Evidential Dempster Shafer-based CNN Architecture for Fetal Planes Detection from 2D Ultrasound Images Leveraging Fuzzy-Contrast Enhancement and Explainable AI

by

Rafeed Rahman
21166033

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

Department of Computer Science and Engineering
Brac University
January 2023

Declaration

It is hereby declared that

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

Student's Full Name & Signature:

29/1/23

Rafeed Rahman
21166033

Approval

The thesis/project titled "Evidential Dempster Shafer-based CNN Architecture for Fetal Planes Detection from 2D Ultrasound Images Leveraging Fuzzy-Contrast Enhancement and Explainable AI" submitted by

1. Rafeed Rahman (21166033)

The prerequisite for the degree of M.Sc. in Computer Science as of Fall 2022 has been acknowledged as satisfactory in partial fulfillment as of December 1, 2022.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

29/01/2023

Examiner:
(External)

Dr. Ashis Talukder, Ph.D.
Associate Professor
Department of Management Information Systems (MIS)
University of Dhaka

29.01.2023

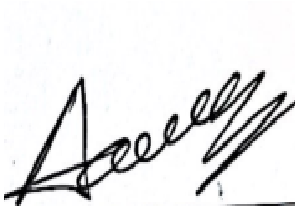
Examiner:
(Internal)

Dr. Muhammad Iqbal Hossain, PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Examiner:
(Internal)

Dr. Md. Ashraful Alam
Assistant Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)



Dr. Amitabha Chakrabarty
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

In order to systematically find specific signs of development of the fetus that are present in ultrasound images, automated picture categorization in large-scale retrospective assessments can be useful for ultrasound image processing and interpretation utilizing machine learning. The advancement of automated diagnosis while preserving accuracy has greatly benefited from the use of Deep Learning architectures in medical picture analysis. The objective of this work is to precisely identify fetus planes from ultrasound images. A dataset of 12400 images is used to train the models. This study showed the effect of enhanced ultrasound image quality using FUZZY LOGIC in fetal plane detection. A proposed Evidential Dempster-Shafer Based CNN Architecture is used and the outcome is compared with PReLU-Net, SqueezeNET, and Swin Transformer. Dempster-Shafer ensures that all the pieces of evidence are properly analyzed and the classification output can be a range containing belief and plausibility rather than a single probability. Thus it handles the uncertainty. The results of each classifier were noteworthy with the proposed Evidential Classifier's accuracy reaching 83%. The result is evaluated in terms of training and testing accuracies. Moreover, the Lime Algorithm and the Gram Cam are used to examine the classifier's decision-making process to incorporate explainability.

Keywords: Evidential Dempster-Shafer CNN; Swim Transformer; Fuzzy Logic Contrast Enhancement

Acknowledgement

My deepest gratitude goes out to Professor, Dr. Md. Golam Rabiul Alam who has been my supervisor throughout this research effort.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Statement

During pregnancy, women undergo fetal ultrasound testing which provides visual information about the unborn child (uterus). It is a secure method of determining a baby's health. The heart, head, and spine of the unborn child are examined along with other body parts during a fetal ultrasound [39]. Ultrasound uses a device known as the transducer for sending out sound waves and receiving them as well. Sound waves enter through the skin, muscle, bone, and fluids at different speeds as the transducer is moved around the abdomen. The baby acts as a sound reflector, reflecting sound waves back to the transducer. An electronic picture that may be viewed on a computer screen is created from the sound waves by the transducer [39].

During the first trimester, an ultrasound is performed to determine the existence and precise location of the pregnancy, get the number of fetuses and assess the duration that the mother going to carry the child. Additionally, ultrasonography can be used to check for uterine or cervix abnormalities during the first trimester of pregnancy [25].

A standard ultrasound is carried out to evaluate the pregnancy's numerous characteristics in the second and third trimesters, including the fetal anatomy. On average, this test is done between weeks 18 and 20. However, the time of this ultrasound may vary depending on several factors which could make it more difficult to see the fetus [25]. This accuracy is very important in the diagnosis of fetal planes. Manual diagnosis with the naked eye takes considerable time and is prone to human error.

In particular, with the introduction of Deep Learning (DL) [6] and its amazing advancements in image recognition tasks using Convolution Neural Networks, artificial intelligence has had impressive growth over the past ten years [37]. In recent years, CNN has demonstrated its impact in a variety of medical imaging applications. We have seen the use of Deep Learning in Pneumonia diagnosis [21] [35], in the diagnosis of Brain Tumor [16] [17] [30] [14], in the detection of skin cancer [34] [23] [11] etc.

In this study, the abdomen, brain, femur, thorax, and mother's cervix (the entrance in the lower part of the uterus (the womb) that connects to the top of the

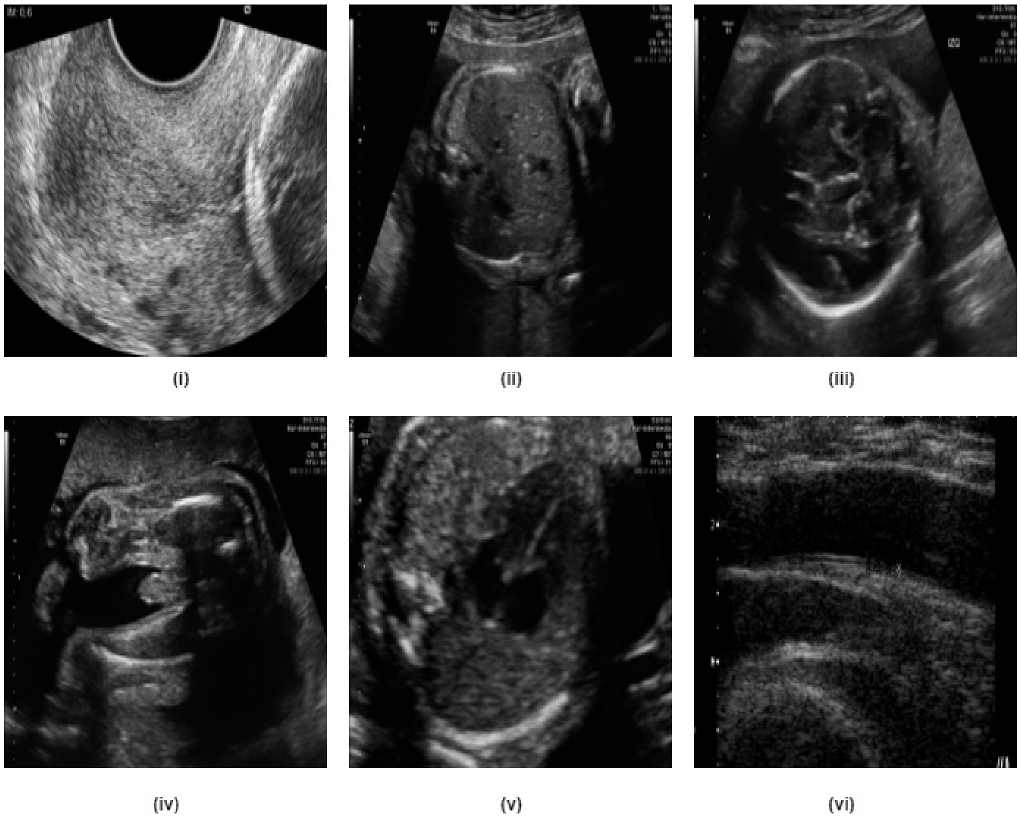


Figure 1.1: Images from Dataset - (i) Maternal Cervix (ii)Fetal Abdomen (iii) Fetal Brain (iv)Fetal Femur (v) Fetal Thorax (vi) Others

vagina) are among the anatomical planes of the fetus on which this research has concentrated (birth canal). Figure 1.1 shows example images from each class of the dataset that this study is focused upon.

1.2 Research Contributions

In this study, the application of Evidential Dempster Shafer-based classifier is seen to detect fetal planes from 2D Ultrasound images. The process begins with image contrast enhancement using Fuzzy Logic based contrast enhancement [12] [4]. The enhanced contrast images is then given as input to the Evidential CNN+DS Layer and the output is analysed in the results section. For contrast, we have also enhanced the image using traditional Histogram Equalization and the final accuracy scores are recorded. Along with this pre-processing, we have used Swin Transformer, SqueezeNET and PreluNet to compare the final accuracies. Finally, the output from each classifiers are also explained using Explainable AI. The following are the research's main contributions:

1. In this study, an Evidentiary classifier called the Dempster Shafer Layer is used in conjunction with a customized designed CNN architecture. In machine learning, the process of classifying new elements by applying a learning set of labeled cases is known as classification. Precision classification, when a

sample is assigned to only one of the five available classes, is a typical classification problem. Unfortunately, a challenging assignment typically happens when there is a lot of ambiguity [38]. Set-valued classification is a potential remedy for this problem. Thus, it can be shown how employing Evidential CNN can help in cases where fetal planes are being detected in ultrasound pictures which are rarely used till now.

2. Along with that, the contrast between utilizing Histogram Equalization and Fuzzy Logic-based contrast enhancement for pre-processing MRI images is also displayed in the results section. The impact of using Fuzzy Logic Contrast Enhancement in ultrasound images as pre-processing is assessed with metrics such as F1 scores, recall, and precision of each classifier.

3. In addition to that, according to the study of related works, Swin Transformer is less frequently used in the ultrasound fetal planes analysis. So an effort is seen in this research to study the impact of the Swin transformer in the detection of planes - Fetal Thorax, Brain, Abdomen, Femur, and Maternal Cervix.

4. The results section shows the performance of the PreLUNet, SqueezeNET, and the Swin Transformer architectures as well as the CNN+DS Layer and other architectures. Due to the Swin Transformer's recent use in medical image processing, it has the potential to significantly contribute to the automation of fetus diagnosis. Understanding the output with the Grad Cam and Lime Algorithm is also demonstrated.

1.3 Paper Outline

The study is organized as follows: Section II includes a literature overview of previous studies in this field. Along with that, Section II contains background studies for each deep learning architecture used in this study and information regarding Dempster Shafer's Theory. Detailed information on picture contrast enhancement and customized CNN + Ds Layer-based Evidential architecture may be found in Section III. Analysis of the models' outcomes is included in the results section, which is the last section. The results section had shown a complete analysis of the outcome of each model in terms of training, validation, and testing accuracy. Additionally, the prediction of each classifier has been scrutinized using in terms of Explainability with Lime and GradCam.

Chapter 2

Related Work

2.1 Literature Review

In the first part of this section, the works related to this research are studied. The section initially contains the summaries of the related works concentrating on their contributions, novelty, and achievements in terms of outcomes and limitations if any. This next subsection of the section provides healthy supporting material to understand the Dempster Shafer Theory, SqueezeNET Architecture, and Swin Transformer. The subsections are details of each classifier.

Using data from both the whole image and the clipped fetal structures, Sridar et al. developed a system to automatically identify 14 distinct features in 2-D prenatal ultrasound pictures [19]. Using the full ultrasound fetal imaging as well as the discriminant areas of the fetal features included in the entire image, their method enhances two feature extractors that have already been trained. Their approach stands out since it incorporates categorization findings from both local and worldwide data and doesn't rely on any prior knowledge. Their investigation produced mean accuracy, precision, and recall scores of 97.05%, 76.47%, and 75.41%, respectively, for a collection of 4074 2-D ultrasound images [19].

In this study, Hong et al. recommended a method of segmentation via the use of a Fully Connected Convolution Neural Network in the case of Cortical planes [27]. They created a novel loss function which they also as a hybrid to improve accuracy using 3D information as well as a test-time augmentation method for multi-view aggregation to improve segmentation accuracy. 52 fetal brain MR images were used in ten evaluations of their approach.

In the following research article [24], a strong deep learning segmentation approach has been designed in order to lessen the laborious human segmentation refining process. A fully convolutional neural network architecture, deep supervision, residual connections, and special deep attentive modules with mixed kernel convolutions are used to demonstrate their methodology. They evaluated their strategy statistically using a range of performance criteria and professional opinions. They quantitatively examined their plan based on a range of performance criteria and expert opinions. The outcomes showed that both a cutting-edge multi-atlas segmentation method and a number of state-of-the-art deep segmentation models are outperformed by their method. Fetuses between the gestational ages of 16 and 39 weeks had their fetal brain MRI pictures reconstructed.

A deep convolutional neural network for multiple tasks is suggested in this study [18] for automatically segmenting and estimating HC ellipses. This network minimizes a compound cost function made up of the MSE of the ellipse parameters and the segmentation dice score. Using an ultrasound dataset of a fetus at different phases of pregnancy, experimental results show that the segmentation results and the produced HC closely match the radiologist annotations. The Link-Net topology, which was initially created for semantic segmentation, was used to create a multi-task deep network in this study. According to their research, multi-task learning performs better than a single-task network and generates more accurate segmentation results.

Radiological technology and ultrasound expertise were needed for the segmentation of anatomical features in ultrasound pictures [28]. The manual segmentation process is time-consuming and frequently depends on the clinician's clinical expertise. They presented an automated method for segmenting and measuring ultrasonic images. For the purpose of fetal biometric information extraction from two-dimensional ultrasound images, they presented a scale attention feature pyramid network (SAFNet). For each level, the feature pyramid is generated by controlling the scale attention module. An auxiliary layer is used to learn how to define object boundaries under close supervision. Additionally, they offered a two-stage framework known as the automatic categorization measurement system (ACMS), which initially identifies the picture type with three labels: head, abdomen, and femur.

The goal of this research [32] is to determine whether deep learning algorithms can be utilized to identify abnormal or unhealthy fetal brain sonographic pictures in conventional axial planes [32]. Images from 10251 good pregnancies and 2529 problematic pregnancies were used in the study from a huge hospital database. Any unusual occurrences were supported by autopsy, follow-up, or newborn ultrasound results. Segmentation accuracy, recall, and DICE scores were 97.9%, 90.9%, and 94.1%, respectively. 96.3% of the categories were correctly categorized overall.

In order to identify structural anomalies and substructures of the heart in fetal ultrasound recordings, the scholars in [36] suggested an architecture called Supervised Object detection with Normal data Only (SONO), based on a convolutional neural network (CNN). To show the likelihood of detection, we created a timeline that resembled a barcode, and we assigned each video an irregularity score.

2.2 Background Studies

Given two functions, f(x) and g(x), the mathematical process known as convolution generates a third function that represents how the form of one is altered by the other. Contrary to the cross-correlation operator, the f(x) or g(x) function for real-valued functions in the convolution is reflected about the y-axis [33].

A convolutional layer is CNN's primary building block. It requires training to understand the parameters of the many filtration systems (or kernels) it comprises. Often, the filters are less in size than the original image. By combining with the image, each filter creates an activation map. For convolution, the filter is stretched across the image's height and width, and on each cutover, the quantitative comparison between each filter element and the feed was taken as shown in Fig. 1. Convolution layers are frequently used in medical image processing applications [9] [13].

The field of computer vision has seen the substantial application of deep learning. Using single-shot detectors, squeeze-and-extinction deep convolution neural networks, augmentations, and multi-task learning, the authors of the study Early detection of Pneumonia [26] present a computational technique for recognizing pneumonia areas. In the next subsection, we have reviewed Dempster-Shafer Theory and Evidential Classifier,

2.2.1 Dempster-Shafer Theory And Evidential Classifier

The Dempster-Shafer Theory (DST) is an empirical mathematical theory [2]. [Shafer, 1976], an update of [Dempster, 1967], contains ground-breaking research on the topic. The Dempster-Shafer theory can be thought of as a discrete-space expansion of probability theory because it distributes probabilities to sets rather than to mutually exclusive singletons. DST evidence can be connected to several potential events, such as collections of events. The Dempster-Shafer model is replaced with the conventional probabilistic formulation whenever there is sufficient data to allow assigning a probability to each specific occurrence.

The task of predicting the class of a new sample using a learning set of labeled cases is referred to as classification in machine learning. Precision classification, where a sample is placed into just one of the five potential classes, is the most typical classification issue [38]. Unfortunately, when there is a lot of uncertainty, a difficult assignment frequently results in misclassification.

When there is too much uncertainty to establish a precise categorization, set-valued classification, which is characterized as a potential solution to the issue, is the assignment of a new observation into a non-empty subset of classes. Set-valued classification enables classifier caution, a better representation of categorization uncertainty, and ultimately lower error rates.

A mass function confirm $(\emptyset) = 0$ and

$$\sum_{A \subseteq \Omega} m(A) = 1 \tag{2.1}$$

for a finite set $\Omega = \{\omega_1, \ldots, \omega_k\}$ Given a specific piece of evidence, for any A, m(A) might be viewed as the conviction that one is ready to commit to A.

$$bel(A) = \sum_{B \subseteq A} m(B) \tag{2.2}$$

$$pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \tag{2.3}$$

The quantity pl(A) represents the amount of belief that might possibly be expressed in A, while the bel(A) is taken as a general evaluation of one's faith that the hypothesis is indeed true [20] [1] [3][27,28,29].

$$m1 \oplus m2(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{\sum_{B \cap C \neq \emptyset} m_1(B) m_2(C)} \tag{2.4}$$

In the publication [38], Tong et al. developed a unique classifier for set-valued classification by combining a CNN with the Dempster-Shafer (DS) theory. Extraction of high-dimensional data from the input data is the initial stage in the convolutional and pooling layers of the evidentiary deep-learning classifier. Dempster's rule is then used to convert the features into mass functions, which are then combined in a DS layer.

Given a specific piece of evidence, for any A, m(A) might be viewed as the conviction that one is ready to commit to A. The next step is set-valued classification using mass functions in a presumptive utility layer. It provides an end-to-end learning technique for collaborating on network parameter modifications. A method for choosing incomplete multi-class actions is also offered [38].

2.2.2 SqueezeNET

A CNN architecture called SqueezeNET employs 50 times fewer parameters than AlexNet while still being equally accurate. The following advantages accrue to a CNN model with fewer parameters: Improved effectiveness of distributed training, fewer costs associated with distributing new models to customers, and Embedded and FPGA deployment that is feasible [7].

Certain techniques are used by the Squeeze Net model to reduce the majority of parameters. Which are: putting 1x1 filters in place of the 3x3 filters and Three-by-three filters with fewer input channels subsequent downsampling at the network [7]. They reduce the input channel count to 3x3 filters [43]. It is best to downsample late in the network so that convolution layers have large activation maps in order to maximize accuracy on a constrained parameter budget [43]. Convolution neural

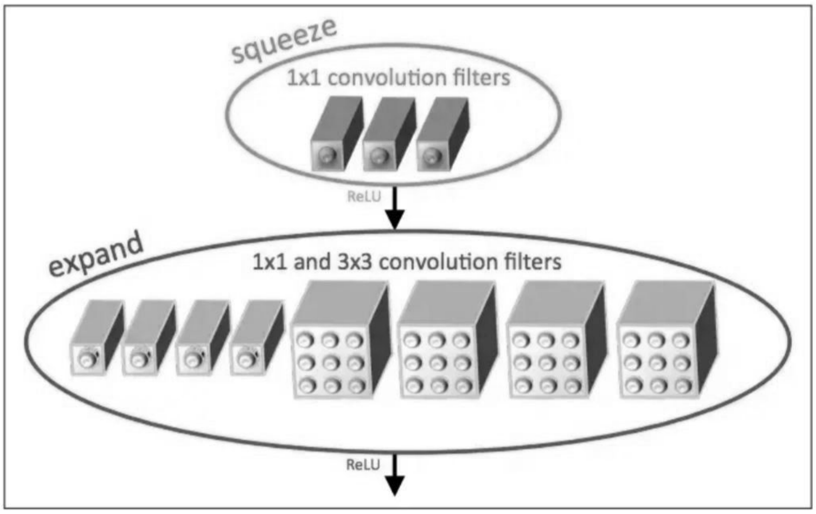


Figure 2.1: The components of the Fire Module of SqueezeNET Model

networks, most notably SqueezeNET, utilize construction blocks called Fire Modules as part of their architecture as in Figure 2.1. In order to create a Fire module, a squeeze convolution layer—which contains only 1x1 filters—feeds into an expand layer, which combines 1x1 and 3x3 convolution filters [43].

In the design of SqueezeNET, eight Fire modules (fire2-9) are placed before a single convolution layer (conv1) and a ninth convolution layer is placed after that (conv10). From the beginning to the end of the network, there are consistently more filters per fire module. SqueezeNET places its pooling a little later in accordance with Strategy 3 [7] after layers conv1, fire4, fire8, and conv10.

There are various applications where scholars have shown the use of SqueezeNET architecture. In automatic vision-based systems, the model recognition (MMR) of cars is crucial. The SqueezeNET architecture is used in a study [15] to present a unique deep-learning method for MMR. Vehicle picture frontal views are first extracted and given as input for training and testing purposes into a deep network. O MMR system is more effective thanks to the use of the SqueezeNET design, a variation of the vanilla SqueezeNET, with bypass links between the Fire modules.

The suggested model achieves a 96.3% recognition rate at the rank-1 level with an efficient time slice of 108.8 ms, according to testing results using the large-scale vehicle datasets that we have gathered.

The researchers in another study demonstrated how an AI-based framework can outperform earlier [31]. The SqueezeNet light network design is optimized for the COVID-19 diagnosis using Bayesian optimization additive. Because of its improved performance over existing network designs and higher COVID-19 diagnosis accuracy, the proposed network makes use of a supplemented dataset and fine-tuned hyperparameters.

Also, the impact of SqueezeNET is seen in the field of Agriculture where the diseases of Leaves are detected. This study [10] concentrated on developing a squeeze net-based model to categorize seven different tomato plant illnesses on leaves, including healthy leaves using Keras. The Vegetable Crops Research Institute (Balitsa) in Lembang, West Java provided the tomato plant leaf images that were utilized in this investigation. This study's automated disease detection of tomato plants using images of their leaves has an average identification accuracy of 86.92

Hence it showed the application of SqueezeNET in multiple areas. In our case, we used SqueezeNET architecture for the detection of Fetal Planes in 2d Ultrasound images. The performance is compared with the outcome of the CNN+Ds Layer.

2.2.3 Swin Transformer

The common MSA module (multi-head self-attention) is changed to one based on moveable windows in Swin Transformer, leaving the remaining layers alone. A 2-layer MLP with GELU nonlinearity and a shifted window-based MSA module make up a Swin Transformer block. Prior to each MSA module, MLP, and residual connection, the LayerNorm (LN) layer and the module are applied, correspondingly [37]. The architecture of the Swin Transformer is shown in Figure 2.2.

Due to disparities between language and vision, such as the stark contrasts in visual element magnitude and the higher clarity of pixels in images compared to words in the text, Transformer struggles to switch between the two [37]. They proposed a hierarchical Transformer whose representation is computed using shifted windows to accommodate for these differences. With cross-window communication still possible, the shifted windowing strategy increases efficiency by limiting self-attention computation to non-overlapping local windows. This hierarchical architecture is capable of representing a range of scales with linear computing costs with respect to image size.

Due to these features, the Swin Transformer may be applied to a wide range of vision tasks, such as semantic segmentation, object recognition, and object detection. All MLP designs also profit from the hierarchical architecture and shifting window method.

In the following study, [40], SwinMR, a brand-new Swin transformer-based technique for quick MRI reconstruction, was introduced. An input module (IM), a feature extraction module (FEM), and an output module made up the entire network (OM). The FEM was made up of a cascade of residual Swin transformer blocks (RSTBs) and 2D convolutional layers, while the IM and OM were both 2D convolutional layers. A succession of Swin transformer layers made up the RSTB (STLs). Instead of performing the multi-head self-attention (MSA) of the original transformer in the entire picture space, the shifted windows multi-head self-attention (W-MSA/SW-MSA) of STL was done in shifted windows.

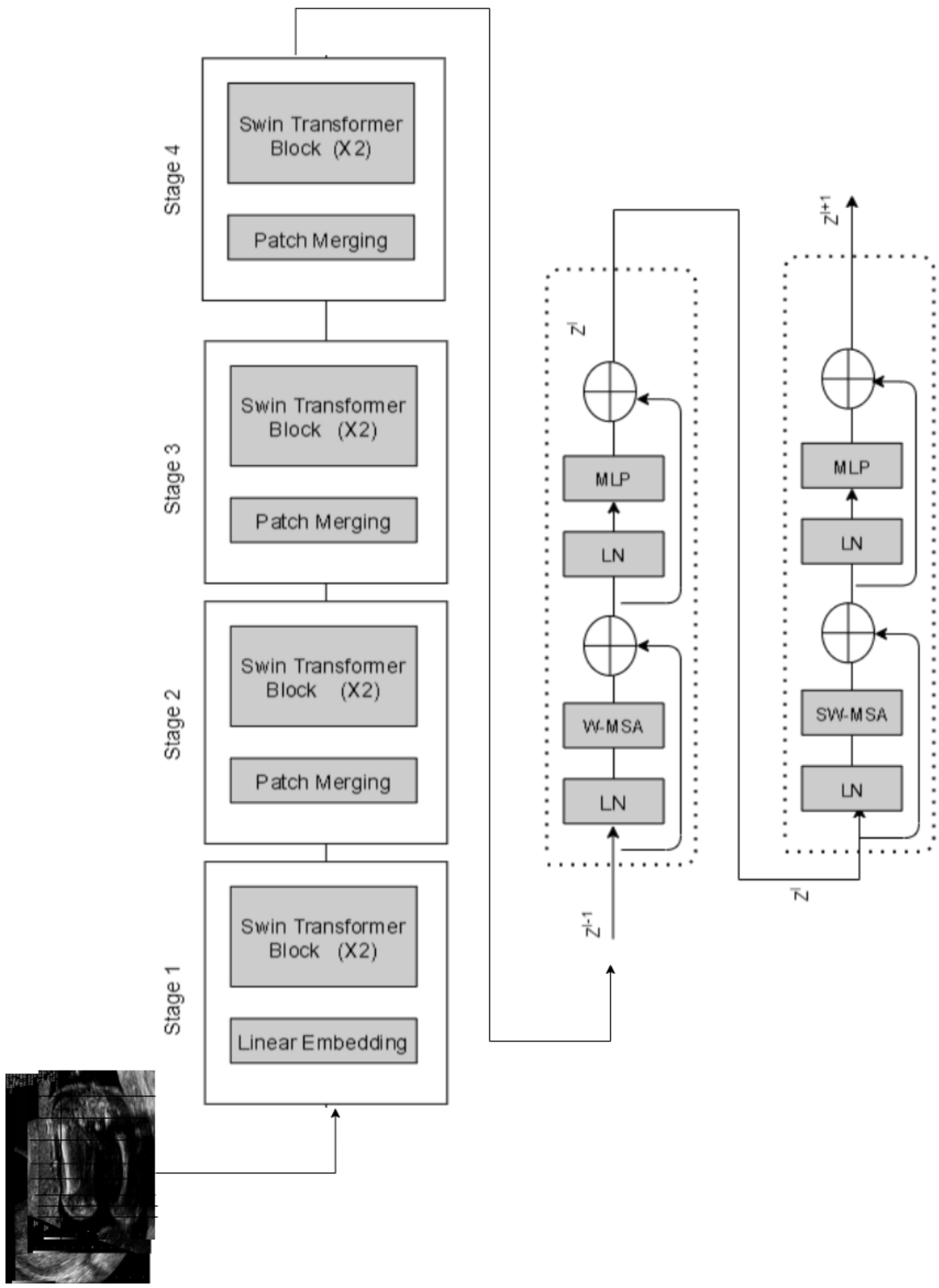


Figure 2.2: Architecture of the Swin Transformer

2.2.4 PreLUNET

PReLUNet has been introduced in the paper [5]. The authors proposed a generalized version of the conventional rectified unit called the Parametric Rectified Linear Unit (PReLU). With almost no additional computational effort and no over fitting risk, PReLU enhances model fitting. The contrast between PreLU and Relu is displayed in Figure 2.3

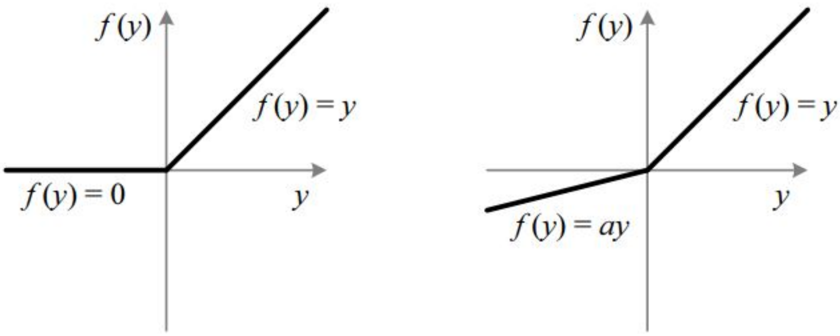


Figure 2.3: Difference between Relu and PreLU

They put out a fresh generalization of ReLU that they refer to as the Parametric Rectified Linear Unit (PReLU). This activation function increases accuracy at a very low additional computing cost while adaptive learning the rectifier parameters. Second, they investigate the challenge of training very deep corrected models. The scholars derived a theoretically valid initialization strategy by explicitly modeling the nonlinearity of rectifiers (ReLU/PReLU), which aids in the convergence of extremely deep models that were trained entirely from scratch, such as ones with 30 weight layers.

$$f(y_i) = \begin{cases} y_i & \text{if } y_i > 0 \\ y_i a_i & \text{if } y_i >= 0 \end{cases} \tag{2.5}$$

Here, ai is a coefficient that regulates the slope of the negative component, and yi is the input of the nonlinear activation function on the ith channel.

Chapter 3

Methodology

The top-level layout of the proposed Evidential Dempster Shafer-based Fetal Planes Analysis from 2D ultrasound images (EDS-FPA) system has been presented in Figure 3.1. The proposed system has 4 prime modules: Dataset collection (2 Dimensional Ultrasound images were used [22], dataset pre-processing using Fuzzy Logic Contrast Enhancement and Histogram Equalization, formulation of training and testing dataset to train the proposed Evidential CNN+Ds Layer architecture, Swin Transformer, PreLUNet, and SqueezeNET and finally visualizing the output in the results section and incorporating the Explainability with Lime and GradCam.

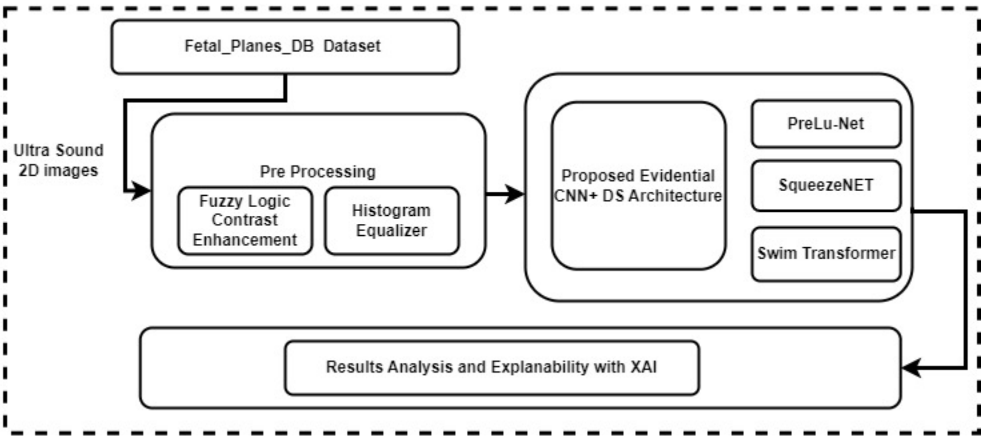


Figure 3.1: Top level layout of the proposed system

3.1 Image Contrast Enhancement using Fuzzy Logic

Fuzzy Logic Contrast Enhancement [12][4] [29] is used to improve the contrast. In the following subsection section, we will be discussing the methodology followed to increase the enhancement using fuzzy Logic.

3.1.1 Method to increase Enhancement using Fuzzy Logic

Figure 3.2 shows the overall process to increase the contrast of an image using Fuzzy Logic. Initially, there is an RGB to CIELAB transformation of the input image [29]

[12] [4], with progress on the L channel. Based on pixel intensity and M value, the

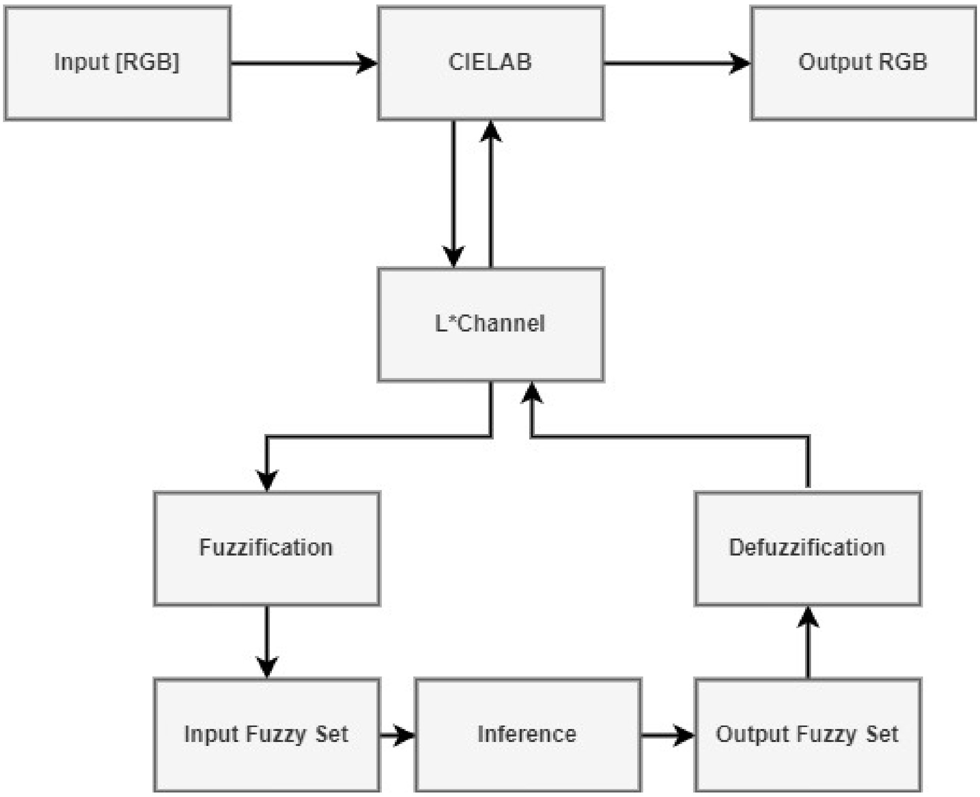


Figure 3.2: Image Contrast Enhancement process using Fuzzy Logic

fuzzification step determines the degree of each class membership for each individual pixel. Gaussian Function is used in the membership functions.

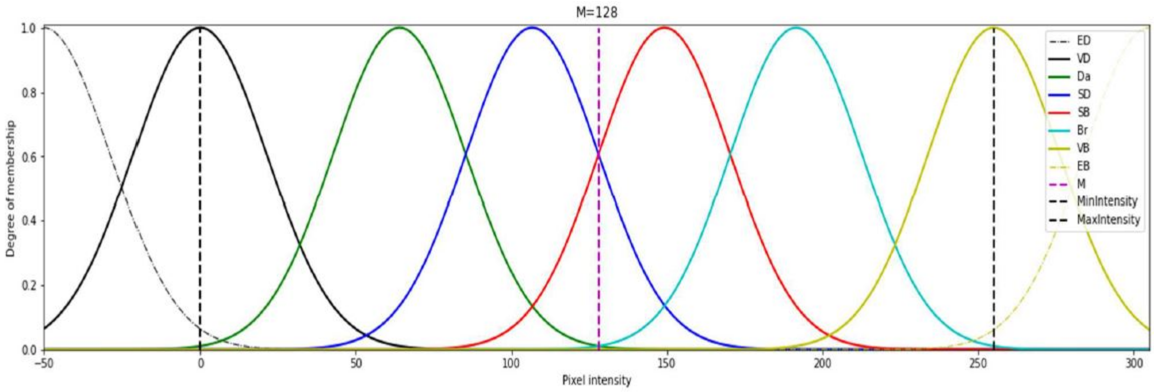


Figure 3.3: Function representing Membership of each class

The papers [12] [4] had shown the use of Fuzzy Logic to improve the contrast of images. This has a significant positive impact on the accuracy of our deep-learning algorithms. Defuzzification: Determine the centroid value of each output fuzzy set for each pixel. Image enhancement is essentially a technique that raises the image's quality and perceptibility. It is crucial in areas like remote sensing, surveillance, and medical imaging, among others [12]. The rule set is applied as shown in Figure 3.5. Using the rule set, the output fuzzy set from the input pixel intensity is

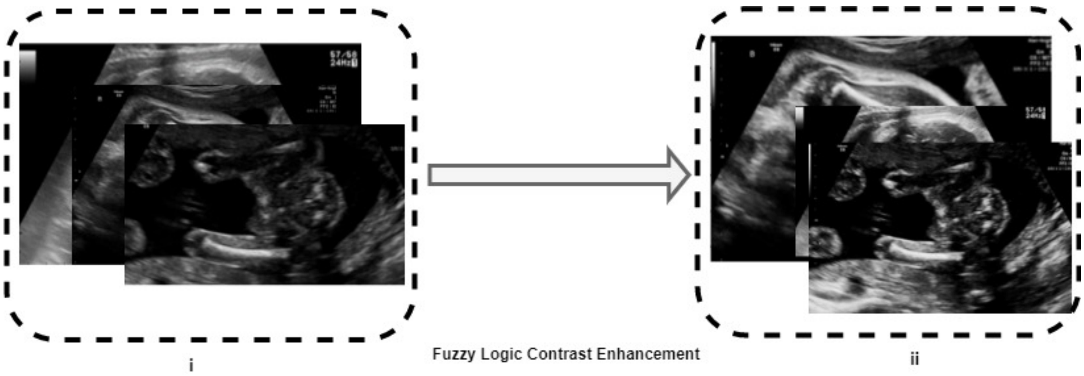


Figure 3.4: Contrast Enhancement Procedure using Fuzzy Logic. (i) represents the actual images from the dataset without any Fuzzy Contrast Enhancement and (ii) shows the image as an outcome of Contrast Enhancement

determined.Convert the resulting image from CIELAB to RGB, then merge the adjusted L channel with the original AB channels. Figure 3.4 shows the outcome of the enhanced image.

3.1.2 Histogram Equalization

Another approach to enhance the image i.e. Histogram Equalization is also used. To enhance contrast in photographs, Histogram equalization [HE] is a computer image processing method. By substantially extending the intensity range of the image, it achieves this by effectively spreading out the most common intensity values.

When an image's useful data is represented by close contrast values, this method typically results in an increase in the image's overall contrast. This enables regions with low local contrast to acquire a higher contrast [44] [42]. The classifiers

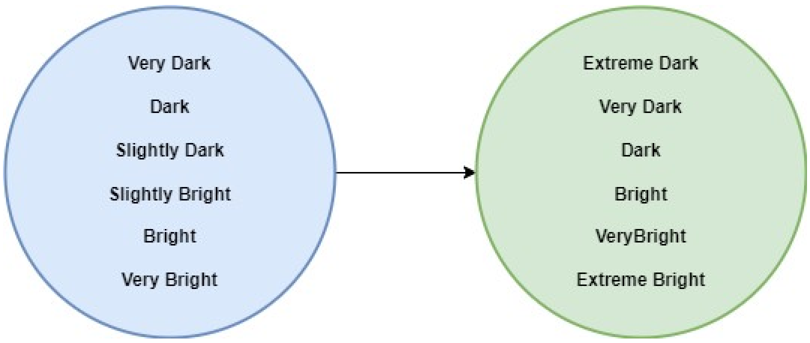


Figure 3.5: Rule Set for transformation. Using the rule set, the output fuzzy set from the input pixel intensity is determined.

CNN+DST, SqueezeNET, and PreluNET were then fed the resulting images of enhanced contrast as an array. In the findings section, the accuracy for training and validation is shown for each image contrast enhancement technique. Additionally, test data were used to evaluate the classifiers. The output is also visualized using Explainable AI methods.

3.1.3 DataSet

Burgos-Artizz et al [22] developed the open-access dataset which contains 12400 images of ultrasound. Several operators and ultrasound devices collected ultrasound images from two different hospitals. A maternal-fetal doctor with extensive experience manually tagged each photograph. Images are broken down into six classes: the mother's cervix (which is frequently used for preterm screening), the four most popular fetal anatomical planes (the abdomen, brain, femur, and thorax), and a miscellaneous category that includes any other less common image plane. Figure 3.6 shows the number of images in each of the three main classes - Maternal Cervix, Fetal Brain, Fetal Thorax, Fetal Abdomen and Fetal Femur.

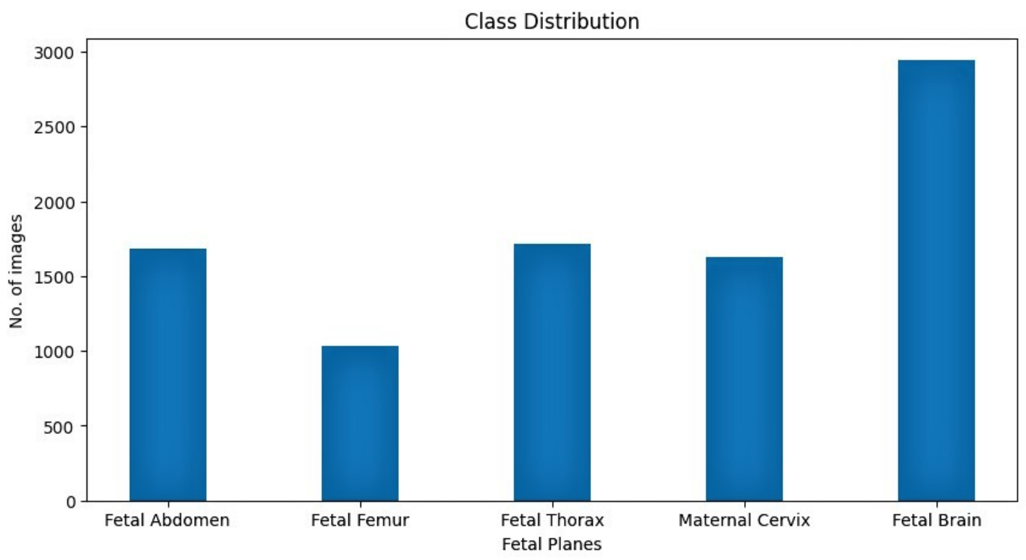


Figure 3.6: Dataset Class Distribution

3.2 Proposed CNN+DS Layer Evidential Model

After being improved in terms of contrast, the images were utilized to train the models. CNN + DS was initially trained using the photos. The training model is depicted in Figure 3.7. Tong et al. proposed the evidential deep learning classifier for set-valued classification, a new classifier built on the Dempster-Shafer (DS) theory and deep CNN. In this classifier, high-order features are extracted from the raw data using a deep CNN. The features are then loaded into a DS layer that is distance-based in order to create mass functions [38]. In our study, we used an evidential

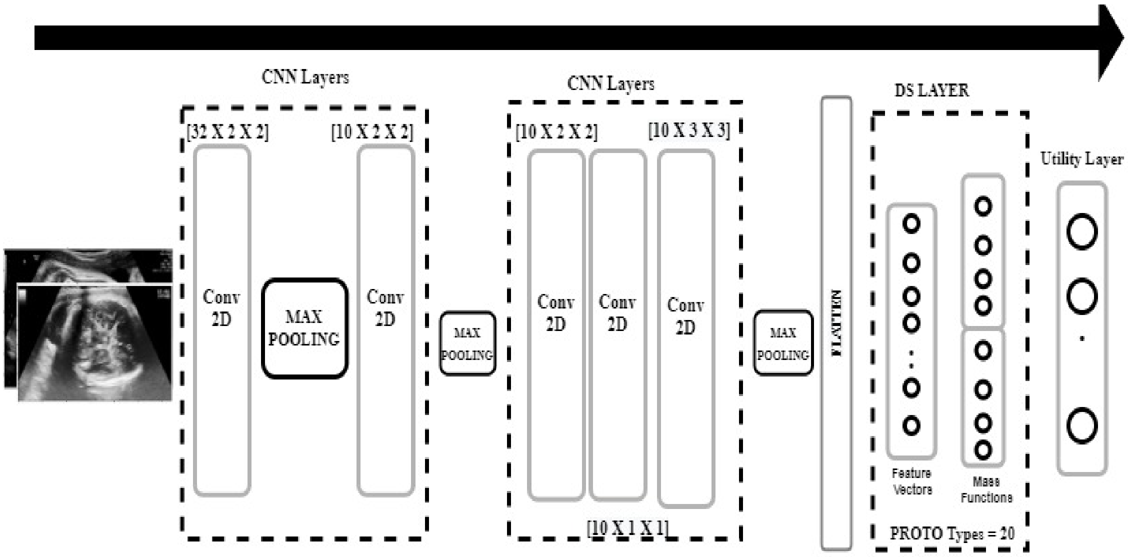


Figure 3.7: Evidential CNN+DS Classifier

convolutional neural network (E-CNN) classifier constructed on the DST framework, in accordance with the research of Tong et al. [38]. However, a customized CNN architecture was initially used to get features from the Images. Then transformed into simple mass functions, and then aggregate them using Dempster's rule with a specifically created CNN architecture. Given that this is a distance-based classifier, the E-CNN classifier splits test samples into various groups based on how closely an input vector resembles the prototypes in the model [38].
75% of the dataset was used for training, and 25% was used for testing. The NumPy array of image pixels is then utilized as input for the first convolution layer, which has 2 X 2-dimensional filters. The number of filters is 32. Prior to MaxPooling2D, the layer is followed by Batch Normalization.

The output is then input into another Convolution layer, which used Batch Normalization and Max Pooling in the same manner. It is followed by three convolution layers of 10 X A X A each, with A equal to 1 for the first two CNN Layers and 3 for the third layer and the result is max pooled again before flattening. After flattening, the output is again given input to DS Layer with prototypes = 20. The necessary libraries for the DS layer are available online [45]. Finally ds_layer.DS3_normalize is given as input to utility_layer_train [45].

The model is trained for 25 epochs with Categorical Cross Entropy as the loss function. An input sample is propagated through a number of steps, just like in probabilistic CNN, to extract latent features crucial for classification. According to [38], the output vector is ready to be given as input to the DS layer. The DS layer receives the feature vector produced in Step 1 and transforms it using Dempster's rule into mass functions.

$$m = m(\{\omega_1\}, \ldots, \{\omega_m\})m(\Omega)^T \tag{3.1}$$

For the Swin Transformer, PreLUNet, and the DST layered CNN, all of the photos were resized to 150*150. Squeeze net's input size, however, was 32*32. The models underwent 25 epochs of training. The training images were split into two groups: 70% were utilized for training, and the remaining 30% were used for validation—a step that is crucial for the analysis. Additionally, a few photos were kept apart for testing. In the findings section, each model's outcomes were discussed.

Chapter 4

Results and Discussion

In this chapter, the effectiveness of the proposed Evidential CNN+DS Layer architecture is assessed. This chapter compared the proposed architecture's performance to those of PreLUNet, SqueezeNET, and Swin Transformer, three other State of the Art architectures. This section compares the classifier's performances when the images are pre-processed using conventional histogram equalization and fuzzy logic contrast enhancement. Training and Validation Accuracy, Testing Accuracy in terms of F1 Score, Recall, and Precision are the metrics used to assess the performance of prediction. The classifiers' computation time contrast is also noted. GradCam and the Lime Algorithm are then used to observe the explainability of the classifiers.

4.1 Performance Evaluation Measures

The accuracy, sensitivity, specificity, and F-Measure are examples of well-known performance measures that have been used to assess the results. The number of true positives (TP), false positives (FP), true negatives (TN), and false negatives are used in the calculation of these measures (FN). The number of correctly identified negative and positive samples are denoted by TN and TP, respectively. FN and FP represent the proportion of positive and negative samples that were misclassified.

The classification scheme's overall effectiveness is judged by its accuracy. This is how it can be calculated:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

A classifier's sensitivity is its capacity to identify positive class sequences. The calculation is as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.2}$$

Precision and recall are used by F-Measure to calculate classification accuracy.

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN} \tag{4.3}$$

$$F - Measure = 2X \frac{PrecisionXRecall}{Precision + Recall} \tag{4.4}$$

4.2 Experimental Results

The proposed Evidential CNN+DS model was trained with ultrasound images enhanced with Fuzzy Logic Contrast Enhancement. The performances are evaluated in this subsection.

4.2.1 Performance evaluation of the proposed model

Figure 4.1 depicts the CNN+DS Evidential Architecture training and validation accuracy curve for fuzzy logic contrast enhancement and histogram equalization enhancement. In the scenario where the visual contrast was raised using fuzzy logic, training accuracy climbed gradually from 0.3689 to 0.8354 after 35 training epochs. A similar pattern could be seen in the validation accuracy, which had a maximum value of 0.8364. In the case of Histogram Equalization, the maximum training accuracy achieved is 0.8723, and the maximum validation accuracy is 0.8714.

Hence a similar outcome is seen in the case of DS layer-based CNN when both HE and Fuzzy contrast although accuracies are better for HE by a small margin. The confusion matrix in Figure 4.2 was created based on the results of testing the model using testing data. The results showed most of the photos in each class were correctly categorized. The higher density of the diagonal section in the confusion matrix proved that. However, there were very few images that were wrongly labeled.

For fuzzy logic contrast enhancement and histogram equalization, the SqueezeNET training and validation accuracy curve is depicted in Figure 4.3. In the situation where the visual contrast was improved using fuzzy logic, after training for 25 epochs, the training accuracy climbed gradually from above 0.3527 to 0.9103. With a maximum value of 0.8565, the validation accuracy displayed a similar pattern. The maximum values for training accuracy and validation accuracy in the case of Histogram Equalization are 0.9515 and 0.8825 respectively.

The confusion matrix in the form of a stacked bar chart in Figure 4.4 was created based on the results of testing the model using testing data after 25 epochs of

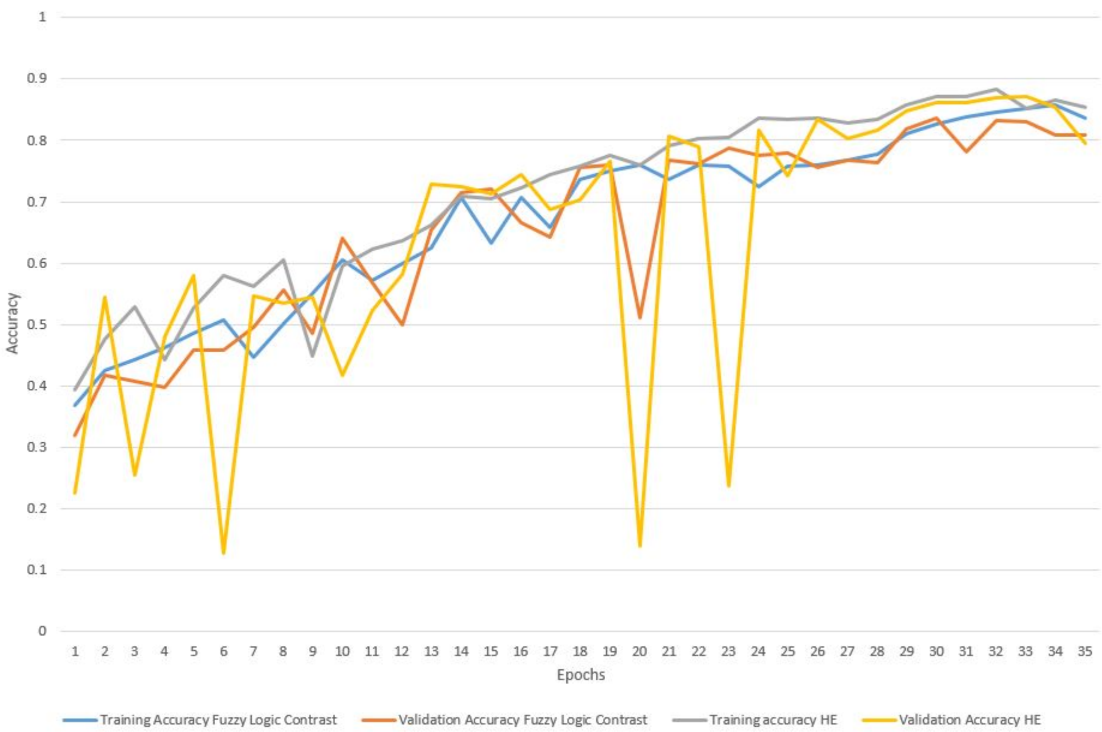


Figure 4.1: Training and Validation curves for Evidential CNN+DS Classifier of Fuzzy Contrast Enhancement and Histogram Equalization

Figure 4.2: Confusion Matrix of Evidential CNN+DS Classifier

model training. The results showed that while there were a few misclassifications, most of the photos in each class were correctly categorized. This is indicated by the large regions of the bar chart with the same color as the color of the particular class indicating the correct classification for those number of images.

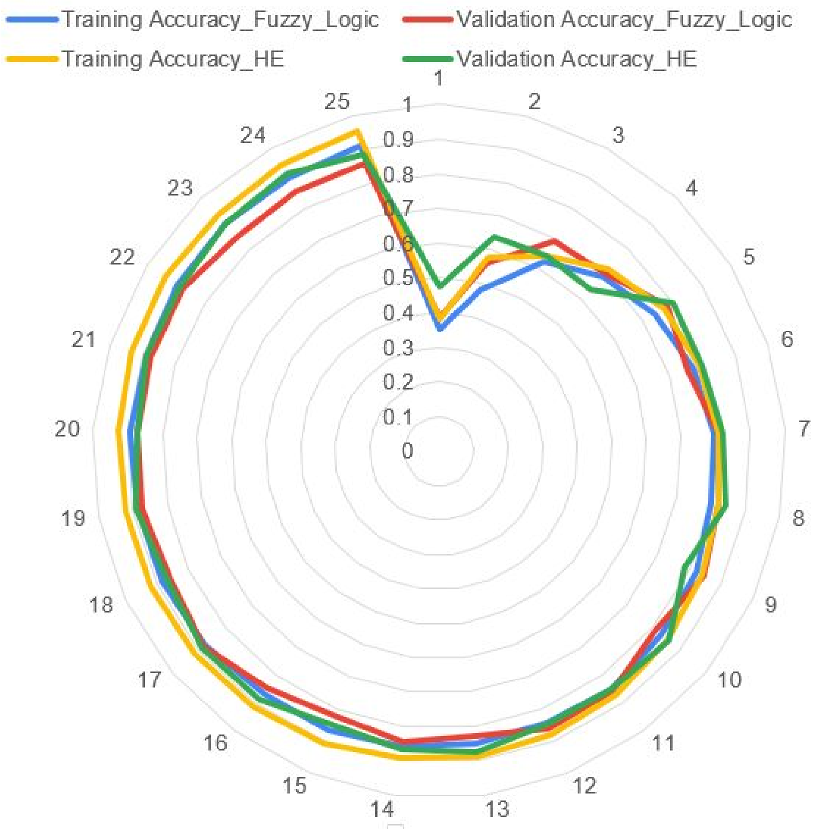


Figure 4.3: Training and validation accuracies of SqueezeNET for Fuzzy Logic and Histogram Equalization

The PreLUNet's maximum training and validation accuracy for both Fuzzy Logic Contrast Enhancement and Histogram Equalization are displayed in Figure 4.6 After training for 25 epochs, the training accuracy increased progressively from above 0.6654 to 0.9553 in the scenario when the visual contrast was increased using fuzzy logic. Despite some dips between the 16th to 18th epochs.

The validation accuracy showed a similar pattern, with a maximum value of 0.8559. For training accuracy in the case of HE, the maximum training accuracy value is 0.9521, and the validation accuracy is 0.8768. When the results of Histogram Equalization and Fuzzy Logic Image Contrast Enhancement are examined, it can be seen that the classifier worked better when images undergo histogram equalization. Figure 4.7 shows the classification accuracy in terms of Stacked Bar chart of the PreLUNet when tested with testing data.

Figure 4.8 showed the accuracy curves for Swin Transformer. The graphs showed an increase in the accuracies during both training and validation for HE and Fuzzy Logic Contrast Enhancement. The training accuracy reached 0.8890 and validation accuracy 0.8106 in the case of Fuzzy Logic Contrast Enhancement and the training

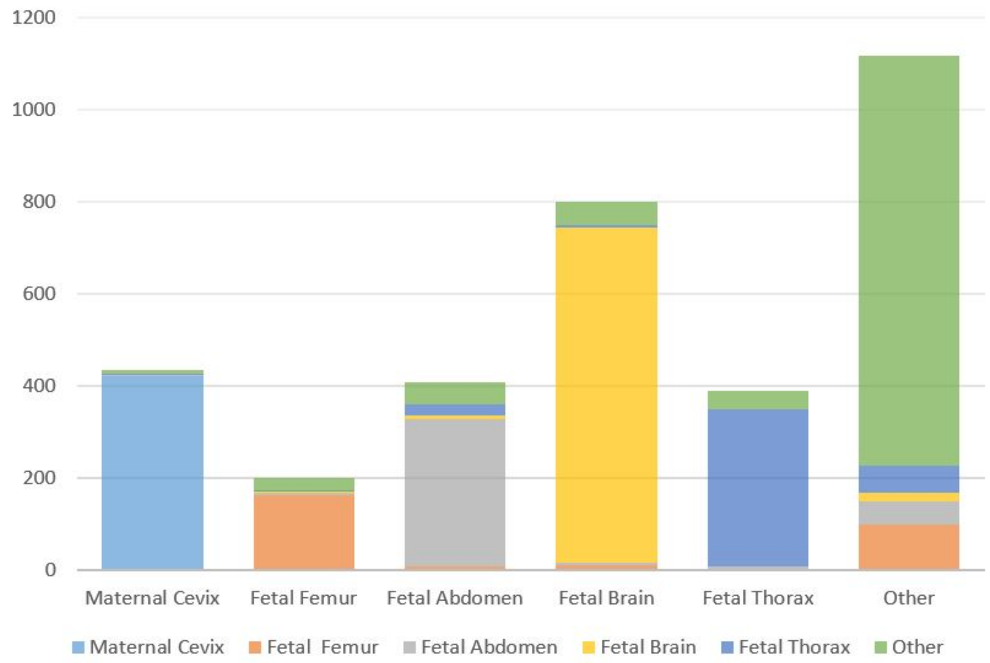


Figure 4.4: Stacked Bar chart representing predictions of SqueezeNET with testing Data

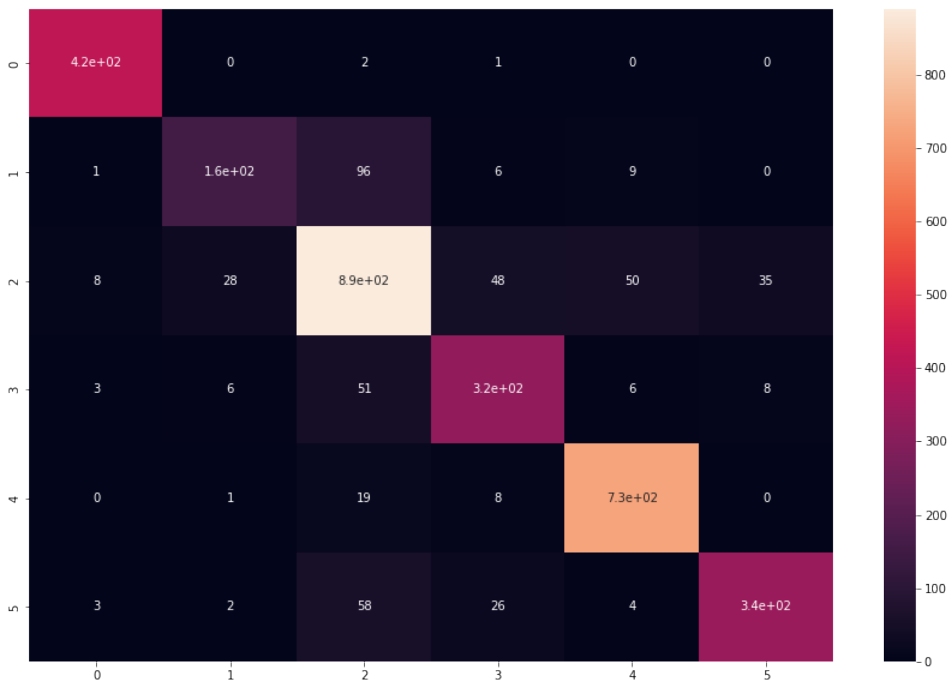


Figure 4.5: Confusion Matrix of SqueezeNET

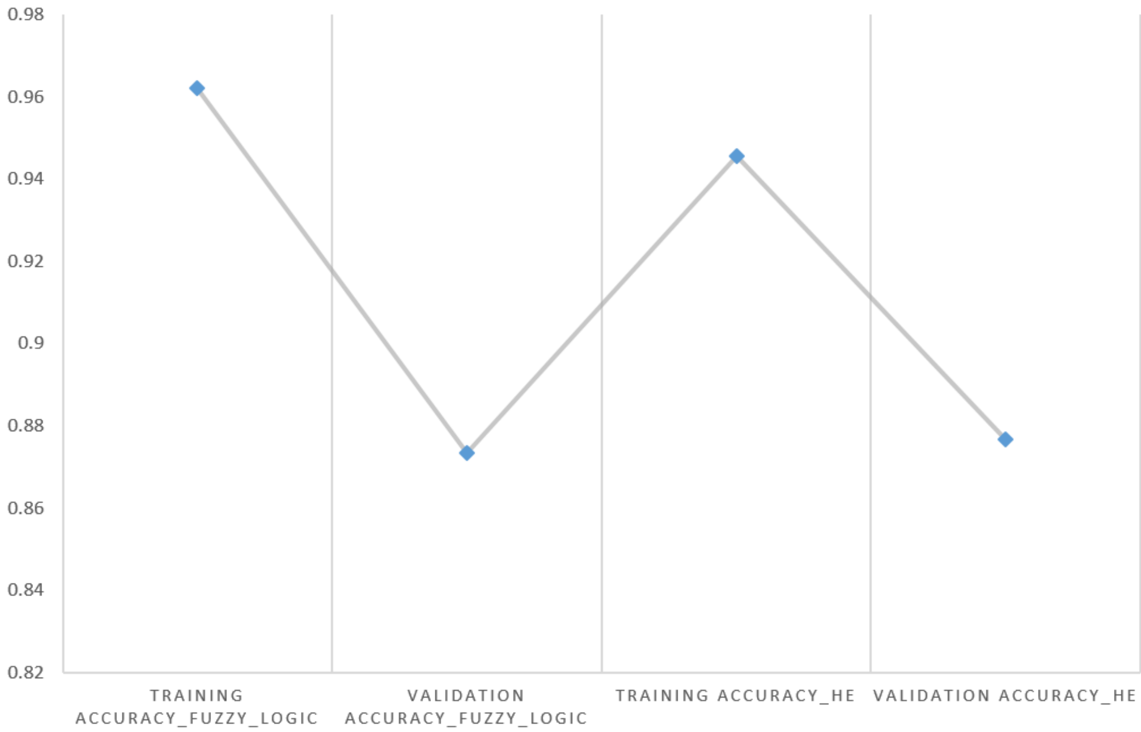


Figure 4.6: Maximum Training and Validation accuracy of PreluNET of Fuzzy Logic Contrast Enhancement and Histogram Equalization

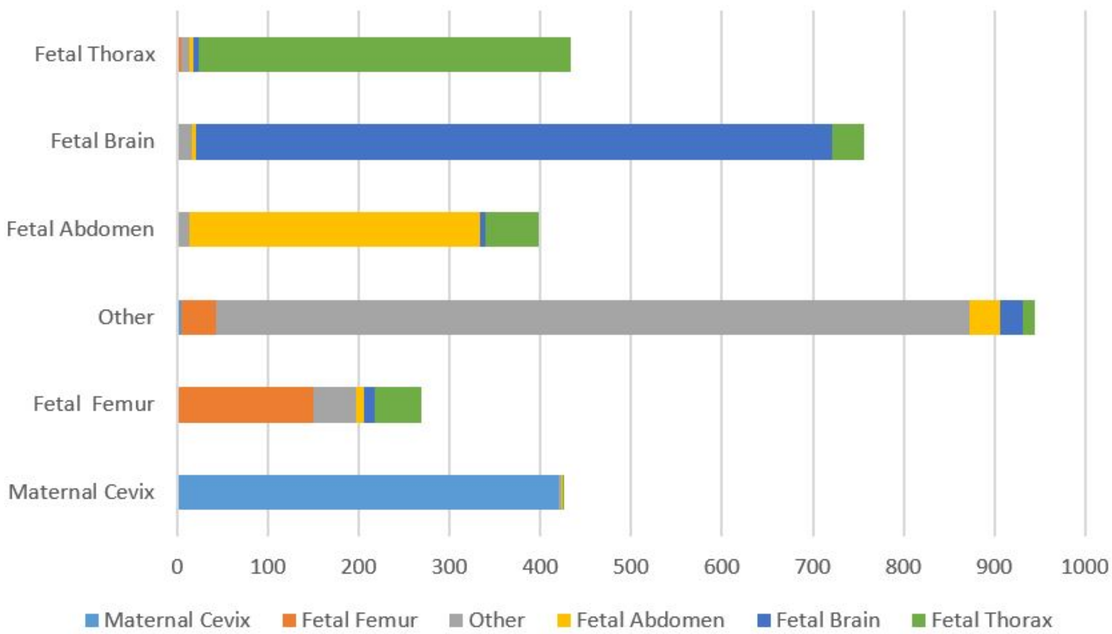


Figure 4.7: Stacked Bar-chart representing classification performance of PReLUNet with testing data

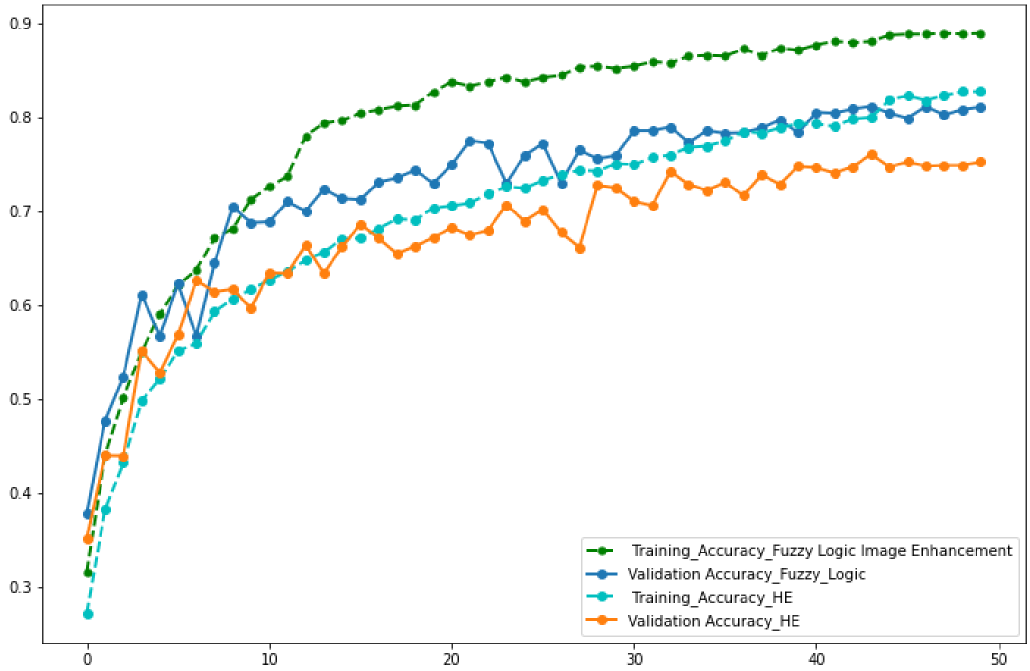


Figure 4.8: Training and validation accuracies of Swin Transformer for Fuzzy Logic and Histogram Equalization

accuracy reached 0.7979 and validation accuracy 0.7471 in the case of Fuzzy Logic Contrast Enhancement. To summarize the outcome from each classifier is reviewed in Tables 4.1 and 4.2 that corresponds to the maximum training and validation accuracy of each classifier contrasting the impact of Fuzzy Logic and Histogram Equalization based image enhancement.

Table 4.1: Maximum Training and Validation Accuracy using Fuzzy Logic Based Contrast Enhancement

Category	Models	Maximum Training Accuracy	Maximum Validation Accuracy
	Evidential CNN+DS Model	0.8354	0.8364
	SqueezeNET	0.9103	0.8565
Fuzzy Contrast Enhancement	PreLUNET	0.9553	0.8559
	Swin Transformer	0.88903	0.8106

Table 4.2: Maximum Training and Validation Accuracy using Histogram Equalization

Category	Models	Maximum Training Accuracy	Maximum Validation Accuracy
Histogram Equalization	Evidential CNN+DS Model	0.8723	0.8714
	SqueezeNET	0.9515	0.8825
	PreLUNET	0.9553	0.8559
	Swin Transformer	0.7979	0.7471

The classification report in Tables 4.3 and 4.4 again proved how the architectures performed when tested with testing data. A total of 3337 testing images were used.

The F1 scores, precision, and recall values for each classifier were displayed in the classification report in the figure. When using Fuzzy Logic Contrast Enhancement on the CNN+DS layer with the Maternal Cervix class, the precision is 0.96, and when using Histogram Equalizer, it is 0.88.

The recall for the maternal cervix is 0.99 for CNN+DSLayer Evidential Classifier, 0.99 for fuzzy logic contrast enhancement, and 0.99 for HE. How many test photos from each class were used as indicated on the support. With precision levels of 0.56 for CNN+DSLayer for Fuzzy Logic Contrast Enhancement and 0.81 for Histogram Equalizer, the findings for the Fetal Thorax class are a little less impressive. The

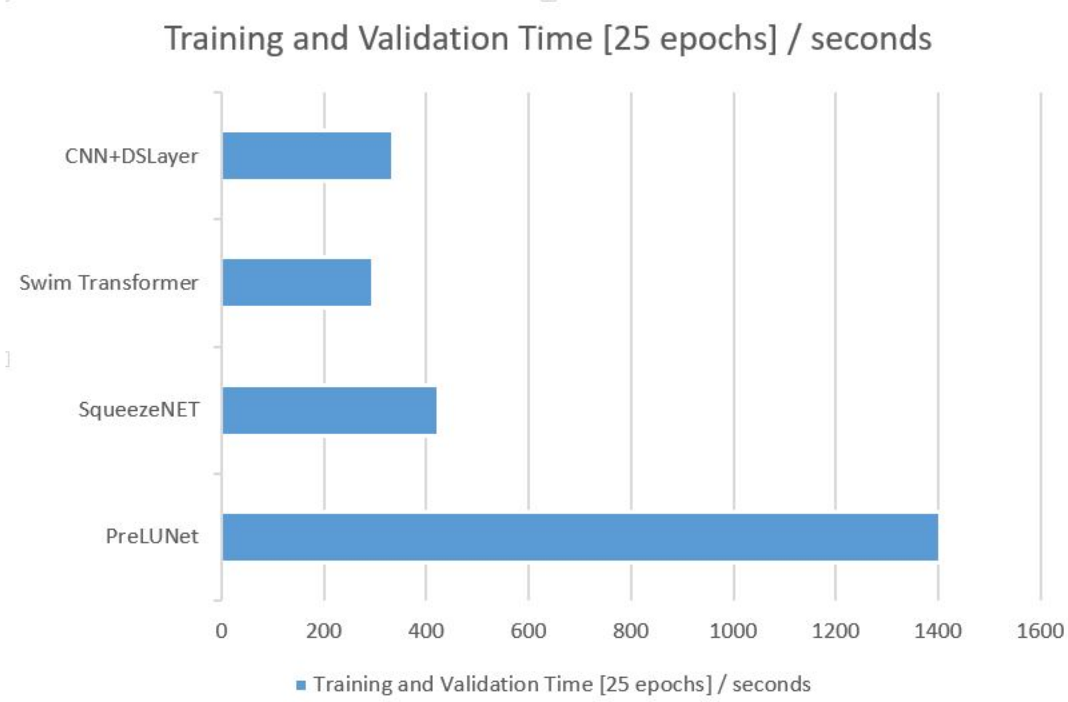


Figure 4.9: Computation time (25 epochs) for training and validation

classifiers were also compared by the amount of time consumed to run 25 epochs while getting trained and validated. To get an idea of training and validating time for a constant number of epochs, time for 25 epochs of the CNN+DS Layer is considered. The graph in Figure 4.9 represented that Swin Transformer took the shortest time i.e. 296.516s. CNN+DS Layer took 334.369s to get trained for 25 epochs. The highest time taken by PreLUNet is 1404.51s.

4.2.2 Explaining the Outcomes of the models with LIME Algorithm and GRAD CAM

XAI is a general term that refers to methods, equations, and tools that advance our knowledge of how AI works. Typically, AI solutions are "black boxes," which means the developer is unable to explain the results of the algorithm. As a result of XAI, developers, decision-makers, customers, and other stakeholders can better understand an AI forecast [41]. Understanding the behavior of our machine learning model is more important.

Local Interpretable and Model-independent LIME is referred to as explanations [8]. LIME is capable of detecting text, images, and tabular data using classifiers. The scholars introduced LIME, a ground-breaking explanation technique that accurately

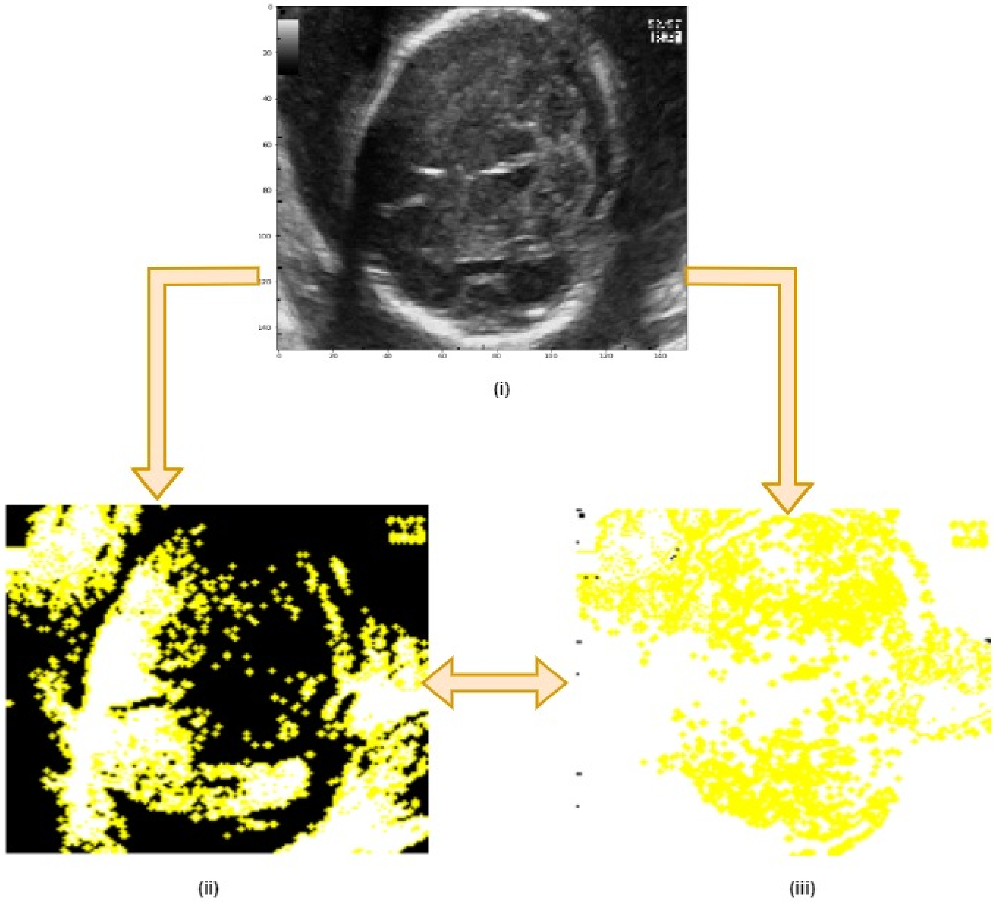


Figure 4.10: Mask representing Super Pixels in Lime Algorithm (i) Actual Test image that the proposed Evidential CNN+DS will classify. (ii)-(iii) Super pixels generated keeping the only positives that triggers decision on and off respectively

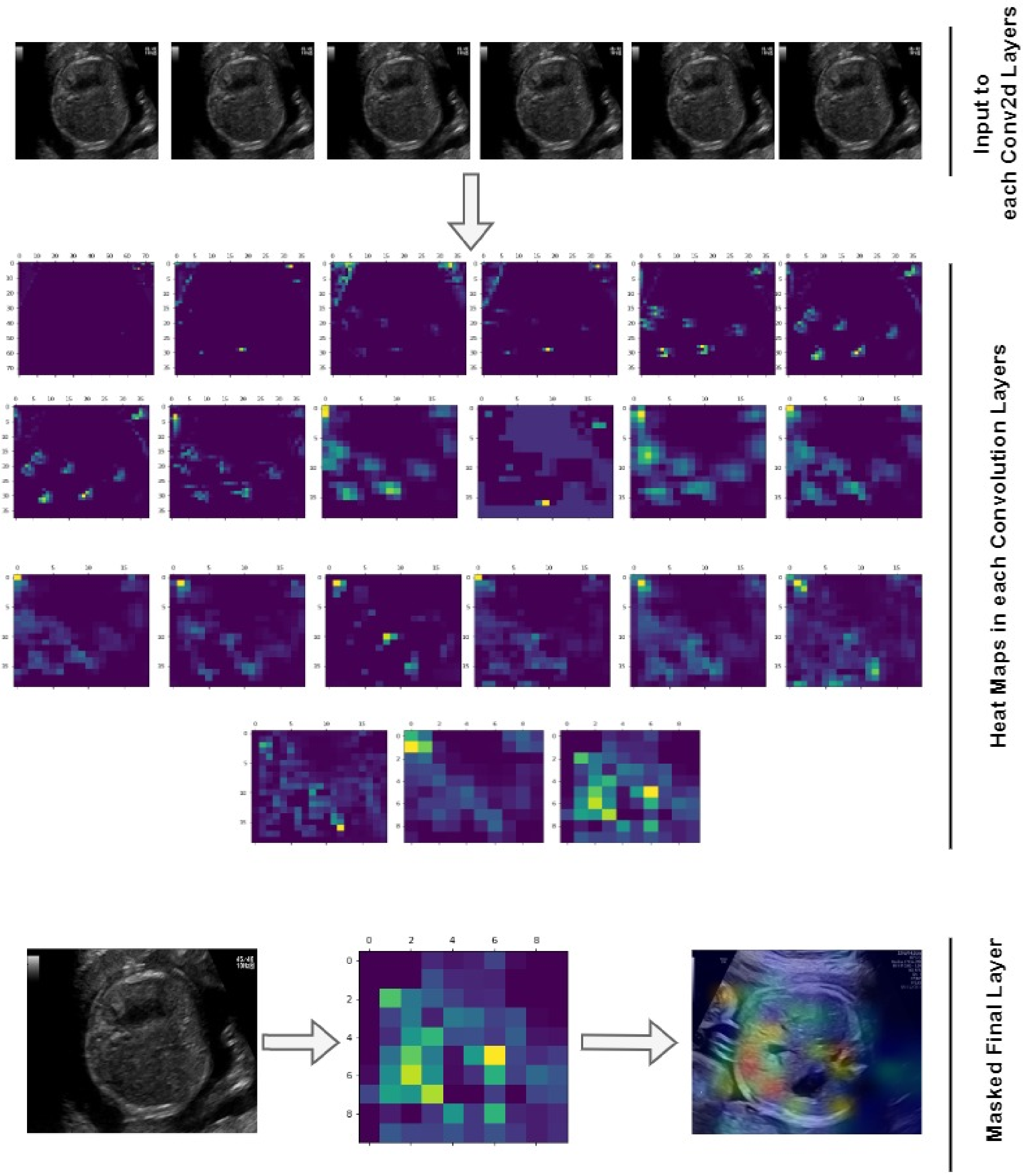


Figure 4.11: Heat Maps from Conv2d_1 layer to Conv2d_22 layer

and understandably explains any classifier's predictions by building an interpretable model locally around the prediction.

Initially, a test image is selected to be given as input to the trained Evidential + CNN Classifier. The prediction is then analyzed using Lime Algorithm. By using pip install lime, the lime is installed in Google Colab. Using the LimeImageExplainer function, we extracted information from the lime image. All that is needed is to use the explain_instance_function on the explainer object we previously built. The parameters were: *images* indicates which image LIME should describe, *classifier_fn* indicates which prediction function should be used to classify the image, and *top_labels* indicates how many labels LIME should display. If it is 3, the remaining labels will not be displayed and the top three labels with the highest probabilities will be. *Num_samples* is used to set the number of fictitious data points that LIME will generate that are similar to our input.

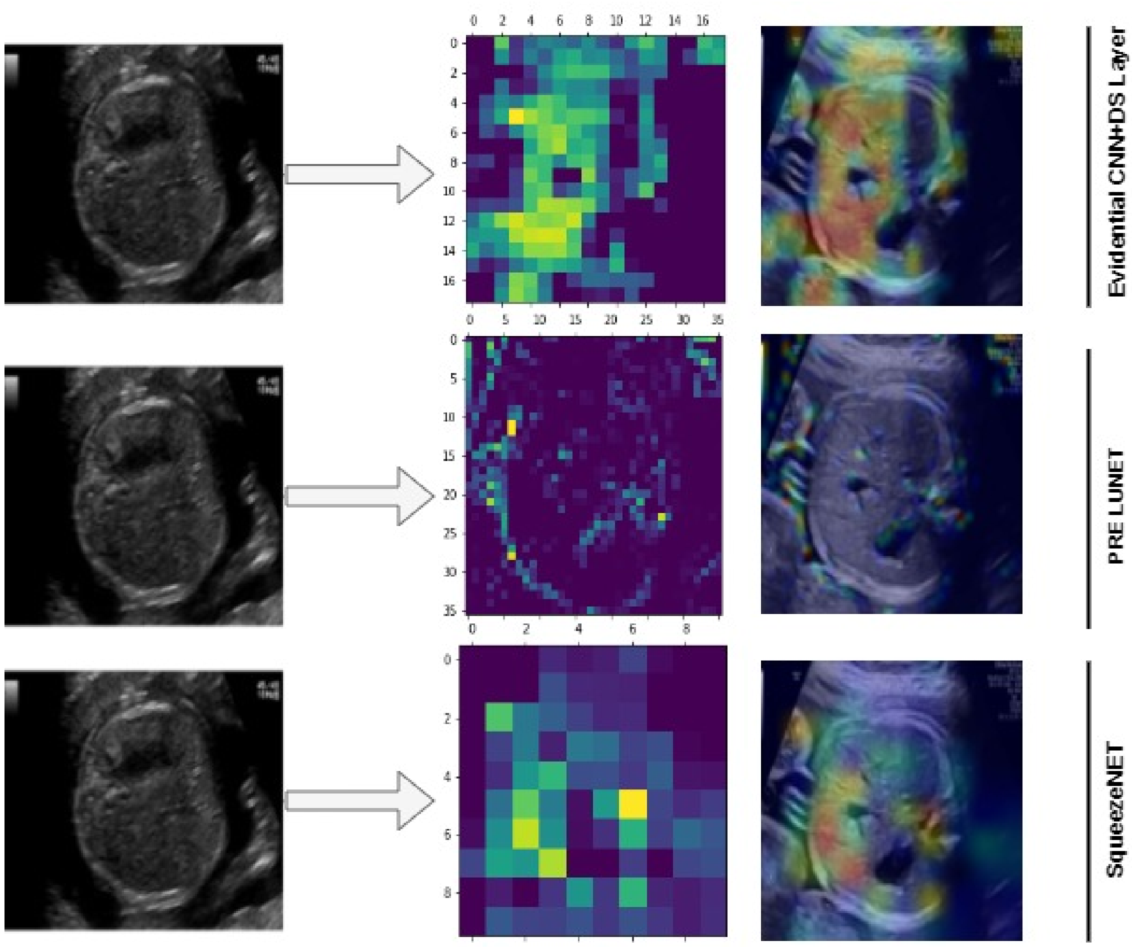


Figure 4.12: Heat Maps from the last convolution layer of each classifier

The explanation object provides a get_image_and_mask() method that accepts predicted labels for the previously parsed 3-d image data and returns a pair of (image, mask) tuples, where the image is a 3d NumPy array and mask is a 2d NumPy array that may be used with skimage.segmentation.mark boundaries. The features in the image that were used to make the prediction are represented in the returned image using the associated mask. The output is shown in Figure 4.10

In case of explaining the outcome using GradCam, when the classifier is given the image of the fetal abdomen as input, Figure

Table 4.3: Classification Report of Each Classifiers

Category	Models	Metrics	Maternal Cervix	Fetal Femur	Fetal Abdomen	Fetal Brain	Fetal Thorax	Other
Fuzzy Logic Based Contrast Enhancement	PreLUNet	Precision	0.97	0.79	0.85	0.94	0.60	0.90
		Recall	0.99	0.56	0.80	0.92	0.95	0.78
		F1 Score	0.99	0.66	0.84	0.93	0.73	0.83
		Support	421	271	397	754	436	1058
	SqueezeNET	Precision	0.97	0.81	0.78	0.91	0.89	0.80
		Recall	0.99	0.59	0.81	0.96	0.79	0.84
		F1 Score	0.98	0.68	0.80	0.94	0.83	0.82
		Support	421	271	397	754	436	1058
	CNN+DS Evidential Classifier	Precision	0.97	0.81	0.78	0.91	0.89	0.80
		Recall	0.96	0.78	0.78	0.92	0.56	0.89
		F1 Score	0.98	0.71	0.49	0.93	0.95	0.78
		Support	421	271	397	754	436	1058
	Swin Transformer	Precision	0.99	0.70	0.83	0.70	0.68	0.71
		Recall	0.92	0.49	0.30	0.86	0.75	0.80
		F1 Score	0.95	0.58	0.44	0.77	0.71	0.75
		Support	421	271	397	754	436	1058

Table 4.4: Classification Report of Each Classifiers

Category	Models	Metrics	Maternal Cervix	Fetal Femur	Fetal Abdomen	Fetal Brain	Fetal Thorax	Other
Histogram Equalization	PreLUNet	Precision	0.98	0.78	0.83	0.94	0.59	0.89
		Recall	0.99	0.56	0.80	0.92	0.95	0.78
		F1 Score	0.99	0.66	0.83	0.92	0.95	0.84
		Support	421	271	397	754	436	1058
	SqueezeNET	Precision	0.98	0.77	0.93	0.98	0.90	0.80
		Recall	1.00	0.77	0.84	0.91	0.84	0.88
		F1 Score	0.99	0.77	0.88	0.95	0.87	0.84
		Support	421	271	397	754	436	1058
	CNN+DS Evidential Classifier	Precision	0.88	0.77	0.75	0.98	0.81	0.68
		Recall	0.83	0.61	0.74	0.69	0.78	0.88
		F1 Score	0.89	0.73	0.82	0.79	0.79	0.77
		Support	421	271	397	754	436	1058
	Swin Transformer	Precision	0.99	0.73	0.67	0.55	0.77	0.81
		Recall	0.98	0.71	0.49	0.93	0.95	0.78
		F1 Score	0.98	0.47	0.50	0.69	0.74	0.74
		Support	421	271	397	754	436	1058

Chapter 5

Conclusion

This study demonstrated how to identify the fetus planes in ultrasound pictures using a Dempster Shafer-based evidential CNN classifier. This paper also illustrates the application of the Swin Transformer in the analysis of medical images. Once the inputs were improved with Histogram Equalization and Fuzzy Logic Contrast, the classifiers functioned extremely accurately. In the findings section, the outputs from each enhancement are contrasted. All four of the classifiers—CNN+DS, Swin Transformer, SqueezeNET, and PreLUNet—performed admirably. The Dempster Shafer Layer is an Evidentiary classifier that is used in this study together with a specially created CNN architecture.

In machine learning, classification is the action of applying a learning set of labeled cases to a new sample in order to categorize it. Precision classification is a typical classification problem in which a sample is assigned to just one of the five potential classes. Unluckily, assignments that are difficult are frequently those that have a lot of uncertainty.

A possible solution to this issue is set-valued classification. As a result, it can be seen how using Evidential CNN, which has only recently been utilized to locate fetal planes in ultrasound images, can be helpful. Second, the comparison between using Histogram Equalization and contrast enhancement of MRI images based on fuzzy logic is also shown in the results section. Metrics like F1 scores, recall, and precision of each classifier are used to evaluate the effects of using fuzzy logic contrast enhancement in ultrasound images as pre-processing. The examination of the ultrasound fetal planes uses the Swin Transformer less frequently.

Therefore, this study makes an effort to investigate the role of the Swin transformer in the identification of the fetal thorax, brain, abdomen, femur, and maternal cervix. The performance of the CNN+DS Layer, PreLUNet, SqueezeNET, and the Swin Transformer topologies is displayed in the results section together with that of other architectures. The Swin Transformer has recently been used for medical image processing and has the potential to dramatically advance fetal diagnosis automation. Additionally, the Grad Cam and Lime Algorithm is used to explain how to understand the output.

As fetal brain imaging can be further divided into Trans-thalamic, Trans-cerebellum, and Trans-ventricular classes, there are more areas to investigate. Therefore, the purpose of this study is to continue classifying the images using UNET architecture segmentation to group the brain images into the aforementioned classes. However, if we focus on a few limitations, then first of all the dataset was very large and hence Colab free version was not sufficient. However, the issue was resolved using Google Colab pro which provided sufficient memory.

Bibliography

[1] G. Shafer, *A mathematical theory of evidence*. Princeton university press, 1976, vol. 42.

[2] K. Sentz and S. Ferson, “Combination of evidence in dempster-shafer theory,” 2002.

[3] A. P. Dempster, “Upper and lower probabilities induced by a multivalued mapping,” in *Classic works of the Dempster-Shafer theory of belief functions*, Springer, 2008, pp. 57–72.

[4] G. Raju and M. S. Nair, “A fast and efficient color image enhancement method based on fuzzy-logic and histogram,” *AEU-International Journal of electronics and communications*, vol. 68, no. 3, pp. 237–243, 2014.

[5] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[6] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size,” *arXiv preprint arXiv:1602.07360*, 2016.

[8] M. T. Ribeiro, S. Singh, and C. Guestrin, “” why should i trust you?” explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

[9] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, “Medical image analysis using convolutional neural networks: A review,” *Journal of medical systems*, vol. 42, no. 11, pp. 1–13, 2018.

[10] A. Hidayatuloh, M. Nursalman, and E. Nugraha, “Identification of tomato plant diseases by leaf image using squeezenet model,” in *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, IEEE, 2018, pp. 199–204.

[11] K. M. Hosny, M. A. Kassem, and M. M. Foaud, “Skin cancer classification using deep learning and transfer learning,” in *2018 9th Cairo international biomedical engineering conference (CIBEC)*, IEEE, 2018, pp. 90–93.

[12] S. Joshi and S. Kumar, “Image contrast enhancement using fuzzy logic,” *arXiv preprint arXiv:1809.04529*, 2018.

[13] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," *Classification in BioApps*, pp. 323–350, 2018.

[14] G. Hemanth, M. Janardhan, and L. Sujihelen, "Design and implementing brain tumor detection using machine learning approach," in *2019 3rd international conference on trends in electronics and informatics (ICOEI)*, IEEE, 2019, pp. 1289–1294.

[15] H. J. Lee, I. Ullah, W. Wan, Y. Gao, and Z. Fang, "Real-time vehicle make and model recognition with the residual squeezenet architecture," *Sensors*, vol. 19, no. 5, p. 982, 2019.

[16] S. Sajid, S. Hussain, and A. Sarwar, "Brain tumor detection and segmentation in mr images using deep learning," *Arabian Journal for Science and Engineering*, vol. 44, no. 11, pp. 9249–9261, 2019.

[17] M. Siar and M. Teshnehlab, "Brain tumor detection using deep neural network and machine learning algorithm," in *2019 9th international conference on computer and knowledge engineering (ICCKE)*, IEEE, 2019, pp. 363–368.

[18] Z. Sobhaninia, S. Rafiei, A. Emami, *et al.*, "Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning," in *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, IEEE, 2019, pp. 6545–6548.

[19] P. Sridar, A. Kumar, A. Quinton, R. Nanan, J. Kim, and R. Krishnakumar, "Decision fusion-based fetal ultrasound image plane classification using convolutional neural networks," *Ultrasound in medicine & biology*, vol. 45, no. 5, pp. 1259–1273, 2019.

[20] Z. Tong, P. Xu, and T. Denoeux, "Convnet and dempster-shafer theory for object recognition," in *International Conference on Scalable Uncertainty Management*, Springer, 2019, pp. 368–381.

[21] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, "Pneumonia detection using cnn based feature extraction," in *2019 IEEE international conference on electrical, computer and communication technologies (ICECCT)*, IEEE, 2019, pp. 1–7.

[22] X. P. Burgos-Artizzu, D. Coronado-Gutiérrez, B. Valenzuela-Alcaraz, *et al.*, "Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[23] J. Daghrir, L. Tlig, M. Bouchouicha, and M. Sayadi, "Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach," in *2020 5th international conference on advanced technologies for signal and image processing (ATSIP)*, IEEE, 2020, pp. 1–5.

[24] H. Dou, D. Karimi, C. K. Rollins, *et al.*, "A deep attentive convolutional neural network for automatic cortical plate segmentation in fetal mri," *IEEE transactions on medical imaging*, vol. 40, no. 4, pp. 1123–1133, 2020.

[25] *Fetal ultrasound*, Nov. 2020. [Online]. Available: https://www.mayoclinic.org/tests-procedures/fetal-ultrasound/about/pac-20394149.

[26] T. Gabruseva, D. Poplavskiy, and A. Kalinin, "Deep learning for automatic pneumonia detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 350–351.

[27] J. Hong, H. J. Yun, G. Park, *et al.*, "Fetal cortical plate segmentation using fully convolutional networks with multiple plane aggregation," *Frontiers in neuroscience*, vol. 14, p. 591 683, 2020.

[28] P. Liu, H. Zhao, P. Li, and F. Cao, "Automated classification and measurement of fetal ultrasound images with attention feature pyramid network," in *Second Target Recognition and Artificial Intelligence Summit Forum*, SPIE, vol. 11427, 2020, pp. 661–666.

[29] Nguyenvlm, *Fuzzy logic - image contrast enhancement*, Jun. 2020. [Online]. Available: https://www.kaggle.com/code/nguyenvlm/fuzzy-logic-image-contrast-enhancement.

[30] T. Saba, A. S. Mohamed, M. El-Affendi, J. Amin, and M. Sharif, "Brain tumor detection using fusion of hand crafted and deep learning features," *Cognitive Systems Research*, vol. 59, pp. 221–230, 2020.

[31] F. Ucar and D. Korkmaz, "Covidiagnosis-net: Deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (covid-19) from x-ray images," *Medical hypotheses*, vol. 140, p. 109 761, 2020.

[32] H. Xie, N. Wang, M. He, *et al.*, "Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal," *Ultrasound in Obstetrics & Gynecology*, vol. 56, no. 4, pp. 579–587, 2020.

[33] G. Capobianco, C. Cerrone, A. Di Placido, *et al.*, "Image convolution: A linear programming approach for filters design," *Soft Computing*, vol. 25, no. 14, pp. 8941–8956, 2021.

[34] M. Dildar, S. Akram, M. Irfan, *et al.*, "Skin cancer detection: A review using deep learning techniques," *International journal of environmental research and public health*, vol. 18, no. 10, p. 5479, 2021.

[35] P. Gupta, "Pneumonia detection using convolutional neural networks," *Science and Technology*, vol. 7, no. 01, pp. 77–80, 2021.

[36] M. Komatsu, A. Sakai, R. Komatsu, *et al.*, "Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning," *Applied Sciences*, vol. 11, no. 1, p. 371, 2021.

[37] Z. Liu, Y. Lin, Y. Cao, *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 012–10 022.

[38] Z. Tong, P. Xu, and T. Denoeux, "An evidential classifier based on dempster-shafer theory and deep learning," *Neurocomputing*, vol. 450, pp. 275–293, 2021.

[39] *Fetal ultrasound*, Jan. 2022. [Online]. Available: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/fetal-ultrasound.

[40] J. Huang, Y. Fang, Y. Wu, *et al.*, "Swin transformer for fast mri," *Neurocomputing*, vol. 493, pp. 281–304, 2022.

[41] T. Leers, *The explainable AI boom: Why is XAI important? And why now? - Tim Leers — dataroots.io*, https://dataroots.io/research/contributions/why-xai-and-why-now, [Accessed 05-Dec-2022].

[42] *Medium — kyawsawhtoon*, https://medium.com/@kyawsawhtoon/a-tutorial-to-histogram-equalization-497600f270e2., [Accessed 06-Dec-2022].

[43] *Medium — towardsdatascience.com*, https://towardsdatascience.com/review-squeezenet-image-classification-e7414825581a., [Accessed 05-Dec-2022].

[44] *Medium — towardsdatascience.com*, https://towardsdatascience.com/histogram-equalization-5d1013626e64., [Accessed 06-Dec-2022].

[45] tongzheng1992, *E-cnn-classifier/libs at main · tongzheng1992/e-cnn-classifier*. [Online]. Available: https://github.com/tongzheng1992/E-CNN-classifier/tree/main/libs.