantly better results. After testing all of the approaches with a variety of scenarios and obstacles, such as low contrast images, obscure brain areas, diffuse brain boundaries, overlapping brain and skull boundaries, complicated brain structures, and extremely small regions in the bottom and top slices, this has a great result on average of J= 86.7 percent and D = 92.8 percent based on 200 cases. Furthermore, this research has focused on the human placenta, which has a much faster turnaround time. They employed MRI to obtain a clear segmentation image of the placenta, and it was dependent on network accuracy criteria. For training purposes on Titan X, around 50 trained data sets were used on the U-net network. This research project included local and worldwide data, which had not been done before in earlier studies. Still, here is a condensed version of the dataset, which does not guarantee that it will be accurate.

The first trimester is crucial for determining the fetus's growth, hence an ultrasound during this time is also necessary to determine the fetus' gestational age. A single 3D ultrasound scan to detect and measure several fatal anatomies in a first-trimester fetal assessment is offered in the research effort [14] to uplift the valuable image of the fetal. Furthermore, semantic segmentation of the fetus is performed using deep learning and image processing, which is also employed as a standard orientation for biometric observation. They used a V7-3 transducer with a voxel resolution of 0.33mm x 0.33mm x 0.33mm and a Philips HD9 ultrasound machine (Bothell, WA 98021, USA) to collect data. Furthermore, a total of 65 datasets were employed to identify training and test sets (44 among them is training set and 21 is test set). The study includes physical detachment of the head and abdomen in order to visualize these times extremely precisely. In addition, a deep learning and image processing algorithm was used to create a crystal image of the first trimester. To train the networks and test the technique, the VGG-16 network was employed. In these research areas, however, all FCNs used stochastic gradient descent (SGD) and softmax loss. It was begun with 10-3 and then gradually decreased to achieve the desired outcomes. The results demonstrate that it has a 98.9% accuracy rate, which is far higher than earlier research. The goal of the project was to develop a completely automated biometry system that would use 3D fetal ultrasound throughout these crucial early trimesters. This 3D ultrasound will be employed in fetal segmentation sooner or later, however, there are limitations in global and local data, which will have an impact on research.

According to the research work [18], we require a prenatal ultrasound, which is one of the most common examinations during pregnancy that measures the Fetal Head

circumference and determines the gestational age, monitors the fetus' growth status, and infers the newborn's condition. The dataset we used was the HC18 challenge, which included 999 three-dimensional ultrasound images and comments. Every 2D ultrasound photo has a length of 540 x 800 pixels, with pixel lengths ranging from 0.052 to 0.326 mm. One of the challenging aspects of this dataset is that the ultrasound images were collected from all three trimesters of pregnancy. There were 165, 693, and 141 prenatal ultrasound photos taken in the first, second, and third trimesters, respectively. The fetal head circumference is difficult to measure because the ultrasound image of the skull is hazy and faulty. They employed a deep learning U net architecture method to segment the fetal skull for fetal HC measurement so that the physician could make an accurate measurement of the fetus's head and make a further prognosis. The researchers used a U-Net with Encoders, Dilated Convolution Modules, Skip Connections, and SE Blocks and encoders in this scenario. The approach they employed for fetal HC measurement was the fetal skull border and fetal skull. Due to its remarkable performance in biological image processing, the network was modified to use U-Net. Then, following the last encoder, it added dilated convolution layers and Squeeze-and-Excitation (SE) blocks for the U-Net skip connections to segment fetal skull boundaries and merge fetal skulls in 2D ultrasound pictures. The model is built on the HC18 grand challenge dataset, which includes 2D ultrasound images from several trimesters of pregnancy. Researchers obtain 2D pictures of prenatal ultrasound images in this experiment. The length of fetal HC can be estimated using the segmented results. Fetal HC is important for calculating gestational age and due date, as well as monitoring fetal growth status during the pregnancy. It is critical to obtain a correct fetal HC measurement. As a result, they outperformed U-Net, with higher Dice scores and more precise HC measurement estimates. Despite the fragmentary and unclear HC in the ultrasound images, their model can predict a complete fetal skull border and skull, as well as point to useful information. In fetal skull segmentation, U-Net can get a very high Dice score, while the mean absolute difference is lower than in fetal skull border segmentation. To some extent, this work is reliable and efficient, but there are some flaws, such as little features that have always been overlooked in this research and do not specify in-depth knowledge of image segmentation. Furthermore, only global data has been used so far for picture segmentation, which is one of the key problems because research is limited. Instead of having numerous flaws, this model is 96 percent correct, implying that it performs better than other networks. The model also accurately predicts fuzzy images with a 3 mm lower mean difference.

The authors incorporate Gestational age (GA), an explanatory evidence of head circumference (HC) and biparietal diameter (BD) and those are two measurements of fetal growth, in their research effort [17]. (BPD). This paper educated and evaluated proposed networks using a dataset that was freely available. This data was collected at the Radboud University Medical Center's Department of Obstetrics in Nijmegen, the Netherlands. The actual observation was made for the purpose of measuring the head circumference, and the dataset was eventually made public through the Grand Challenge HC18. In this presentation, the BPD and HC are automatically measured using a deep learning-based approach based on fetal head segmentation from ultrasound images. The analyzer in this picture analysis ensures that deep learning techniques are used to produce more efficient algorithms than ex-

isting conventional algorithms various image segmentation and image enhancement applications. The pixel classifier and fetal skull detector were both used to segment this fetal head. Based on the segmentation of the fetal head from ultrasound pictures, this research [17] proposes a deep learning-based technique for measuring BPD and HC. The proposed approach in this study solves ultrasound segmentation concerns such as speckle noise, inherent ultrasound image artifacts, and difficulty when there is a lack of amniotic fluid in the images. So, they utilized MFP-U-net, a previously described effective convolutional network architecture for left ventricle segmentation from echocardiography images.. To segment the fetal head from ultrasound images, the researchers used an effective convolutional network architecture called multi-feature pyramid U-net (MFP-U net), which was previously proposed for left ventricle segmentation from echocardiography images. The suggested network addresses the U-major net's flaw, which is that in the segmentation process, it ignores all semantic strengths. HC18 challenge dataset ultrasound images were used to train and test the MFP-U-net (fetal head circumference challenge). Several measures, including Dice similarity coefficient (DSC) and Hausdorff distance (HD), were used to measure the accuracy of anatomical segmentation that has been required. Dice similarity coefficient (DSC) and Hausdorff distance (HD) were used for measuring the data, and the correlation analysis was performed using Pearson's test. They both had Bland-Altman and correlation graphs, that assessed how well predicted and manually driven parameters agreed, with HC having the best consistency between manual and automatic measurements. The issue is that this experiment used a short dataset, and there isn't enough information on how the technique was carried out. Again, there aren't enough details concerning the fetal ailment, and the work they're doing isn't well-supported.

2.2 Background Study

2.2.1 U-Net Architecture

Olaf Ronneberger et al. proposed and developed U-NET architecture to process and perform segmentation on biomedical images. The "U" shaped architecture is symmetrical consisting of two major parts. The left one is denoted as the contracting path or encoder, is used to frame the context of the images by a typical convolution process. This encoder is basically a combination of convolution and max-pool layers. Here, each process consists of two successive convolution layers with a kernel size of 3x3, followed by a max-pooling layer along with ReLu activation functions. As the convolution process increases the depth of the image, the number of channels changes in each process. This orientation is repeated multiple times.

Another part of this architecture is denoted as the expansive path which is used to upsize the image to its original size via transpose convolution up-sampling technique. The novelty of the u-net is seen in this part. In this part, the feature map is up-sampled by the convolution process, using a kernel size of 2x2. Then the up-sampled image is concatenated with the cropped feature maps of the same level contracting part. The reason for combining the information from previous layers is to get a more precise prediction. To meet the prediction requirements, the image

is reshaped by using a convolution of kernel size 1x1. So, this architecture does not have any dense layers. Hence, images of any dimension can be fed as inputs to this architecture. Thus, by using the context from a larger overlapping area, the contextual information is propagated along the u-shaped network allowing the segmentation of objects. [22]

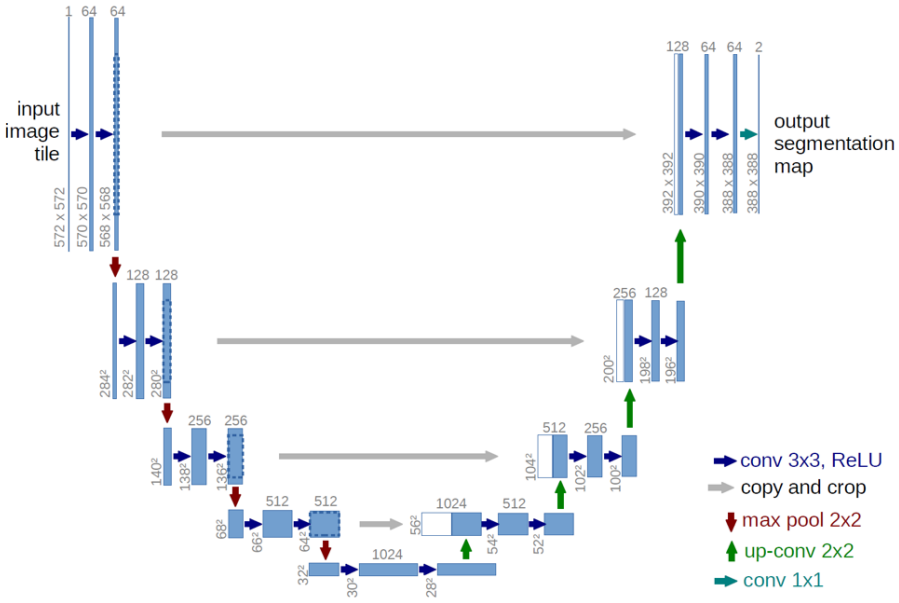


Figure 2.1: U-Net Architecture

In figure 2.1, blue boxes represent feature maps consisting of multiple channels. The number written on the top of the boxes represents the channel numbers. At the lower edge of each box, the height and width are represented. Each white box denotes the copied featured maps from the contraction part and different colored arrows represent various operations as denoted in the picture above.

In figure 2.1, the latter two blue bars in each step denote the usage of two convolution layers in a row. The output tensors produce 9 different outputs where the first four outputs represent the outputs of downsampling layers and the later 4 refer to the outputs of upsampling layers and the last output is the final output with one-dimensional matrix only. Max Pooling Layers' output tensors produce only 4 outputs overall. Typical convolution layers and max-pooling layers are used in the contraction path on the left. The size of the image is reduced as the depth is steadily increased in the contraction part. For example, starting from a width of 128, height of 128 and depth of 3, it gets reduced to a height of 8, width of 8 and depth of 256.

On the right side, transposed convolutions are used along with the typical convolution process. This path is denoted as the expansion path. The file size steadily increases in this path, while the depth gradually decreases. For example, starting at a height of 8, width of 8 and depth of 256 and going up to a height of 128, width of 128 and depth of 1. Gradual usage of up-sampling helps to achieve precise localization.

For achieving more precise locations, at each step of the expansion path, the feature maps from the same level contraction paths are concatenated. Therefore, a horizontal arrow from leftward to rightward in each layer was shown in the above figure which basically denotes the concatenation of them.

After each process of concatenation, two regular convolutions are used consecutively

to facilitate the process of generating even more precise results. The name U-NET is thus taken in accordance with the symmetric U-shape of the design.

The overall relationship can be denoted as follows:

Input ⇒ Contraction Path ⇒ Expansion Path ⇒ Output [12]

2.2.2 Why Ultrasound Image?

Ultrasound is one of the most used methods for prenatal diagnosis, and it is frequently used for monitoring pregnant women, making it a very important test during pregnancy. This approach is considered one of the fundamental imaging modalities in use since it is rapid, affordable, non-invasive, radiation-free, low-risk, and real-time [1]. Ultrasound imaging is milder, more dominant, and less expensive than other imaging technologies like computed tomography (CT) and magnetic resonance imaging (MRI). Moreover, despite the fact that MRI is safe and capable of producing images with superior anatomical definition, medical experts restricted its usage to women under the age of 20 weeks [15]. It is also said to be ineffective in terms of screening for general abnormalities. The most common approach for fetal growth monitoring is 2D ultrasound. However, 3D US has earned notoriety as a result of its speedy acquisition. Furthermore, the scanning quality of both 2D and 3D US is the same. Nonetheless, finding the ideal 2D viewing plane in a 3D scan requires talent [1].

2.2.3 Biometric Parameters

Fetal segmentation is the method of separating several portions of the fetus, such as the fetal brain, abdomen, and femur. These are then used to determine fetal biometrics parameters such as fetal head circumference (HC), biparietal diameter (BPD), fetal abdomen circumference (AC), femur length (FL), humerus length (HL), and crown-rump length (CRL). During pregnancy, estimating gestational age (GA), fetal size, baby size, weight, and due date require these parameters. While measuring the gestational age of a fetus aged between 8 and 12, CRL provides the most exact result. Following that, HC is the most accurate method for determining GA after 13 weeks [8]. These parts must be identified in order to screen for prenatal disorders. However, because US pictures are operator-dependent, they have a number of aberrations such as boundary ambiguity, noise, attenuation, low signal-to-noise ratio, reverberations, motion blurring, acoustic or sound shadows [10]. As a result, these artifacts lead to erroneous observations and interpretations [10]. As a result, identifying anatomical structures, discrepancies, and measuring errors becomes more challenging.

2.2.4 Why Autonomous Segmentation?

The direction of the fetus, the US machine, maternal tissue, specialists' experience, and so on are all aspects that contribute to the failure to identify desirable structures. As a result, for effective biometric parameter measurement, autonomous detection is required. To eliminate intra-observer variability caused by manually calculating

biometric measures, an autonomous method is implemented to reduce measurement time, variability, and the number of steps required. Furthermore, an autonomous approach can extract the Region of Interest (ROI) with expected and desired results [2]. As a result, this method can easily and precisely segment and quantify fetal sections. This strategy quickly improves workflow efficiency.

Chapter 3

Research Methodology

3.1 System Model

The proposed system's purpose is to perform fetal segmentation using U-Net architecture to determine various fetal biometric parameters. In order to do so, the model requires designing a process that takes 2D and 3D ultrasound images as inputs, systematic extraction of necessary features from those images and performing segmentation on the images in terms of three major parts of the fetus: head, abdomen and femur. After the segmentations have been performed, ellipses, lines or polynomial curves are fit into the segmented region based on the image type.

The proposed method aims to perform segmentation on the basis of the approximation of ellipse and bar-like structures to find the geometric profile of the fetal head and abdomen respectively [4]. U-Net is a dedicated network used to localize and differentiate the borders via performing classification and segmentation on every single pixel. This system will be used to perform segmentation on input images for estimating the gestational age and evaluating the growth of the fetus using head circumference (HC), abdominal circumference (AC) and femur length (FL). The implementation procedure can be divided into two tasks that will be performed in a sequential manner: the first one is the detection and recognition of edges and the second one is the process of object fitting. Then, the segmented regions will be used in determining different biometric measurements such as HC, AC and FL.

This model consists of the following major stages:

Data Pre-processing:

→ **Annotation and Mask Generation:** Among the images of the collected datasets, 2D head datasets were annotated by the experts. The next step is to generate masks of these images. Again, the rest of the images were not pre-annotated by the experts. Hence, these images will be annotated manually using annotation tools.

→ **Resizing:** Most of the images are of large pixel values which are not fully appropriate to be fed to the proposed architecture. Hence, the images will be resized to smaller pixel values.

→ **Normalizing:** The resized images will be normalized to get a further improved version of the raw images.

Splitting: After completing all the necessary pre-processing steps, the images will be split into training, validation and testing data.

Training: The raw images and the corresponding masked images separated for training will be fed to the to the model for fitting. The model will start learning the procedure of segmentation.

Segmentation of Testing Data:

The testing data will be fed to the model. Using the U-Net architecture, the testing images will be segmented in accordance with their respective category. For instance, data that are used here consist of head, abdomen and femur images of fetuses. After training the model with annotated images, the model will predict the segmentation of each of these parts of the fetus in accordance with the category of testing images.

Fitting and Measurement:

After the segmentation is complete, an ellipse or line or polynomial curve will be fit to the segmented region based on the image type. Head circumference and abdomen circumference will be measured by ellipse fitting, on the other hand, femur length will be measured by line fitting. Eventually, using the length of the femur, gestational age will be estimated.

Figure 3.1: The Work Flow of Proposed U-Net Based Fetal Segmentation Model

3.2 Data Collection

Data must be analyzed to gain a thorough understanding of the field in which we work, make predictions, and manage operations in a way that adds value. It becomes significantly easier for us to make informed decisions when we collect facts and information. It verifies the issues around the research gap. As a result, data gathering is an unavoidable aspect of any research project.

It can be difficult to collect annotated data for processing prenatal photographs as input data. We needed 2D and 3D ultrasound photos of the fetal head, abdomen, and femur for our study. Finding such a large amount of data was difficult because most hospitals do not maintain old records. We went to numerous hospitals and were eventually able to collect our data from three of them. This dataset contains 1334 2D fetal head images of pixel size 700 by 550 ranging from 0.051 to 0.43 mm. In order to make our model learn correctly and predict the accurate result, we need to feed it with annotated images, thereby, we annotated all the images from a radiologist. The other dataset from the second hospital contains 711 2D fetal abdomen images ranging from 0.037 to 0.418 mm and 1040 2D femur images ranging from 0.0491 to 0.427 mm. Then the remaining dataset contains a total of 316 3D data, amongst which 99 images consist of the fetal head of pixel size 320 by 440 approximately, 103 are of the fetal abdomen with a pixel size of 400 by 600 and the remaining 114 images are of femur with a pixel size of 320 by 440 approximately.

3.3 Data Pre-processing

The dataset used for processing the segmentation task contains third-trimester images of the fetal head, femur and abdomen. The most important part before moving on to feeding the model with data is to pre-process the data appropriately.

3.3.1 Annotation and Masking

Our pre-processing procedure started with generating mask images of the data. 2D images of the abdomen and femur and 3D images of the head, abdomen and femur were not annotated and thereby, we had to manually annotate them using an annotation tool that generates JSON data including the information of the polygon or ellipse. Later, we had to generate mask images from the JSON Mask data. We used the method fillPoly() of OpenCV for achieving this objective.

The 2D head dataset contains annotated data of the raw images. To generate mask images of the data, we extracted the contours of the annotated images so that the contours can be used to fit and fill the ellipse. In order to perform ellipse fitting, we considered the concept of the Hough Transform method and used OpenCV both for finding contour and fitting ellipse. Figure 3.2 depicts the thorough comparison between the annotated image and generated mask image from that.

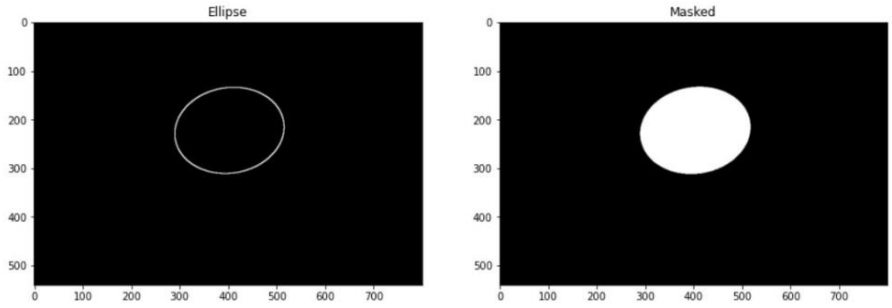


Figure 3.2: Annotated Image vs. Generated Mask Image

3.3.2 Converting to Grayscale

We have a total of 3401 fetal data (both 2D and 3D) with different pixel sizes. Among these 3401 images, 2D images have 1 channel representing grayscale images and 3D data have 3 channels representing colored images. In order to maintain consistency among all the data, we converted all of them to grayscale images using the bgr2gray method of OpenCV.

3.3.3 Resizing

As most of the pixels of our data are larger than 500 x 500 pixels which might over-fit our U-Net model, we have resized the image into 256 x 256 to overcome this problem. The final dimension of the image data after channel conversion and resizing becomes 256 x 256 x 1 which will be used for the rest of the segmentation task procedure. We have used the method transform.resize() of the Scikit image library for resizing the data.

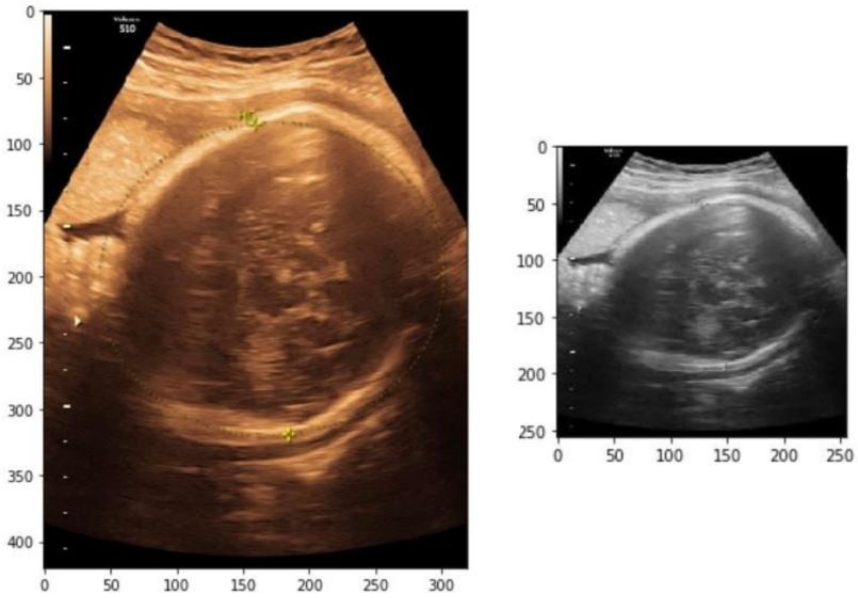


Figure 3.3: Raw Image vs. Resized Grayscale Image

3.3.4 Normalization

For the last stage of data augmentation, we created a slightly improved version of the raw data to ensure a better representation than the previous ones. As the raw data and masked data are of different data types, the model would fail to train and predict them for the inconsistency of the data type and will not be able to map the raw data with their corresponding masked data. Therefore, it is necessary for all of them to have the same data type. In order to achieve that, we divided all the training and testing raw data by 255 to normalize them and store them in an array as a floating-point value. In case of masked data, we normalized them by dividing the data by their mean value and stored them in an array containing the values 0.0 and 1.0 as a floating-point number where 1 will represent the white-colored highlighted area and 0.0 will represent the dark background.

3.4 Train-Test Split

We have a total of 3401 fetal data among which 3085 are 2D and the remaining 316 are 3D data. Amongst these 3401 images, we split them into two categories for training and testing purposes. We reserved 80% images from each of them for training the model and the remaining 20% images for testing and predicting the result. Amongst the 80% of the data that we reserved for training the model, we extracted 25% from them for validation purposes in order to ensure that the training and validating data is being mapped properly. We have used the method train_test_split() to split validation data from training data. The training dataset would be used to train our model and evaluation would be done using the test dataset, which is unseen data for the trained model.

The model is fitted and the parameters are estimated using the training dataset. By evaluating the performance of the fitted model, validation dataset indicates whether the training dataset is causing the problem of model overfitting. So, validation dataset is employed to optimize the model parameters via tuning. Performance evaluation of the final model requires the testing dataset. As the testing data were unseen to the model during the training phase, these testing dataset is capable of providing an unbiased estimation of the final model.

Chapter 4

Implementation and Performance Evaluation

This section highlights the implementation of the proposed model for autonomous fetal segmentation. Model designing and training the dataset are two major sub-processes of segmentation. Our proposed U-Net architecture is capable of creating pixel-wise masks in an autonomous manner for different objects that are present in an image. This section also demonstrates the result of running the implementation of the proposed model for fetal head, abdomen and femur image segmentation. The resulting images after running the model have been added here.

4.1 Model Implementation

As mentioned earlier in the data pre-processing step, the height of the images is 256, the width is 256 and 1 channel has been used. Initially, we started with 128 units and a "3 x 3" matrix, then we gradually doubled the number of feature channels in every down-sampling step like "256 x 3", "512 x 3", "1024 x 3" respectively. However, instead of using 2 convolution layers in every step, we used 3 in order to make our model a bit complicated according to the condition of our dataset so that our model becomes capable enough to learn integrating features of the data. Moreover, at the bottom step, we have used 5 convolution layers instead of 2 as per the above-mentioned model description for exactly the same reason. Furthermore, we used a total of 3 upsampling and downsampling steps instead of 4. Now for the down-sampling layers, we used a "2 x 2" matrix for max-pooling operation and used Batch Normalization before that in every layer to convert the matrix values to a floating point number of each batch in order to maintain the consistency throughout every batch in case the value range goes beyond 0 and 1. This is how downward-sampling works. Then the expansive path performs up-sampling operation using a "3 x 3" matrix and every up-sampling layer uses Transpose of Conv2D and a "2 x 2" matrix before concatenation. The concatenation layers similarly have 3 convolution layers as down-sampling steps. The number of feature channels gets halved by this process gradually like "512 x 3", "256 x 3", "128 x 3" respectively. As per the architecture shown below, the image size is reduced in order to decrease the depth size in the encoder part or the contraction path. We gradually dropped 50% data in every layer of both up-sampling and down-sampling in order to avoid overfitting. In this whole modeling process, the ReLu activation function has been used in each layer

so that the model does not suffer from vanishing gradient problem and the padding has been kept as "same". In the final layer,each 128-component feature vector was mapped using sigmoid activation function so that the output layer produces matrix containing 0 and 1 only as we have normalized our data previously and lastly "1 x 1" convolution has been achieved. Furthermore, Adam optimizer has been used to optimize the model. The learning rate was set to 0.0001. The batch size has been set to 5 with 150 epochs and 15 steps per epoch. The following figure 4.1, visulaizes the modified and implemented architecture to have a clear overview,

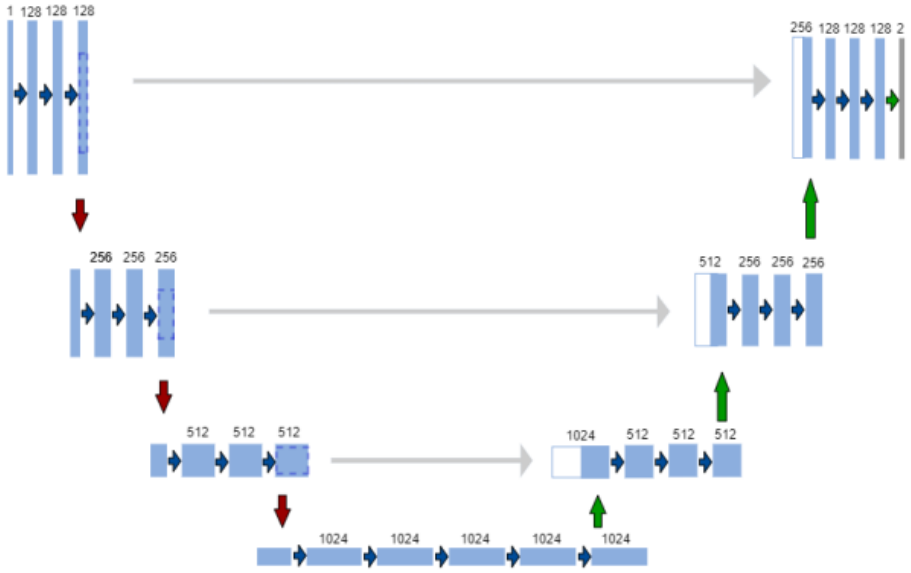


Figure 4.1: Implemented Model Architecture

4.2 Performance Metrics and Loss Function

In order to evaluate our model's performance, we decided to implement the following metrics besides accuracy because of its own integrating characteristics dedicated to giving the best result in case of semantic segmentation tasks since accuracy only aim to deliver result based on the pixel percentage of the image which is oftentimes misleading in semantic segmentation because it will only focus on predicting the negative parts and end up with a large number of false positives instead.

4.2.1 Dice-coefficient (F1 Score)

Dice coefficient is one of the most frequently used performance metrics in the realm of semantic segmentation for quantifying the performance of the model. It is also

known as F1 Score or Sorensen-Dice index. This statistical tool is used for measuring the similarity between two sets of data. It measures the similarity between the mask image and the validated data generated by the autonomous model by a pixel-wise comparison. It can be calculated by dividing the total size of the two images from the size of the overlapped area of the two segmentations. Hence, the formula requires the overlap area and the total number of pixels in both raw and mask images [19] and the formula goes as follows -

$$\text{Dice Coefficient} = \frac{2 \times |A \cap B|}{|A| + |B|}$$

Here, in the equation, the numerator part, $|A \cap B|$ refers the overlapping area of two images whereas the denominator part, $|A| + |B|$ refers to the total number of pixels in both segmented and ground truth images.

Along with measuring the number of true positives, the false positives found by the model are penalized by dice-coefficient. This evaluation metric ranges from 0 to 1 and 0 denotes that there is not any overlap and 1 denotes that there is a perfect overlap. Therefore, the higher the value of dice coefficient, the better the model performs.

4.2.2 IoU (Intersection-Over-Union)

Intersection-Over-Union or Jaccard Index is another most common semantic segmentation evaluation metric. It is quite similar to the previously mentioned dice coefficient metric which also measures the overlaps between two pixels. IoU is denoted by the overlapped area between the images that is predicted from validation and the ground truth or mask image divided by the area of union between those [19]. The formula goes as follows -

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}$$

Here, from the above equation, it is perceived that the numerator part, $|A \cap B|$ represents the overlapping area, whereas, the denominator part, $|A \cup B|$ represents the union area, in short, the surrounded segmented and ground truth area.

Similar to dice coefficient, this performance metric again ranges from 0 to 1, 0 indicating no overlap of predicted and ground truth images and 1 indicating a perfect overlapping of them [19] For multi-class segmentation, as we are trying to achieve by feeding 3 different classes which are fetal head, abdomen and femur, we need to use the mean IoU of the image which is measured by averaging the IoU of each class. In this case also, the larger value of IoU refers to the better performance of the model.

4.2.3 Dice Loss

Previously, binary cross entropy has been used as a loss function which calculates the average per-pixel loss discreetly without having any knowledge of the characteristics of the near pixels whether they belong to the class that we are looking for or not. In consequence, binary cross-entropy doesn't consider its global environment and monitors the loss at its most minimal. Therefore, evaluating the goodness of the model by monitoring the loss of the prediction for semantic segmentation using binary cross-entropy will provide a poor result. In order to avoid such a scenario, we used dice loss which originates from Dice Coefficient [16]. As discussed earlier, the dice coefficient measures the overlapping area of the predicted pixel and ground truth pixel and finds the similarities between them, therefore, the non-overlapped area is the loss of our model. Moreover, dice coefficient considers pixels of both local and global scale which is efficient as the loss information is measured correctly. The formula of dice loss goes as follows,

$$\text{Dice Loss} = 1 - \text{Dice Coefficient}$$

4.3 Model Fitting & Training

While training the data, we have used MSI Radeon RX Armor Graphics Card with 8 GB memory for the most optimized time possible from our u-net model. As mentioned earlier, we provided augmented data which has a pixel size of 256 x 256 with 1 channel referring to a grayscale image. The validation split which is used to predict the fit of the model to the expected output is set to 0.25. Epoch, which refers to training the model using all training data in one cycle including forward and backward pass together, is arranged to 150 with a step size of 15 and each epoch is made of 5 batches. We have implemented the model using Keras in 5 workstations of 31,031,685 total trainable parameters.

4.4 Evaluating the Model

Following the model fitting and training, we had to evaluate the model for our newly modified U-Net architecture by a comparative demonstration of its performance against other image segmentation methods. Image segmentation works on the pixel level of an image to detect the object from the background by separating the image into different classes. For instance, image segmentation of femur detects the pixel density of femur length into one class and the rest of the background into another class including all other parts of the fetus as one element. Identically, for head and abdomen circumference, the pixels within the round circle are classified as one content to understand the shape and boundaries of different fetuses' HC and AC. While different shapes and borders are trained in the hidden layers of the proposed U-Net model, it produces a general shape and boundary using the architecture.

While training the proposed model, 80% of the data is used in the training phase and 20% data is kept for testing purpose. With an enormous number of images, sometimes it's possible not to have large images in both the test set and the training dataset and it can lead to overfitting or underfitting. Therefore, we used 'validation split' to understand if our model is training in the right direction by using the validation set data after every epoch. This U-Net model has used a 25% validation split to evaluate the training phase to tune or enhance the hyperparameter and configurations from time to time. In addition to the validation split, we gained validation loss which is the metric to understand the overall appearance of the U-Net architecture to the validation on the validation phase by calculating the error from each phase and it is calculated after every single epoch. We received an average of 13% validation loss throughout the time of validation set usage and it led the model to understand the femur, abdomen and head circumference in the training phase.

For evaluating the model, measurements that are looked over are accuracy, dice coefficient, IoU, dice loss and etc. Initializing with accuracy, accuracy is the standard that explains how the image segmentation model performs among all classes of equal importance. The result is calculated in percentage to show the accurate pixel representation and in this model, we have acquired training accuracy = 95% or .95 and validation accuracy = 92% or .92 from our fetal head circumference, Abdomen and femur length images. The figure below shows the graph of the training and validation accuracy progress throughout the epoch,

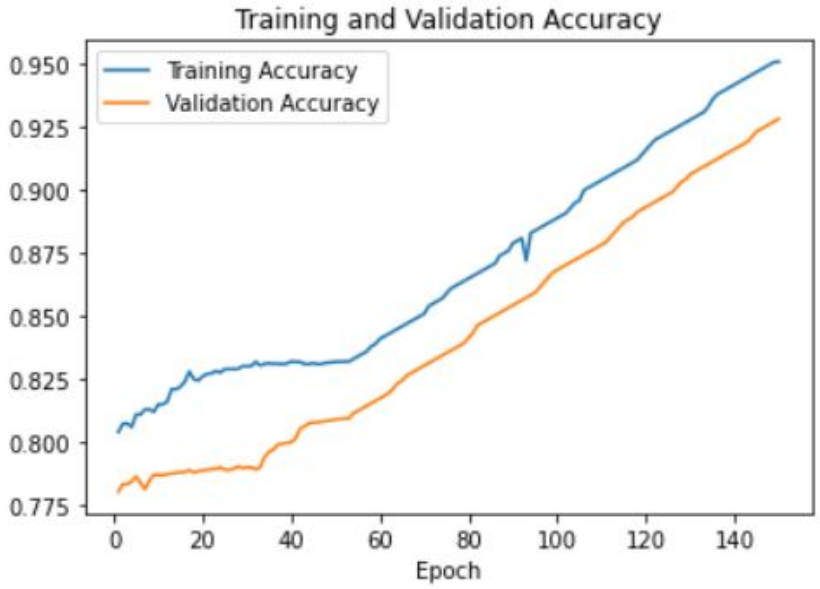


Figure 4.2: Training and Validation Accuracy

Here, Figure 4.2 presents both training accuracy with the blue line and validation accuracy with the yellow line. Training accuracy begins with 0.800 from epoch 0 and with an increment graph line, it ends at epoch 150 holding a final value of 0.95. In the training line, it was increasing till 20 epochs and has a decrease till 60 epochs, after that it has again increased till epoch 150. Moreover, there is a downfall near 90 epochs and still the graph has increased the value more than validation accuracy.

In the validation accuracy, the graph has started with value 0.775 and including an increment graph, it has gained a final value of 0.925. The validation accuracy was quite low at first and it has taken a leap near 40 epochs and kept a similar increment rate with training accuracy. The graph shows a closer value of Training accuracy and validation accuracy for difference of 0.025.

Along with accuracy, this model has used dice coefficient to get even more perfect results among images. As discussed earlier that dice coefficient is a statistical tool used predominantly for semantic segmentation that checks the similarity between predicted segmentation and the correlated original value. The value of dice coefficient also known as dice similarity coefficient is between 0 to 1, 0 is for no similarity between segmentation and mask and 1 is for the complete overlap between the segmentation and binary mask. From the model of this paper, we have gained training dice coefficient = 84% or 0.84 and validation dice coefficient = 74% or 0.74 for medical image segmentation of fetal development. Both Training and validation score is shown in the following graph,

Figure 4.3: Training and Validation Dice Score

Figure 4.3 shows the increasing value of training score with the blue line and validation score with the yellow line with respect to the increasing value of the training and validation dice score. The initial value starts from 0.3 for both dice scores and until 20, both of the values are quite similar with slight differences in values. After reaching 20 epochs, the dice scores have a little dissimilarity in the increasing line and when it reaches 40 epochs there is a sudden drop in the validation dice score which comes back to the usual increasing rate later and here both of the dice scores are 0.65. After 40 epochs the training dice score has a higher increasing rate than the validation dice score and this gap is more significant after 100 epochs. Until 150 epochs the validation dice score reaches 0.74 and

training dice score reaches 0.84. The graph displays an expected small scale variance between both dice scores while getting the utmost dice score from the model.

Intersection over Union (IoU) is another procedure used in this model to calculate the accuracy of the object detector of a specific image. To implement Intersection over Union, firstly the algorithm needs the ground truth boxes which contain the manually labeled images of femur length, head circumference and Abdomen and these are labeled with a python-based application named "label me" for 2D and 3D images. Secondly, it requires the predicted bounding box from the u-net model of this paper and it's a square box for femur length and round box for both head circumference and abdomen. This model of U-net has obtained a training IoU = 72% or 0.72 and validation IoU = 62% or 0.62 for all 3 fields of implementation and it describes that the similarity between the predicted and obtained images are properly achieved by the proposed U-net model.

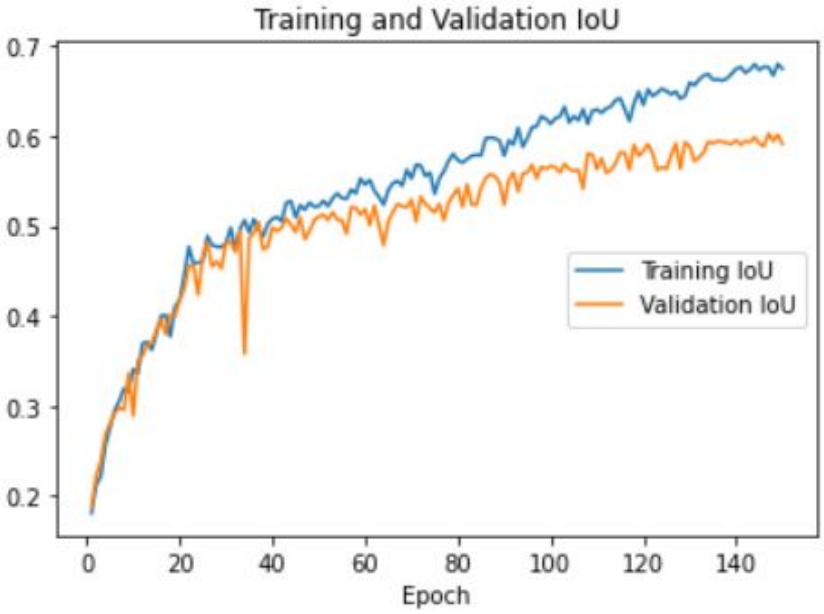


Figure 4.4: Training and Validation IoU

This Figure 4.4 presents the Intersection over Union of both the Training and Validation phase starting from value 0.1 and epoch is 0. This graph has a resemblance to the previous dice coefficient and dice loss graph and so training and validation IoU has almost overlapping lines till 20 epochs and after that has a slight difference till 40 epochs. As the epochs increase, Both of the IoU values increase, including some fluctuations in numerous epochs. This graph has counted 150 epochs just like the previous ones and at the final point the validation IoU has value 0.62 and the training IoU has value 0.72 indicating likeliness of difference with previous measurement metrics.

From the Dice Coefficient, the model obtains dice loss to get a fast and accurate convergence of the model for image segmentation. Dice Loss specifically looks for a data imbalance problem which is a very commonly occurring problem in medical

image segmentation. It solves the problem between foreground and background by creating a balance between generating hard examples and easy examples because easy examples provide less loss function resulting in a less optimal output.

This model for both 2D and 3D image segmentation has shown 0.16 training dice loss and 0.26 validation loss in both 2D and 3D images of femur length, head and abdomen circumference. As dice loss is completely related to dice coefficient it will change if the dice coefficient with more training images.

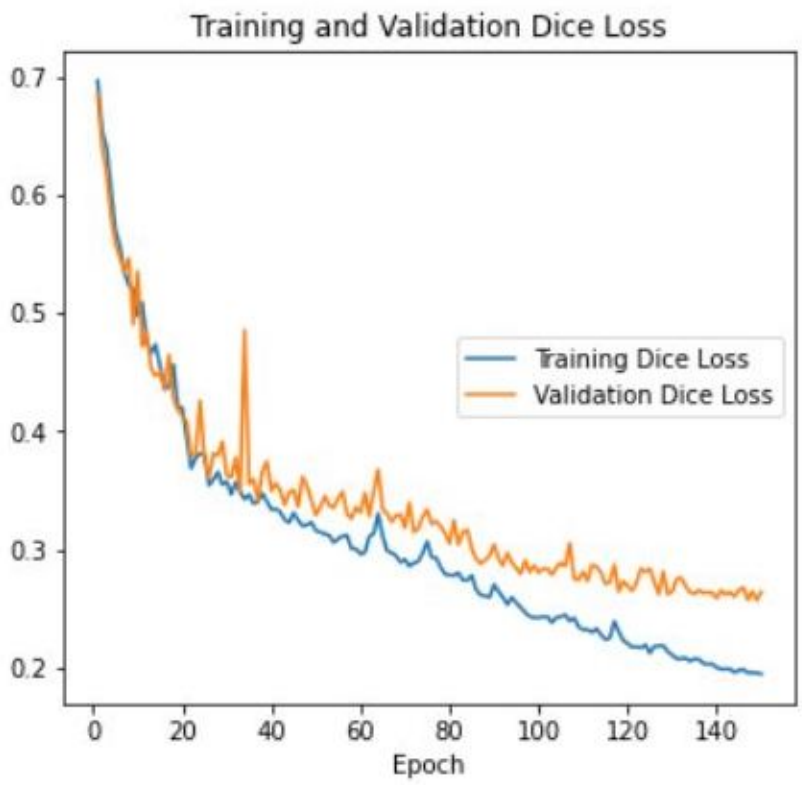


Figure 4.5: Training and Validation Loss

Figure 4.5 conveys the decreasing training dice loss with the blue line and validation dice loss in the yellow line. In the beginning, the dice loss was 0.7 for both and it decreased to 0.35 in the 20 epochs. Just like the dice scores, there are similarities in decreasing between the lines until 20 epochs and after that, the continuous training dice loss is fairly lower than the validation dice loss. Finally, the training dice loss is 0.16 and validation dice loss is 0.26 till the last epoch 150. The dice scores decrease significantly in the first 20 epochs and then it decreases slowly creating a difference between both training and validation scores by creating a complete opposite graph of the dice score.

The higher the number of the epoch, the more likely the model has a higher value in all increment measurement metric graphs. Also, the training line always has an exceeding value than the validation line and it has a little less fluctuation than the validation values. All of the metrics have closer values and it indicates that this is a valid model performing the separation of the object from the background.

Therefore, it can be said that our model is well trained and validated to show a segmented version of 2D and 3D fetus medical images of femur, head and abdomen.

4.5 Circumference & Length Measurement

4.5.1 Head & Abdominal Circumference Measurement

Measurement of head circumference determines the likeness of problems or issues regarding fetal health and regulates the monitoring process. For instance, hydrocephalus or water on the brain can be diagnosed by the rapid growth of the head. Again, excessively slow development of the head could be a sign of a smaller head than expected, which is denoted by the term microcephaly.

On the other hand, amid pregnancy, abdominal circumference (AC) refers to the estimation process of measuring the abdomen circumference of the fetus. AC is another indicator of fetal development which demonstrates whether the growth is normal in terms of size and weight.

There are some steps that we followed to measure the fetal head and Abdomen circumference. First, we found the head and abdomen contour, then we detected the center points using the Hough transform method, then we used ellipse fitting to detect the long and short axis [24].

a. Finding Contour: FindContours, offered by OpenCV, serves the purpose of finding contours in images. To ensure proper detection of contours, images ought to be processed beforehand. We used the .findContour() method of the OpenCV library to find contours. This helped us detect the borders of the fetal head and abdomen and localize them easily in an image.

b. Hough Transform Method: Detection of the center point is presumably the main step in ellipse detection. Determining the midpoint of the line in the image joining two points having tangents parallel to each-other is a generally utilized strategy. On the off chance that these two points are on the same ellipse, indicates the midpoint to be the ellipse center. This rule can be utilized as the base of the HT process. For fitting ellipses and determination of the center point, we utilized a hough transform method. [24]

c. Ellipse Fitting: After detecting the center, we determined the long and short axis of the ellipse using that center. We used the .fitEllipse() method of the OpenCV library to fit the ellipse and get the x,y coordinates along with long and short axes.

Then we used the values of the long and short axis in the respective formulas to measure head circumference [24] and abdominal circumference. [5]

Head Circumference Formula:

$$\mathrm{h} = \frac{(a-b)^2}{(a+b)^2}$$

$$\mathrm{HC} = \pi(\mathrm{a+b})(1+ \frac{3h}{10+\sqrt{4-3h}})$$

Abdominal Circumference Formula:

$$\mathrm{AC} \approx \pi\left[3(a+b) - \sqrt{(3a+b)(a+3b)}\right]$$

Index	EHC(cm)	HC(cm)	Error
1	4.3	4.43	0.13
2	6.39	5.681	0.71
3	7.42	6.875	0.55
4	7.61	6.9	0.71
5	6.547	5.981	0.566
6	6.08	6.98	0.9
7	6.93	6.784	0.146
8	5.8	6.28	0.48
9	5.36	6.21	0.85
10	5.182	6.231	1.049
			Average: 0.6091

Table 4.1: Estimated Head Circumference Result

In Table 4.1, we listed the estimated and actual values of the first 10 predicted images of head circumference. We compared our estimated values with the original one and could see small differences between the values of EHC and HC, that's why we got errors. It can be observed from the table that errors vary from image to image. For example, in the image no. 4, the estimated head circumference is 7.61 cm, and the actual head circumference is 6.9 cm. So we got an error of (7.61- 6.9) = 0.71. However, we got an average error of 0.6091 overall in the context of measuring head circumference.

Index	EAC(cm)	AC(cm)	Error
1	17.16	16.56	0.6
2	17.56	18.3	0.74
3	17.23	16.56	0.67
4	16.77	15.91	0.86
5	13.93	14.51	0.58
6	14.23	13.48	0.75
7	13.3	13.95	0.65
8	28.01	29.3	1.29
9	13.23	14.85	1.62
10	15.71	15.37	0.34
			Average: 0.776

Table 4.2: Estimated Abdominal Circumference Result

In Table 4.2, we listed the estimated and actual values of the fetal abdomen circumference from the first 10 predicted images of abdominal circumference. Here also we could see small differences between the values of EAC and AC. We got small errors in the case of every image. For example, in the image no. 9, the estimated abdomen circumference is 17.23 cm, and the actual abdomen circumference is 16.56 cm, so we got an error of (17.23- 16.56) = 0.67. Here, we got an average error of 0.776.

4.5.2 Femur Length Measurement

Femur length, the largest of the long bones, least movable and easiest to image. It is as accurate as BPD in the prediction of gestational age. Previously, characterizing the presence of dwarfism required the measurement of the femur length (FL). Now it is also proven to be one of the greatest estimators of gestational age. To measure the femur length, [11] we followed the following steps,

a. Finding Contour: Just like the previous one, we used the .findContour() method from the OpenCV library to get contour points which helped us detect the borders of the femur length and localize them easily in the image.

b. Line Fitting: After detecting the contour points, we used the .boundingRect() method of OpenCV in order to perform line fitting based on the concept of Hough Line method and eventually got x, y coordinates along withthe rectangle's height and width. The width obtained from the line fitting is the length of the femur.

Index	EFL(cm)	FL(cm)	Error
1	6.96	6.39	0.57
2	7.2	7.37	0.17
3	3.88	4.1	0.22
4	4.155	4.41	0.14
5	4.8	4.77	0.03
6	1.59	1.68	0.09
7	3.76	4.11	0.35
8	6.52	7.03	0.51
9	6.55	6.8	0.25
10	3.68	3.41	0.27
			Average: 0.26

Table 4.3: Estimated Femur Length Result

In Table 4.3, we have shown the estimated and actual values of the fetal femur lengths which we collected from the first 10 testing images of femur length. Just like the previous two, after measuring the femur length from the predicted segmented image and comparing them with the actual value collected from the hospital, we noticed few errors more or less in every image, because of the differences between the estimated and actual femur length. For example, in case of image 5, the EFL is 4.80 and FL is 4.77, so the error is (4.80-4.77) = 0.03. Eventually, We got an average error of 0.26.

4.6 Gestational Age Measurement

We use gestational age to figure out the duration of pregnancy. Gestational age is calculated in weeks starting from a woman's last menstrual cycle to the current date, indicating the period of time of fetal growth inside the mother's uterus [6]. Between 14 weeks of pregnancy, femur length (FL) measures can be used to correctly predict gestational age. In the third trimester, most observers believe the FL and BPD assessments are equally accurate. Although there is debate about the accuracy of the FL prior to 26 weeks of pregnancy, FL-based gestational age prediction is the most accurate in the second trimester and least accurate near term [11].

After determining the femur length from the rectangle, we used the formula below to calculate the gestational age [11], which only requires one parameter, the femur length,

$$\text{EGA} = 1.863 + 6.280\text{FL} - 0.211\text{FL}^2$$

Lastly, by using the estimated femur length values from the predicted segmented image by following the previously mentioned formula listed in Table 4.3, we calculated

the gestational ages using the formula given above of those first 10 predicted images in the category of femur length and compared them with their actual gestational age given by the hospital and listed them in Table 4.4. We can see that in the case of image 10, EGA is 22.12 and GA is 20.5, so the error is (22.12-20.5 = 1.62). Here, we got an average error of 1.59 overall.

Index	EGA(weeks)	GA(weeks)	Error
1	35.35	33.0	2.35
2	36.14	37.5	1.36
3	23.05	23.2	0.15
4	26.07	24.4	1.67
5	27.14	26.0	1.14
6	11.31	15.0	3.69
7	22.49	23.2	0.71
8	33.84	36.0	2.16
9	33.94	35.0	1.06
10	22.12	20.5	1.62
			Average: 1.59

Table 4.4: Estimated Gestational Age Result

Chapter 5

Results & Discussion

5.1 Training Result Analysis

To have a handful of research, we need to train the data with a suitable model which will give us some feedback about our research growth. After training our model with the data and evaluating them with the proper performance metrics, we eventually came a long way to gain a training accuracy of 0.95, dice coefficient of 0.84, IoU of 0.72, dice loss of 0.16 along with validation accuracy of 0.92, dice coefficient of 0.74, IoU of 0.62 and dice loss of 0.26. A thorough description and comparison between them are discussed below.

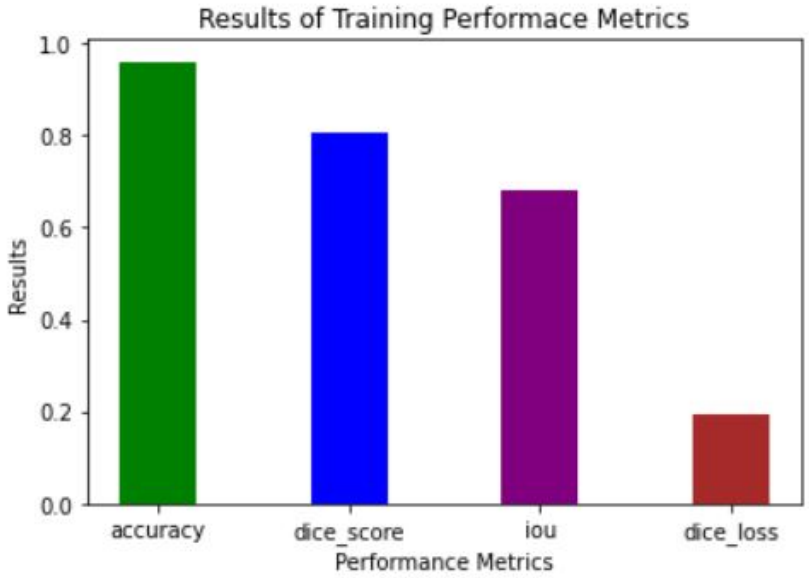


Figure 5.1: Bar Chart of Training Performance Metrics

From Figure 5.1, we can observe the gradual increase of our training performance and end with the result of accuracy as 0.95, dice score as 0.84, iou as 0.72 and dice loss as 0.16. Though the accuracy is higher than dice coefficient and IoU according

to the graph, however, having a result of 0.84 as a dice-coefficient is quite promising. On the other hand, IoU also ends with a decent value of 0.72. Overall, the model loses values nearly around 0.16 which can be considered as a prosperous output.

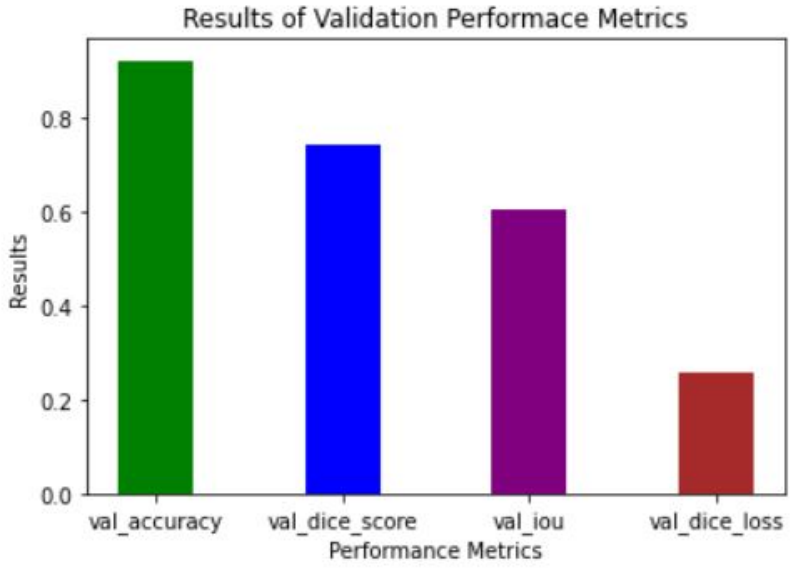


Figure 5.2: Bar Chart of Validation Performance Metrics

From Figure 5.2, it is quite evident that our model was capable enough to learn properly as it eventually acquired a validation accuracy of 0.92, dice score of 0.74 , iou of 0.62 and dice loss of 0.26. If we compare the validation chart with the training chart it can be easily observed that our model becomes weaker to learn complicated features of the data nearly to the end of the epoch and therefore, couldn't raise its validation performance at the end, consequently, it couldn't stand out with their respective training results. As a result, it ended up with a validation accuracy of 0.92, dice coefficient value of 0.74, IoU of 0.62 and a dice loss of 0.26 whereas the corresponding training values were 0.95, 0.84, 0.72 and 0.16 respectively. Therefore, a gap between them became bigger near the end of the epoch, which can be visible from Figure 5.4.

5.2 Segmented Image Analysis

Previously, we had created ground truth images of their respective annotated images, pre-processed our data by normalizing and resizing, designed our model in accordance with our dataset and trained our model by splitting our dataset of 80% by 20% for both training and testing purpose, validated among the training 80% by splitting it to 75% and 25%, eventually, fed the training and validation data with their corresponding mask or ground truth image to our model which facilitates the training process, thereby, ended up getting a favorable result. After training our

model, it is required to make our model predict the testing dataset, perform segmentation of them, visualize the predicted image and compare it with their ground truth which is discussed in the latter part of this section. Before analyzing the result, it is necessary to mention that, a generated predicted image often segments a few incorrect areas considering that as a region of interest, however, that is being discovered as a noise later. These noises, later on, become an obstacle while estimating circumference & length as the wrong area produces incorrect results from what we are expecting. Therefore, in order to suppress these noises, we implemented a simple thresholding after performing prediction on test data. The threshold value we are considering is 0.5 which represents that pixels which are containing value equal to or more than 0.5 are the region of interest we are looking for, below that value are considered as noise in our scenario.

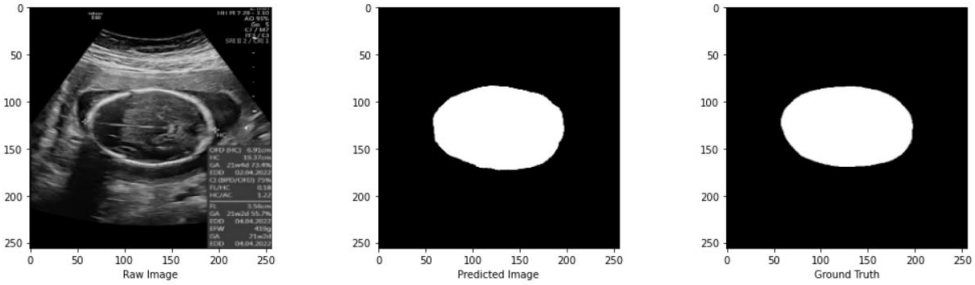


Figure 5.3: Segmented Image of Head Circumference

Figure 5.3 presents the raw image, predicted image and ground truth image of the fetal head circumference consecutively one after another in a row. The boundaries of the predicted image are a bit scattered, especially there is a tiny mismatch at the top right corner and the bottom left corner of the predicted image if we compare it to their ground truths. These could be because boundaries of head circumferences vary from image to image yet our model predicted the shape properly. However, both of the images are 90% similar in oval shape and that is sufficient for detecting a fetal head circumference. Therefore, considering the reference predicted image as a standard one is quite beneficial as it produces a slight error-prone and nearly 100% accurate head circumference value.

Figure 5.4 portrays the raw image, predicted image and ground truth image of the fetal abdominal circumference sequentially in a row. The ground truth is very close to a perfect oval shape with a slight extension in the right most corner. It is evident from Figure 6.4 that the predicted image does have a shape of oval, however, the boundaries are uneven in comparison with the ground truth image. Predicted

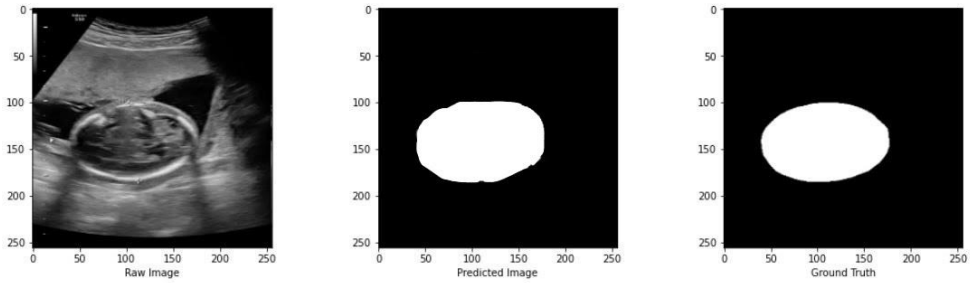


Figure 5.4: Segmented Image of Abdominal Circumference

image has an increased portion in the topmost left and right corner. Despite having uneven boundaries, the predicted image is mostly closer to the mask image and the additional tiny noise can be ignored as the reference predicted images were used to estimate abdominal circumference which was pretty accurate. Therefore, this predicted image shows that every abdominal circumference of fetus does not have a perfect oval shape and it will create different shape depending on the fetus's abdominal growth and size.

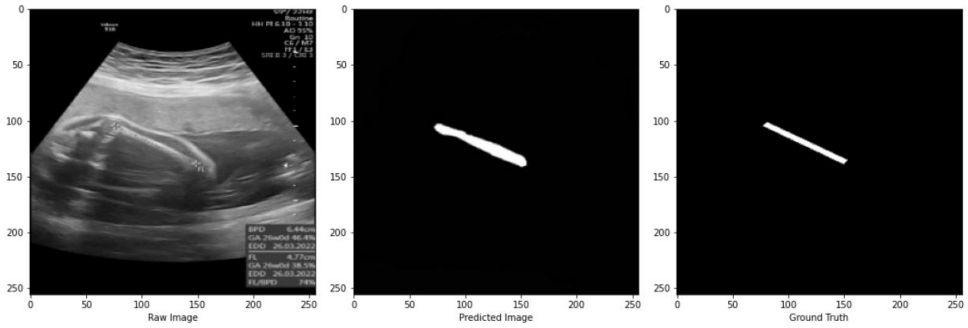


Figure 5.5: Segmented Image of Femur Length

Here Figure 5.5 displays raw image, predicted image and ground truth image of the fetal femur length in a row. The ground truth image has a shape of rotated rectangle which originated from the raw image. The predicted image also manages to show a rotated rectangular shape but with uneven boundaries unlike a straight line as per ground truth image. Moreover, it is moderately broader than the ground truth image and the rightmost part at the end mostly has a circular shape. As femur size changes very quickly in the fetus last trimester, the shape and size also vary in every image and it is considered to be the reason for the predicted image having curving edges at the rightmost portion. Eventually, presuming the reference predicted fetal femur as satisfactory is good enough as the model conducts closely an accurate estimation of femur length and lesser error-prone gestational age from the femur length.

5.3 Circumference, Length & Gestational Age Results

We took the first ten images from each category (Fetal head, fetal abdomen and fetal femur). Then we followed the above-mentioned techniques and formulas to calculate the head and abdominal circumference, femur length and gestational age. We compared the estimated and original values and observed the errors, and table 4.1, 4.2, 4.3 and 4.4 prove that our results are quite satisfactory as the errors are really minor. This shows that our gestational age values are quite accurate too. We got an average error of 0.6091 in the calculation of head circumference, an average error of 0.776 in case of abdominal circumference, an average error of 0.26 in femur length measurement and an average error of 1.59 in gestational age measurement. As these errors are not large, we can easily avoid them and hopefully use them later on with real life experiments as our calculations are pretty accurate.

Chapter 6

Epilogue

6.1 Why U-Net Architecture & Modified Version?

U-net architecture, which originated from CNN, can be explained by making only a few tweaks to the CNN architecture. The original purpose of creating U-Net architecture was to perform biomedical image segmentation. The architecture of the model falls into two core modules which are encoding and decoding networks. A DNN's final result was the only significant part in Classification based problems, however, semantic segmentation facilitates not just pixel-level identification but also a technique to visualize the distinguishing features which were learned at every downsampling stage of the encoder onto the pixel space. This is the primary motivation for using the U-net design.

However, our modified U-Net is designed in such a manner that it is dedicated to learn complicated and intricate features. Instead of using 2 convolution layers in each upsampling and downsampling step, we used 3 in order to make our model a bit complex, thus, increasing the depth of the model. Starting with a unit of 128 instead of 64 will avoid overfitting. As the original U-Net with smaller number of layers failed to learn deeper and complex characteristics of an image, thus, producing noise in the predicated image to a great extent and sometimes were even unable to predict segmented area, therefore, we replaced it with a new version of deeper U-Net model which is significantly better and predicts even more properly than before.

6.2 Limitations

With our own version of the U-Net model, which is better than the original one, we managed to achieve a decent result in every performance metric such as having a training dice score of 0.84, accuracy of 0.95 and iou of 0.72. However, our model still lacks in few areas that need to be taken care of which can easily be inferred by observing the gap between training and validation loss, shown in figure 4.5. Though we managed to minimize the loss in the training phase while still losing 0.16 data, yet it is a bit unsatisfactory to lose a data of 0.26 in the validation phase. The gap gets bigger from the latter half of the epoch till its end. In consequence, despite the fact that our model still learns properly and a rapid improvement is noticed in the training phase but fails to learn at the same pace in the validation phase, thus

improving gradually onwards. This signifies that our model falls into the trap of an overfitting scenario at the end of the validation phase. According to our assumption, the dataset that we are feeding to our model is not large enough in accordance with the complexity of our model. Therefore, an increment in the dataset might resolve our problem by reducing the gap between training and validation loss and give better results than now.

6.3 Future Work

At the end of our research, we were unable to minimize validation loss and thus ended up with a decent but not the best result for example a validation dice coefficient of 0.72. In future, we are aiming to solve the problem that our model is confronting currently which is the scenario of overfitting and unable to segment data extensively. Therefore, we are optimistic to improve our model by decreasing the layers in such a manner that the model performs better with current data instead of manually increasing the dataset. However, we are also expecting to extend our research by working with more biometric parameters which could easily solve our current problem as the dataset itself will get increased significantly and a larger dataset would require a complicated model to learn intricate features. Therefore, it is assumed that the gap between training and validation loss will get reduced in our future addition.

Henceforth, we have focused on segmenting the fetal head, abdomen and femur along with the computation of head and abdominal circumference, femur length and gestational age by injecting the value of femur length so far. In the near future, we want to extend our boundaries in order to work in a diversified manner. We are expecting to work with even more biometric parameters, for instance, Biparietal Diameter (BPD), Crown-Rump Length (CRL), Humeral Length (HL) etc. We are optimistic to segment these additional biometric parameters along with the current ones and figure out the best methodologies to estimate their dedicated values in order to compute gestational age more accurately since BPD is one of the standard biometric parameters which is known to predict gestational age accurately from 14 to 20 weeks. In contrast, CRL is also used to predict the length of an embryo accurately during early pregnancy. Looking more into this case, if we consider working with more biometric parameters mentioned right above, it is presumed that we can monitor fetal development explicitly since we could even predict newborn's size, weight and due date of labor alongside gestational age. Implementing these features in our current model could add prosperous value to our research.

6.4 Conclusion

Ultrasound imaging is one of the reliant mediums for measuring and interpreting head circumference, abdominal circumference, femur length etc. which are some of the most important biometrics regarding fetal development. It has turned into a general phenomenon for diagnosing prenatal anomalies. By estimating gestational

age and monitoring fetal growth using ultrasound imaging, anomalies during the prenatal phase can be detected for further diagnosis. Fetal segmentation is one of the major significant steps in the estimation process of gestational age and other biometric parameters. Due to the presence of various interventional facets, segmentation of fetal parts for prenatal diagnosis using ultrasound images is strenuous [20]. The whole process of segmentation is still dependent on the sonographer. The results of measurements are influenced by the sonographer's expertise regarding the segmentation procedure. Moreover, manual segmentation requires a lot of time. So, an automated system for ultrasound image segmentation can be a stepping stone in the field of determining prenatal biometrics. With the advent of artificial intelligence and neural networking, the automation process can be further modernized. Many types of research have been performed regarding the automation process of fetal segmentation but the accuracy, time requirement and procedure can be optimized further.

In this paper, a fully autonomous fetal segmentation procedure of 2D and 3D ultrasound images using U-Net architecture has been proposed. Because of the symmetric pattern observed in biomedical images, U-net is often utilized for image segmentation. U-net is capable of leveraging the skip connections and decreasing complexity in computation efficaciously [18]. The encoders are subject to extracting features from images and for extracting more semantic features, decoders are used to cascade corresponding feature maps from the encoder part [18]. Our system will take in 2D and 3D ultrasound images collected from different healthcare organizations as inputs. These images will be segmented and an ellipse, circle, or line will be fit to the segmented region based on the images. Different biometric parameters can be determined by using these shapes. During the testing phase, the performance of the model will be evaluated via comparison experiments and verification. Thus, by using this architecture, ultrasound images can be segmented and later can be used for the identification of different biometrics of fetuses. These biometrics are the determining factors of fetal health. Any anomalies can be detected by observing these biometric parameters.

Thus, an autonomous fetal segmentation procedure using U-Net architecture will facilitate the determination of fetal biometric parameters effectively and in an optimized manner, resulting in the detection of anomalies regarding fetal health. This will allow the providers to make appropriate clinical decisions accordingly. We hope that our system will offer a potent solution to various risks during pregnancies and can be utilized in a variety of clinical settings.

Bibliography

[1] M. Yaqub, R. Napolitano, C. Ioannou, A. Papageorghiou, and J. Noble, "Automatic detection of local fetal brain structures in ultrasound images," *Proceedings - International Symposium on Biomedical Imaging*, pp. 1555–1558, May 2012. DOI: 10.1109/ISBI.2012.6235870.

[2] C.-W. Wang, "Automatic entropy-based femur segmentation and fast length measurement for fetal ultrasound images," Jun. 2014, pp. 1–5, ISBN: 978-1-4799-5846-7. DOI: 10.1109/ARIS.2014.6871490.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," vol. 9351, Oct. 2015, pp. 234–241, ISBN: 978-3-319-24573-7. DOI: 10.1007/978-3-319-24574-4_28.

[4] L. Zhang, X. Ye, T. Lambrou, W. Duan, N. Allinson, and N. Dudley, "A supervised texton based approach for automatic segmentation and measurement of the fetal head and femur in 2d ultrasound images," *Physics in medicine and biology*, vol. 61, pp. 1095–1115, Jan. 2016. DOI: 10.1088/0031-9155/61/3/1095.

[5] J. Jang, P. Yejin, B. Kim, S. M. Lee, J.-Y. Kwon, and J. Seo, "Automatic estimation of fetal abdominal circumference from ultrasound images," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–1, Nov. 2017. DOI: 10.1109/JBHI.2017.2776116.

[6] M. Saii and Z. Kraitem, "Determining the gestation age through the automated measurement of the bi-parietal distance in fetal ultrasound images," *Ain Shams Engineering Journal*, vol. 9, Nov. 2017. DOI: 10.1016/j.asej.2017.08.008.

[7] J. Cerrolaza, M. Sinclair, Y. Li, *et al.*, "Deep learning with ultrasound physics for fetal skull segmentation," Apr. 2018, pp. 564–567. DOI: 10.1109/ISBI.2018.8363639.

[8] T. Heuvel, D. de bruijn, C. de Korte, and B. Ginneken, "Automated measurement of fetal head circumference using 2d ultrasound images," *PLOS ONE*, vol. 13, e0200412, Aug. 2018. DOI: 10.1371/journal.pone.0200412.

[9] M. Sinclair, C. Baumgartner, J. Matthew, *et al.*, "Human-level performance on automatic head biometrics in fetal ultrasound using fully convolutional neural networks," Apr. 2018.

[10] M. Sinclair, J. Martinez, E. Skelton, *et al.*, "Cascaded transforming multi-task networks for abdominal biometric estimation from ultrasound.," May 2018.

[11] F. A. Hermawati, H. Tjandrasa, Sugiono, G. Sari, and A Azis, "Automatic femur length measurement for fetal ultrasound image using localizing region-based active contour method," *Journal of Physics: Conference Series*, vol. 1230, p. 012 002, Jul. 2019. DOI: 10.1088/1742-6596/1230/1/012002.

[12] H. Lamba, *Understanding semantic segmentation with unet*, 2019. [Online]. Available: https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47.

[13] A. Rampun, D. Jarvis, P. Armitage, and P. Griffiths, "Automated 2d fetal brain segmentation of mr images using a deep u-net," Nov. 2019.

[14] H. Ryou, M. Yaqub, A. Cavallaro, A. Papageorghiou, and J. Noble, "Automated 3-d ultrasound image analysis for first trimester assessment of fetal health," *Physics in Medicine and Biology*, vol. 64, Aug. 2019. DOI: 10.1088/1361-6560/ab3ad1.

[15] Z. Sobhaninia, S. Rafiei, A. Emami, *et al.*, "Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning," vol. 2019, Jul. 2019, pp. 6545–6548. DOI: 10.1109/EMBC.2019.8856981.

[16] S. Du, *Understanding dice loss for crisp boundary detection*, 2020. [Online]. Available: https://medium.com/ai-salon/understanding-dice-loss-for-crisp-boundary-detection-bb30c2e5f62b.

[17] M. Ghelich Oghli, S. Moradi, N. Sirjani, *et al.*, "Automatic measurement of fetal head biometry from ultrasound images using deep neural networks," Oct. 2020, pp. 1–3. DOI: 10.1109/NSS/MIC42677.2020.9507932.

[18] D. Qiao and F. Zulkernine, "Dilated squeeze-and-excitation u-net for fetal ultrasound image segmentation," Dec. 2020. DOI: 10.1109/CIBCB48159.2020.9277667.

[19] E. Tiu, *Metrics to evaluate your semantic segmentation model*, 2020. [Online]. Available: https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2.

[20] Y. Yang, P. Yang, and B. Zhang, "Automatic segmentation in fetal ultrasound images based on improved u-net," *Journal of Physics: Conference Series*, vol. 1693, p. 012 183, Dec. 2020. DOI: 10.1088/1742-6596/1693/1/012183.

[21] S. K. Zhou, D. Rueckert, and G. Fichtinger, *Handbook of Medical Image Computing and computer assisted intervention*. Academic Press, an imprint of Elsevier, 2020.

[22] N. Siddique, P. Sidike, C. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. PP, pp. 1–1, Jun. 2021. DOI: 10.1109/ACCESS.2021.3086020.

[23] B. Bila, *Ultrasound: 9 ways high frequency sound waves provide important checks on a developing fetus*, 2022. [Online]. Available: https://www.bila.ca/2018/02/the-importance-of-ultrasound-in-pregnancy/.

[24] J. Zhang, C. Petitjean, and S. Ainouz, "Segmentation-based vs. regression-based biomarker estimation: A case study of fetus head circumference assessment from ultrasound images," *Journal of Imaging*, vol. 8, p. 23, Jan. 2022. DOI: 10.3390/jimaging8020023.