AN INNOVATIVE MECHANISM FOR SEGMENTATION TECHNIQUES ON FETAL BRAIN MRI IMAGES

A Thesis submitted

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN
COMPUTER SCIENCE AND ENGINEERING

By

N.SURESH KUMAR
17SCSE301019

Supervisor

Dr. AMIT KUMAR GOEL
Professor

SCHOOL OF COMPUTING SCIENCE & ENGINEERING
GALGOTIAS UNIVERSITY
Plot No 2, Sector 17-A Yamuna Expressway
Greater Noida, Uttar Pradesh
INDIA

NOVEMBER, 2021

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis, entitled "AN INNOVATIVE MECHANISM FOR SEGMENTATION TECHNIQUES ON FETAL BRAIN MRI IMAGES" in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy in Faculty of Computer Science and Engineering and submitted in Galgotias University, Uttar Pradesh is an authentic record of my own work carried out during a period from January 2018 under the supervision of Dr. Amit Kumar Goel, Professor, School of Computing Science and Engineering, Galgotias University.

The matter embodied in this thesis has not been submitted by me for the award of any other degree or from any other University/Institute.

N.SURESH KUMAR
17SCSE301019

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

Dr. AMIT KUMAR GOEL
Supervisor
SCSE

School of Computing Science & Engineering
Galgotias University

CERTIFICATE

This is to certify that **Mr.N.Suresh Kumar (Reg. No. 17SCSE301019)** has presented his pre-submission seminar of the thesis entitled **"AN INNOVATIVE MECHANISM FOR SEGMENTATION TECHNIQUES ON FETAL BRAIN MRI IMAGES"** before the committee and summary is approved and forwarded to School Research Committee of Computing Science & Engineering, in the Faculty of Engineering & Technology, Galgotias University, Uttar Pradesh.

Dean – SCSE

Dean – PhD & PG

The Ph.D. Viva-Voice examination of **N.SURESH KUMAR** Research Scholar, has been held on _____

Supervisor

External Examiner

APPROVAL SHEET

This thesis/dissertation/report entitled "**AN INNOVATIVE MECHANISM FOR SEGMENTATION TECHNIQUES ON FETAL BRAIN MRI IMAGES"** by **N.SURESH KUMAR** is approved for the degree of Doctor of Philosophy.

Examiner Supervisor Chairman

DEDICATION

"This research is dedicated to my father K. Natarajan and my mother N. Bhuvaneswari who sacrificed their comfort and resources to make sure what I am today. Special dedicated to my wife Kiruthika, Daughter Harshada and Son Kavin for encouraging me all the way at the expense of their quality time. Additionally, dedicated to my brother Kirubakaran, Sister-in-law Baby Shalini and Son Nivin for encouraging and trusting me all the time."

ACKNOWLEDGEMENT

Working as an Assistant Professor and doing research for the degree of Ph.D. in Galgotias University was quite magnificent and challenging experience for me. In all these years, many people directly or indirectly contributed in shaping up my career. It was hardly possible for me to complete my doctoral work without the precious and invaluable support of these personalities.

I would like to give my small tribute to all those people. Initially, I would express my sincere gratitude to my supervisor Dr. Amit Kumar Goel, Professor, School of Computing Science and Engineering for his valuable guidance, enthusiasm and overfriendly nature that helped me a lot to complete my research work in a timely manner.

I must owe a special debt of gratitude to Hon'ble Chancellor and Hon'ble Vice-Chancellor, Galgotias University for their valuable support throughout my research work.

I express my sincere thanks to Dr. Munish Sabharwal, Dean School of Computing Science & Engineering and Dr. Naresh Kumar, Dean PG & PhD for their guidance and moral support during my research work and all faculties of School of Computing Science & Engineering who helped me a lot in my course of research work and all those who stood behind me.

Nothing is possible without the constant support of my family. I would like to convey my deep regard to my parents for their wise counsel and indispensable advice that always encouraged me to work hard for the completion of my research work.

N.SURESH KUMAR

ABSTRACT

In the medical field, the intensifying use of Artificial Intelligence and Machine Learning tends to lower the time and accuracy in detecting illnesses in their early stages. AI will have a 10.4 percent influence on the Indian economy in 2030, amounting to 0.9 trillion dollars. More than 80 anomalies in a child's fetal development have been discovered yet. Machine Learning in clinical imaging brings in an exciting new era of reengineered and rethought clinical capabilities. Deep Learning makes it simpler for doctors to feel like they're wandering in the park, but with a few more desirable resources. Brain segmentation of fetal MRI is a new work on the completely automated treatment has been published. Automatic brain segmentation methods established for adult MRI cannot be used for studying fetal brain development, since the geometry and tissue morphology of the fetal brain are substantially different. In this thesis, two approaches for segmenting the brain and two methods to localize the fetal brain and its abnormalities are performed.

It takes more time to manually identify the lesion tissues in the fetal brain. A convolutional neural network is a form of neural network that is most frequently applied to image processing problems. A computer recognizes artifacts in a picture and utilizes convolutional neural networks that are so essential in deep learning and artificial intelligence today.

Object Detection is one of the thought-provoking tasks to be accomplished in computer vision. The basic aim of this initial phase of the study is to detect objects and to interpret the type of object. The research work includes the identification and description of clinical items. The model is trained in combination with photographs and images using YOLO v3 object detection techniques to identify the medical objects. The model is programmed in the cloud system to do well in a short time of preparation. The work summarizes the model that has been developed with the potential to recognize and detail unknown objects,

whether static or moving in a real-life context. The program not only shows knowledge in the form of text but also spells out text in an artificial voice to help you easily understand the object. The model has an accuracy of 98.62% to detect the objects in the clinical practices.

Classification and object identification is the second objective of this research work. The darknet yolov4 is used, to perform the classification, and region of interest detection with the best accuracy scores. The model is trained with the Tesla GPU and obtained the results of the existing techniques in the field of fetal brain classification and localization. The accuracy of 97.92% and precision percentage of 96.70 is achieved in the research work.

The Fetal Brain is one of the vital organs that must be properly and regularly targeted by an obstetrician in order to track the child's growth. More than 76 abnormalities can be discovered during the first tri-semester of pregnancy. It is feasible to avert and save a child's life if the flaws are found during the child's early diagnosis. The optimal answer to this challenge will be automatic sematic segmentation using a U-Net architecture and deep learning techniques. The Precise Slice Enhancement (PSE) Algorithm is used to perform the segmentation on the fetal brain efficiently and secondly, Optimal Semantic Blend (OSB) Algorithm was implemented to enhance the results with different innovative representation for insight to the fetal brain. To train the model, 47 sets of MRI images with various defects were employed. The model is fed an image of T2-MRI images as input. As a help to the doctor, a better segmented fetal brain image with mask, structure, and an unique enhanced image will be shown. With the interpretation, the physician will be able to spot fetal anomalies.

The next phase in the research is to categorise and segregate the anomalies in fetal brain MRI images. The raw images given are used to assess the location of normal and lesion tissue. The model will be able to perform localization, segmentation, and augmentation of the fetal brain using the Improved Semantic Blend Segmentation Algorithm, as well as

treat two critical abnormalities such as encephalocele and arteriovenous malformation. For each input, the model has been trained to segment the Region of Interest (ROI) in an average of 7.2 seconds. There are more opportunities for slipups and mix-up, which increases the risk. Using a rapid cloud environment, the Machine Learning method is used in fetal MRI scans to discover and pinpoint fetal brain anomalies. The art of finding the abnormalities is a significant step in the field of medicine to eradicate or reduce the impact of the fetal brain problem. In this study, the classification and localization of Encephalocele and Arteriovenous Malformation from fetal anomalies is obtained and evaluated using a machine learning algorithm to determine the size of the lesion tissue and classify the kind of abnormalities. This way of processing the classification and detection of irregularities will allow the physician to prepare the diagnosis straightforwardly. Participate With all-existing procedures, accuracy and time are comparably lower. As this Tesla P100 GPU is employed in the cloud environment, the output is more convenient and accessible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
CNN	Convolutional Neural Networks
COG	Centre of Gravity
CPU	Central Processing Unit
CT	Computer Tomography
DL	Deep Learning
DNN	Deep Neural Network
EI	Enhanced image
FCN	Fully Convolution Networks
FPS	Frames per Second
GPU	Graphical Processing Unit
HOG	Histogram of Oriented Gradients
LM	Label Map
ML	Machine Learning
MM	Digital Mammography
MR-CNN	Mask Region based Convolution Neural Networks
MRI	Magnetic Resonance Imaging
OH	OneHot
OSB	Optimal Semantic Blend

PEL	Precise Epic Localization
PSE	Precise Slice Enhancement
R-CNN	Region based Convolution Neural Networks
ReLU	rectified linear unit
ROI	Region of Interest
RPN	Region Proposal Network
SBS	Semantic Blend Segmentation
SI	Segmented Image
SVR	Slice-to-Volume Reconstruction
TF	Tensor Flow
TNF	True Negative Fraction
TPF	True Positive Fraction
TPU	Tensor Processing Unit
WM	Weight Map
WSI	Whole Slide Tissue Image

LIST OF PUBLICATIONS

JOURNALS:

[1] N. Suresh Kumar and T. Kumar, “A Civilized Method to Fetal Brain Segmentation with U-Net Architecture using Optimal Semantic Blend Algorithm,” Int. J. Emerg. Technol., vol. 11, no. 2, pp. 187–191, 2020. **(SCOPUS Indexed)**

[2] N. S. Kumar, A. K. Goel, and T. Kumar, “An Intuitive Framework to Segment the Fetal Brain Abnormalities using Improved Semantic Blend Segmentation Algorithm,” Int. J. Curr. Res. Rev., vol. 13, no. 10, pp. 165–169, 2021, doi: 10.31782/IJCRR.2021.131013. **(SCOPUS Indexed)**

[3] S. Vijayalakshmi and N. S. Kumar, “A Review on Fetal Brain Structure Extraction Techniques from Human MRI Images,” Int. J. Comput. Sci. Eng., vol. 6, no. 4, pp. 239–242, 2018.

[4] N. Suresh Kumar and Amit Kumar Goel, “A Lucid Machine Learning Algorithm for Detection and Localization of Fetal Brain Abnormalities,” International Journal of Emerging Technology and Advanced Engineering. [Communicated] **(SCOPUS Indexed)**

CONFERENCES:

[5] N. Suresh Kumar and K. G. Amit, “An Optimized Approach to Clinical Object Identification using YOLO v3 in the Cloud Environment,” in 2021 IEEE International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021, Mar. 2021, pp. 856–859, doi: 10.1109/ICACITE51222.2021.9404592. **(SCOPUS Indexed)**

[6] N. S. Kumar, A. K. Goel, and S. Jayanthi, “A Scrupulous Approach to Perform Classification and Detection of Fetal Brain using Darknet YOLO v4,” in 2021 IEEE International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021, Mar. 2021, pp. 578–581, doi: 10.1109/ICACITE51222.2021.9404656. **(SCOPUS Indexed)**

CHAPTER 1

INTRODUCTION

1. Introduction

1.1 Medical Imaging

Medical imaging states the practices and procedures used to capture and generate images of the human body for medical resolutions, such as medical treatments and diagnostics, or medical sciences [1]. Medical imaging is a most treasured tool in medicine. Magnetic Resonance Imaging (MRI), Computer Tomography (CT), Digital Mammography (MM), and other imaging techniques are excellent ways of non-invasively mapping an object's anatomy. MRI [2] and CT [3] modalities are commonly used imaging techniques in the diagnosis of lesion tissues.

In this day and age, the huge amount of rising for the medical needs, the service of computers to ease and process the digital images for the physicians to interpret them [4]. Physicians now have access to high-quality pictures because to advances in imaging methods like computer tomography (CT) and magnetic resonance imaging (MRI). The MRI images provide very high precision and accuracy when compared to the CT images. The Fetal Brain abnormalities can be addressed efficiently by using MRI images. Meanwhile, medical staff has far more comprehensive treatment of a large number of images. In general, segmentation is an important stage in this process. Manual segmentation, on the other hand, takes a long time and results are not reproducible or variable across observers and inter-observers.

1.1.1 Digital Image Processing

The digital image is a 2D area of quantized intensity values or numbers so digital image processing [5] [6] means how to manipulate these numbers. The Digital image is the 2D area of numbers that have x and y coordinates. Suppose, visual quality or noise needs to be improved for an image, then the numbers have to be manipulated. So, manipulation of these numbers is mainly image processing. It removes the noise or improves the brightness or contrast in the image. In the image equation, the light is converted into an electrical signal, it produces an analog signal which is an analog image. The analog image can be converted into a digital image by the process of sampling and quantization. The sampling needs to be performed along the x-direction and sampling along the y-direction. It is called spatial sampling.

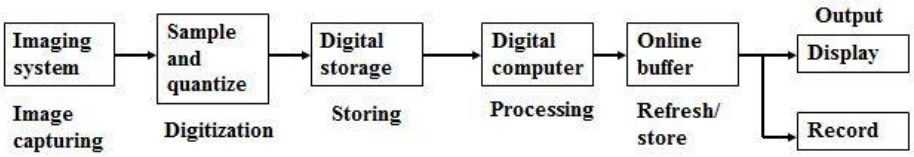

Figure 1.1: Steps for Digital Image Processing

The processing of digital pictures using digital hardware, most often a computer, is referred to as digital image processing. The manipulation of the quantized intensity value, the manipulation of the numbers which represents the typical digital image processing sequence in the above figure 1.1.

The light is coming from the source and it is reflected by the object and the imaging system converts the light photon into an electrical signal to get an analog image. Analog is a continuous function of time. To convert the analog signal or image to digital, the sampling and the quantization have to be processed. The digital image can be stored in digital storage. It can be processed and displayed on digital computers.

1.1.2 Data Acquisition

Image sensors are available for image acquisition and after processing the digital image is acquired. The digital image can process by image processing hardware and image processing software so the process image can be displayed and it can be stored. The light is coming from the source and it is reflected by the surface and optics is mainly the camera. When the light is reflected by the surface, the image is obtained in the image plane as represented in figure 2 below.

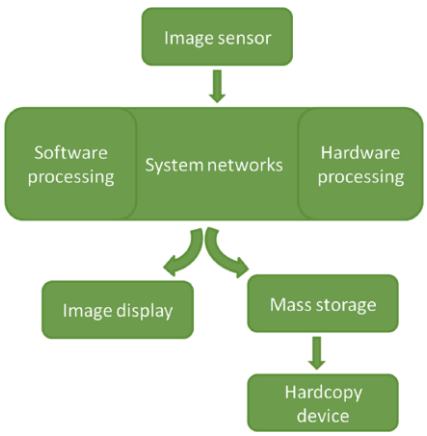

Figure 1.2: Data Acquisition

The sensor will convert the light photon into an electrical signal, the pixel intensity depends on the amount of light coming from the surface so if I consider one pixel at this point, this pixel value or intensity value depends on the amount of light reflected by the surface. The pixel value depends on the surface property. The surface property is called the reflectance property of the surface which is the amount of light reflected by the surface. This process constitutes form the typical image formation system.

Computer vision depends on the physics of imaging, which includes the camera, light, and colour. The use camera to acquire the image of the human organ and analyze the position of the camera. The measurement of light is a significant step in image acquisition. The amount of light emitted by the source and reflection received from the surface is the measurement of the light, which is termed radiometry. The final important characteristics are the colours, the underlying mechanism determines the efficiency of the results of the image acquisition.

1.1.3 Imaging Modalities

An X-ray is a form of radiation it penetrates the body to produce an image. physiotherapists commonly use X-rays [7] to evaluate a possible fracture-dislocation and bone infection. X-ray has many benefits it can be performed quickly they are readily available and cost-effective. The main limitation of X-ray is it produces a 2-dimensional image of a 3-dimensional object, so a minimum of 2 different views is often required to evaluate the site of injury.

CT scan fires several X-rays from different angles it can produce a detailed cross-sectional image of the bone blood vessel and soft tissues. Physiotherapists commonly use, CT scans to evaluate complex regions of the body these include the spine, pelvis, ankle, and foot. CT scan [8] has many benefits it can detect small fractures that may have been missed on an X-ray. It can also produce a 3-dimensional image which makes it easier to locate the site of injury. CT scan has many limitations it is relatively expensive and compared to the plain X-ray it can expose the person to a greater amount of radiation.

Ultrasound creates images by using high-frequency sound waves. Physiotherapists commonly use ultrasound to evaluate soft tissues such as the muscle-tendon and

ligament. Ultrasound [9] has many benefits it does not expose the person to any radiation. They are readily available cost-effective and most importantly they can be performed during movement, this is particularly helpful in the evaluation of shoulder impingement. The main limitation of Ultrasound is, it cannot penetrate the tissues and produce an image of deep structures such as the meniscus and the ACL.

MRI uses strong magnets and radiofrequency waves to produce an image. Physiotherapists commonly use MRI to evaluate the meniscus, labrum, spinal, disc joint, and soft tissues. MRI [10][11] has many benefits it does not expose the person to any radiation it provides detailed information about the body structures. This enables us to make an accurate assessment of soft tissue injuries. MRI has many limitations; it takes a lot longer to perform it is relatively expensive and it cannot be performed in people with certain metal implants as some metals can heat up or attract to the magnet. In summary X-ray and CT scans are often used to evaluate bony injuries such as fracture MRI and ultrasound are often used to evaluate soft tissue injuries including the muscle, tendon, and ligament.

1.1.4 Types of images

In general, in our daily life, many images are addressed, some are natural and some are synthetic. Natural images are being taken for the natural objects with the help of cameras and while synthetic images are generated by using computer programs. The images can be categorized broadly into four categories. Types of images are based on the attribute, based on the colour's images, based on dimensions, and based on data types [12]. In each and all category contains certain types of images that are associated with it. Based on attributes, two possible types of images are classified, raster images and vector graphics. Based on colour; binary image, grey image, colour image, and pseudocolour. Accordingly, various types of images exist under every category as stated in figure 1.3.

The first type of image categorization is based on attributes; there are two types: raster image and vector graphics image. The raster images are pixel-based and the quality of raster images is dependent on the number of pixels [13]. The operations on raster images are like enlarging images or sometimes reducing image quality as well. In vector graphics, it uses basic geometric attributes like lines, circles, triangles, etc., to describe an image. The notation of resolution is practically not present in the graphics.

In raster images, the pixels are visible while in vector graphics no notation of pixels is practically present.

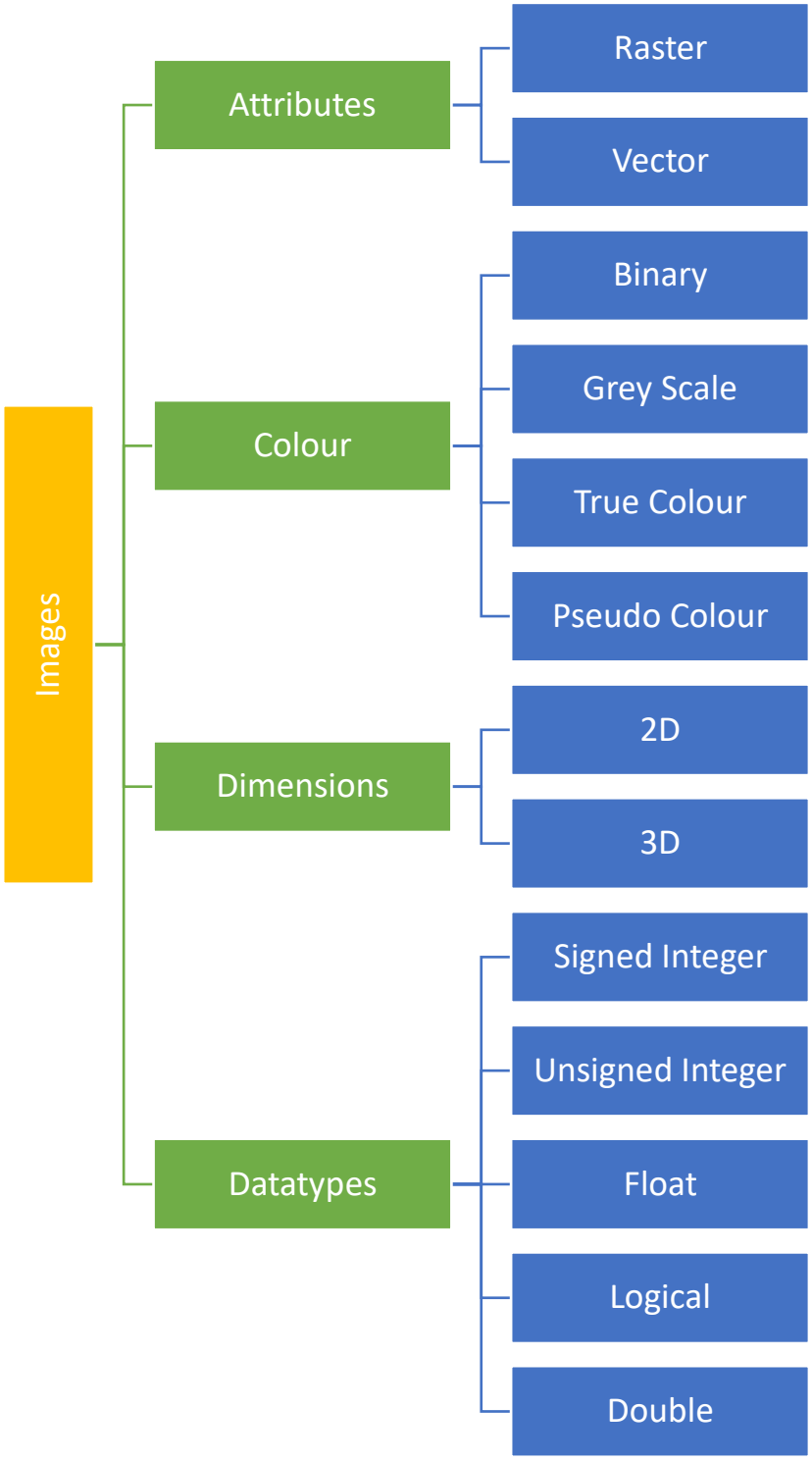

Figure 1.3: Types of Images

The second type of image categorization is based on colour, it can be further classified into 4 subcategories, like binary image, grayscale image, true-colour image, and pseudo colour image. Monochrome images are binary and grayscale images, whereas true colour images depict the whole spectrum of accessible colours. These are fake colour images in the case of faux colour photos. A binary image is often referred to as a bi-level image. Pixels can assume two values either 0 or 1. so in the case of a binary image, only one bit is sufficient to represent the pixel value. Binary images are created from grayscale photographs by a threshold process in which the image's pixel value is compared to the threshold value; if the image's pixel value is greater than the threshold value, the image is considered 1; otherwise, the image is considered 0.

Grayscale pictures feature various shades of grey in between black and white; if an image has 8 bits, then 28 equals 256; hence, grayscale may depict a total of 256 levels. Because the human visual system can distinguish only 32 grey levels. So, out of 256, one level will be a complete Black Label and another will be a pure white label. Black Label can be represented by level 0 and white level by level 255. So, starting from 0 to 255 total, 256 levels are available in between the black and white various grey shades. The majority of medical imaging is grey scale, including x-rays, CT scans, MRIs, and ultrasounds.

True colour images are also known as full-colour images, which are represented by the full range of available colours. It uses 24 bits to represent all the colours so they can be considered as three-band images. Eight bits for one band, eight bits for another band, and the remaining eight bits for the third band. So, the number of possible colours will be 28 coming out of 256 x 256 x 256 = 1,67,77,271. True colour photographs allow for such a wide range of colour tones. The faux colour pictures are fake colour images in which the colours are intentionally created depending on data interpretation. Colour pictures are particularly prevalent in the medical field; Doppler colour images are an example of pseudo colour images. Colour pictures, such as true colour images, are very commonly utilized in image processing. For example, a Remote Sensing image comprises of 3 to 11 bands in an image, thus this information is outside the human vision range. As a result, the usage of pseudo-colour graphics became necessary.

The next category of images is based on the dimensions, generally of 2D images and 3D images. Digital images are 2D rectangular arrays of pixels but if in the 2D one

additional dimension is being added which is known as depth or the additional dimension can be of any other characteristics as well. Then, the type of image is known as a 3D image. The next category of images is based on the data type, binary image, grayscale image, and colour image. The binary image is a 1-bit image that depicts black or white pixels. Grey images maybe 1- or 2-bytes images. Those many possible Shades will be there, in case of one byte 8-bit number of possible grey Shades will be 256, while in the case of two bytes number of possible Shades will be 216 which is coming out 65536 colour images. Colour images are used mostly 24 bit and sometimes it uses 32 bits also to represent the colour and intensity values.

Negative numbers and floating-point numbers are expressed in this way. As a result, signed and unsigned integer types are used to handle negative numbers. The first bit determines whether an integer is positive or negative in various data formats. Based on the first piece of information, it will decide if the number is positive or negative. All values from 0 to 2^{n-1} are represented by the unsigned integer with n bits.

Floating-point data is stored in a particular notation. For example, 1230 can be represented as 0.1230 x 104, where 0.123 is known as significant and power is known as an exponent. So, there are many floating-point conventions also in images.

The final category of images is based on domain-specific; it is based on domain and applications where such images are encountered. There are two types of images, range images, and multispectral images. Range images are also known as depth images because while all other types of images are known as intensity images. In computer vision, range pictures are encountered, where the pixel value represents the distance between the object and the camera. The other form of image is a multispectral image, which includes the infrared and ultraviolet areas. A spectrum is a multispectral image that includes the infrared and ultraviolet regions. It is mostly encountered in Remote Sensing applications, which can take different bands of visible or infrared regions. So, depending upon the application areas, images can be classified.

1.1.5 Convolutional Neural Networks for Medical Image Analysis

Humans' experts need to transfer knowledge to machines for the learning process with minimal human effort especially in areas that require expertise like medicine [14]. The complementary skills of humans in knowledge abstraction and machines in data

processing as a synergistic process in which, humans can provide some action and machine provides processing extracts information presents that extra information to the human. Then, the human can be guided based on the information to select the max action. The human is the center of the process of interacting with a classifier, interacting with a decoder, or interacting with a feature extractor [15]. So, the pipeline that takes the image extracts features and classifies them is usually used for object detection identification, and the next pipeline extracts features for the decoder is used for object segmentation.

The backpropagation [16] has brought many interesting problems for deep learning, as deeper as the network then you have many problems that appear with backpropagation for example that makes it difficult to fix so to improve the parameters of the filters in the first layers.

The user will interact with a sequence of convolutional layers so the idea of the future structures is represented by a sequence of convolutional layers. A sequential self-convolutional layer essentially transforms the feature space they put images into another feature space and again and again. so, this feature space can be thought of as a global feature space. The feature space to get when you concatenate the output of each filter in the convolutional layer as a single feature vector. The idea of this process is actually to map images from different classes into different subspaces of these global feature spaces. It manages to do that; the classifier will not have any problem in separating the classes [17].

1.2 Artificial Intelligence

According to Nilsson, the purpose of AI work is to create robots that can execute jobs that would ordinarily need human intellect. Whether it copies the way human intelligence works is not specified. But it performs tasks that would require intelligence. So, what is involved if some AI system has to work, that requires intelligence for a human being to do the same work [18]. This would require interaction with the real world. That is, the system should be able to perceive, understand and act. It would involve things such as speech recognition, image understanding. The next thing would be reasoning and planning involving modelling the external world, planning and decision making, and dealing with unexpected problems and uncertainties. Thereafter, any intelligence system should be able to learn and adapt.

That is, should have internal models which should be always updated. For example, a baby learns to categorize and recognize animals. Such work would therefore involve a huge number of disciplines contributing and interacting with each other.

Artificial intelligence [19] is one hydrated technology for the 21st century that can bring a phenomenal changing every segment of computer automation as shown in figure 1.4. AI systems are designed to work independently and intelligently like the human brain, an AI-based system would be a combination of one or more deep different technical areas. AI is a wild branch of computer science and it's a concept inspired by the human brain. The easiest way to understand the concept of AI is to correlate the system's behavior with humans. let's begin the exercise of understanding the nature of the system in three steps to design and develop an AI model [20].

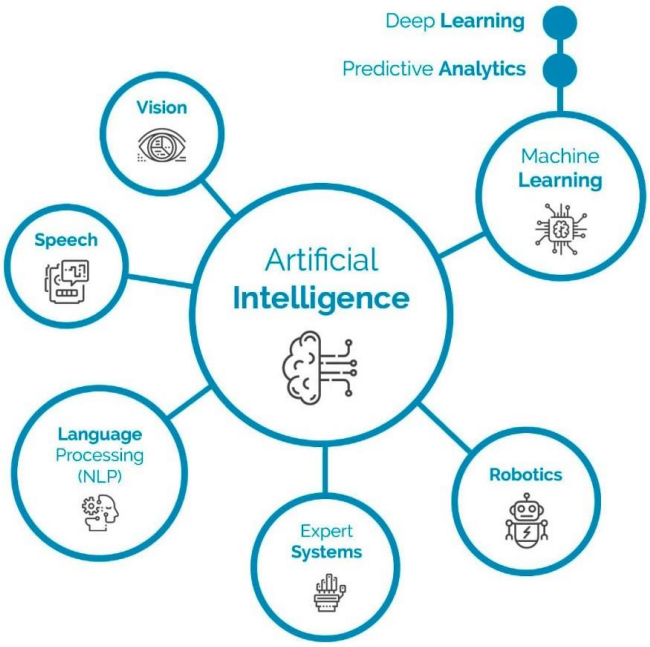

Figure 1.4: Artificial Intelligence

First, understand the specific natural brain, cognitive skills and map them with technology blocks. Initially, the focus needs to be given to the feature mapping. Humans can see and understand the things around them; this can be achieved using computer vision techniques. Humans can also speak and understand the context using any language; this is the area of speech recognition and natural language processing or NLP together these learnings are coming under statistical-based learning. Humans can see things around and understand the surroundings, this is the area of computer vision.

A human can also move around fluently; this is the area of robotics. It is a collaboration of various technologies like computer vision and NLP or speech recognition with robotic engineering, most important feature of the human brain is the ability to view and analyze the patterns from experience. Machines can also do pattern recognition technologies; this is the area of machine learning. As human brains are achieving their processing capability using neurons or a network of neurons. For example, the brain takes the input from body organs and processes the data using similar to the activation functions behind. The processing ability of these functions will be more and more efficient when we get more experience.

An AI system uses the appropriate algorithm of activation functions and these models are built for addressing any specific problems. we can keep adding more and more functions or more algorithms to handle complex tasks; keep adding such models in a network can handle complex scenarios in real life. Appropriate modeling and networking of systems like this can achieve cognitive skills to the machines this is the area of the neural network [21]. If the network is deeply constructed, it can deliver high performance, which is the field of deep learning. Deep Learning (DL) can improve and automate the AI machine's efficiency over some time automatically [22]. As mentioned earlier, the two categories of machine learning which are symbolic and statistical if We instruct the network to scan pictures structurally from top to bottom and left to right, are convolutional Neural networks. This technique is used when we need to train a machine to identify specific images from a group. Humans can they're limited past for example, what is performed during the past few days in the office or at home? This sort of feature can be built using a recurrent neural network.

Next, would be the design and develop the right model for the system. Supervised learning is a method to train the algorithms with a set of data that also contain its answers, for example, getting the mission to identify a person and provide the identity as well in the data. If your algorithm with the data and the missions are to figure out the patterns that are called unsupervised learning for example using the patient's historic data and the social data about diseases the missions are asked to identify the potential chances of getting into any specific disease. Reinforced learning conducted through a set of data with the target to achieve the missions are asked to try it out with all the trial and error to achieve the target [23].

When human intelligence is compared with artificial intelligence [24], human intelligence is based on our cognition or experience in our daily lives. AI as it states is artificial, the decisions that it makes are based on the model's data are fed into the machines for decision making but the decisions of AI are free from bias it does not incline to a specific person whom you like to make decisions [25]. Humans on the other hand can do multitask while playing a game, simultaneously listen to music but artificial intelligence machines can do only those tasks that they are programmed. There are two types of artificial intelligence in general, which are strong AI [26] and weak AI [27]. Strong AI includes the capability of understanding a vast scope of activities with generalized human cognitive abilities so it is assumed that these machines possess cognitive abilities like how the humans would and take positions on their own so they will be able to find a solution without human intervention as of now. Weak AI which is also called narrow AI is the machines that are designed to perform specific tasks, for example, the smart speaker Alexa, Robo, and self-driving cars.

1.2.1 History of AI

The evolutions of Artificial Intelligence over different period makes us understand the exquisiteness of AI. At the beginning of AI in 1943 with the McCulloch and Pitts, Boolean circuit model of the brain, which is about the neural model. In the 1950s, the initial promise with Turing's interesting and seminal paper on computing machinery and intelligence. Samuel's checker program on machine learning, which was the first program of machine learning. The Samuel's checker program running on an IBM machine, playing checkers with a human, could defeat a human and be very popular then. It is a stepping stone towards the computational growth of artificial intelligence. In early AI programs like Newell and Simon's logic theorists. These were very initial promising enterprises in the area of AI. The enthusiasm continued for the decade 1955 to 1965. With the Dartmouth meeting came the term artificial intelligence [28]. Very shortly thereafter, Newell and the General Problem Solver was created by Simon. Thereafter, there was the Geometry Theorem Prover and John McCarthy came up with the programming language called LISP. However, very soon the reality dawned. There was a realization that many artificially intelligent problems are intractable. People soon realized the limitations of existing neural network methods and even people left neural networks as something that would not make any great impact. However, between 1969 and 1985, people were starting to realize that domain knowledge is something that

could change the tilt. The growth of the knowledge-based system was accelerating. There are many success stories, such as the success of rule-based expert systems like DENDRAL and MYCIN. However, they were too fragile in practise and did not scale effectively. Machine learning, on the other hand, has been on the increase since 1986. Neural networks have resurfaced in prominence, although with a slew of new features. There were mature advances in machine learning, both in terms of algorithms and applications. Beginning of 1990, people started to talk of uncertainty; the role of uncertainty in machine learning and AI in general. They looked at Bayesian networks for knowledge representation. In 1995, the turning point was when artificial intelligence started to be accepted as science. During this time, learning, reasoning, and knowledge representation were all integrated. Vision, language, data mining, and a variety of other real-world challenges have all benefited from AI technologies. In the history of AI; till 1980-85, AI research was involved in its initial euphoric concepts. Beyond 1985, machine learning took over the area of AI roughly. This period of AI, no longer concerned with the general AI problem and can be categorized as narrow AI; AI with very focused attention - problems that are focused and do not talk of a general AI concept. It is expected to make more progress in machine learning, requires to apply intelligence to any problem; near human-level intelligence and would go back to our general AI problem.

1.2.2 Disciplines on Contribution to the Growth of AI

One of the disciplines that have made a large contribution to the growth of AI is philosophy; If one of these logic theories is correct, it has contributed to the process of thinking, the mind as a physical system, the foundations of learning, and the rationality of language. The next area that has contributed hugely to the growth of AI is mathematics; involving formal representation and proof theory; algorithms, computational decidability, tractability. The other area that has contributed to the growth of AI is statistics and probability which involve modeling uncertainty and learning from data. Of course, economics had a huge impact on AI in theories such as theories of utility and decision theory. The other area that has hugely influenced AI is the growth of the area of neuroscience or neurons as information processing units [29]. These developments in this area have hugely moved the course of AI. Psychology and cognitive science, have made a huge impact in the area of AI. This study looks at how people act, perceive, interpret, and portray information. Our understanding of these

processes allows us to develop artificial intelligence systems that can do the same thing. The growth of cognitive science has also been pushed by the growth of AI, particularly machine learning. The field of computer engineering is the next field of study that has made AI feasible and pushed the boundaries of AI. Fast computers have had a significant influence. Then there's control theory, which discusses how to build systems that optimise an objective function over time, which has propelled AI forward. One area which has made a huge impact in the way artificial intelligence systems are developed over the years is the area of linguistics; particularly, knowledge representation and grammar.

1.3 Computer Vision

The core objective of computer vision is to build a machine, which can process and interpret images just like a human visual system does, it may be a compliment of biological vision. When the similarity between the human visual system and the computer vision system is compared. Human Visual System has eyes to see images, videos, or any objects and after that, it is processed in our brain to make intelligent decisions. In case of the computer vision, cameras are represented as eyes, which may be single or multiple cameras for an image or the video equation. Then, the image pre-processing is performed with pattern recognition and the artificial intelligence algorithms for making the decisions [30].

To improve the visual quality of an image, image processing techniques are applied to remove noises. In the case of computer graphics, the input is a model, from the model the image is generated. In computer vision, the input is an image and output is, interpretation of the image or video. The image processing, input is an image and output is also an image. An analysis image and synthesis image are taken. Using the analysis image, the interpretation or some model is generated then it is termed as computer vision. Alternatively, synthesis contains the model and can able to generate the image from the model, then it is called synthesis image. In this case, analysis means computer vision and synthesis means computer graphics. The fundamental steps of computer vision include image acquisition using cameras, which is the input image. The pre-processing is performed on the input to improve the visual quality of an image or remove the noises in the image. Image segmentation means the partitioning of an image into a connected homogeneous region is adapted in the next step of computer

vision. Segmentation extracts the features from the image, based on these features the classification is performed.

1.4 Convolutional Neural Networks (CNN)

CNNs are a type of artificial neural network that differs from ordinary neural networks in that they accept pictures as input. They are designed to design to work on images, mostly to handle computer vision problems like an artificial network. These networks have weights, neurons, and bias units. The weights in these CNNs [31] [32] are estimated by optimizing an appropriate objective function because the input to these networks are images. A typical medical image will have hundreds of slices going through particular anatomy and addressing them manually, labelling each of these voxels by hand is a very tedious and error-prone task so this can serve as a huge support for radiologists [33] who can look at this kind of images every day for interpreting them.

The CNN is made up of a series of hidden layers, which are either convolutions or pooling. It consists of convolution and pooling layers alternated with a sequence of fully connected layers leading to a classification layer, similar to an artificial neural network.

1.5 Mask Region-based Convolution Neural Networks (Mask R-CNN)

Deep learning and neural networks have made the most exciting advances in the field of computer vision. It includes any difficulty with an image or camera input as part of the computer vision category. Self-driving vehicles, fMRI analysis, Mars exploration Rovers, face recognition systems, object detection, and augmented reality are just a few examples of technological advancements. There are two phases: for object detection, we utilise an architecture similar to that of faster CNNs, and for semantic segmentation, we use fully convolutional networks, as illustrated in Figure 5.

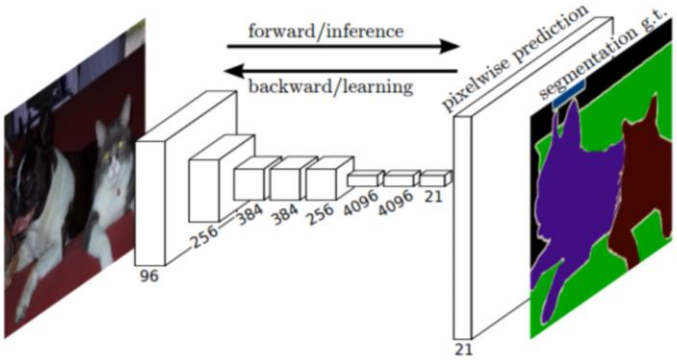

Figure 1.5: Fully Convolutional Network

These minor distortions could be something as simple as the rotation of an image. Conceptually, the Mask RCNN is similar to the faster RCNN. Mask The object mask is also produced by RCNN utilising pixel to pixel alignment. For each region of interest, this mask is a binary mask. Much overhead isn't incurred when computing this mass as it is done in parallel with the bounding box creation and classification [34].

FCIS is a different framework that uses semantic segmentation and object detection to categorize box and mask objects in an image and does so quickly. FCIS, on the other hand, makes systematic mistakes on overlapping instances and produces excessive edges, showing that it is having trouble with the core challenges of segmenting instances. Instance segmentation is object detection with semantic segmentation [35]. Mask RCNN is an architecture to achieve instance segmentation it combines faster RCNN with fully convolution networks (FCN). Masked RCNN uses ROI align which preserves the spatial orientation of features and leads to no loss of information.

1.6 Machine Learning

Machine Learning (ML) is building useful models from data. The models hopefully that improve with increasing volumes of data. In most fields of science and engineering, increasing volumes of data are being collected because of improved measurement technologies, storage, and computational techniques. In reality, ML is built on foundational advances in mathematics and computer science that have been happening over decades in particular optimization and regression. There are dozens of hundreds of other techniques in optimization and regression to build models from data that are not

neural networks and still have a very important place in kind of the suite of techniques in machine learning.

Machine Learning (ML) gives us statistical tools for analysing data. It also comprehends that specific information. The first is called supervised machine learning, the second is called unsupervised machine learning, and the third is called reinforcement learning, which is also known as semi-supervised machine learning.

Supervised Learning, contains the label data. This labeled data is prepared from the passed data and with the help of this kind of data. It uses label data as training material and from these label datasets, the model learns the association between input and output as unsupervised learning every piece of data and the data consist of an input and output in case of object detection the input as a picture and the output as the name of the object. So, we feed our model a large number of examples to Train on. After the training process when the Model identifies the object in any picture, can detect it because it trained on a huge number of examples. An example of supervised learning and healthcare applications in the real world can be seen in the classification of lung nodules and also recognition of body organs from medical images.

In the second category, Unsupervised Machine Learning here will not be having any labeled data that means in the dataset. It is used unlabeled data as training material for the machine learning model [36]. Clustering is the most common use of this method, which is used to divide big datasets into smaller groups based on similarities and common features between each piece of data and the data set. Unsupervised learning can also be used for anomaly detection, in which the model attempts to identify rare and suspicious items within the dataset that do not fit into any of the dataset's subgroups, and as an example of unsupervised learning in the real world, in the prevention of heart diseases through data clustering and classification. There are different clustering techniques like K-Means [37], Mean shift [38], DBSCAN [39], etc., Machine learning we will be solving clustering techniques and it does, based on the similarity of that data, it will try to group that data and there are some mathematical concepts like Euclidean distance used inside the computation.

The datasets used to train the model in a semi-supervised learning machine learning technique comprise both labeled and unlabeled data. This technique is very useful to use within datasets that contain a huge number of unlabeled data and fewer number of

labeled examples. So, it can be used to acquire a sufficient amount of label data to be used in the training of supervised learning models. Because acquiring appropriately labeled datasets is very hard especially in the case of the healthcare field.

The data will be labelled in reinforcement learning, and later on, certain parts of the data will not be labelled, so the computer or machine learning model will learn slowly by looking at previous data, and it will be learning as soon as new data comes up in the new environment. The most significant part is, it needs to have data and it also provides some statistical tools to analyze and explore the data [40].

1.6.1 Applications of Machine Learning

Healthcare data contains data from medical records medical images genomic data smartphone data wearable devices such as smartwatches and sensors social media internet and environmental data. All these types of data and the huge amount of data being collected by healthcare providers make processing and extracting insights from the data a very tough problem. So, machine learning systems are used to process and analyze these data and instead of traditional methods and the applications of these machine learning systems in healthcare can be categorized into four sectors, diagnosis, prognosis, clinical workflow, and treatment.

Prognosis is the process of preventing the advancement of disease in clinical practice by recognizing symptoms and science linked with it, as well as assessing whether the symptoms and science worsen, improve, or remain stable over time. Based on all these predictions, Machine learning models have been utilised in the classification and identification of specific malignancies such as brain tumours, and the prediction of the chance of survival has some intriguing applications and medical prognosis. To assist and prognosis these situations by predicting whether the tumor or nodule is malignant or not, new possible applications of machine learning and the field of prognosis are being researched as part of research efforts aimed at enabling customized treatment. Machine learning-based approaches are used to extract clinical characteristics from electronic health records to aid in the diagnostic process, whether it's for electronic health record interpretation or medical picture analysis. A semi-supervised learning model, for example, is used for the extraction of ignored information from unstructured health data, and machine learning is used for the diagnosis of diabetes from EHRs,

among many other examples, including mortality prediction and the production of length of stay duration and hospitals.

Machine learning models employ medical image analysis such as MRI, CT, ultrasound, and other forms of medical pictures to offer functional and anatomical information about various human organs and aid in the identification and diagnosis of various problems. Image augmentation for damaged medical pictures is a critical pre-processing step that has a direct impact on the diagnostic process. Because of the common artifacts and noises that can be found in medical images. Different machine learning models are being used an enhancement and the noising of medical damages such as convolutional denoising auto-encoder that was used for the noising of chest radiographs and also generative adversarial networks that have and cleaning motion artifacts and multi-shot MRI image.

The second task of machine learning models and medical image analysis is detection to identify specific disease patterns and abnormalities like tumors or fractures and medical images. A common clinical strategy This technique is typically performed by a skilled medical practitioner and requires a substantial amount of time and effort, although machine learning algorithms have shown promise in this area. For example, convolutional neural networks were used to diagnose and classify cancer, as well as to identify mitotic and breast cancer images [41]. The third task is tissue and organ segmentation, which allows for quantitative analysis of normalities such as measuring the shape and volume of cancer tumors and brain images. The fourth task is image reconstruction, which is the process of generating interpretable images from more data acquired by imaging sensors. Machine learning models are increasingly being used in research and medical image generation, and numerous deep learning models, such as convolutional neural networks and autoencoders, have been used to reconstruct MRI and CT images.

The next important stage is medical image registration, which is the process of mapping input pictures concerning a reference image. This has a wide range of possible applications, including medical image analysis, however it is limited in terms of image restoration techniques and real-world applications. Medical image retrieval, the enormous amount of data and medical images, and hospitals are extremely difficult to manage using traditional methods, so machine learning models are being used to

facilitate the process of comparison and classification, as well as to reduce the search space by all looking for relevant images and data. In the process of treatment, natural language processing models are being used to produce and generate textual reports to narrate the findings in a given medical image. This approach to helping and the efficiency and effectiveness of the treatment process and also machine learning has been applied in real-time health monitoring which is very important in critical and dangerous cases. Patients' data is collected via wearable devices and smartphones, then transferred to the cloud for analysis and interpretation using machine learning and deep learning models, with the results returned to the device along with suitable suggestions. Clinical workflow is the final use of machine learning in health care, and it summarises the impact of all previous applications. Natural language processing models have been employed in anything from sickness prediction and diagnosis to report preparation.

1.7 Deep Learning

Deep learning is a collection of techniques, or a subset of machine learning, that has been increasingly applied in the industry to solve issues in computer vision, natural language processing, and speech recognition [42]. In our daily life, a lot of different tools and plug-ins on your smartphones that use this type of algorithm. The reason it came to work very well is primarily, the new computational methods. Deep learning is computationally expensive and needs to find techniques to parallelize the code and use GPUs specifically to graphical processing units, to be able to compute, the computations in deep learning [43]. The second part is the data available has been growing after, after the Internet bubble, the digitalization of the world. So now people have access to large amounts of data and this type of algorithm has the specificity of being able to learn a lot when there is a lot of data. So, these models are very flexible when more data is fed in. It will be able to understand the salient feature of the data and finally algorithms. People have come up with new techniques to use the data, use the computation power and build models.

1.8 Problem Statement

- To perform fine-grained localization and precise segmentation of the fetal brain from Fetal MRI images.
- To analyze Brain Structure, Size and to highlight the abnormalities to assist the physician.
- To identify the normal and abnormal fetal brain from the given MRI image.
- To identify the fetal brain anomalies in a short period.

1.9 Research Objectives

- To Identify the abnormalities in the Fetal Brain.
- To recognize the internal structure of the lesion tissues.
- To classify the normal and abnormal fetal brain.
- To observe Disease Progression in the abnormal fetal brain, and help the Physician to plan accordingly.
- To highlight and enhance the ROI for interpretation.

1.10 Challenges

- Orientation:
- Un-Uniform Intensity
- Motion
- Computational Time

1.11 Organization of Thesis

The chapter-wise organization of the thesis is as follows:

Chapter-1: Presents an introduction to medical imaging and the impact of artificial intelligence in the field of medicine.

Chapter-2: Discusses and presents a detailed literature review about the existing object detection, localization, and segmentation approaches and also sorts out the research directions.

Chapter-3: Discusses and comprehends various aspects of proposed object detection which involves the identification of medical objects.

Chapter-4: Delineates different experimentation results and Comparison Analysis appropriate to comprehend the fetal brain localization in the cloud environment.

Chapter-5: Discusses and presents various aspects of proposed fetal brain segmentation algorithms.

Chapter-6: Delineates and comprehends the improved fetal brain segmentation algorithms which enhance the interpretation of the results.

Chapter-7: Presents the fetal brain anomalies detection and localization using the cloud environment.

Chapter-8: Draws the Conclusion and discusses the possibilities for future research.

CHAPTER 2

LITERATURE SURVEY

2. Literature Survey

2.1 Image Processing

Md. Shakowat Zaman Sarker et al., [44] debated the morphological algorithm of the watershed. Images are analyzed using the pre-and post-processing watershed method. At first, the Gaussian filter and threshold value are analyzed in the pre-processing phase to produce segments of the watershed segments that lead to the following stage of the region merging. In addition, the morphological region merging based on RAL is employed to handle the false segments. It transforms all pixel intensity from 38 to 255 using the thresholding procedure. It eliminates dark and light areas, filters noise in the backdrop of the image. Compared to other current algorithms the region merging approach is superior [45]. In their experimental technique, the authors have effectively minimized the excessive segmentation.

The detectors recently presented will recreate the original image and attempt to correspond to the region of interest. Because the model is trained by certain object size. The author, Kaiming He et al., [46] said, the same object size can't be expected as a test image. The difficulty is therefore resolved by trying to rescale the items in the picture. This approach has been characterized by the authors as a scale adaptive training, which may improve the accuracy of the object detection and dramatically increase the data set size. Adaptive Scale Training One or two areas of interest overlap in the input frame, which in the projected output picture generates more than one limiting box [47].

Sébastien Tourbier [48] used the template to slice block matching on fetal brain segmentation to achieve better results. The block extraction and dimension reduction are first applied to the input picture. The unique slice-to-template brain extraction technique is used on the input to automatically determine the ROI. The healthy and mild pathological patients are used for brain extraction and brain mask refining. This method was computationally intensive and would not fit all of the input image's orientations.

The technique was adopted by Amir Alansary [49] and his team to extract the features of the segmented pixels and produce a superpixel graph. This method works by analyzing the superpixel graph to classify a given fetal brain. It will then tell us if the brain is non-brain or brain.

2.2 Object Identification

In the process of finding things in the image, the Faster R-CNN [50] and the YOLO [51] are the masterpieces. If the two-stage object identification detectors are employed, the accuracy improves by greater height. Backbone networks are utilized for the classification task involving network-based mapping of classes. The objective is to recognize items that decide whether objects come from classes backgrounds. The spatial position and the degree of precision for each object case are analyzed to an image (Stethoscope, Lab Coat, Sharps containers, ECG unit).

Preceding 20 years ago, P. Viola and M. Jones [52], [53] presented the detection of objects and challenges. In this philosophy of object detection, the object detection tried by these writers is still a milestone. The high quality of object detection with a 700 MHz low-configuration Pentium 3 computer is registered and called a VJ detector pride for this path to confront. The authors have employed 200 features to the detection rate of 95% to identify the face of 24 x 24 pixels of the input face frontal image. Additionally, they have achieved the false positive rate of one in 14084. The classifier has the capability of performing the classification of the faces at the rate of 0.7 seconds per input. The VJ detectors are inquisitive about the identification of human faces. In this connection, the sliding window technology [54] has been applied and the whole image was scanned to find the area of interest from the input.

When object detection is executed, the orientation of the image will be a huge problem. The model is trained by the picture of the face, for example. However, if the image is shown upside down as an input. Then the model can't detect the face. A technique suggested by H. Zhu, X. et al, [55] solves the problem. To provide the input data to the model with another direction, the authors have employed a data augmentation approach [56]. Flip, height, and shear are the basic data increase methods that the writers use to achieve superior analytic results.

The object Identification problem has been addressed by the writers of Papagesorgiou et al. [57]. A) Greedy selection; this approach recognizes the interaction value between items, which produces the bounding box with the greatest score between objects overlaid. b) Bounding Box Aggregation: this technique takes full account of all objects and maps an object's bounding. b) Many well-known researchers, including the VJ detector, have used this approach, and c) NMS learning: NMS as a filter for training

and re-finding all rough detections as part of a network is the basic idea of this technique. NMS as a filter is considered [58].

In object detection technology, YOLO is the next major milestone. You only look once. In the year 2015, YOLO was launched. This exceeds all the previous approaches and methods that have been developed to detect and locate the items in this image. YOLO divides the whole image into segments and utilises a convolutional neural network to detect the segment's region of interest. This approach shows the rapid and high-precision identification of the input ROI. In recent years, this approach has been published with several variants: YOLO v2 in 2017 [59], YOLO v3 in 2018 [60], and YOLO v4 in 2020 [61].

The extremely tiny capacity of YOLO V5 is around 27 MB only, it was released later in 2020. But, due to its uniqueness, YOLO v5 was very controversial after its release. However, the most exact results in object identification, localization, and classification have been obtained for all these versions.

The region-based neural networks (R-CNN) provide the next successful method to object detection. The region of interest is selectively searched [62]. The items recognize d are configured and fed via the coevolutionary neural network. The method gathers the properties of the recognize d items and will later be utilized to identify the objects using the support vector machines. Furthermore, RCNN Fast was developed in comparison to RCNN to improve efficiency. It offers the settings for the Bounding Box model and the detection of methods in a single configuration. By comparing the RCNN, the accuracy has been 10 percent increased and ROI detection is 200 times faster. Few pre-trained detectors are used to explore hidden objects quickly and reliably. These designs have several unique qualities and are successful with outstanding outcomes, for example, VGG Net [63], Alex Net [64], Google NET, Res Net, etc.

2.3 Classification

Bernhard Kainz, etc., [65] concentrated on the T2 images and proposed a fully automatic brain voxel classification technique. The algorithm has achieved the accuracy of 97% precise classification of the fetal brain in the work. The Nvidia CUDA [66] was implemented and utilized cross-validation to select the preferred descriptor-treated voxel set, and eventually achieved the descriptor-sized outcomes from 3 to 4

millimeters. K Somasundaram et al., [67] analyzed the T1 weighted MRI images of human head scans. The input image of size 256 x 256 is used initially and passed to the point spread function to perform the deconvolution to obtain the binary image. The eroded image was extracted from the binary image and the largest connected areas are combined to acquire the dilated image. The brain mask is finally matched with the input image to get the resultant segmented final brain portion of the brain image. The Jaccard and dice indices of 0.9092 and 0.9531 respectively were obtained as the results in the proposed work. The result is superior to the current market approaches and has minimal inconveniences with the form and size specified.

For the study, which used a region-based segmentation approach, the Omneya Attallah et al. categorized the images based on their classification. The data were then used to formulate a classification system for the brain with the accuracy of 95.6%.

A new method to identify tumors at their underlying phase using MRI images was suggested by B.Devkota et al. [68]. The median filter is utilized to execute the pre-processing stages and division of four distinct types of brain tumors using the approach of mathematical morphology reconstruction to get findings with 92% accuracy. Xia T et al. [69] have reported the patch-based classification approach on histopathological images, with a training size of 4196 at a precision of 84.3%. Le Hou et al. established the technique for classification called the Whole Slide Tissue Image (WSI) [70] and categorized lung and non-small cell carcinoma. In addition, a classifier is needed to characterize a target subject. The additional classes and the possible interpretations are more hierarchical, semantic, and insightful for visual recognition. Object identification can be grouped, either by identifying individual examples in two categories by identifying individual instances in large classes. In the first kind, occurrences of a single item are detected, a corresponding dilemma in essence [71].

Instances of these present objects are to be described using the second type. Historically it was a single sort of identifier or a couple of distinct categories that concentrated most of the work in object detection. The scientific group has begun to create object detecting systems in the last several years, using analogies, for broad uses when the target spectrum is large. It mirrors the capacity of man to track things. The DPM [72] is one of these classification systems by combining object components into a scalable model with deformation. Extreme deformation processing cost in DPM, with the

assistance of Carefully built low-level functions from a graphical model Decompositions, are paired with kinematically influenced pieces

2.4 Localization

K. Somasundaram et al. claimed that it is a timely but efficient approach to get the best outcomes by employing the center of gravity (COG) technique [73] to localize the fetal brain. The high-intensity pixels were removed from the fetal brain to obtain the binary image. The morphological opening filter is used on the binary image to process the hole filling method, which uses the strategy of flood filling algorithm. Finally, the brain mask is obtained from the algorithm to segment the fetal brain. He has done this using less useful computing approaches. After the algorithm is displayed with the input, the time needed to locate the embryonic brain is less than one minute.

In a coarse segmentation method, an author has employed the Bottom-Up Strategy to locate the embryonic brain. The P-Net [74] is utilized to carry out automated segmentation of the region and then to generate a reconstructed fetal brain with an over-the-top resolution. A model of a Bayesian network [75] consisting of a computationally costly mathematical basis technique. This is a useful strategy for estimating uncertainty using many models.

K.Somasundaram et al. have introduced a technique to locate the fetal brain in the MRI images using the center of gravity (COG) method [73]. A circle is drawn around the region of interest and within the region, the high-intensity pixels are removed gradually. It is converted as the binary image and then a morphological opening filter is used to obtain a better result to visibly see the ROI. In addition to that, the hole filling technique is espoused to the obtained image. Finally, the brain mask is generated to perform the fetal brain segmentation.

2.5 Segmentation

The P-Net architecture is used for 2D image segmentation by Guotai Wang et al., [76] The convolutional neural network is implied on the brain to perform the segmentation of the whole brain and tumor. Fetal Brain, placenta, fetal lungs, and maternal kidneys are taken as the ROI for the research work. The training is performed with the fetal brain and abnormalities to prepare the trained CNN model. The testing is executed with the user-provided bounding box to obtain the localization and segmentation of the

brain. The initial segmentation is achieved using the foreground probability method. The annotated photos are put into the CNN model that has already been trained. The weight map analysis function is used to fine-tune the segmented picture using the weighted loss function, yielding the refined result of segmented brain anomalies. The BRATS dataset from the year 2015 [77], were used for their research work. The NVIDIA Graphical Processing Unit (GPU) was used to accelerate the performance of the training [78].

The deep cut method was introduced by Martin Rajchl et al., [79] which is an extended version of the GrabCut technique [80]. The fetal brain and lungs are used as the region of interest in their work and achieved the mean accuracy of 94.1%. The original image was localized with the bounding boxes and DeepCut algorithms applied to obtain the patches on the ROI. The naive learning approach is used to produce a segmented sketch above the ROI, which follows the DeepCut from the bounding boxes technique to optimize the result to obtain the segmentation of the brain. The naive approach and deep cut used the CNN model with iterative dense CRF to obtain the features of the object. In addition to the received results, DeepCut from pre-segmentation and fully supervised CNN segmentation is implied to achieve the enhanced segmented ROI.

Morteza Pishghadam et al. [81] presented a unique methodology for automated MRI fetal brain extraction utilizing a variational level set method. In this study, T2 weighted pictures from 105 photos with 8 individuals were used to perform fetal brain abstraction. Initially, the fetal eyes were identified using MRI scans as a novel method for localizing the fetal brain. The ROI is then highlighted by matching the fetal eyes with the fetal brain using the variational level set approach. The dice and Jaccard coefficients are both 99.56 percent and 96.89 percent, respectively. The cortical interface of the 90-dga sheep brain was segmented by Rosita Shishegar et al., [82]. In their work, the outer brain mask is computed and Subcortical developing fiber tract is detected using the MRtrix software program for streamlining the extraction. The background levels are streamlined to obtain the inner cortical mask and outer cortical surface region. To confirm the precision of the segmentation findings, the dice ratio was computed. The dice values between 0.87 and 0.92 for internal and external cortical faces are very consistent with manual segmentation results. The method presented carries out the semi-automatic approach [83]. The outcomes of semi-automatic

segmentation and automated segmentation [84] are then demonstrated using the fetal sheep brain.

The multiple sclerosis of MRI images is segmented using the automatic method by Shahab Aslani et al., [85]. For 3D volumetric data, a deep end-to-end two-dimensional convolutional neural network is developed utilizing slice-based segmentation. The system can encode information from different modalities and features a multi-branch downsampling approach for CNN. The model was trained using two distinct datasets: 37 patients from NRU's private clinical data and 14 patients from the ISBI 2015 public dataset. Because the dataset had fewer MRI pictures, the authors were subjected to overfitting. The transfer learning approach is adopted to reduce the overfitting reasonably.

2.6 Abnormalities

In a multi-modal image scenario, José Dolz, et al. [86] used 3D Convolutional Neural Network hyper-density connected network. T1 and T2 pictures are processed independently by HyperDenseNet and linked densely. In cerebrospinal fluid, grey matter, and white matter, the dice values were 0.957, 0.920, and 0.901, respectively.

The world-beating diagnosis tool during pregnancy is Magnetic Resonance Imaging (MRI). Since the movement of the fetus will be frequent inside the womb, the clarity of the digital image will be accurate only by the MRI method. The images are obtained in slices and later it is reconstructed into a volume by the slice-to-volume reconstruction (SVR) method [87]. Khalili [88] states the fully convolutional network for fetal brain MRI images using U-Net architecture with a learning rate of 0.0001 through the adam optimization method. The results were obtained regardless of age and orientation of the given input image. They have achieved the result of dice coefficient ranging from 0.98 to 0.99.

To improve the accuracy of the outcome, Deep Learning [DL] is used. When compared to Machine Learning, Deep Learning can create neural networks with a large number of layers, resulting in the extraction of a hierarchy of complex properties from the input, and it is cognitively powerful enough to handle large amounts of data during model training. The Graphical Processing Unit (GPU) joins hands with the DL framework to increase the speed of the training and efficient handling of data. Guotai Wang et al.,

have proposed the Uncertainty-Guided Interactive Refinement framework using MG-Net [89]. Input stack of Slices is given as input to the MG-Net to perform the initial segmentation with a single network and a forward pass. This method is comparatively competent for interactive segmentation.

The localization is the significant step in identifying the lesion tissue in the MRI images. The bottom strategy is used to segment the fetal brain with a pixel-level prediction method [90]. The details of the size and shape are not fed into the model to perform the coarse segmentation in this work, which is a great advantage to save time. The Seg-Net architecture is used to conduct fetal brain segmentation, which requires relatively little memory. The classification of the brain is performed using the Histogram of Oriented Gradients (HOG) method [91]. The input image is given into the algorithm to compute the gradients and transform the spherical coordinates. Then, the orientation of the brain is verified and normalized the contrast on the spatial blocks of the image. Finally, the HOGs detection window method will predict the input as brain or non-brain.

The Brain is located in the MRI images using various techniques by different authors. Template matching [92] is one of the simple methods, which is implemented to identify the shape and position of the region of interest in the given input. This method receives the stack of images to the system and matches the template of the brain and re-orients the ROI based on the template. The above step is repeated to make all the images in the stack to be aligned as per the template. Secondly, they have adopted six degrees of freedom to estimate the motion [93], and the optimization metric is analyzed using the normalized cross-correlation. designations.

The segmentation of melanoma detection is performed using YOLO v4 [94]. Morphological operations are applied to the raw images to detect skin cancer, which enhances the image by sharpening it to build the machine learning model. The bounding box is utilised to classify infected and non-infected tissues using the active contour segmentation technique. The accuracy was measured using the dice and Jaccard method, with results of 1 and 0.989 respectively.

CHAPTER 3

MEDICAL OBJECT DETECTION

3. Medical Object Detection

3.1 Introduction

Objects detection is the significant step to identify the unknown objects which are perceived by the human. Artificial intelligence, in conjunction with a machine learning algorithm, can assist in locating and identifying items in the real world. In the work, the clinical objects are taken as the region of interest such as Stethoscope, Lab Coat, Sharps containers, ECG unit, and so on [95]. When some developers decided to focus on seeing faces, car numbers, street signs, studying distant and clinical images, and so forth. Each of these "everlasting" challenges is illuminated by the efforts of numerous meetings of architects and scholars. As modern specialized arrangements get exorbitantly pricey, the task of computerizing the creation of product apparatuses for dealing with academic challenges is planned and thoroughly unraveled. In the field of image handling, the required toolkit should assist the research and recognition of images of previously obscure substance and ensure the persuasive development of usage by standard developers.

The Windows toolbox also encourages the development of interfaces that address a range of related concerns. Item acceptance is a grouping of related PC view commands that include exercises like object separation in advanced pictures. Picture arrangement activities include things like predicting the class of one class in a picture. The term "item restriction" refers to identifying the region of at least one class in an image and creating a bounding box around their degree. Thing detection is performed by combining these two tasks and constraints to characterize at least one object in a photograph. When a customer or professional mentions the phrase "object recognition," they typically mean "object identification." It may be difficult for beginners to distinguish several relevant computer vision tasks.

Thus, we can recognize these three computer vision undertakings, initially starting with the Image Classification [96] which Foresees to sort or class an object in a picture. Input is taken as a picture that comprises of one object or more than one object, for example, a pen on the desk or more stationary items on a desk. The output of the process is the label for the class. The model will depict one or more numbers that are mapped to class labels. The next, computer vision task is to perform object localization, this is accomplished by identifying the proximity of the objects in an image and

revealing their presence using a bounding box on the top of the object. The last task which is addressed in the research work is object detection, which provides the details of whether the object which is expected is available in the real world or the input picture.

3.2 Method

3.2.1 Object Detection

Object detection is frequently utilized in everyday life in a variety of applications. Object detection's major goal is to recognize and anticipate known and unknown items. There are various applications are available such as, identifying the number plate on the car, detecting the face reactions, and so on.

Object detection aims to recognize all kinds of objects within a known or specified category, such as people, autos, or faces in an image. Only a few examples of the item are displayed in the graphic; nevertheless, there are an enormous number of potential places and scales at which they might exist, which must be investigated in some way. Every point in the picture has some form of posture data associated with it. This can be as straightforward as the article's area, scale, or the degree of the item described by a bounding box. The postural data are becoming progressively dotted, with criteria for clear or straight correction in a range of various situations. For example, the regions of the eyes, nose, and mouth may be included in a face indication recognition if they are indicated with a box.

The localization is marked using the bounding box which works based on the following equation (1),

$$Bounding\ Box = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad \dots (1)$$

The clinical or medical objects [97] which are taken as the region of interest (ROI) will be highlighted with the help of the bounding box. In some cases, a picture may contain more than one ROI, which triggers one or more boxes in an output image. The edges of the objects are identified using the innovative algorithm and generate the layer of the boundary with a line. Figure 6 shows the representation of the bounding box, which provides a layer of the rectangle shape based on the ROI object.

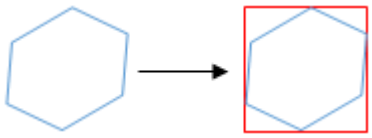

Figure 3.1: Localization of objects using Bounding Box

The major challenge in this process is the variable output dimension that is induced by the variable number of artifacts that can be found in any given input image [98]. Every general machine learning task involves a fixed input and output dimension for the model to be educated. Another significant challenge to the widespread implementation of object tracking systems is the need for real-time (30 Frames per Second) detection while being precise. The more complex the model is, the longer it takes to be inferred; and the less complex the model is, the less precise the model is. The trade-off between precision and efficiency must be selected as per the application.

Tensor flow is a data flow that serves several industries and a diversified open-source programming library. This math library is representative and is mostly used for AI implementations, including neural networks, etc. For research and growth, Google was sponsored. Google Mind generates a Tensor flow to use Google on the inside. It is downloaded from ApacheLicense2.0 on November 9, 2015. The tensor flow was launched on 11 February 2017. Although the infusion runs on a single gadget, tensor flows can be run in various CPU (Central Processing Unit) and GPU modes. Tensor flow is a second-age Google Cerebrom system (with discretionary CUDA and SYCL augmentations for universally useful figuring on illustration handling units). TensorFlow is usable in different phases, such as 64-bit Linux, macOS, Windows, Mobile and Android, and iOS. Tensor flow engineering facilitates the fast organization of computing across several phases (CPUs, GPUs, TPUs [99]) and from workspaces-server classes to versatile edge gadgets.

Tensor flow equations are distributed as static data flow diagrams. The term Tensor flow is derived from the operation of neural networks in MDI clusters referred to as tensors.

In addition to a broad range of important scientific skills to work on these displays, NumPy is Python's programming library, which offers support for large, multidimensional clusters and matrices. Initially, Jim Hugunin made NumPy's predecessor, Numeric, a list of engineer obligations. In 2005 Travis Olphant translated NumPy into quantitative by consolidating processing highlights Num exhibit with

expansion modifications. NumPy has a lot of trends and is open-source programming [100]. SciPy contains modules for some enhancements, direct polynomial math, coordination, interjection, unique capacity, FFT, sign and picture preparing, Tribute solvers, and different undertakings regular in design. The NumPys Stack component, which includes resources such as Matplotlib, Pandas, and SymPy, etc., and an increasing number of logical libraries, is SciPy, which greatly abstracts the NumPies cluster object. The NumPy stack has similar implementations, such as MATLAB, Octave, and Scilab. Besides, the NumPy stack is repeatedly referred to as the SciPystack. The SciPy library has been distributed under the BSD license and is supported and supported by an open architecture network. It is also sponsored by NumFOCUS, an organization devoted to the advancement of reproductive and open science. OpenCV is a programming library with mainly PC visionary capabilities are created by Intel, Willow Carport will be reinforced by Itseez later. The library is a cross-story BSD license open-source library.

3.2.2 YOLO version 3

YOLO is an abbreviation for "You Only Look Once," and it detects objects using Convolutional Neural Networks. YOLO is capable of detecting several items in a single picture. This implies that, in addition to predicting item classes, YOLO identifies the positions of these objects on the picture. YOLO processes the entire image using a single Neural Network. This Neural Network divides an image into sections and assigns probability to each. Then, depending on the likelihood, YOLO estimates the number of Bounding Boxes that will cover specified sections of the image and chooses the best ones. In object detection and localization, YOLO version 3 incorporates the following terminology: Convolutional Neural Networks, Residual Blocks, Skip connections, Up-sampling, Leaky ReLU activation function, Intersection over Union, Non-maximum suppression.

The architecture of YOLO version 3, uses the convolutional layers and it originally, consists of 53 convolutional layers that are also called Darknet-53. But for detection tasks, the original architecture stacked with 53 more layers give us 106 layers of architecture for YOLO version 3. In the Darknet framework, 106 layers were loaded when the Yolo 3 architecture is initialized. The detections are performed at three different layers: 82, 94, and 106. This current version 3 includes some of the most important features, such as Residual Blocks, Skip Connections, and Up-sampling. The

batch normalisation layer and the Leaky ReLU activation function follow each convolutional layer. Additional convolutional layers with stride 2 are utilised to down-sample feature maps instead of pooling layers. Because the use of additional convolutional layers to down-sample feature maps prevents the loss of low-level features that the pooling layer simply excludes, the use of additional convolutional layers to down-sample feature maps prevents the loss of low-level features that the pooling layer simply excludes. As a result, capturing low-level features helped to improve the ability to detect small objects. On the images, pooling excludes numbers, but convolution takes into account all numbers. The input is a collection of pictures of the shape (n, 256, 256, 3), where n is a number of images. The next two numbers are width and height. The last one is several channels – red, green and blue. Increasing the resolution of input might improve the model's accuracy after training. In the current research work, an input of size 256 by 256 is assumed for the fetal brain localization and segmentation. These numbers are also called input network size. Images for input can be any size, but they must be scaled before being fed to the network to fit the input network's size. There is a possibility to experiment aspect ratio by adjusting parameters when training and testing in the original Darknet framework, in TensorFlow, Keras, or any other framework.

Throughout the network, YOLO version 3 detects at three different sizes and three different places. The independent detection spots are layers 82, 94, and 106. The Network downsamples the input image by the following factors at each of those locations: 32, 16, and 8 are the three numbers. These three figures are known as the network's stride, and they indicate how the output at three different points in the network is less than the input. For example, if we consider stride 32 with an input network size of 416 by 416, the output will be 13 by 13. As a result, the output for stride 16 will be 26 by 26 and for stride 8 will be 52 by 52. The 13 by 13 detector detects large things, the 26 by 26 detector detects medium objects, and the 52 by 52 detector detects little objects. Without leaving a reminder, the Network's input must be divisible by 32. Because if it holds true for 32, then also holds true for 16 and 8. In the detecting kernels, it creates YOLO version 3 output, which is applied one by one at three separate sites in the network. 13 by 13, 26 by 26, and 52 by 52 are 1 by 1 convolutions applied to downsampled input pictures. As a result, the spatial dimensions

of the generated feature maps will be the same. The following equation is used to calculate the depth of the detecting kernel.

$$[\boldsymbol{b} * (\mathbf{5} + \boldsymbol{c})]$$

where b = 3.

"b" is the number of bounding boxes that each cell of the generated feature map can predict. YOLO v3 predicts three bounding boxes for each feature map cell. As a result, "b" equals 3. Each bounding box contains the following 5 + c attributes: The center coordinates of the bounding box; the width and height of the bounding box; the objectness score; and the list of confidences that correspond to each class of the bounding box.

YOLO v3 predicts 3 bounding boxes for every cell of the feature map, which are termed as grid cells in the YOLO v3. Each cell, in turn, predicts an object through one of its bounding boxes if the center of the object belongs to the receptive field of this cell and this is the task of YOLO v3 while training: identify this cell that falls into the center of the object. Again, this is one of the feature map's cells produced by detection kernels. When YOLO v3 is training, it has one ground truth bounding box that is responsible for detecting one object. It defines to which cell the ground truth bounding box belongs and does the first detection scale, where 32 as the stride of the Network is considered. This grid now represents produced output feature map. When all cells, that the ground truth bounding box belongs to, are identified, the center cell is assigned by YOLO v3 to be responsible for predicting this object and the objectness score for the cell is equal to 1. This is another one of the associated feature map cells that is in charge of ROI detection. To anticipate bounding boxes, YOLO version 3 use anchors or priors, which are pre-defined default bounding boxes.

These anchors are eventually utilised to determine the true width and height of the anticipated bounding box. There are a total of 9 anchor boxes in use. Each scale has three anchor boxes. It means that by employing three anchors, each grid cell of the feature map may predict three bounding boxes at each scale. To calculate these anchors, k-means clustering is applied in YOLO v3. They are grouped according to the scale at three separated places at the Network. YOLO v3 calculates offsets to predetermined

anchors to forecast the true width and real height of the bounding boxes. This offset is also known as the log-space transform, and it uses the sigmoid function to estimate the centre coordinates of the bounding boxes. The width, height, and centre coordinates of the anticipated bounding box are calculated using the following formulae. The anticipated bounding box's centre coordinates, width, and height are represented by b_x, b_y, b_w, and b_h. The Network's outputs after training are t_x, t_y, t_w, and t_h.

$$\boldsymbol{b_x = \sigma\,(t_x) + \ c_x}$$
$$\boldsymbol{b_y = \sigma\,(t_y) + \ c_y}$$

$$\boldsymbol{b_w = p_w * \ e^{tw}}$$
$$\boldsymbol{b_h = p_h * \ e^{th}}$$

Consequently, this cell and its neighbors have an object score of nearly 1, whereas corner cells have an objectness score of almost 0. In other words, the objectness score represents the probability that this cell is a center cell responsible for predicting one particular object and appropriate bounding box containing the object inside. The distinction between an objectness score and class confidences is that class confidences describe the likelihood that the discovered object belongs to a specific class. The chance that the bounding box includes an item is represented by the objectness score. The following is a mathematical representation of the objectness score.

$$P\_object * IoU = \sigma\,(t\_0) = P\_0$$

Where Item is a predicted probability that a bounding box includes an object, and IoU denotes the intersection of the predicted and ground truth bounding boxes. The result is sent into a sigmoid function, which returns values between 0 and 1.

3.3 Proposed System

The python library is used and set up the YOLO v3 environment for building the model. Initially, the dataset set is prepared with a set of 25 classes containing all the medical-related objects. The dataset preparation includes the data augmentation techniques, such as flip, brightness, and shear. This step increased the size of the dataset and improved the performance of the accuracy. The dataset is loaded in the python program with the Train, Test, and Validate folder. The training folder contains

70% of the dataset which undergoes the training and features are extracted. The model is trained using the GPU in the cloud environment. The Google Colab is used to train the model efficiently in a shorter period. The Tesla P100 GPU was used to train the model with all the 25 classes of labels. The cloud environment-based machine learning practice provides a handy way to reuse and share the code anywhere and anytime.

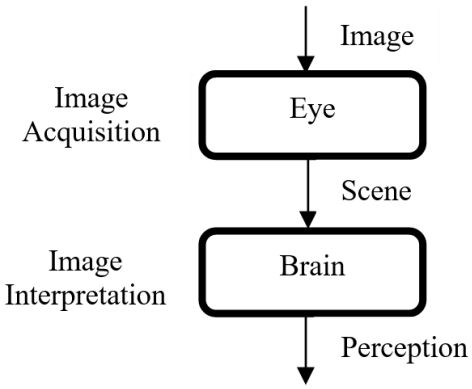

Figure 3.2: Proposed Framework for Object Detection

Figure 3.2, shows the architecture of the proposed framework. The acquisition of the image is the acquisition of the items in the model to perform the classification. Objects are defined by a qualified model and the image mark is positioned at the top of the objects. In addition to defining the objects in the input and the model, the python library is learned to spell out the name of the label found in the input.

The model is extracted after the completion of the research work and saved as a backup file on the computer for later use. This mechanism provides the option to reuse the best model by adding the additional classes to the existing model.

3.4 Results and Discussion

The model is trained with 100 epochs in the GPU for 3 hours and achieved the best accuracy results. First, the model was trained with 10 classes using CPU but it takes more than 18 hours to obtain the accuracy of 98.46%.

Figure 3.3: Achieved Results for Object Detection and Classification

The capability of identifying and detailing us about the unknown objects whether static or moving in the real-life environment. The application not only displays information in the form of text but also spells out the text in an automated voice to help comprehend the object easily. Name display above the mobile screen with audio that spells out the text. In figure 3.3, the objects are identified using the object detection algorithm. As the research work, mainly focuses on the medical object's classification, which identifies the lab coat, stethoscope, and lotion.

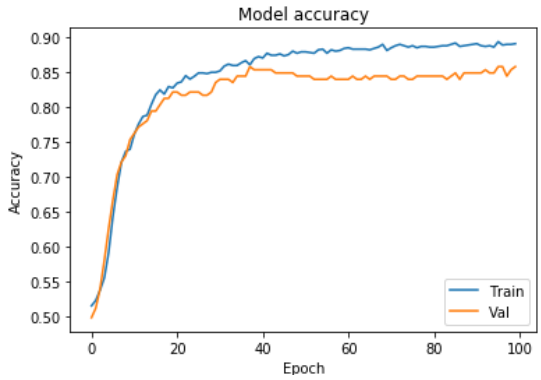

Figure 3.4: Model Training and Validation Accuracy

Figure 3.4, shows the learning growth and accuracy of the trained model.

CHAPTER 4

FETAL BRAIN LOCALIZATION

4. Fetal Brain Localization

4.1 Introduction

Machine learning, which teaches machines to be smarter and it is a special type of artificial intelligence. Deep learning is a subdivision of machine learning. Originally developed in the 1960s, deep learning is a modern reincarnation of artificial neural networks. It is a group of simple, layered training units working together to solve complicated tasks or model them. In general, with smaller data sets and reduced computation, other approaches typically perform well, which is what we had in the 1980s and '90s. But with broader data sets and greater model sizes and more computing power, we found that neural networks run much more easily.

A typical neural network is provided by an input layer, hidden layers, and an output layer. While these data are calculated by the hidden layers, the input layer encompasses inputs in different forms. The output sheet then supplies the product of the measurements and extractions. In each of these levels, some neurons are connected to neurons in the previous layer, and each neuron has its weight. As they perceive knowledge as spatial, convolutional neural networks function differently. Instead of neurons being bound to each neuron in the previous layer, they are only attached to neurons close to them and all have the same weight. The simplification of the relations implies that the network preserves the spatial structure of the data collection. This suggests that the network does not assume that the entire image comprises a Region of Interest (ROI). The word coevolutionary refers to the filtering process that takes place in this kind of network. The feeding of a complex image into it is clarified by a convolutional neural network, so it can be properly analyzed and understood. Several layers are made up of a convolutional neural network [101] inside a normal neural network. There are a few layers that make it unique, the convolutional layer and the pooling layer. It will, however, also include a layer of a rectified linear unit (ReLU) and a completely connected layer, like most neural networks. The ReLU layer functions as an initialization device as the data travels through each layer of the network, preserving non-linearity. Without it, the data fed into each layer would lose the dimensionality that needs to be retained. The entirely connected sheet, meanwhile, assists in the classification of the data collection to be carried out. The convolutional layer functions

and generates a complicated feature map by adding a filter over a set of image pixels. We've had the second pooling sheet. It downsamples or lowers the sample size of a given feature map. If it eliminates the number of criteria to be managed by the network, it also makes it much simpler to manage. The success of this is a pooled feature diagram. To do this, there are two methods: complete pooling, which takes the maximum input of the transformed function given, or average pooling, which only takes the average. These steps involve the extraction of features from which, according to its own mathematical rules, the network generates a representation of the image details. The entirely related layer is used to perform the classification task.

Medicine is a human operation, literally. The narrow AI is the sort of AI that focuses specifically on medication, which is to order a machine to perform a very narrow task that a human does not do [102]. In reality, pathology has enormous possibilities and transformations in the work culture. As an example, these latest technologies would have a great deal of potential to boost patient care, an algorithm that will immediately classify cancer cells [103] with the specific location and emphasize just what type of cancer it is, say in a lymph node, breast cancer. This model or algorithm would be more than the average breast pathologist, which would be a huge improvement in quality or specificity that is useful for patient care and could do this much more accurately than a human pathologist could do [104]. It is a very boring, time-consuming task that, by using this AI, can be addressed to save pathologists' time. A major gap down the lane is seen by the difference between doing models and the thorough stuff an average radiologist does. A pathologist with an additional degree in AI model technology and facilities from the new clinical application would be able to seamlessly change the workflows of the patient's healthcare in the ongoing growth [105].

4.2 Method

The goal of Yolo V4 is to build a fast-functioning object detector that is also equipped for parallel processing for production systems. It had to be better in a variety of aspects, as compared to the present practices. It is super-fast, high quality, and provides convincing results for object detection in terms of accuracy. Item detectors usually consist of several components: Input-This is where the picture is entered. Next, variants of VGG16, Resnet-50, Darknet52 or ResNext50, can be the backbone, which refers to the network that takes the image as input and extracts the feature map. The neck and

head are backbone sub-sets that maximize the discriminability and robustness of the function using FPN, PAN, RFB, etc., and the forecast-managing head. For a single-stage detector such as Yolo and SSD, this may either be used for dense prediction or FRCNN and Mask RCNN, a two-stage detector also known as Sparse Prediction.

The option of architecture is the mechanism that can conjure up a suitable entity detector to be created. For the backbone, based on theoretical logic, there was an alternative between CSPResNeXt50, CSPDarknet53, and EfficientNet B3 [106], and several CSPDarknet53 [107] neural network tests were found to be the most optimal model. The Yolo V4 concept, the CSPDarknet53 Spatial Pyramid Pooling block, also known as SPP, was used. Since the receptive region is significantly enhanced, it distinguishes the most important context characteristics and produces practically no decrease in the speed of network traffic. As a form of parameter aggregation for various detector levels from different backbone levels, PANet, also known as the Path aggregation network, is also used, and this was used instead of the FPN, also known as the YOLOv3 Feature Pyramid Network. Eventually, they chose the head, Yolo V3 [60], as Yolo V4's head. Different classifier training features Different training features of the detector Different backbones and pre-trained training weightings of the detector Different minibatch sizes of detector training Different training features of the detector Since there are several features that they had to test, particularly in the bag of freebies and specials. So the approach they used was to use a methodology called ablation analysis to test every feature. An ablation analysis is when you manually remove parts of the input to see which parts of the input are relevant to the network performance.

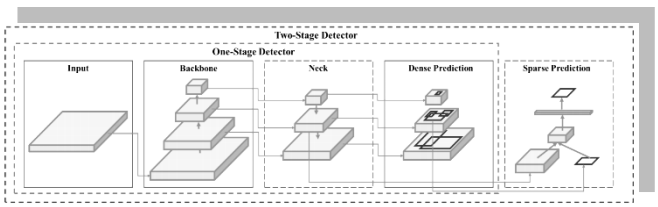

Figure 4.1: Yolo V4: Object Detection.

Normally, it looks like a table like this with the observations on the right-hand side. Speaking of results, if we look at how YOLOv4 relates to others, you would be very impressed. But to ensure that we compare each other with oranges and apples. Fig 4.1. depicts the steps involved in the object detection process in the Yolov4. Separate GPU architectures are used for inference cycle checking to test broadly accepted GPU

architectures as competing models. The comparative GPU architectures used were the Maxwell, Pascal, and Volta architectures. YOLO V4 is comparable to the fastest and most reliable in terms of both speed and accuracy. This analysis uses a state-of-the-art detector that on MS COCO datasets is faster in terms of Frames per Second (FPS) and more reliable than all available alternative detectors. On a traditional 8-16GB VRAM GPU that is readily accessible, Yolo v4 can be educated and used. The Yolo V4 is checked with a broad variety of features and the best ones are selected to improve both the classifier and the detector efficiency.

4.3 Proposed System

The Darknet device is used to train the tiny Yolo V4 for fetal brain images. The object detection technique is implemented from the darknet method to distinguish the fetal brain and the abnormalities from the source picture. The main advantage of this approach is that we can use the deep learning method to optimize the model and train on some kind of photograph to get the best results to identify the area of interest (ROI). In the scientific work, the Fetal Brain is concerned with ROI. The main objective is to perform the classification and to the fetal brain, which is taken as the region of interest (ROI) in the research work. Figure 4.2 below represents the fetal brain localization architecture. Initially, the dataset is prepared with the annotations in the darknet supported format to perform the object detection and classification using yolo4. There are 3 classes are used in the system such as Fetal Brain Coronal, Fetal Brain Axial, Fetal Brain Sagittal. The data augmentation technique was planted in the dataset to increase the size of the dataset. With the increase in the size of the dataset, the model to predict the ROI with great accuracy. There are a total of 283 images in the size of 256 x 256 is taken as the input to perform the data augmentation. It performs the shear, flips, exposure, cut-outs, cut mix, and noise technique and increases the number of images in the base dataset. The data augmentation has increased the size of the dataset with the ratio of 1:6.75 which generates the 1910 images in total. The whole dataset is split into test, train, and validate categories. The train folder contains 80% of the dataset. The test folder is given 20% of the entire dataset.

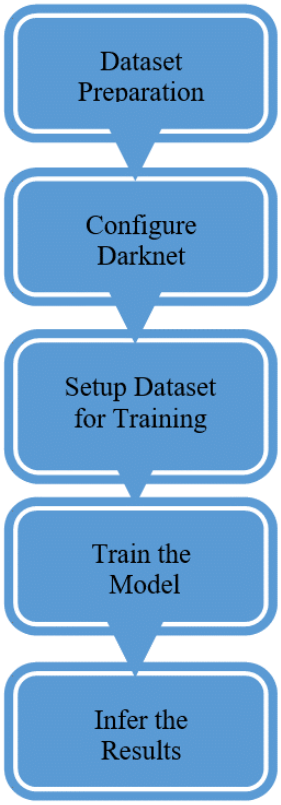

Figure 4.2: Proposed Architecture for Object Localization

The next significant big step is, to configure the environment to train the model with the prepared dataset. The Tesla T4 GPU is used with a clock rate of 1.59 GHz and the size of 16 GB memory used to perform the training model. The computing capability is set to 75 since the GPU is Tesla T4 which provides high performance and faster training capacity. The darknet is loaded with the necessary weights and conv 29. Now, the dataset is fed into the model and the total number of classes is defined. In our work, 3 classes were taken and the corresponding annotation is provided as the input to the model. The obj.names file will contain the label names which are placed along with the train, test, and validation folder. The filter is prepared according to equation (1), which computes the 18 filters for our work.

$$\text{Number of filters} = (\text{Number of classes} + 5) \text{ x } 3$$

The batch size of 64 and subdivision of 24 with the linear activation are defined for the model. The next step is to train the model to predict the fetal brain with its orientation. The CUDA allocates the average memory of 52.8 MB during the training, whenever the

training GPU requires the memory. On an average of 3.26 seconds, 1000 iterations are performed and weights are stored in the backup repository to reuse. The weights are stored in 3 variants, such as best, last, first 1000 for future purposes. The model can be re-used with greater accuracy by performing more training with the increased dataset and enabling the option to add a variety of different classes to the existing model, which provides the support of enhancement features to the work.

4.4 Results and Discussion

The output is achieved with the 3 different functions of boxes, classes, and scores. The detection_boxes function is used to identify the fetal brain in the given input image. The detection_classes is used to categorize the class of the fetal brain. The detection_scores step will provide the accuracy of the class matched on the given input image. The cuDNN of 7.6.5 and OpenCV 3.2.0 are used to perform the classification in this model. The following figure 4.3 shows the classification of the fetal brain according to orientation. The classification is performed with fetal brain sagittal, axial and coronal orientation which is achieved at an accuracy of 97.92 %. The localization of the fetal brain was obtained for the physician to understand the growth, size, and abnormalities. This model can able to provide the results by feeding the raw MRI image to the system and able to achieve results of less than 0.4 Seconds.

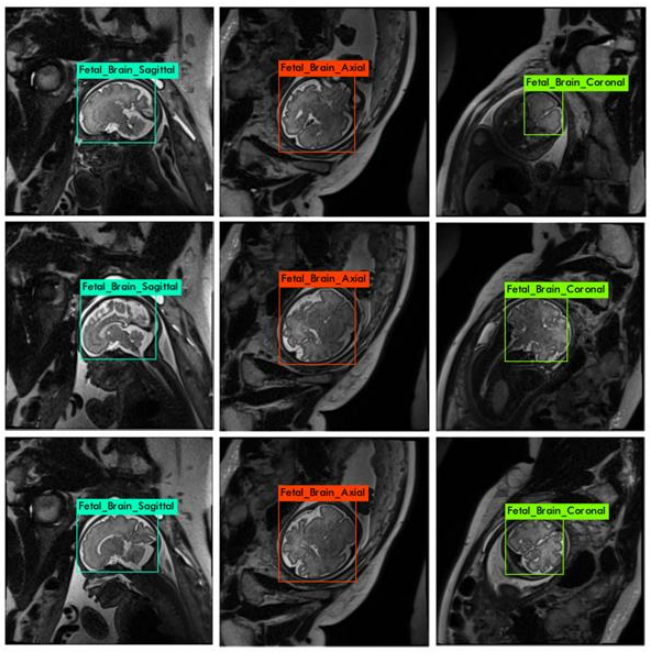

Figure 4.3: Classification of Fetal Brain with Sagittal, Axial, Coronal Orientation

Below table 4.1 explain the obtained paradigm after the achievement of the results out of the model.

Table 4.1: Results of Fetal Brain Localization

Computational Paradigm	Achieved Results
Dice	98.59%
Jaccard Index	96.47%
Accuracy	97.92%
Precision	96.70%
Recall	98.65%

Table 4.2: Accuracy Comparison

Authors	Accuracy
Taleb et al.	84.2
Martin Rajchl et al.	94.1
Keraudren et al.	93
Our Proposed Method	97.92

CHAPTER 5

FETAL BRAIN ABNORMALITIES DETECTION AND LOCALIZATION

5. Fetal Brain Abnormalities Detection and Localization

5.1 Introduction

Artificial intelligence has been around for decades; the word was created in 1956 at Dartmouth College, but it has only recently gained traction due to advances in computer power, access to algorithms, libraries, research, and data. The wind starts in the year 2011, and in 2013 many technical giants start booming into AI. From 2015 through the middle of 2019, the growth curve has shifted. Artificial Intelligence (AI) is a broad term that encompasses a variety of topics. Machine learning is a subdivision of AI, while deep learning is a subclass of Machine Learning (ML).

The core objective of computer vision is to build a machine, which can process and interpret images in the video just like a human visual system does, it may be a compliment of biological vision. When the similarity between the human visual system and the computer vision system is compared. Human Visual System [108] has eyes to see images, videos, or any objects and it is processed in our brain to make intelligent decisions. In case of the computer vision, cameras are represented as eyes, which may be single or multiple cameras for an image or the video equation. Then, the image pre-processing is performed, with pattern recognition and artificial intelligence algorithms are applied for making the decisions. Neural Networks is a very powerful and expressive, machine learning architecture to learn arbitrary input/output functions, with enough training data. A neuron is the functional unit of an input/output node or, to be more exact, a neuron. This node receives an input signal U and performs a mathematical operation to produce output Y. This is multiplied by a constant or added to a constant, or it can be more sophisticated, so often people use sigmoidal function z, these are called activation functions, and if U is small, the output is just 0 and if U is large, regardless of how large it is, the output might be 1 and then there's some smooth activation function from 0 to 1.

The artificial neuron is mathematically represented by,

$$z = f\left(\sum w_i x_i + c\right)$$

Where x_i and w_i are the input and weights respectively which are linearly combined with an added bias offset of c.

When the neuron, that unit and start to stack it either in series or in parallel or both. The two neurons are placed next to each other and do a more complicated function. The neurons can be added further to make the structure more complex to build up an Artificial Neural Network (ANN). The neural network with nodes and edges describes the topology of how it's connected. Since it's artificial so neural networks are built up out of these building blocks and the different nodes can have different activation functions. They can add up linear combinations of their previous layer and if keep stacking more and more of these layers so each layer is doing some kind of sequential processing and by adding a lot of these then result as a Deep Neural Network (DNN) [109].

The convolutional neural networks or CNNs are the ones that have gained big popularity in the community and, if you go down most people would be associating a deep neural network or anything to do with deep learning with the CNN. It is mainly used in image recognition [110]. The basic idea of CNN is, it takes a mask from the given input image and slides it across the image doing local computations in local patches which pull out the edges or features. When it is run through a convolutional layer and pulls out these edges and features, then is given as an input for another convolution layer and another convolutional layer. The CNN layers [111] are stacked and processed. Once the features are extracted, then downsize the images using max-pooling or other methods. The reason to downsize the image is to apply the filter to the larger area of an image. The features at a later stage are usually more semantic and can be interpreted by humans such as the head of the bird or eyes of the bird finally we need to flatten the vectors into one dimensional vector and perform the image classification. Finally, apply the soft max for multiple class. The output layers dimension is equal to the number of classes in the data set.

The designing and implementation of these architectures are becoming extremely simple because of the explosion of open-source software by Google and Facebook and others. The Tensorflow and PyTorch and Keras, are these incredibly powerful environments where you can design neural network architectures and then train them with training data to build these very powerful and expressive models so it's really neat and also increasingly easy to design and use.

Machine learning has created a computer software that can learn from experience robotically. Machine learning is the fastest-growing technical discipline at the intersection of statistics and computer science, and it is a key component of artificial intelligence. Machine learning approaches were implemented in Science, Engineering, Technology, and Commerce, resulting in data-intensive decision-making in fields as diverse as healthcare, manufacturing, education, financial forecasting, agriculture, and police, among others. The first is the difficulty of generating computer programmes that learn from experience; the second is the task of establishing numerical computational information-theoretic principles that govern learning in computational systems in particular and in any organisation in general. The goal of machine learning is to investigate the issues. Working on means and goals to advance the extremely practical computational systems that have been filled in many applications, as well as underlying the theory of learning.

Deep learning is a set of techniques, say a subset of machine learning and it's one of the growing techniques that have been used in the industry specifically for difficulties in computer vision, NLP, and speech recognition. In our daily life, a lot of different tools and plug-ins on your smartphones that use this type of algorithm. The reason it came to work very well is primarily, the new computational methods. Deep learning is computationally expensive and needs to find techniques to parallelize the code and use GPUs specifically to graphical processing units, to be able to compute, the computations in deep learning. The second part is the data available has been growing after, after the Internet bubble, the digitalization of the world. So now people have access to large amounts of data and this type of algorithm has the specificity of being able to learn a lot when there is a lot of data. So, these models are very flexible when more data is fed in. It will be able to understand the salient feature of the data and finally algorithms. People have come up with new techniques to use the data, use the computation power and build models.

5.2 Methods

5.2.1 Convolutional Neural Network

The object detection using CNNs performed by taking similar kinds of the same object (K) in the different perceptions, for example, varieties of cars are taken to perform the classification. The training is performed with these as the input to the CNN to learn a

car classifier, so it is a standard cross-entropy loss of CNN. The different sliding windows of multiple scales, say 100000 sliding windows potentially from the image and each of those sliding windows is given as the input to the CNN assuming that they all are of the same size or can be brought to the same size. At test time, an image is fed into the model to identify whether the K (car) is available inside the image and if it is positive. The CNN classifies each window in the original image as belonging to a K or ~K. In this approach, non-maximum suppression is included to improve the performance but before getting the list of all possible bounding boxes where K could be present, consider a bounding box with the highest-class signal. The IOU of all the other bounding boxes is calculated and the object K is bounded using a rectangular box. The object detector will locate the object, for our example, the car is located precisely in the image with a bounding box.

The loss function is defined in equation 1, which illustrates the linear combination of confidence and location loss. The confidence loss, which uses categorical cross-entropy to determine how confident the network is in the computed bounding box [112].

$$L(x, c, l, g) = \frac{1}{N}\left\{L_{conf}(x, c) + \alpha L_{loc}(x, l, g)\right\}$$

The L2 norm is used to determine how far away the network's anticipated boundary boxes are from the ground truth ones.

5.2.2 Region-Based Convolutional Neural Network

The R-CNN or Region-Based Convolutional Neural Network was proposed in 2014 by Girshek and the team. This approach has a first stage where it uses a method known as selective search which is an image processing method that doesn't use neural networks. But using a selective search approach it proposes several regions that could have objects; each of these regions is then evaluated by a CNN to perform classification as well as do bounding box regression to get the correct localization inside those region proposals. The selective Search module uses a graph-based image segmentation or a mean shift-based image segmentation to get an initial set of region hypotheses. So, the hypotheses are hierarchically grouped based on region similarity measures like colour, texture, size, region filling, and so forth.

Once you get the region proposals, this particular method selects 2000 such region proposals by combining and qualifying different subregions to get somewhere around 2,000 region proposals. Each of these regional proposals is wrapped to the same size because it is given as the input to the CNN. Since CNN can only take a fixed input size as its input dimension, as it is the nature of the architecture. The feature maps are acquired out of CNN that goes through a fully connected Network. The next step is the softmax component, which classifies each of these regions into one of the known classes that are the ROI, for a while performing detection. Finally, it also takes the feature maps through a different fully connected layer to perform a bounding box regression.

5.2.3 YOLO v4

YOLO v4 features superior accuracy and an even faster detection speed when compared to all currently existing detectors. The fourth version was recently introduced in April 2015 by Alexei Bukowski, as seen in fig 5.1; the major purpose of this method was to create a super-fast object detector with great accuracy.

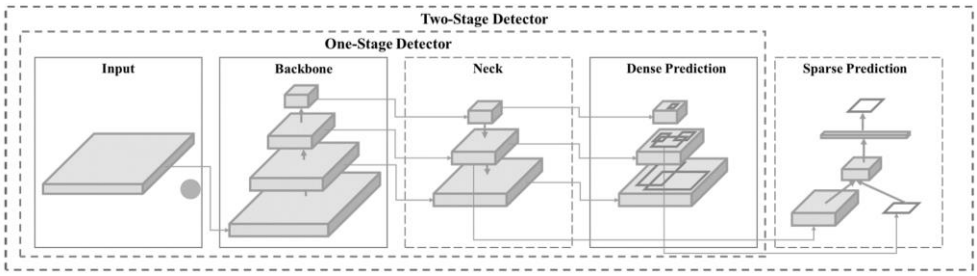

Figure 5.1: YOLO v4 Framework

Object detector topologies are typically made up of numerous components. The input is the image, followed by the backbone and neck, which take the image as input and extract the feature maps using a DNN, and ultimately it makes and handles these predictions using an object detector such as yellow or SSD. The optimized object detector uses the Yolo v3 architecture as the end-of-chain object detector. A single Graphical Processing Unit can be used to train the model (GPU). The bag of freebies and bonuses is an added feature of the YOLO v4 over the YOLO v3.

The YOLO v4 architecture's backbone can be CSPRESNEXT50, CSPDARKNET53, or EFFICIENTNET B3. When compared to the all-existing model, the best architecture

for predicting the object is CSPDARKNET 53. It is much more efficient to organize context features with no loss in network operation processing time. The YOLO v3 serves as the head for the YOLO v4. Since the YOLO v4 is considered one of the best object detectors, it can support parallel computing.

By altering the training approach, these freebies will help to increase the model's accuracy. Plugin modules and post-processing approaches are among the specialisations, which may raise the inference cost somewhat but considerably improve accuracy.

5.2.4 Dataset

The Prenatal Development dataset of 159 subjects is collected from OpenNEURO public dataset supported by Stanford and National Science Foundation [113]. The extracted dataset has been transformed and is appropriate for building a machine learning model to identify, localize, and classify anomalies in the foetal brain.

5.3 Proposed System

The YOLO v4 network architecture is imparted to perform the detection, classification, and localization of the fetal brain and its abnormalities. Throughout this work, the fetal brain and abnormalities are considered as the region of interest.

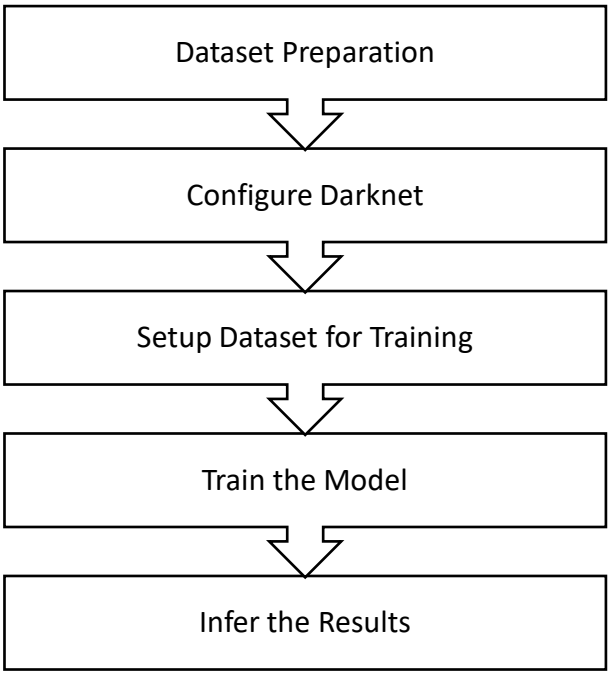

Figure 5.2: Proposed System for Fetal Brain Detection, Classification, and Localization

The main objective of the work is to perform the detection of fetal brain abnormalities from the given MRI image using Precise Epic Localization (PEL) algorithm. The fig. 5.2 represents the process of the proposed system. The raw data after the data acquisition step is taken as the input to the system.

 # Pseudocode: Precise Epic Localization (PEL) Algorithm

Start

Pre-processing

Load Dataset (.nii format) from the Local Disk

Convert from .nii to .jpg

Resize to 256 x 256

Convert Train Label

Clean the dataset

Processing Step

Train the Model

Localize the Fetal Brain

Classify healthy and unhealthy ROI

If: Unhealthy

Localize and Enhance the Abnormality

Else:

Denote the Orientation of the ROI

Post-processing

Enhance the Localized Image for interpretation

Stop

The well-known format of .nii is used to perform the detection and classification of the ROI. The fetal volumes of various axes are taken to train the model. Initially, the dataset is prepared with the fixed size of 256 x 256, and ready to be fed into the CNN. The dataset was prepared by implementing the data augmentation method to increase the number of images. It performs the shear, flips, cut-outs, cut mix, exposure techniques and elevated the count to 1910 images in the dataset, which is of the ratio 1:6.75 from the initial dataset.

The cloud environment is organized for performing the training of the model. The model was executed in the Tesla T4 GPU [114] with the size of 16 GB memory capacity. The GPU has a competence of 1.59 GHz of clock rate with a computing capability of 75. In the next step, the training and testing data are equipped with an 80:20 ratio respectively. There are two significant abnormalities are taken as the classes for the training of the model with three planes of fetal brain images. The obj.names will contain all the classes that need to be predicted from the given image. The batch size of 64 with the linear activation function is adopted to the model. The training is started in the GPU to build the model for classification of the abnormalities with 1000 iterations in 3.26 seconds and weights are stored for reusability.

5.4 Results and Discussion

The detection_classes function contains the Fetal Brain with all the three planes of axial, coronal, and sagittal are considered as three different classes and two critical abnormalities are added in the classes. If the given input of MRI image contains the normal and healthy fetal brain. It will be marked with a bounding box with Fetal_Brain + Plane (Sagittal, Coronal, and Axial) shown in Fig 5.3 below. The abnormal brain containing fetal abnormalities such as Arteriovenous Malformation, Encephalocele is marked with the name of the detected abnormality with the bounding box. The detection and localization of the fetal brain and abnormalities are represented in Fig. 5.4.

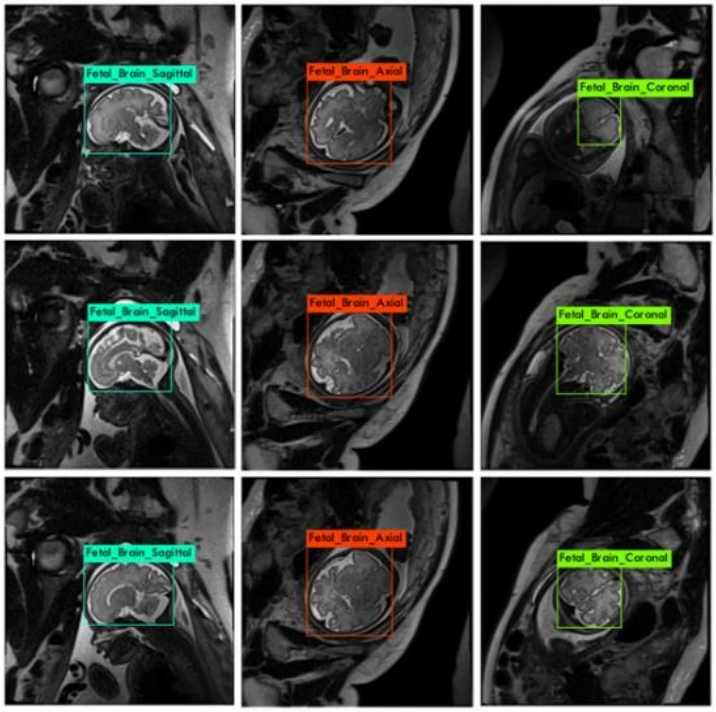

Figure 5.3: Detection and Localization of Fetal Brain with three planes

The enhancement of the abnormalities is highlighted by using the filter to analyze the lesion tissues in MRI images. The different colours can provide the intensity of the elements present in the region of interest.

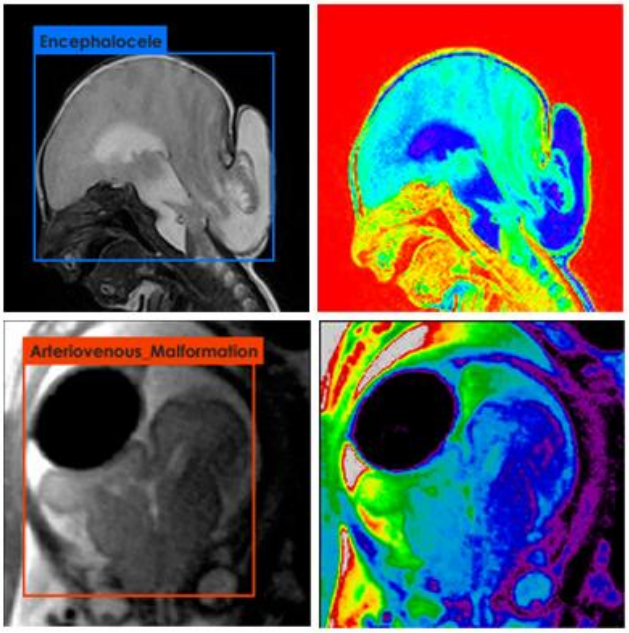

Figure 5.4: Detection and Localization of Fetal Brain Abnormalities

The physician can easily insight the impact of the problem and stages of the abnormalities with this method. The accuracy of 97.27% is achieved by this model. The Dice and Jaccard Index is computed to be 96.42% and 98.58% respectively. The output is received in less than 0.5 seconds from the time of giving the raw input to the model. The detailed achieved results are mentioned in table 5.1 below.

Table 5.1: Achieved Results for Proposed Model of Fetal Brain Abnormalities Detection and Localization

Computational Paradigm	**Achieved Results**
Architecture	Yolo v4
Dice	98.58 %
Jaccard Index	96.42 %
Accuracy	97.27%
Sensitivity	98.63 %
Specificity	99.40 %

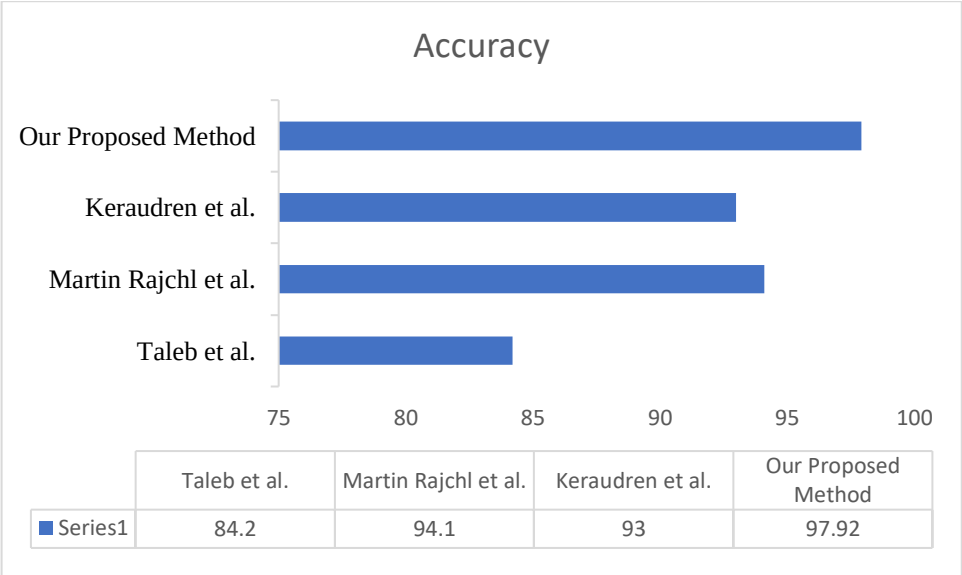

Figure 5.5: Accuracy Comparison

CHAPTER 6

FETAL BRAIN SEGMENTATION

6. Fetal Brain Segmentation

6.1 Introduction

Image segmentation is the process of portioning the image with the Region of Interest (ROI). In our research work, The Fetal Brain is taken as the region of interest and it is detected, located, and segmented using the machine learning algorithm. The image segmentation can perform on any image like flower, cat, satellite images, or medical images. The application of the interest will differentiate the method and technique to perform the image segmentation algorithm.

Image segmentation is a process of partitioning a digital image into multiple reasons and then extracting the meaningful reason, which is known as ROI. The reason of interest will be varying based on the application, consider the flower is separated from the background of the tumor is segmented from the given computer tomography (CT) or magnetic resonance image (MRI) images, which is highly helpful for the doctor to track the progress of the disease. In our work, the system can able to assist the physicians to understand the comprehensive structure and lesion impact on the fetal brain. A single universal segmentation algorithm cannot be adopted for all the applications of image processing. It tries with the various algorithms to identify, which is giving us the better result as it is a hit and trial method.

In medical imaging, the goal of a doctor is to analyze the tumor or any lesion tissue in a CT or MRI image. The doctor will try to identify the nature of the tumor awareness present it, here the tumor is considered as the ROI for this application. Similarly, Iris is well-thought-out as the Roi for an eye image. The success and failure of the extraction of ROI ultimately influence the success of image processing applications.

Image segmentation algorithms are classified into two approaches, similarity principle, and discontinuity principle. The similarity principle is also known as the region approach. The objective is to group the pixels based on the common property to extract a coherent region. The discontinuity principle is also known as a boundary approach. Consider, when drawing an image of a ball, there must be a certain boundary. Once it is observed closer, to the behavior, then the pattern and components can be extracted at the inner portion of the boundary and the outer portion of the boundary accordingly. It is imagined as a boundary because it is an abrupt change, so the boundary principle is

also known as a discontinuity principle. Another goal is to isolate areas with different attributes, such as intensity, colour, and texture. Consider the example of the ball; the intensity of the pixel will vary. It means mostly abrupt changes in intensity among the region, giving us the extraction of the edges.

An image can be partitioned into many smaller sub-regions so if you say the image is denoted by R. The original images have been divided into various components, say R_1, R_2, R_3, and R_4. The region R_2 is subdivided into four sub-components as R_{21}, R_{22}, R_{23}, R_{24} shown in figure 6.1.

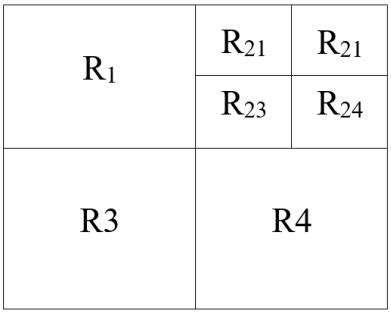

Figure 6.1: Regions and Sub-Regions

The boundary between the regions can be obtained using the intensity of the pixel in the region. In the below Fig. 6.2.

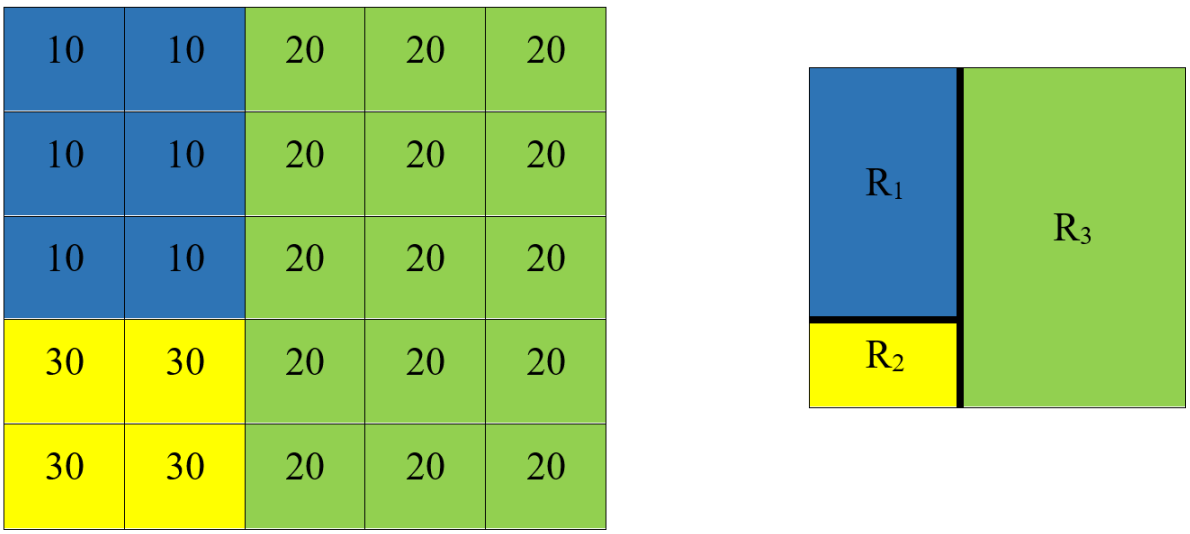

Figure 6.2: Regions with Pixel Intensity

The image with 5 rows and 5 columns is taken in the above image. The values in the corresponding boxes are the intensity level of the desired pixel.

In columns, 3, 4, and 5 all the values are the same while in columns 1 and 2 up to the third row all the pixel values are the same as 10. Similarly, the intensity is the same in the 4^{th} row and 5^{th} row for columns 1 and 2. Based on the similarity principle, the regions are grouped based on their intensity level. The matrix, as shown with the different colours to recognise consisting of all the pixels having values 10, 20, and the other contains the value 30. The boundaries will separate the region with simple and clear, which helps to perform the analysis accurately.

Let R represents the whole image and segmentation is partitioning R which is the given image into n subgroups with is denoted as R_i and I varies from 1 to n. If the sum of the regions is combined to obtain the original image, it can be mathematically written as,

$$U_{i=1}^{n} R_i = R$$

R1, R2, and R3 do not share any common property which means it must be null, mathematical it can be written as,

$$R_i \cap R_1 = \emptyset \ \ (\forall\, i \ and \ j) : i \neq j$$

$P(R_i)$ = TRUE for i = 1,2,3, … n

$P(R_i \cup R_j)$ = FALSE for i ≠ j

Predicate (P) or a set of predicates like intensity or it can be any other characteristics any other attributes of the image. It can be a colour, grayscale value, texture, or any other parameter.

6.2 Segmentation Algorithm

In the classification of image segmentation algorithms, there are various approaches to classify the images available in fig. 6.3. It can be performed in two categories based on the user interaction or based on the pixel relationship. The user will choose the method or approach to perform the overall process to obtain the ROI is termed as the user interaction segmentation method. Second, one is based on the pixel relationship, which estimates the correlation in the neighborhood of the pixels.

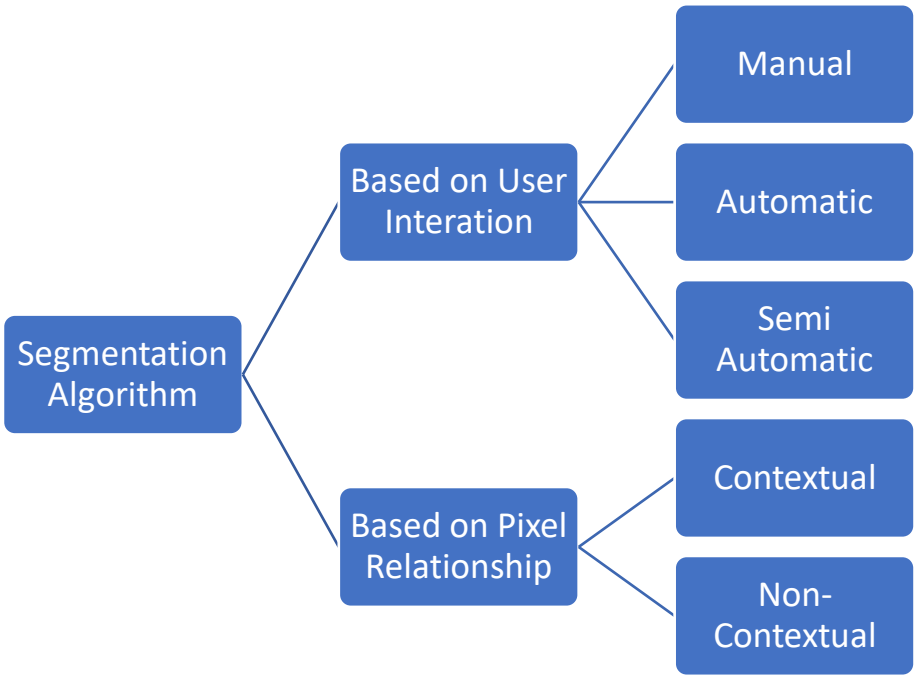

Figure 6.3: Classification of Image Segmentation Algorithms

The manual method object of interest is observed by an expert. The expert is the user who traces its ROI boundaries with the help of the software. It is completely based on the expert or observer who decides to address the segmentation on the image. Many software systems and tools which are available, helped in tracing the boundaries and extracting them to perform machine learning. But boundary tracing is a subjective process, variations may exist from the opinion of the expert to other experts. In this approach, the result validation is not being done. But in the case of the automatic segmentation method, the results are evaluated and validated.

Automatic segmentation is commonly preferred because it is being done which is used for the structures of the objects without the intervention of any human observer. It means they are preferred for every task that needs to be carried out even for a large number of images. A manual operator or a human observer can perform the task but for a smaller one.

In an automatic segmentation method, it can be performed at a very high rate so there is a requirement to utilize the automatic segmentation algorithm and it will carry out the result very quickly for a large number of images [115]. The semi-automatic segmentation method is in between the manual and automatic approaches. It is a combination of manual and automatic algorithms. In this method, human intervention

or an observer is required at the initial stage. The human observer is supposed to provide the initial seed point. These are provided by the human observer which indicates the ROI and the extraction process is carried out automatically.

The next broad category of the segmentation algorithm is based on the pixel relationship based on the similarity and discontinuity and it has two sub-categories, such as, contextual and non-contextual. the contextual segmentation algorithm is also known as a region-based or global segmentation algorithm. This algorithm groups the regions based on the common properties in general. Pixels are grouped based on some sort of similarity which exists between them. Similarity may be based on the colour, grey values, and texture.

The next segmentation algorithm is based on a non-contextual pixel relationship. The non-contextual is also known as pixel-based or local. Here the process is to identify the discontinuities which are present in the images and that will help us to segment the image and these kinds of discontinuities may be observed as lines or edges. It can be simply grouped into a region based on some global-level properties.

6.3 Methods

The U-Net architecture [116] is a system for segmenting medical images. It is made up of 47 MRI pictures with a total of 884 characteristics. The input picture was a segmented mask [117] with a dimension of 227 by 227 pixels.

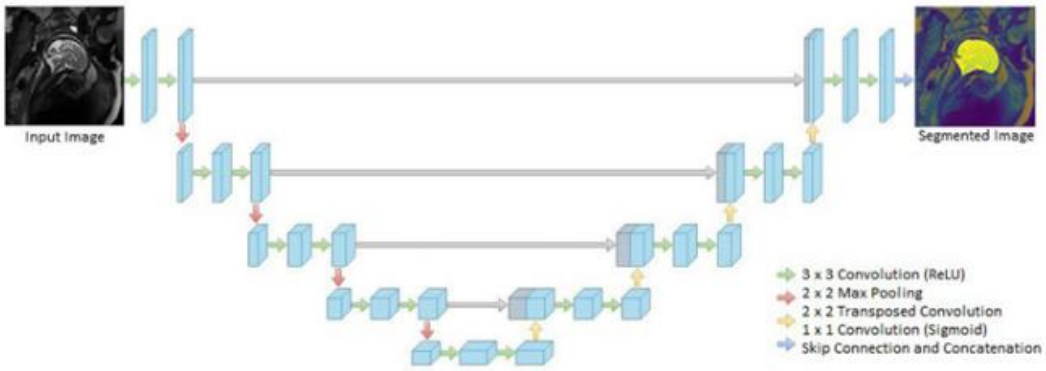

Figure 6.4: Fetal Brain Segmentation using U-Net Architecture

Expect Ki is the Kth layer's result. This vector is regularly produced by a planning AI comprising of a CNN and a non-linear activation function from the yield of the previous layer Ki in a U-Net architecture.

$$K_i = A_i\ (K_i - 1)$$

Connection follows the U-net Architecture example of iteratively linking every element yield in a feed-forward manner. The U-Net architecture was created with the purpose of segmenting biological pictures in mind [118]. U-Net is a completely convolutional network since it is comprised entirely of convolutional layers with no recurrent or thick layers. The U-Net architecture is named by the form of the architecture, which is shaped like a 'U'. The design is created using an encoder and a contracting/downsampling path, which is then followed by a layer with the bottleneck and a 3rd escalating/upsampling path.

In the downsampling process, there are two 3x3 unpadded convolutions employed, followed by a ReLU and having a stride of 2, a 2x2 max pooling operation. We square the number of feature channels at each stage of the downsampling process. The feature map connects the entire upsampling path, which consists of 2x2 convolution. The next step is to add two 3x3 convolution layers, each followed by a ReLU. The Skip connection, which serves as a link between the downsampling and upsampling paths, is an important part of the U-Net design. To get local information, the feature map created during the downsampling process is employed. In the upsampling process, it creates a full image by converting local information to global information.

A 1x1 convolution matches the feature vector to the target classes in the architecture's final layer. The localization information is collected by the downsampling channel, which is an advantage of the U-net design. To predict possible segmentation, the position and context of the fetal brain MRI are used.

6.4 Proposed System

The raw data from MRI scans are saved in the .nii format. The itk package loads the Dicom volume from the file path and saves it in an array using the itk reader. Using the nibabel library, each slice of the NIfTI-1 Data Format is translated to an hdf5 file format. The dataset is pulled from the folders, and the hdf5 file's location is supplied.

The Precise Slice Enhancement (PSE) approach is employed on the model represented in figure 6.5. It begins by reading the chosen route in order and placing all photographs in an array, labelling each one. In order to acquire cumulative value, normalisation (x) will determine the minimum (min val) and maximum (max val) values (C).

$$Cummalative\ Value\ C = \frac{(x - min\_val)}{(x - max\_val)}$$

Using the Histogram Equalization approach, the Cumulative values are used to flatten the input image. The OneHot (OH) function will be expanded along all three axes to increase the dimension required to identify the embryonic brain. The ROI in the supplied input picture is mapped using the Weight Map (WM) function

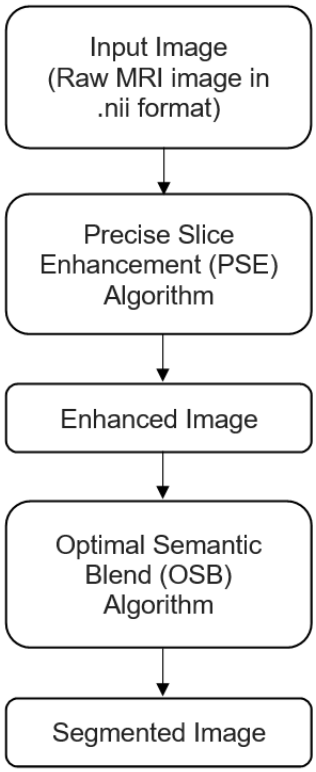

Figure 6.5: Architecture for Fetal Brain Segmentation

 # Pseudocode: Precise Slice Enhancement (PSE) Algorithm.

```
Start
# Input Section
        Volume V = "/Dataset/Test_Images/"
        Read Image (I) and Mask (M) in an Array
        I = down_sample (I)
        Read Features from read Image Array
# Processing Section - Precise Slice Enhancement Algorithm (Step 1)
        Normalize (N) the Image and obtain Cumulative Image
        Find Z_Min and Z_Max
        Analyze the Features with the given Input Image (Image)
# Post Processing Section - Precise Slice Enhancement Algorithm (Step 2)
  Enhanced Image EI = Histogram (Image)

# Display Output Section
  Display Enhanced Output Image
End
```

Three separate variables are obtained once the weight map is analysed. a) The argmax function is used to create the label for the picture on the axis=3 axis. b) In the presented picture, the canny edge detection technique is employed to morph the image segment using the binary dilation approach. b) For the specified image form, a weight map is constructed. Finally, to achieve image downsampling, utilise the softmax function. Finally, to achieve image downsampling, utilise the softmax function.

The softmax $p_c(x)$ represented using the following equation,

$$p_c(x) = \exp \frac{\left(a_c(x)\right)}{\sum_{c'=1}^{K}\left(a_{c'}(x)\right)}$$

where $a_c(x)$ signifies feature channel c activation at pixel location x. The maximum-function approximation is p_c, and the number of classes is K. $a_c(x)$ for the c with the highest activation, $p_c(x)$ 1 for the remaining k, and $p_c(x)$ 0 for the c with the lowest activation, $p_c(x)$ 0. Weight Map will apply the label and perform semantic segmentation on the provided input picture by displaying it in a variety of colours. A batch of images is fed into the model using the Precise Slice Enhancement (PSE) method.

Finally, the Optimal Semantic Blend (OSB) Method was applied to improve the PSE algorithm's image quality (EI). The OSB algorithm receives the upgraded fetal brain image and segments the foetal brain region with a distinct colour.

 # Pseudocode: Optimal Semantic Blend (OSB) Algorithm

```
Start

# Input Section

  Read Enhanced Image (EI) and Image Mask (IM)

# Processing Section

  Verify the dimension of EI and IM

  SI = Blend EI and IM

# Display Output Section

  Display Segmented Image (SI)

End
```

6.5 Results and Discussion

The PSE technique is used to find the foetal brain, and then a mask is constructed using the test image given. The segmentation findings from the raw input file (.nii format) are shown in Figure 6.6, It assesses the shape and structure of the foetal brain MRI given. The picture has been improved using the histogram tool, which has highlighted the dense tissues in the ROI. To amplify and underline any anomalies on the test sample, the OneHot (OH) and Weight Map (WM) functions were applied. The ROI's shape and size are assessed in order to aid the physician in making a more precise diagnosis. By disregarding the remainder of the ROI image space, the OSB algorithm gives the foetal brain picture an unique (yellow) colour, allowing for more detailed examination and highlighting of foetal brain anomalies. There are 17 test images in all. In 19.23 seconds, the nii format may be subdivided automatically.

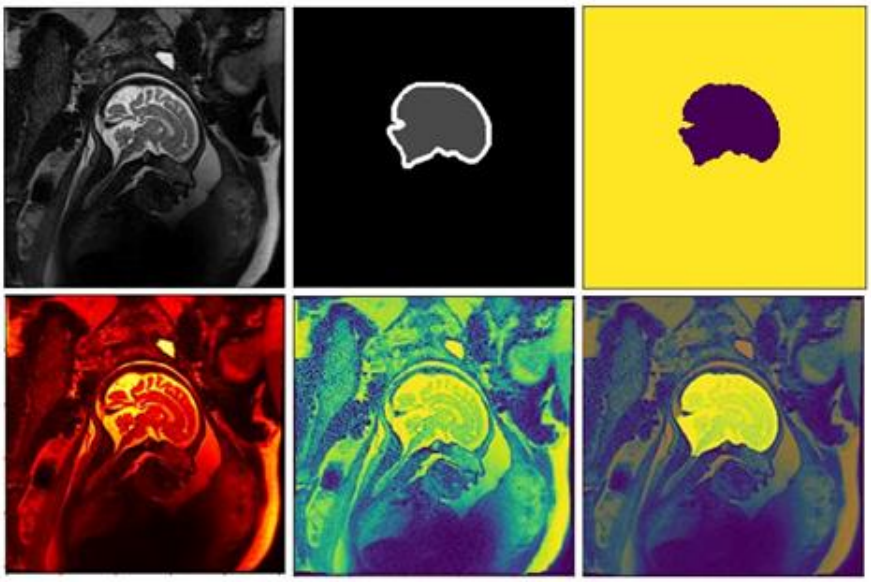

Figure 6.6: Fetal a) Raw Input b) Boundary of the Fetal Brain c) Shape of the brain) Enhanced Image e) Enhanced Image with Histogram f) Segmented Results

The suitable ROI mask was determined by True Positive Fraction (TPF) as sensitivity using Eq. (4.1), and the remaining region in the input picture was determined by True Negative Fraction (TNF) as specificity using Eq. 4.2. Eq. 4.3 and Eq. 4.4 were used to get the Dice (D) and Jaccard (J) Coefficients, respectively.

Eq.4.1 True Positive Fraction TPF is given by,

$$\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Eq.4.2 True Negative Fraction TNF is given by,

$$\frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}$$

Eq.4.3 Dice Coefficient D is given by,

$$\frac{2\,|\text{True Positive}|}{2\,|\text{True Positive}| + 2|\text{False Negative}| + 2\,|\text{True Negative}|}$$

Eq.4.4 Jaccard J is given by,

$$\frac{2\,|\text{True Positive}|}{2\,|\text{True Positive}| + 2|\text{False Positive}| + 2\,|\text{False Negative}|}$$

When compared to alternative designs, the Dice score rose by more than 7%. The model was trained using a Graphics Processing Unit (GPU) NVIDIA 1050 to boost efficiency and minimise training time. This technique saves around 3 hours when compared to using a CPU.

Table 6.1: Achieved Results of Improved Fetal Brain Segmentation Algorithm

Computational Paradigm	**Achieved Results**
Architecture	U-Net
Dice	98.58 %
Jaccard Index	96.42 %
Sensitivity	98.63 %
Specificity	99.40 %
Time	4.45 seconds / Iteration

Table 6.2: Computational Results during Training

Computational Paradigm	Min	Max	Cur
Accuracy for Training	0.882	0.929	0.929
Accuracy for Validation	0.918	0.925	0.928
Mean Squared Error for Training	0.011	0.019	0.011
Mean Squared Error for Validation	0.011	0.013	0.011

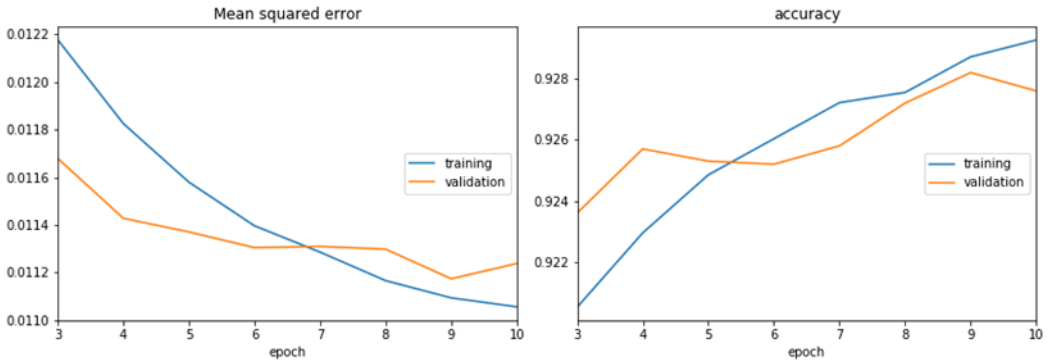

Figure 6.7: Mean Square Error and Accuracy

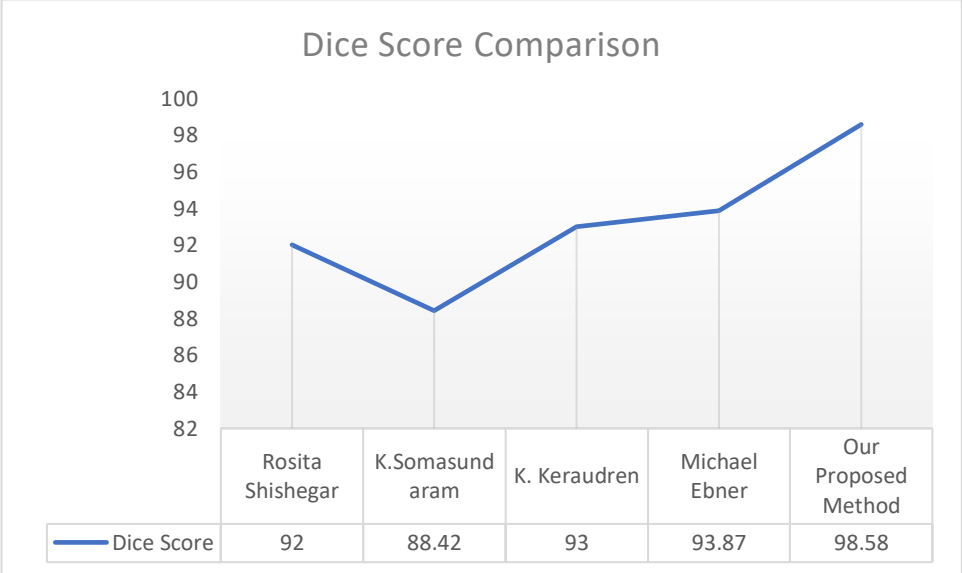

Figure 6.8: Dice Score Comparison

Tables 6.1 and 6.2 show the percentage and computational outcomes attained throughout training, respectively. The full training graph of mean squared error and accuracy is shown in figure 6.7. Images from the first three semesters of MRI The training set consists of 47 fetal brain volumes, each with 52 slices, while the test set consists of 15 volumes. The model has been trained using 823 features and can segment the different problematic fetal brain structures along all three axes.

CHAPTER 7

FETAL BRAIN ABNORMALITIES SEGMENTATION

7. Fetal Brain Abnormalities Segmentation

7.1 Introduction

The radiologists are tasked with increasing and completing the outstanding duty of collecting additional patient information for electronic health records. It will take more time to perform the manual segmentation of the differences from the norm in the photographs based on the information acquired. The computational framework's features aid the pathologist in reducing the number of errors made by a significant margin.

With the aid of growing technology, automatic segmentation is the ideal option to handle this issue. Machine Learning (ML) has made tremendous strides in recent years, conquering every area where time consumption must be lowered and accuracy must be raised. ML will let physicians to operate more efficiently and precisely, resulting in better patient care results [119]. From image capture through interpretation and sharing of results, ML has the potential to be a valuable contribution to the full image interpretation lifecycle. A drive is being made to develop new procedures and technologies to more effectively assess and treat illnesses that impact a large portion of the population. In this decade, ML protection medical services applications may be used to identify between various illnesses or any lesion, which is generating critical innovation in the health care business. In the medical profession, ML and modern IT technologies will be used to create data-driven or evidence-driven machines with a human-optimized workflow.

Machine Learning has a significant role to play in improving the medical industry with enhanced early illness and abnormality prediction. Medical imaging segmentation is an important part of the diagnostic procedure [120]. It will be more laborious, incredibly mind-boggling, and prone to high error rates and delays in conclusion if the person manually evaluates the photos one by one for the aberrant lesion. Perhaps the treatment is risky and has unintended consequences for the patient's health. Using Machine Learning Algorithms, Automatic Segmentation is a precise approach to discover irregularities.

Fetal abnormalities affect three out of every 1000 women [121]. The major goal of this study is to find, segment, and improve Fetal Brain Abnormalities using Fetal MRI

images, which are used as a Region of Interest (ROI). Various deep learning algorithms have been developed to increase MRI image enhancement and identify anomalies with the advent of profound learning. Magnetic Resonance Imaging, or MRI for short, is a unique technique that allows for a more accurate interpretation of foetal brain scanned data.

The MRI pictures are used as a source of data for our model. The picture enters the room in the .nii format, which is the scanning machine's direct output. The .nii photos are fed into the model and machine learning methods are applied to them. The conclusion created will aid the clinician in determining the lesion's form, size, severity, and, more importantly, the foetal brain abnormalities. This methodology eliminates human-caused inference errors and reduces the amount of effort required during an evaluation.

7.2 Method

The deep convolutional network for biomedical image segmentation mainly employs the U-Net Architecture. The U-Net is dedicated to biomedical photos alone. As demonstrated by these little arrows, the U-Net design will be similar to all other convolutional networks in that it will consist of a plethora of distinct jobs. The input picture is fed into the network first, followed by data generation in every feasible method until the segmentation map is retrieved at the end. Each blue box on the x and y size scale denotes a multi-channel function and the number of channels incorporated.

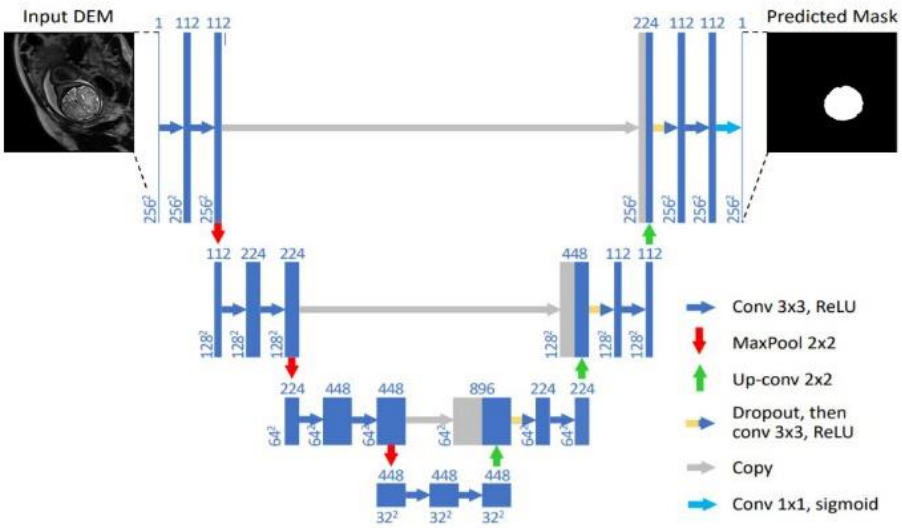

Figure 7.1: U-Net architecture

In the majority of the tasks, convolutions are utilised, followed by a nonlinear activation function. A 3x3 convolution is used, followed by a nonlinear activation function. A crucial design decision is that it employs the valid section of the convolution, which means that for a 3 by 3 convolution, a 1-pixel border is lost, allowing large images to be handled in individual tiles afterwards. The next operation in the U-Net is max-pooling, which decreases the XY size of the element map, as shown by the downward arrow.

Each channel is followed independently by the max-pooling algorithm, which generates the maximum activation from each 2x2 window to the next feature map.

$$f(x) = ln\,(1 + e^{x}) \ldots (1)$$

$$f(x) = \frac{ln(1 + e^{x})}{k} \ldots (2)$$

After each max-pooling procedure, the number of highlighted channels increases by a factor of two. In general, the arrangement of convolutions and max-pooling processes generates a spatial contraction, which gradually enhances throughput while reducing wear. This is where the traditional categorisation ends. At the present, all features that alter a single yield vector have an additional extension method for constructing a high-resolution segmentation map. This expansion path is made up of a sequence of up convolutions that are linked to the high-resolution characteristics of the contracting path. In the output segmentation map, there are two distinct channels: foreground and background class. Because of the random convolutions, the map is smaller than the input image. The following is the definition of the Softmax Output layer,

$$\emptyset\,(S_q) = \frac{e^{Sq}}{\sum_{i}^{\infty}\left(e^{S_j}\right)} \ldots (3)$$

Where S denotes the input vector.

7.3 Proposed System

The foetal brain MRI picture is fed into the system, and the images are saved in the T2 imaging standard of .nii or.nii.gz format. In this study, 94 volumes of healthy fetal brains and 23 volumes of unhealthy foetal brains were employed. The proposed procedures for performing the Improved Semantic Blend Segmentation (SBS)

Algorithm followed by the Optimal Slice Enhancement (OSE) Algorithm are shown in figure 7.2.

i) The dataset preparation for localization and segmentation is part of the pre-processing procedure. Using the python module med2image, the pictures are first converted from .nii to.jpg format. This library will take all of the .nii files from the source folder and convert them to JPG files, which will then be stored in the destination folder.

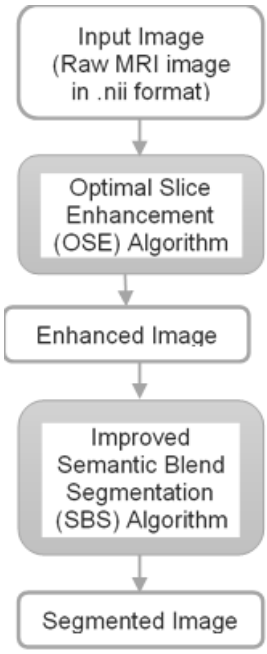

Figure 7.2: Steps to obtain the segmented region of interest

ii) The Tensor Flow Record (TF Record) file must be produced in the processing phase using the stated classes. Fetal Brain (Region of Interest) and Unhealthy Fetal Brain with different abnormalities will be the classes in this study. The TF Record will be created from the.csv files containing the training photos, and the same will be done for the testing images containing the test. The labelmap (LM) file will be prepared in concurrently with the csv file. The ROI will be localised using the U-Net architecture.

The training and testing record directory has been configured in the model, and it is ready for training. The MRI scans of the foetal brain pictures are separated into two folders: the Healthy Fetal MRI images (Normal) and the Unhealthy Fetal MRI images (Abnormal), which are utilised in the training phase. The training will begin with the model being trained on the NVIDIA GPU 1050 to learn the ROI image attributes.

Machine Learning is used to train foetal brain pictures utilising collected image attributes. The training will begin with the model being trained on the NVIDIA GPU 1050 to learn the ROI image attributes. Machine Learning is used to train foetal brain pictures utilising collected image attributes.

The trained model is used to forecast anomalies based on the input. The model is fed the test picture in order to determine if it is healthy or harmful. If you're unwell, what kind of anomalies have you discovered? Three separate courses were used to train the model. The healthy fetal brain is depicted in Class 1, whereas Encephalocele and Arteriovenous Malformation are addressed in Classes 2 and 3. In each of these categories, the model will highlight the physician, the type of aberration discovered, and the impact of the lesion in the supplied input.

Pseudocode: Improved Fetal Brain Abnormality Segmentation

Start

Pre-processing

Load Dataset (.nii format) from the Local Disk

Convert from .nii to .jpg

Resize to 256 x 256

Convert Train Label from .xml to .csv and generate tfrecord

Clean the dataset from Outliers

Processing Step

Train the Model

Localize the Fetal Brain

Classify healthy and unhealthy ROI

If: Unhealthy

Segment and Enhance the Abnormality

Post-processing

Apply the Improved Semantic Blend Segmentation Algorithm to obtain the segmented image

Enhance the Segmented Image

Stop

iii) The segmentation of the foetal brain is started in the Post-Processing Step by feeding the input picture to the Improved Semantic Blend Segmentation Algorithm. To achieve the segmented picture as the result, the algorithm will blend the processed image into the selected LM file. 3 As a result, the output will be improved.

7.4 Results and Discussion

The algorithm receives the raw input image as input and processes the segmented improved image as output. The whole incremental process of the system output is depicted in Figure 7.3. In Fig. 7.3 (a), the algorithm with the best perspective for analysis selects a slice of the MRI foetal brain, which is originally in the format. nii. The label of the embryonic brain is retrieved from the input picture in Fig. 7.3 (b).

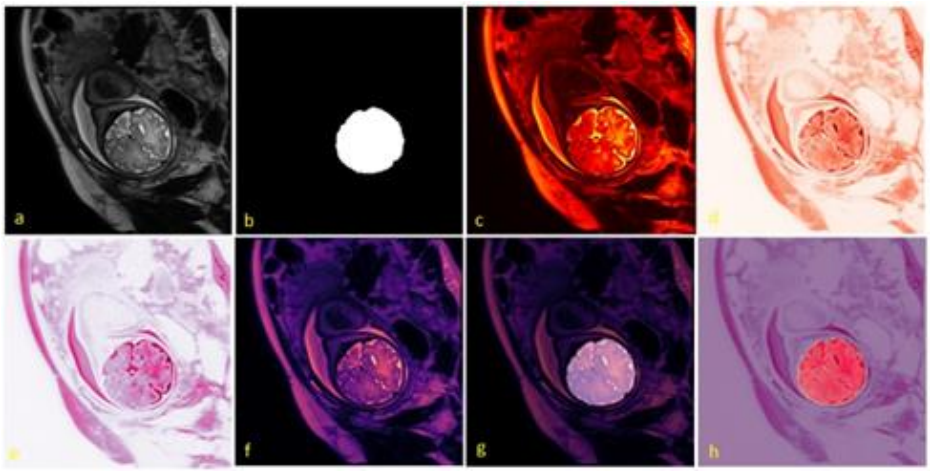

Figure 7.3: Healthy Fetal Brain Segmentation and Enhancement

The improved picture of the foetal brain is processed in Fig. 7.3 (c, d, e, f) to produce a better visualisation analysis for the clinician to discover anomalies. The segmentation of the embryonic brain and its enhanced representation is shown in Figure 7.3 (g, h).

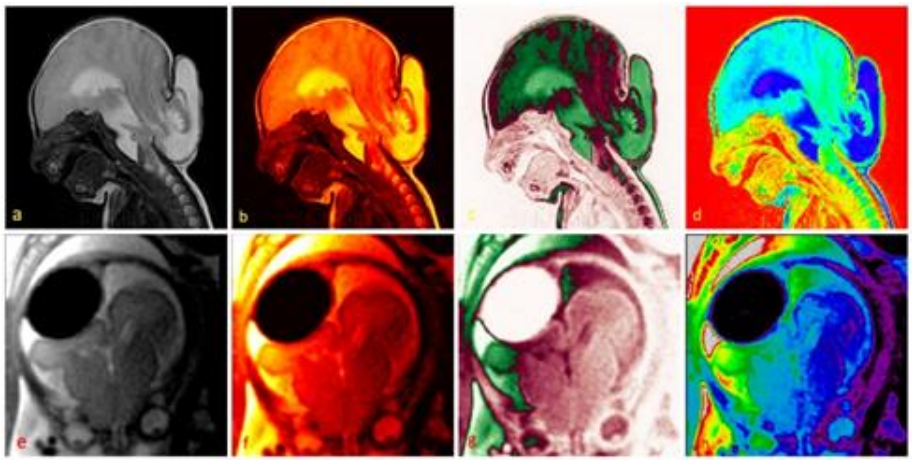

Figure 7.4: Fetal Brain Abnormal Cases Segmentation: i) Encephalocele (a, b, c, d) and ii) Arteriovenous Malformation (e, f, g, h)

Figure 7.4 shows an interpretive picture of the Class 1 category of the embryonic brain, which is generally devoid of abnormalities. Figure 6.4 processes classes 2 and 3 of encephalocele and arteriovenous malformation, respectively, providing a comprehensive resource and perspective for clinicians to study the patient's illness development and make decisions, which is incrementally better than earlier techniques.

Table 7.1: Batch wise Time analysis in seconds

Case	Batch 1	Batch 2	Batch 3
1	5.39	6.81	7.89
2	5.37	6.47	7.44
3	5.35	6.23	7.11
4	5.35	6.08	6.64
5	5.34	5.94	6.15
6	5.38	5.88	5.98
7	5.49	5.85	5.92
8	5.44	5.82	5.55
9	5.3	5.68	5.19
10	5.3	5.75	4.96
11	5.39	5.81	4.81
12	5.3	5.77	4.62
13	5.36	5.76	4.5
14	5.33	5.71	4.31
15	5.38	5.73	4.55
16	5.38	5.71	5.29
17	5.36	5.78	5.7

All of the pictures were created automatically after receiving input from a local disc, which is far faster than manual MRI image processing. The system's most important feature is that it reduces the time it takes to segment foetal MRI pictures.

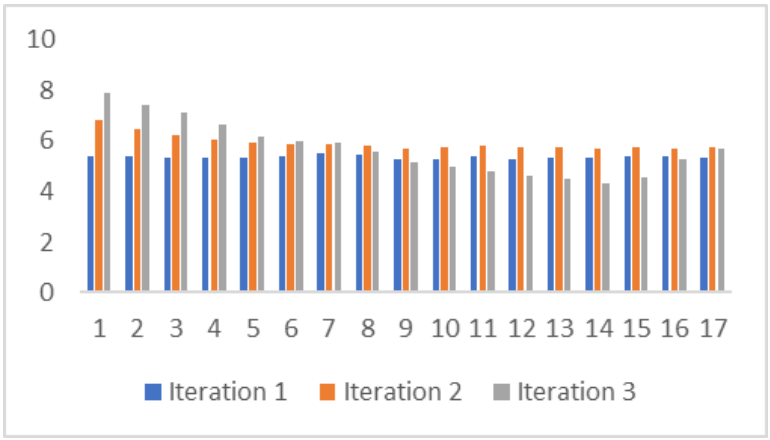

Figure 7.5: Segmentation Time per Case for 3 Batches

In 318 seconds, the anomalies in a batch of 17 MRI instances are located and segmented using the model in Table 7.1. The model takes an average of 18.74 seconds per example, which is extremely fast when compared to other models.

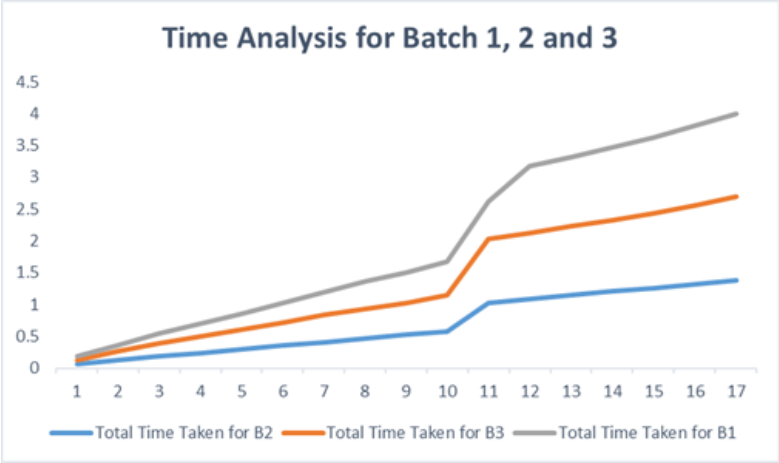

Figure 7.6: Time Comparison graph for 3 Batches

The bar chart and graph in Figures 7.5 and 7.6 respectively interpret the individual case time taken and relate it to the rest of the batch's volume. It took 5 minutes and 18 seconds to finish the entire batch of 17 MRI volumes of pictures.

The following bar chart in Fig 7.7 interpret the dice score comparison analysis with the creamy existing proposed methods.

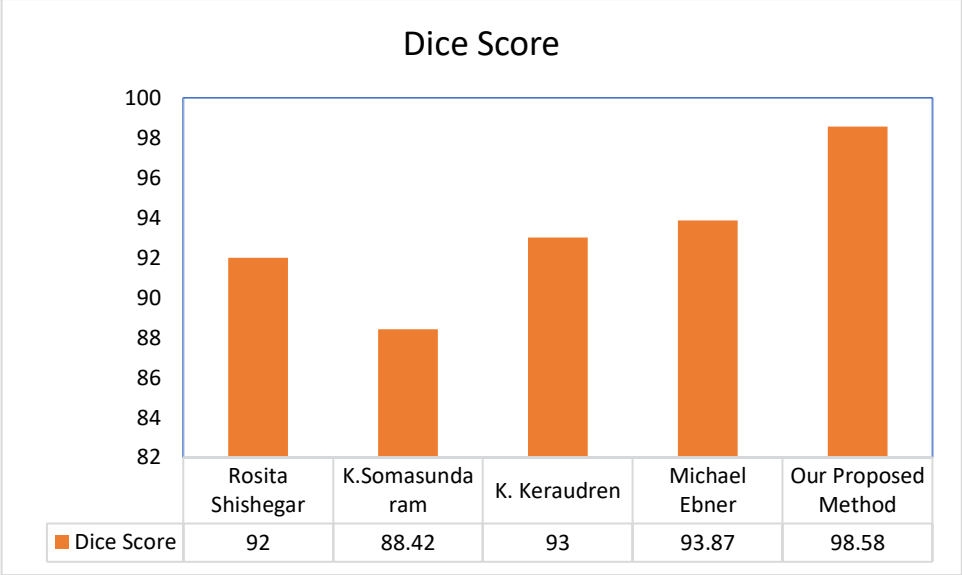

Figure 7.7: Dice Score Comparison

CHAPTER 8

CONCLUSION

8. Conclusion

The fetal brain segmentation is mainly addressed in the research work to save the life of the child. The abnormalities can be identified in the earlier stages, could help to prevent the effect of the harmfulness to the child's health. The various techniques were analyzed during the literature survey state, merits and demerits in the processing power, computational time, and lack of interpretation possible with the resultant output. The numerous direction and inputs are collected during the literature study. For example, Automatic segmentation is a significant method to work with fetal brain images which are more efficient on the MRI images, and to reduce fetal motion artifact, it is possible to use thicker cuts for more eminent image quality. The various challenges like size, texture, pattern, orientation, low light, and unusual intensity, computation hardware, etc., are faced during the development of the model to perform the fetal brain segmentation. The problem statement includes detecting, locate and segmenting the fetal brain and its abnormalities, which is taken as the core subjective problem for the research work. In this work, various machine learning techniques were proposed to address all the issues faced during the challenge of fetal brain segmentation. So, the work is developed into 5 different phases to gradually provide the solution for the core objective of the research.

Phase one starts with the hands-on experience of object detection with machine learning. The purpose of this work is to establish a method for the recognition of medical objects in 2D images. This work discusses experimental studies on different approaches to the recognition and classification of 25 different classes. The model is trained in the Tesla P100 GPU for 3 hours in the cloud environment and achieved an accuracy of 98.46%. The complete coding is performed in the Google Colab platform to attain the results at a faster rate and accessible from all over the world. During this investigation, the object recognition system developed was tested with the benchmark datasets of COIL100, Caltech101, ETH80, and MNIST, the results outperform all the existing methods.

The next phase of the research work includes the detection and classification of the fetal brain. The model is built using the Darknet YOLO v4 architecture in the cloud environment and achieves an accuracy of 97.92%. The challenges faced during the work are orientation, size, shear, dataset pre-processing. In the process of constructing

the model in the darknet, YOLO v4 outperforms all the current techniques with greater precision and efficiency. The testing is carried out on the Tesla T4 GPU, which requires approximately 9 hours of training. Initially, the model was educated in the CPU, which takes more than 1 day and the process could not be completed. For the machine learning algorithm, thus, training with nominal GPU would be easier as per the knowledge obtained from this project. Segmentation must be prepared for the same dataset in the future, where it is possible to segment the precise position of the fetal brain.

The core implementation phase of the work is performed to segment the fetal brain from the given input of MRI images. It is performed by the machine learning model and achieved the results with the dice coefficient of 98.58% and the Jaccard index of 96.42%. The Precise Slice Enhancement (PSE) algorithm and Optimal Semantic Blend (OSB) Algorithm is adopted in this work to achieve the fetal brain segmentation successfully. The accuracy is comparatively higher than the existing techniques. As the physician only needs to provide the raw image of MRI images in the .nii format as the input to the algorithm and the model will perform the automatic segmentation and assist them with handy interpretation results with the enhanced highlighted output. The test image of 3.2% gets failed in some challenging cases (including motion) with our algorithm.

As the improved version of the work, the classification, localization, and segmentation are performed by this model. The classification informs whether the given input is the normal or abnormal fetal brain. This is the first milestone that delivers significant data about the given input. The localization is performed on the next step to depict the ROI position, with its meticulous shape and size. If the brain is abnormal the lesion tissue is emphasized in the phase. In the final phase, Segmentation of the fetal brain is performed by using the Improved Semantic Blend Segmentation Algorithm, which provides the fetal brain and its abnormalities. This model was a tower of strength to physicians, who work with fetal brain segmentation.

The detection and localization of the fetal brain's healthy and unhealthy categories of images are highlighted in this work. The YOLO v4 framework is used in the cloud environment outperforms all the existing techniques by saving computational time and space. The GPU has provided a great provision in achieving the results of accuracy,

97.93%. Initially, the training of the same model was performed in the CPU in an offline mode, which was time-intensive and required a lot of processing resources. As a result, training the model on the cloud is the best option to develop a model capable of detecting and classifying an object. A significant challenge confronted in the fetal brain localization of the ROI is the pre-processing of the image for the machine learning dataset. In the future, the model is extended to classify the major number of fetal abnormalities with more accuracy. The fetal brain exploration with accurate detection and localization is a countless aid for all the physicians to save the life.

In the future, segmentation is performed with various tissues in the fetal MRI images and CT images which helps to diagnose any abnormalities in detail and be interpreted easily using deep learning techniques. The Challenging cases with motion-affected MRI images have to be addressed in the pre-processing to increase the efficiency of the model accuracy. The Model should be capable of performing all the detection, classification, localization, and segmentation of the fetal brain in one framework. The work can be extended to perform all the same with all the significant organs in the fetus to provide the complete monitoring of the pregnancy duration.

References

[1] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *Radiographics*, vol. 37, no. 2, pp. 505–515, Mar. 2017, doi: 10.1148/rg.2017160130.

[2] E. J. R. Van Beek *et al.*, "Value of MRI in Medicine: More Than Just Another Test? HHS Public Access," *J Magn Reson Imaging*, vol. 49, no. 7, pp. 14–25, 2019, doi: 10.1002/jmri.26211.

[3] C. J. Garvey and R. Hanlon, "Computed tomography in clinical practice," *British Medical Journal*, vol. 324, no. 7345. BMJ Publishing Group, pp. 1077–1080, May 04, 2002, doi: 10.1136/bmj.324.7345.1077.

[4] K. Bhatia and H. Lombaert, Eds., *Machine Learning Meets Medical Imaging*, vol. 9487. Cham: Springer International Publishing, 2015.

[5] O. Lézoray, C. Charrier, H. Cardot, and S. Lefévre, "Machine learning in image processing," *Eurasip Journal on Advances in Signal Processing*, vol. 2008, no. 1. Hindawi Publishing Corporation, pp. 1–2, May 29, 2008, doi: 10.1155/2008/927950.

[6] E. C. Lyons, "Digital Image Processing: An Overview," *Computer (Long. Beach. Calif).*, vol. 10, no. 8, pp. 12–14, 1977, doi: 10.1109/C-M.1977.217813.

[7] W. Huda and R. Brad Abrahams, "X-ray-based medical imaging and resolution," *Am. J. Roentgenol.*, vol. 204, no. 4, pp. W393–W397, Apr. 2015, doi: 10.2214/AJR.14.13126.

[8] G. D. Rubin, "Computed tomography: Revolutionizing the practice of medicine for 40 years," *Radiology*, vol. 273, no. 2. Radiological Society of North America Inc., pp. S45–S74, Nov. 01, 2014, doi: 10.1148/radiol.14141356.

[9] A. Carovac, F. Smajlovic, and D. Junuzovic, "Application of Ultrasound in Medicine," *Acta Inform. Medica*, vol. 19, no. 3, p. 168, 2011, doi: 10.5455/aim.2011.19.168-171.

[10] S. Tocchio, B. Kline-Fath, E. Kanal, V. J. Schmithorst, and A. Panigrahy, "MRI evaluation and safety in the developing brain," *Seminars in Perinatology*, vol. 39, no. 2. W.B. Saunders, pp. 73–104, Mar. 01, 2015, doi: 10.1053/j.semperi.2015.01.002.

[11] T. Yousaf, G. Dervenoulas, and M. Politis, “Advances in MRI Methodology,” in *International Review of Neurobiology*, vol. 141, Academic Press Inc., 2018, pp. 31–76.

[12] N. Bonnier, “Attributes of image quality for color prints,” *J. Electron. Imaging*, vol. 19, no. 1, p. 011016, Jan. 2010, doi: 10.1117/1.3277145.

[13] M. Holzer, F. Schumacher, T. Greiner, and W. Rosenstiel, “Optimized hardware architecture of a smart camera with novel cyclic image line storage structures for morphological raster scan image processing,” *2012 IEEE Int. Conf. Emerg. Signal Process. Appl. ESPA 2012 - Proc.*, pp. 83–86, 2012, doi: 10.1109/ESPA.2012.6152451.

[14] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10. pp. 1345–1359, 2010, doi: 10.1109/TKDE.2009.191.

[15] R. Vaila, J. Chiasson, and V. Saxena, “Feature extraction using spiking convolutional neural networks,” Jul. 2019, doi: 10.1145/3354265.3354279.

[16] J. H. -, “Application of Back-Propagation Network in Healthcare Industry,” *Int. J. Adv. Inf. Sci. Serv. Sci.*, vol. 5, no. 4, pp. 278–286, Feb. 2013, doi: 10.4156/aiss.vol5.issue4.36.

[17] M. Zurowietz and T. W. Nattkemper, “An Interactive Visualization for Feature Localization in Deep Neural Networks,” *Front. Artif. Intell.*, vol. 3, Jul. 2020, doi: 10.3389/frai.2020.00049.

[18] A. Hebbar, “Augmented intelligence: Enhancing human capabilities,” *Proc. - 2017 3rd IEEE Int. Conf. Res. Comput. Intell. Commun. Networks, ICRCICN 2017*, vol. 2017-December, pp. 251–254, Dec. 2017, doi: 10.1109/ICRCICN.2017.8234515.

[19] I. E. Suleimenov, Y. S. Vitulyova, A. S. Bakirov, and O. A. Gabrielyan, “Artificial Intelligence: What is it?,” in *ACM International Conference Proceeding Series*, Apr. 2020, pp. 22–25, doi: 10.1145/3397125.3397141.

[20] T. Davenport and R. Kalakota, “The potential for artificial intelligence in healthcare,” *Futur. Healthc. J.*, vol. 6, no. 2, 2019, doi: 10.7861/futurehosp.6-2-94.

[21] R. Zemouri, C. Devalland, S. Valmary-Degano, and N. Zerhouni, “Neural network: A

future in pathology?," *Annales de Pathologie*, vol. 39, no. 2. Elsevier Masson SAS, pp. 119–129, Apr. 01, 2019, doi: 10.1016/j.annpat.2019.01.004.

[22] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological Physics and Technology*, vol. 10, no. 3. Springer Tokyo, pp. 257–273, Sep. 01, 2017, doi: 10.1007/s12194-017-0406-5.

[23] M. Goudbeek, D. Swingley, and R. Smits, "Supervised and Unsupervised Learning of Multidimensional Acoustic Categories," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 35, no. 6, pp. 1913–1933, Dec. 2009, doi: 10.1037/a0015781.

[24] J. He, S. L. Baxter, J. Xu, J. Xu, X. Zhou, and K. Zhang, "The practical implementation of artificial intelligence technologies in medicine," *Nature Medicine*, vol. 25, no. 1. Nature Publishing Group, pp. 30–36, Jan. 01, 2019, doi: 10.1038/s41591-018-0307-0.

[25] D. Gamez, "The Relationships Between Intelligence and Consciousness in Natural and Artificial Systems," *J. Artif. Intell. Conscious.*, vol. 07, no. 01, pp. 51–62, Mar. 2020, doi: 10.1142/s2705078520300017.

[26] G. W. Ng and W. C. Leung, "Strong Artificial Intelligence and Consciousness," *J. Artif. Intell. Conscious.*, vol. 07, no. 01, pp. 63–72, Mar. 2020, doi: 10.1142/s2705078520300042.

[27] J. C. Flowers, "Strong and weak AI: Deweyan considerations," *CEUR Workshop Proc.*, vol. 2287, 2018.

[28] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, "A proposal for the Dartmouth summer research project on artificial intelligence," *AI Mag.*, vol. 27, no. 4, pp. 12–14, 2006.

[29] T. Poggio, "Neuroscience: New insights for al?," *Proc. - 2006 IEEE/WIC/ACM Int. Conf. Web Intell. (WI 2006 Main Conf. Proceedings), WI'06*, pp. 3–5, 2006, doi: 10.1109/WI.2006.123.

[30] Z. Shen, Z. Liu, J. Li, Y. G. Jiang, Y. Chen, and X. Xue, "Object Detection from Scratch with Deep Supervision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, 2020, doi: 10.1109/TPAMI.2019.2922181.

[31] A. Bhandari, J. Koppen, and M. Agzarian, “Convolutional neural networks for brain tumour segmentation,” *Insights Imaging*, vol. 11, no. 1, pp. 1–9, Dec. 2020, doi: 10.1186/s13244-020-00869-4.

[32] R. Yang and Y. Yu, “Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis,” *Frontiers in Oncology*, vol. 11. Frontiers Media S.A., p. 638182, Mar. 09, 2021, doi: 10.3389/fonc.2021.638182.

[33] F. Pesapane, M. Codari, and F. Sardanelli, “Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine,” *European Radiology Experimental*, vol. 2, no. 1. Springer, Dec. 01, 2018, doi: 10.1186/s41747-018-0061-6.

[34] M. A. Malbog, “MASK R-CNN for Pedestrian Crosswalk Detection and Instance Segmentation,” *ICETAS 2019 - 2019 6th IEEE Int. Conf. Eng. Technol. Appl. Sci.*, Dec. 2019, doi: 10.1109/ICETAS48360.2019.9117217.

[35] X. Shen and I. Stamos, “3D Object Detection and Instance Segmentation from 3D Range and 2D Color Images,” *Sensors 2021, Vol. 21, Page 1213*, vol. 21, no. 4, p. 1213, Feb. 2021, doi: 10.3390/S21041213.

[36] L. Sun, W. Shao, and D. Zhang, “High-order boltzmann machine-based unsupervised feature learning for multi-atlas segmentation,” *Proc. - Int. Symp. Biomed. Imaging*, pp. 507–510, Jun. 2017, doi: 10.1109/ISBI.2017.7950571.

[37] C. Yuan and H. Yang, “Research on K-Value Selection Method of K-Means Clustering Algorithm,” *J*, vol. 2, no. 2, 2019, doi: 10.3390/j2020016.

[38] D. Demirović, “An implementation of the mean shift algorithm,” *Image Process. Line*, vol. 9, 2019, doi: 10.5201/ipol.2019.255.

[39] M. Hahsler, M. Piekenbrock, and D. Doran, “Dbscan: Fast density-based clustering with R,” *J. Stat. Softw.*, vol. 91, 2019, doi: 10.18637/jss.v091.i01.

[40] A. V. Bernstein and E. V. Burnaev, “Reinforcement learning in computer vision,” 2018, doi: 10.1117/12.2309945.

[41] S. Reis *et al.*, “Automated Classification of Breast Cancer Stroma Maturity from Histological Images,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 10, pp. 2344–2352, Oct.

2017, doi: 10.1109/TBME.2017.2665602.

[42] C. C. Aggarwal, “Advanced Topics in Deep Learning,” *Neural Networks Deep Learn.*, pp. 419–458, 2018, doi: 10.1007/978-3-319-94463-0_10.

[43] Y. Wu *et al.*, “Elastic Deep Learning in Multi-Tenant GPU Clusters,” *IEEE Trans. Parallel Distrib. Syst.*, 2021, doi: 10.1109/TPDS.2021.3064966.

[44] T. W. Haw and R. Logeswaran, “Morphological based technique for image segmentation,” *Int. J. Inf. Technol.*, vol. 14, no. 1, pp. 413–419, 2009.

[45] C. Kim, “Segmenting a low-depth-of-field image using morphological filters and region merging,” *IEEE Trans. Image Process.*, vol. 14, no. 10, 2005, doi: 10.1109/TIP.2005.846030.

[46] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2016, vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[47] W. Guo, W. Li, W. Gong, and J. Cui, “Extended feature pyramid network with adaptive scale training strategy and anchors for object detection in aerial images,” *Remote Sens.*, vol. 12, no. 5, 2020, doi: 10.3390/rs12050784.

[48] V. Taimouri, A. Gholipour, C. Velasco-Annis, J. A. Estroff, and S. K. Warfield, “A template-to-slice block matching approach for automatic localization of brain in fetal MRI,” *Proc. - Int. Symp. Biomed. Imaging*, vol. 2015-July, pp. 144–147, Jul. 2015, doi: 10.1109/ISBI.2015.7163836.

[49] A. Alansary *et al.*, “Automatic brain localization in fetal MRI using superpixel graphs,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9487, pp. 13–22, doi: 10.1007/978-3-319-27929-9_2.

[50] H. Xiong, J. Wu, Q. Liu, and Y. Cai, “Research on abnormal object detection in specific region based on Mask R-CNN,” *Int. J. Adv. Robot. Syst.*, vol. 17, no. 3, May 2020, doi: 10.1177/1729881420925287.

[51] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified,

real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2016, vol. 2016-December, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[52] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.

[53] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," 2001.

[54] R. Rothe, M. Guillaumin, and L. van Gool, "Non-maximum suppression for object detection by passing messages between windows," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9003, pp. 290–306, doi: 10.1007/978-3-319-16865-4_19.

[55] H. Zhu, X. Chen, W. Dai, K. Fu, Q. Ye, and J. Jiao, *ORIENTATION ROBUST OBJECT DETECTION IN AERIAL IMAGES USING DEEP CONVOLUTIONAL NEURAL NETWORK*. .

[56] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[57] C. Papageorgiou and T. Poggio, "Trainable system for object detection," *Int. J. Comput. Vis.*, vol. 38, no. 1, pp. 15–33, Jun. 2000, doi: 10.1023/A:1008162616689.

[58] R. Dong, D. Xu, J. Zhao, L. Jiao, and J. An, "Sig-NMS-Based Faster R-CNN Combining Transfer Learning for Small Target Detection in VHR Optical Remote Sensing Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, 2019, doi: 10.1109/TGRS.2019.2921396.

[59] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," Dec. 2016, Accessed: Jun. 30, 2021. [Online]. Available: https://arxiv.org/abs/1612.08242v1.

[60] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, Accessed: Jun. 30, 2021. [Online]. Available: http://arxiv.org/abs/1804.02767.

[61] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," Apr. 2020, Accessed: Jun. 30, 2021. [Online].

Available: http://arxiv.org/abs/2004.10934.

[62] K. E. A van de Sande, J. R. R Uijlings, T. Gevers, and A. W. M Smeulders, “Segmentation as Selective Search for Object Recognition.”

[63] M. F. Haque, H. Y. Lim, and D. S. Kang, “Object detection based on VGG with ResNet network,” *ICEIC 2019 - Int. Conf. Electron. Information, Commun.*, May 2019, doi: 10.23919/ELINFOCOM.2019.8706476.

[64] J. Sun, X. Cai, F. Sun, and J. Zhang, “Scene image classification method based on Alex-Net model,” *2016 3rd Int. Conf. Inf. Cybern. Comput. Soc. Syst. ICCSS 2016*, pp. 363–367, Oct. 2016, doi: 10.1109/ICCSS.2016.7586482.

[65] B. Kainz, K. Keraudren, V. Kyriakopoulou, M. Rutherford, J. V. Hajnal, and D. Rueckert, “Fast fully automatic brain detection in fetal MRI using dense rotation invariant image descriptors,” *2014 IEEE 11th Int. Symp. Biomed. Imaging, ISBI 2014*, pp. 1230–1233, 2014, doi: 10.1109/isbi.2014.6868098.

[66] M. Afif, Y. Said, and M. Atri, “Efficient implementation of integrall image algorithm on NVIDIA CUDA,” in *2018 International Conference on Advanced Systems and Electric Technologies, IC_ASET 2018*, Jun. 2018, pp. 1–5, doi: 10.1109/ASET.2018.8379824.

[67] K. Somasundaram and P. A. Kalaividya, “Brain portion segmentation from Magnetic Resonance Images(MRI) of human head scan using Richardson Lucy deconvolution & intensity thresholding,” Feb. 2017, doi: 10.1109/ICSEC.2016.7859865.

[68] D. Cao, Z. Chen, and L. Gao, “An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks,” *Human-centric Comput. Inf. Sci.*, vol. 10, no. 1, pp. 1–22, Dec. 2020, doi: 10.1186/s13673-020-00219-9.

[69] M. Faramarzi, “Road Damage Detection and Classification Using Deep Neural Networks (YOLOv4) with Smartphone Images,” *SSRN Electron. J.*, Jul. 2020, doi: 10.2139/ssrn.3627382.

[70] L. A. Clark, B. Cuthbert, R. Lewis-Fernández, W. E. Narrow, and G. M. Reed, “Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health’s Research Domain Criteria (RDoC),” *Psychol.*

Sci. Public Interes., vol. 18, no. 2, pp. 72–145, Sep. 2017, doi: 10.1177/1529100617727266.

[71] T. Xia, A. Kumar, D. Feng, and J. Kim, "Patch-level Tumor Classification in Digital Histopathology Images with Domain Adapted Deep Learning," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Oct. 2018, vol. 2018-July, pp. 644–647, doi: 10.1109/EMBC.2018.8512353.

[72] A. Kumar, Z. J. Zhang, and H. Lyu, "Object detection in real time based on improved single shot multi-box detector algorithm," *Eurasip J. Wirel. Commun. Netw.*, vol. 2020, no. 1, pp. 1–18, Dec. 2020, doi: 10.1186/s13638-020-01826-x.

[73] K. Somasundaram, S. P. Gayathri, R. S. Shankar, and R. Rajeswaran, "Fetal head localization & fetal brain segmentation from MRI using the center of gravity," *20th Int. Comput. Sci. Eng. Conf. Smart Ubiquitos Comput. Knowledge, ICSEC 2016*, pp. 0–5, 2017, doi: 10.1109/ICSEC.2016.7859866.

[74] H. Elmarakeby, D. Liu, S. H. Aldubayan, and E. M. Van Allen, "Abstract 1637: Biologically informed deep neural network for genomic discovery and clinical classification in prostate cancer," in *Cancer Research*, Jul. 2019, vol. 79, no. 13 Supplement, pp. 1637–1637, doi: 10.1158/1538-7445.am2019-1637.

[75] S. Nie, M. Zheng, and Q. Ji, "The Deep Regression Bayesian Network and Its Applications: Probabilistic Deep Learning for Computer Vision," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 101–111, Jan. 2018, doi: 10.1109/MSP.2017.2763440.

[76] G. Wang *et al.*, "Interactive Medical Image Segmentation Using Deep Learning with Image-Specific Fine Tuning," *IEEE Trans. Med. Imaging*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018, doi: 10.1109/TMI.2018.2791721.

[77] B. H. Menze *et al.*, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015, doi: 10.1109/TMI.2014.2377694.

[78] T. Fukagai *et al.*, "Speed-up of object detection neural network with GPU," in *Proceedings - International Conference on Image Processing, ICIP*, Aug. 2018, pp.

301–305, doi: 10.1109/ICIP.2018.8451814.

[79] M. Rajchl *et al.*, “DeepCut : Object Segmentation from Bounding Box Annotations using Convolutional Neural Networks,” vol. 0062, no. c, pp. 1–10, 2016, doi: 10.1109/TMI.2016.2621185.

[80] H. M. Ünver and E. Ayan, “Skin lesion segmentation in dermoscopic images with combination of yolo and grabcut algorithm,” *Diagnostics*, vol. 9, no. 3, 2019, doi: 10.3390/diagnostics9030072.

[81] M. Pishghadam, “A new approach to automatic fetal brain extraction from MRI using a variational level set method,” pp. 1–9, 2019, doi: 10.1002/mp.13766.

[82] R. Shishegar *et al.*, “AUTOMATIC SEGMENTATION OF FETAL BRAIN USING DIFFUSION-WEIGHTED IMAGING CUES NICTA Victoria Research Laboratory , Australia Signal and Image Processing Institute , Department of Electrical Engineering Systems , School of Health and Biomedical Sciences , RM,” pp. 804–807, 2017.

[83] T. Doshi, J. Soraghan, L. Petropoulakis, D. Grose, and K. MacKenzie, “Semi-automatic segmentation of tongue tumors from magnetic resonance imaging,” in *International Conference on Systems, Signals, and Image Processing*, 2013, pp. 143–146, doi: 10.1109/IWSSIP.2013.6623474.

[84] M. Y. Bhanumurthy and K. Anne, “An automated detection and segmentation of tumor in brain MRI using artificial intelligence,” Sep. 2015, doi: 10.1109/ICCIC.2014.7238374.

[85] S. Aslani *et al.*, “Multi-branch convolutional neural network for multiple sclerosis lesion segmentation,” *Neuroimage*, vol. 196, pp. 1–15, 2019, doi: 10.1016/j.neuroimage.2019.03.068.

[86] J. Dolz, I. Ben Ayed, J. Yuan, and C. Desrosiers, “Isointense infant brain segmentation with a hyper-dense connected convolutional neural network,” in *Proceedings - International Symposium on Biomedical Imaging*, May 2018, vol. 2018-April, pp. 616–620, doi: 10.1109/ISBI.2018.8363651.

[87] K. Keraudren *et al.*, “Automated fetal brain segmentation from 2D MRI slices for motion correction,” *Neuroimage*, vol. 101, pp. 633–643, 2014, doi:

10.1016/j.neuroimage.2014.07.023.

[88] N. Khalili *et al.*, “Automatic extraction of the intracranial volume in fetal and neonatal MR scans using convolutional neural networks,” *NeuroImage. Clin.*, vol. 24, p. 102061, 2019, doi: 10.1016/j.nicl.2019.102061.

[89] G. Wang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, and S. Zhang, “Uncertainty-Guided Efficient Interactive Refinement of Fetal Brain Segmentation from Stacks of MRI Slices,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12264 LNCS, pp. 279–288, 2020, doi: 10.1007/978-3-030-59719-1_28.

[90] M. Ebner *et al.*, “An automated localization, segmentation and reconstruction framework for fetal brain MRI,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11070 LNCS, pp. 313–320, doi: 10.1007/978-3-030-00928-1_36.

[91] A. Serag, J. P. Boardman, R. M. Reynolds, F. C. Denison, G. Macnaught, and S. I. Semple, “Automatic fetal brain localization in 3T MR images using Histograms of Oriented 3D Gradients,” *2016 24th Signal Process. Commun. Appl. Conf. SIU 2016 - Proc.*, pp. 2273–2276, 2016, doi: 10.1109/SIU.2016.7496229.

[92] L. S. V. Thomas and J. Gehrig, “Multi-template matching: A versatile tool for object-localization in microscopy images,” *BMC Bioinformatics*, vol. 21, no. 1, 2020, doi: 10.1186/s12859-020-3363-7.

[93] H. Wang, X. Tong, and F. Lu, “Deep learning based target detection algorithm for motion capture applications,” in *Journal of Physics: Conference Series*, 2020, vol. 1682, no. 1, doi: 10.1088/1742-6596/1682/1/012032.

[94] S. Albahli, N. Nida, A. Irtaza, M. H. Yousaf, and M. T. Mahmood, “Melanoma Lesion Detection and Segmentation Using YOLOv4-DarkNet and Active Contour,” *IEEE Access*, vol. 8, pp. 198403–198414, 2020, doi: 10.1109/ACCESS.2020.3035345.

[95] C. Castaneda *et al.*, “Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine,” *J Clin Bioinform*, vol. 5, p. 1, 2015.

[96] Y. Sun, B. Xue, M. Zhang, and G. G. Yen, “Evolving Deep Convolutional Neural

Networks for Image Classification," *IEEE Trans. Evol. Comput.*, vol. 24, no. 2, 2020, doi: 10.1109/TEVC.2019.2916183.

[97] O. Strachna and O. Asan, "Reengineering Clinical Decision Support Systems for Artificial Intelligence," Nov. 2020, doi: 10.1109/ICHI48887.2020.9374370.

[98] A. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities," *Medical Image Analysis*, vol. 33. Elsevier B.V., pp. 170–175, Oct. 01, 2016, doi: 10.1016/j.media.2016.06.037.

[99] J. Civit-Masot, F. Luna-Perejon, S. Vicente-Diaz, J. M. Rodriguez Corral, and A. Civit, "TPU cloud-based generalized U-Net for eye fundus image segmentation," *IEEE Access*, vol. 7, pp. 142379–142387, 2019, doi: 10.1109/ACCESS.2019.2944692.

[100] S. Van Der Walt, S. C. Colbert, and G. Varoquaux, "The NumPy array: A structure for efficient numerical computation," *Comput. Sci. Eng.*, vol. 13, no. 2, pp. 22–30, Mar. 2011, doi: 10.1109/MCSE.2011.37.

[101] J. Du, "Understanding of Object Detection Based on CNN Family and YOLO," in *Journal of Physics: Conference Series*, Apr. 2018, vol. 1004, no. 1, p. 12029, doi: 10.1088/1742-6596/1004/1/012029.

[102] P. Hamet and J. Tremblay, "Artificial intelligence in medicine," *Metabolism.*, vol. 69, pp. S36–S40, Apr. 2017, doi: 10.1016/j.metabol.2017.01.011.

[103] H. H. Su, H. W. Pan, C. P. Lu, J. J. Chuang, and T. Yang, "Automatic detection method for cancer cell nucleus image based on deep-learning analysis and color layer signature analysis algorithm," *Sensors (Switzerland)*, vol. 20, no. 16, pp. 1–19, Aug. 2020, doi: 10.3390/s20164409.

[104] Z. Ahmad, S. Rahim, M. Zubair, and J. Abdul-Ghafar, "Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review," *Diagnostic Pathology*, vol. 16, no. 1. BioMed Central Ltd, Dec. 01, 2021, doi: 10.1186/s13000-021-01085-4.

[105] S. Sunarti, F. Fadzlul Rahman, M. Naufal, M. Risky, K. Febriyanto, and R. Masnina,

"Artificial intelligence in healthcare: opportunities and risk for future," *Gac. Sanit.*, vol. 35, 2021, doi: 10.1016/j.gaceta.2020.12.019.

[106] D. Zhang, Z. Liu, and X. Shi, "Transfer learning on EfficientNet for remote sensing image classification," in *Proceedings - 2020 5th International Conference on Mechanical, Control and Computer Engineering, ICMCCE 2020*, Dec. 2020, pp. 2255–2258, doi: 10.1109/ICMCCE51767.2020.00489.

[107] C. Y. Wang, H. Y. Mark Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2020, vol. 2020-June, pp. 1571–1580, doi: 10.1109/CVPRW50498.2020.00203.

[108] B. Zhang, "Computer vision vs. human vision," Oct. 2010, pp. 3–3, doi: 10.1109/coginf.2010.5599750.

[109] S. Hayat, S. Kun, Z. Tengtao, Y. Yu, T. Tu, and Y. Du, "A Deep Learning Framework Using Convolutional Neural Network for Multi-Class Object Recognition," in *2018 3rd IEEE International Conference on Image, Vision and Computing, ICIVC 2018*, Oct. 2018, pp. 194–198, doi: 10.1109/ICIVC.2018.8492777.

[110] B. Li, W. Jiang, J. Gu, K. Liu, and Y. Wu, "Research on Convolutional Neural Network in the Field of Object Detection," in *Proceedings of 2020 IEEE International Conference on Power, Intelligent Computing and Systems, ICPICS 2020*, Jul. 2020, pp. 820–827, doi: 10.1109/ICPICS50287.2020.9202194.

[111] D. Cao, Z. Chen, and L. Gao, "An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks," doi: 10.1186/s13673-020-00219-9.

[112] H. Zhai, J. Cheng, and M. Wang, "Rethink the IoU-based loss functions for bounding box regression," 2020, pp. 1522–1528, doi: 10.1109/ITAIC49862.2020.9339070.

[113] S. Rutherford *et al.*, "Automated Brain Masking of Fetal Functional MRI," *bioRxiv*, p. 525386, Jul. 2019, doi: 10.1101/525386.

[114] T. Carneiro, R. V. M. Da Nobrega, T. Nepomuceno, G. Bin Bian, V. H. C. De

Albuquerque, and P. P. R. Filho, “Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications,” *IEEE Access*, vol. 6, pp. 61677–61685, 2018, doi: 10.1109/ACCESS.2018.2874767.

[115] J. Li *et al.*, “Automatic fetal brain extraction from 2D in utero fetal MRI slices using deep neural network,” *Neurocomputing*, vol. 378, pp. 335–349, Feb. 2020, doi: 10.1016/j.neucom.2019.10.032.

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

[117] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11045 LNCS, pp. 3–11, doi: 10.1007/978-3-030-00889-5_1.

[118] A. Rampun, D. Jarvis, P. Griffiths, and P. Armitage, “Automated 2D Fetal Brain Segmentation of MR Images Using a Deep U-Net,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Nov. 2020, vol. 12047 LNCS, pp. 373–386, doi: 10.1007/978-3-030-41299-9_29.

[119] D. Wiljer and Z. Hakim, “Developing an Artificial Intelligence–Enabled Health Care Practice: Rewiring Health Care Professions for Better Care,” *J. Med. Imaging Radiat. Sci.*, vol. 50, no. 4, pp. S8–S14, Dec. 2019, doi: 10.1016/j.jmir.2019.09.010.

[120] G. Currie, K. E. Hawk, E. Rohren, A. Vial, and R. Klein, “Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging,” *Journal of Medical Imaging and Radiation Sciences*, vol. 50, no. 4. Elsevier Inc., pp. 477–487, Dec. 01, 2019, doi: 10.1016/j.jmir.2019.09.005.

[121] P. Bhide, P. Gund, and A. Kar, “Prevalence of congenital anomalies in an Indian maternal cohort: Healthcare, prevention, and surveillance implications,” *PLoS One*, vol. 11, no. 11, p. e0166408, Nov. 2016, doi: 10.1371/journal.pone.0166408.