Ultrasonics 132 (2023) 107017

Contents lists available at ScienceDirect

Ultrasonics

journal homepage: www.elsevier.com/locate/ultras

Demystifying evidential Dempster Shafer-based CNN architecture for fetal plane detection from 2D ultrasound images leveraging fuzzy-contrast enhancement and explainable AI

Rafeed Rahman [a], Md. Golam Rabiul Alam [a], Md. Tanzim Reza [a], Aminul Huq [a], Gwanggil Jeon [b,*], Md. Zia Uddin [c], Mohammad Mehedi Hassan [d,*]

[a] Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh
[b] Department of Embedded Systems Engineering, Incheon National University, Incheon, Republic of Korea
[c] Software and Service Innovation, SINTEF Digital, Oslo 0373, Norway
[d] Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

ARTICLE INFO

Keywords:
PreLUNet
Evidential Dempster–Shafer CNN
Swin Transformer
Histogram Equalization
Fuzzy Logic Contrast

ABSTRACT

Ultrasound imaging is a valuable tool for assessing the development of the fetal during pregnancy. However, interpreting ultrasound images manually can be time-consuming and subject to variability. Automated image categorization using machine learning algorithms can streamline the interpretation process by identifying stages of fetal development present in ultrasound images. In particular, deep learning architectures have shown promise in medical image analysis, enabling accurate automated diagnosis. The objective of this research is to identify fetal planes from ultrasound images with higher precision. To achieve this, we trained several convolutional neural network (CNN) architectures on a dataset of 12400 images. Our study focuses on the impact of enhanced image quality by adopting Histogram Equalization and Fuzzy Logic-based contrast enhancement on fetal plane detection using the Evidential Dempster–Shafer Based CNN Architecture, PReLU-Net, SqueezeNET, and Swin Transformer. The results of each classifier were noteworthy, with PreLUNet achieving an accuracy of 91.03%, SqueezeNET reaching 91.03% accuracy, Swin Transformer reaching an accuracy of 88.90%, and the Evidential classifier achieving an accuracy of 83.54%. We evaluated the results in terms of both training and testing accuracies. Additionally, we used LIME and GradCam to examine the decision-making process of the classifiers, providing explainability for their outputs. Our findings demonstrate the potential for automated image categorization in large-scale retrospective assessments of fetal development using ultrasound imaging.

1. Introduction

During pregnancy, women undergo fetal ultrasound testing, which provides visual information about the unborn child in the uterus. It is a secure method of determining a baby's health, and during a fetal ultrasound, the heart, head, and spine of the unborn child are examined, along with other body parts [1]. Ultrasound uses a device known as the transducer for sending out sound waves and receiving them as well. As the transducer is moved around the abdomen, sound waves enter at various speeds through the skin, muscle, bone, and fluids. The baby acts as a sound reflector, reflecting sound waves back to the transducer, and an electronic picture is created from the sound waves by the transducer, which can be viewed on a computer screen [1].

During the first trimester of pregnancy, an ultrasound is performed to determine the existence and precise location of the pregnancy, the number of fetuses, and to assess how long the mother will carry the child. Additionally, ultrasonography can be used to check for uterine or cervix abnormalities during the first trimester of pregnancy [2]. A standard ultrasound is carried out to evaluate numerous characteristics of the pregnancy in the second and third trimesters, including fetal anatomy. On average, this test is done between weeks 18 and 20. However, the timing of this ultrasound may vary depending on several factors, such as the patient's weight, which could make it more difficult to see the fetus.

* Corresponding authors.
E-mail addresses: rafeedrahmansham2015@gmail.com (R. Rahman), rabiul.alam@bracu.ac.bd (Md.G.R. Alam), rezatanzim@gmail.com (Md.T. Reza), aminul.huq@bracu.ac.bd (A. Huq), gjeon@inu.ac.kr (G. Jeon), zia.uddin@sintef.no (Md.Z. Uddin), mmhassan@ksu.edu.sa (M.M. Hassan).

https://doi.org/10.1016/j.ultras.2023.107017
Received 18 January 2023; Received in revised form 10 April 2023; Accepted 13 April 2023
Available online 22 April 2023

In recent years, Deep Learning (DL) [3] has introduced extraordinary advancements in image recognition tasks using Convolution Neural Networks, and artificial intelligence has shown impressive growth over the past ten years. CNN has demonstrated its value in a variety of medical applications, including Pneumonia diagnosis [4–6], the diagnosis of Brain Tumor [7–9], and the detection of skin cancer [10–12]. In our approach to detecting fetal planes using the Evidential customized CNN+DS architecture, we concentrated on anatomical planes of the fetus, including 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 vagina) (birth canal).

1.1. Research contribution

In this study, we have shown the use of an Evidential Dempster Shafer-based classifier to detect fetal planes from images. The process begins with image contrast enhancement by adopting Fuzzy Logic based contrast enhancement [13,14]. Although it has been used previously to enhance the contrast of natural images, but in this research the enhancement has been shown in MRI images. Fuzzy Logic-based contrast enhancement is adopted for image enhancement. The enhanced contrast images are then given as input to the Evidential CNN+DS Layer and the output is analyzed in the results section. To compare output, we have also enhanced the image using traditional Histogram Equalization and the final accuracy scores are recorded. Along with this Evidential CNN+DS Classifier, we have used Swin Transformer, SqueezeNET, and PreluNet to compare the final accuracies. The key contributions of the research are as follows:

- In this study, an Evidentiary classifier called the Dempster Shafer Layer is used in conjunction with a custom-designed CNN architecture. In machine learning, the process of classifying a new sample by applying a learning set of labeled cases is known as classification. Precised categorization is a typical classification task in which just one of the possible classes is given to a sample. Unfortunately, there is always a lot of ambiguity when a task is difficult [15]. 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.
- Secondly, Histogram Equalization and Fuzzy Logic-based contrast enhancement are used previously for natural images. In this research, the methods are adopted for showing an application to enhance the contrast of MRI images.
- Swin Transformer is less frequently used in the ultra- sound 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. The results section shows the performance of the Evidential CNN+DS Layer, PreLUNet, SqueezeNET, and Swin Transformer 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 Algorithms is also demonstrated.

1.2. Paper outline

The study is organized as follows: Related Works section includes a literature overview of previous studies in this field along with background information of the architectures and concepts. Detailed information on CNN + Ds Layer-based Evidential architecture are present in the methodology section. 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.

2. Related works

In 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).

Using data from both the whole image and the clipped fetal structures, Sridar et al. developed a system to categorize 14 different structures in 2D ultrasound images of fetal [16]. The categorization is done automatically. 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 does not 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 dataset of 4074 2-D ultrasound pictures [16].

The author in the article [17] used Supervised Object Detection with Normal Data based on the CNN. The purpose of that article is to examine structural irregularities and heart substructures in fetal ultrasound images. They made a timeline that resembled a barcode and gave each film an abnormality number in order to visualize the probability of detection. The ability of SONO to automatically recognize each cardiac substructure in fetal ultrasound images demonstrates how helpful it is for spotting changes in the structural integrity of the heart.

The paper [18] contains a strong deep-learning segmentation approach in order to lessen the laborious human segmentation refining process. A fully connected CNN architecture, residual connections, and specially mixed kernel convolutions with deep attention-based modules 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 results demonstrated that their method outperforms many cutting-edge deep segmentation models as well as a cutting-edge multi-atlas segmentation method.

Radiological technology and ultrasound expertise were needed for the segmentation of anatomical features in ultrasound pictures [19]. 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.

2.1. Dempster–Shafer theory and Evidential classifier

The Dempster–Shafer Theory (DST) is an empirical mathematical theory [20]. [Shafer, 1976], an update of [Dempster, 1967], contains ground-breaking research on the topic. Since the Dempster–Shafer theory distributes probabilities to sets rather than to mutually exclusive singletons, it can be seen as a discrete-space generalization of probability theory. 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 [15].

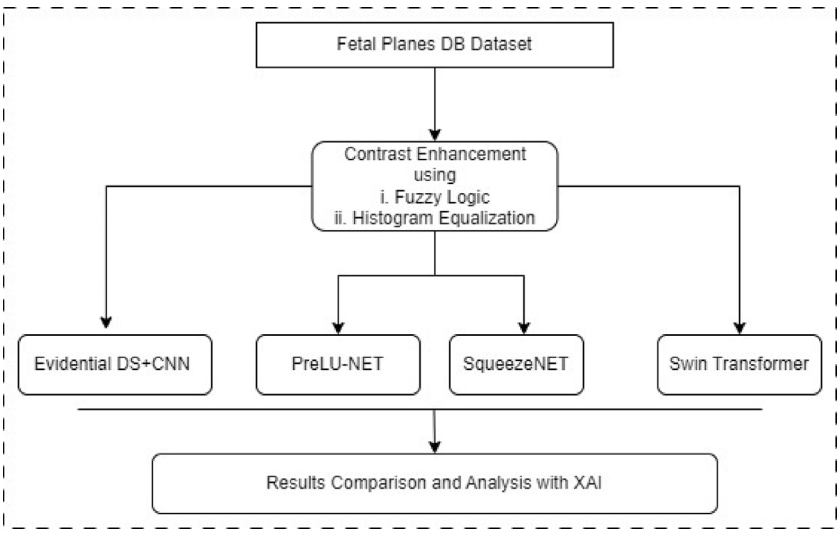

Fig. 1. Research outline.

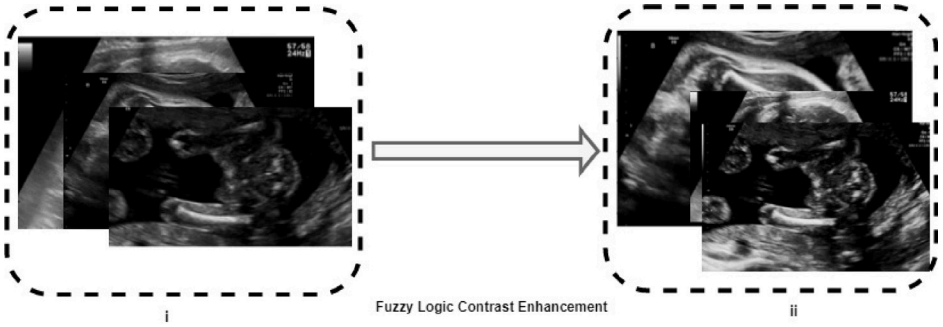

Fig. 2. Contrast enhancement using fuzzy logic — (i) Input image before processing with fuzzy logic contrast enhancement, (ii) outcome of fuzzy logic based contrast enhancement.

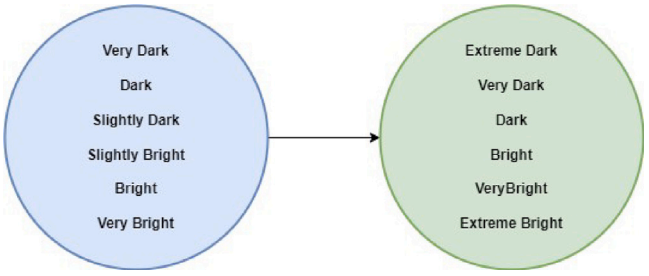

Fig. 3. Rule set for transformation.

Unfortunately, when there is a lot of uncertainty, a difficult assignment frequently results in misclassification. Set-valued classification, which is referred to as a potential remedy for the problem when there is insufficient certainty to construct a precise categorization, assigns a new observation into a subset of classes that is not empty. Set-valued classification reduces error rates because it encourages classifier caution and presents categorization uncertainty more accurately.

In the article [15], Tong et al. constructed a novel classifier for set-valued classification utilizing the Dempster Shafer Based CNN Architecture. The first step is the extraction of high-dimensional features from the input data. Dempster's rule is then used to convert the features into mass functions, which are then combined in a DS layer.

The next step is utilizing mass functions for classification in a set-valued manner in a presumptive utility layer providing an end-to-end technique for collaborating on network parameter modifications. A method for choosing incomplete multi-class actions is also offered [15].

2.2. SqueezeNET

A CNN architecture called SqueezeNET employs 50 times less 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 [21].

Certain techniques are used by the SqueezeNet model to reduce the majority of parameters which are: putting 1×1 filters in place of the 3×3 filters and Three-by-three filters with fewer input channels subsequent downsampling at the network [21]. They reduce the input channel count to 3×3 filters [22]. In order to maximize accuracy on a constrained parameter budget, downsample late in the network so that convolution layers have large activation maps [22]. In the SqueezeNET design, a single convolution layer (conv1) comes before eight Fire modules (fire2-9), and a ninth convolution layer comes after that (conv10). The number of filters per fire module increases consistently across the network from the beginning to the end. Following layers conv1, fire4, fire8, and conv10, SqueezeNET places its pooling a little later in accordance with Strategy 3 [22].

The application of SqueezeNET has been seen in Fetal Brain MRI Dataset [23], in case of Alzheimer Disease Classification [24], a combined approach of SqueezeNET with SVM [25].

2.3. Swin transformer

The traditional multi-head self-attention (MSA) module is changed to one based on moveable windows in Swin Transformer, leaving the

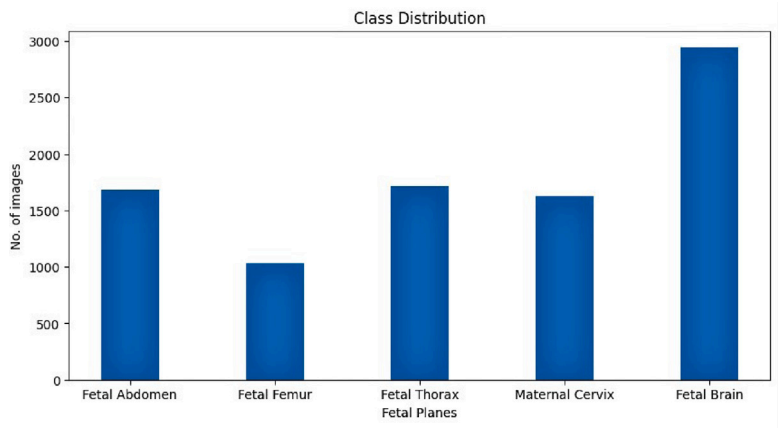

Fig. 4. Dataset class distribution.

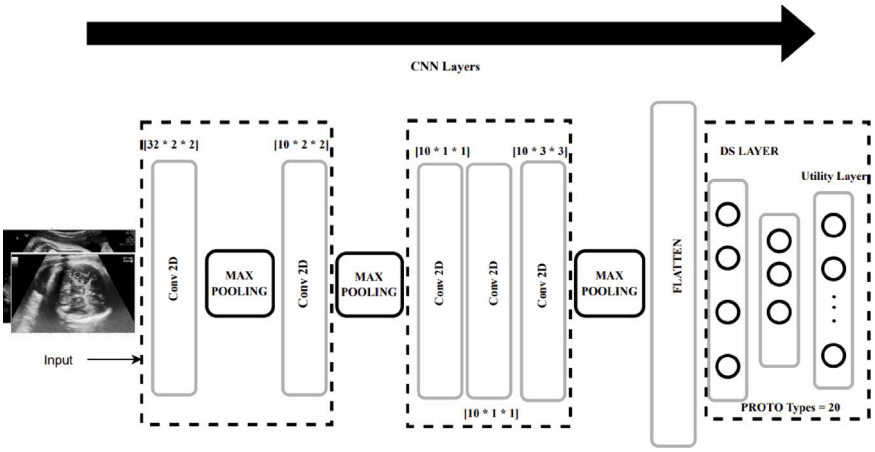

Fig. 5. The architecture of the proposed evidential CNN+DS classifier.

remaining layers alone. A 2-layer MLP with GELU non-linearity 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 [26].

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 [26]. 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.

3. Methodology

The top-level layout of the Evidential Dempster Shafer-based Fetal Planes Analysis from 2D ultrasound images has been presented in Fig. 1. The outline has 4 prime modules: Dataset collection (2 Dimensional Ultrasound images were used [27]), dataset pre-processing using Fuzzy Logic Contrast Enhancement and Histogram Equalization, formulation of training and testing dataset to train the Evidential customized 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.

Fuzzy Logic Contrast Enhancement [13,14,28] is adopted to improve the contrast. Initially, there is CIELAB transformation of the input image [13,14,28], with progress on the L channel. Based on pixel intensity and M value, the fuzzification step determines the degree of each class membership for each individual pixel. Gaussian Function is used in the membership functions.

The papers [13,14] 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 [13].

The rule set is applied as shown in Fig. 3. Fig. 2 shows the outcome of the enhanced image. 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 acquiring a higher contrast [29,30].

The use of multiple histogram-based contrast enhancement for brain MRI pictures has been demonstrated in the journal [31]. Various methods for enhancing MRI brain images, such as global histogram equalization, local histogram equalization (LHE), brightness-preserving dynamic histogram equalization, and adaptive histogram equalization, were examined and compared in another article [32]. In another study, a fuzzy technique is suggested to improve the contrast of brain images obtained through magnetic resonance imaging (MRI) [33].

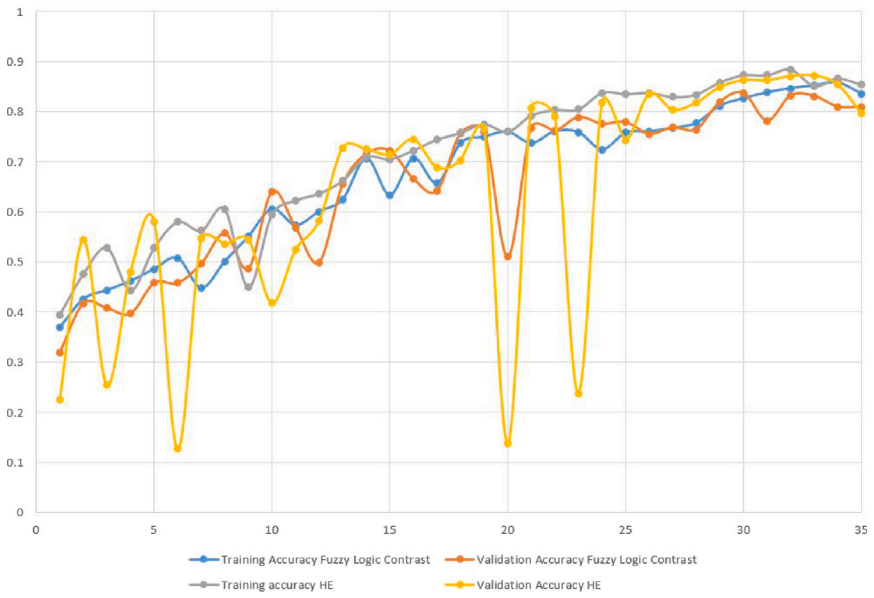

Fig. 6. Training and validation curves for evidential CNN+DS classifier of Fuzzy Contrast Enhancement and Histogram Equalization.

Fig. 7. Confusion matrix of evidential CNN+DS classifier.

The classifiers — Evidential Based CNN+DS Layer, SqueezeNET, Swin Transformer, 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. DataSet

Burgos-Artizz et al. [27] developed the open-access dataset which contains 12,400 images of ultrasound. Several operators and ultrasound devices collected ultrasound images from two different hospitals. A maternal-fetal doctor with extensive experience manually labeled each images. 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. Fig. 4 shows the number of images in each of the five main classes — Maternal Cervix, Fetal Brain, Fetal Thorax, Fetal Abdomen, and Fetal Femur as The Dataset distribution. The remaining images were labeled as others.

3.2. Evidential customized CNN+DS layer model

After being improved in terms of contrast, the images were utilized to train the models. The Evidential Customized CNN + DS Model was initially trained using the images. The proposed Customized CNN + DS layer evidential model is depicted in Fig. 5. 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 25

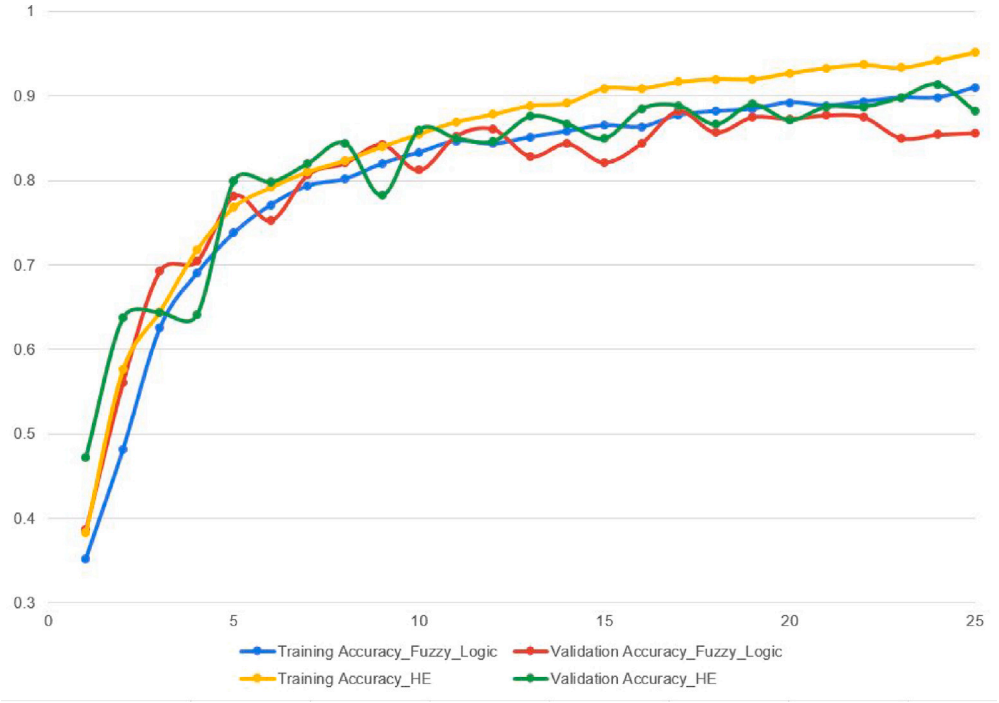

Fig. 8. Training and validation accuracies of SqueezeNET for Fuzzy Logic and Histogram Equalization.

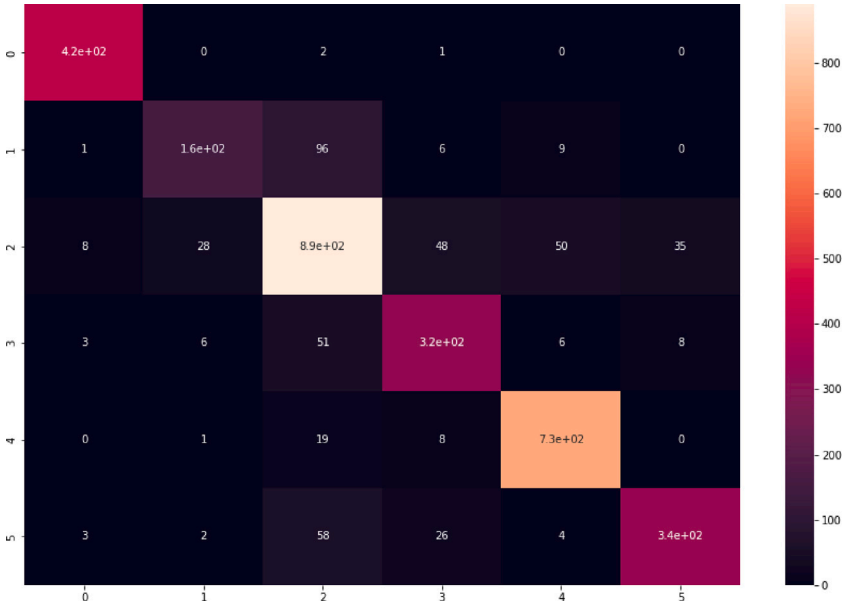

Fig. 9. Confusion matrix of SqueezeNET.

convolutional neural networks (CNN). In that 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 [15]. In our case, we used an evidential convolutional neural network (E-CNN) classifier constructed on the DST framework, in accordance with the research of Tong et al. [15]. However, we customized the CNN architecture, initially, to get features from the images more accurately. Then transformed into simple mass functions, followed by aggregation 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 [15].

The dataset is divided into two parts. The separation occur in a ratio such that 25% images were kept for testing. The NumPy array of image pixels is then utilized as input for the first convolution layer, which has 2 * 2-dimensional filters. The number of filters is 32. Prior to MaxPooling2D, the layer is followed by Batch Normalization. The output is then given as input into another Convolution layer, which uses Batch Normalization and Max Pooling in the same manner. It is followed by three convolution layers of 10 * A * A each, with A equal to 2 and 1 for the first four CNN Layers and 3 for the last layer and the result is max pooled again before flattening. After flattening, the output is again given as input to DS Layer with prototypes = 20. The necessary libraries for the DS layer are available online [34]. Finally ds_layer.DS3_normalize is given as input to utility_layer_train [34].

Fig. 10. Maximum training and validation accuracy of PreluNET of Fuzzy Logic Contrast Enhancement and Histogram Equalization.

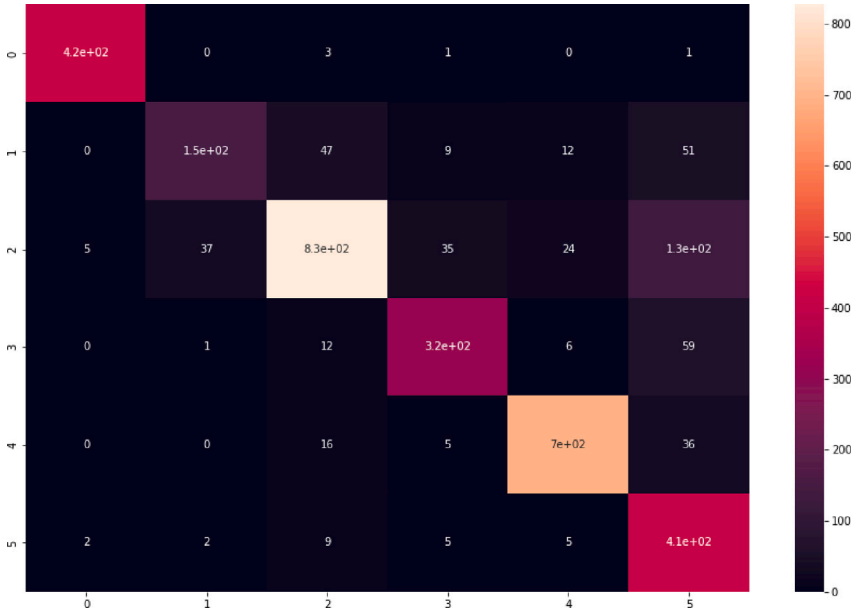

Fig. 11. Confusion Matrix of PreLU NET.

The loss function used is Categorical Cross Entropy and the model is trained for 25 epochs. Same to probabilistic CNN, an input sample is propagated through a number of stages to extract latent properties important for classification. The output vector is prepared to be provided as input to the DS layer according to [15]. The DS layer receives the feature vector produced in Step 1 and transforms it using Dempster's rule into mass functions as it describes the classifier's perception of the sample class's likelihood and measures the degree of uncertainty in the object representation.

For the SqueezeNET, PreLUNet, and the DST layered customized CNN model, all of the images were resized to 150*150. The input size of the Swin Transformer, however, was 32*32. The models underwent certain epochs of training. The training images were split into two groups: 75% were utilized for training, and the remaining 25% were used for validation — a step that is crucial for the analysis. In the findings section, each model's outcomes were discussed.

4. Results analysis

In this part, the effectiveness of the Evidential customized CNN+DS Layer architecture is assessed. This section compared the Evidential customized CNN+DS architecture's performance to those of PreLUNet, Squeeze-NET, and Swin Transformer. This section compares the classifier's performances when the images are pre-processed using conventional histogram equalization and fuzzy logic contrast enhancement separately. Training and Validation Accuracy, Testing Accuracy in terms of F1 Score, Recall, and Precision are the metrics used to

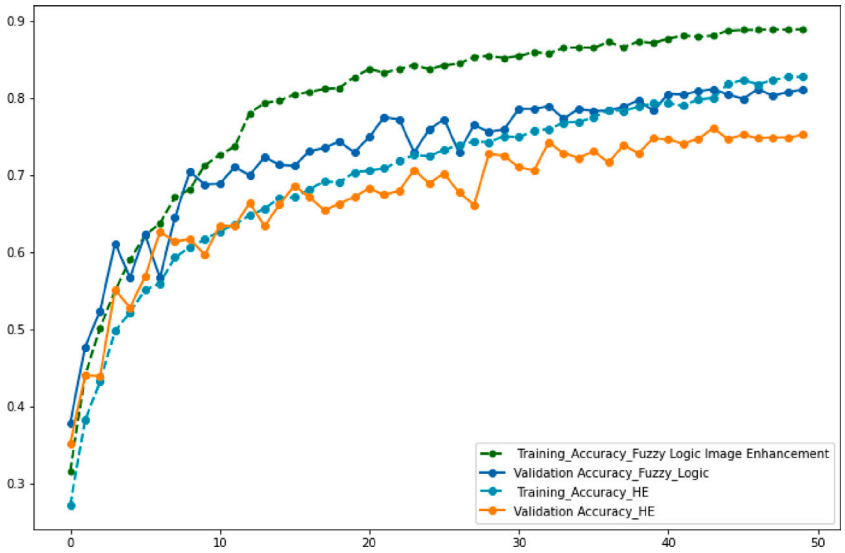

Fig. 12. Training and validation accuracies of Swin Transformer for Fuzzy Logic and Histogram Equalization.

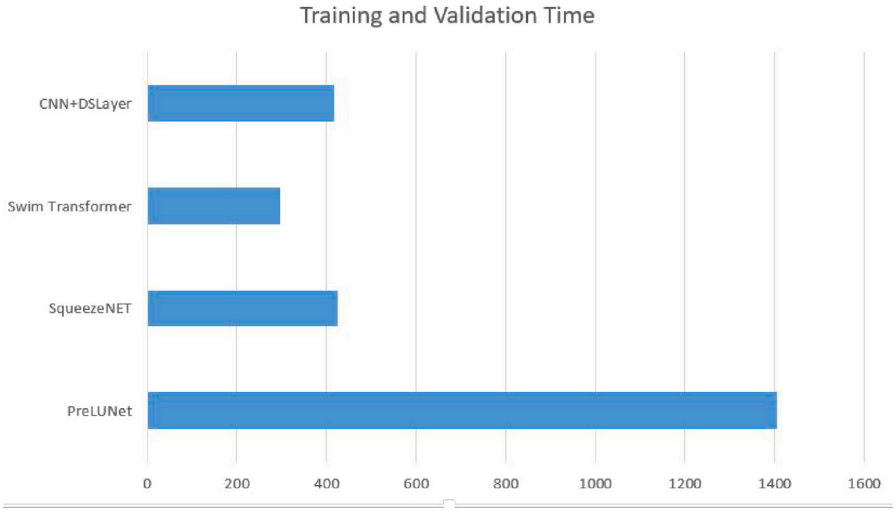

Fig. 13. Computation time: CNN+DS Layer shows computation time for 35 epochs; Other classifiers showed computation time for 25 epochs.

assess the performance of prediction. The classifiers' computation time contrast is also noted.

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 (FN) are used in the calculation of these measures. 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:

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

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

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

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

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

$$F1 = 2 * \frac{PrecisionX\,Recall}{Precision + Recall}$$

4.2. Experiment outcomes

The Evidential customized CNN+DS model was trained with ultrasound images enhanced with Fuzzy Logic Contrast Enhancement. The performances are evaluated in this subsection. Fig. 6 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. The Evidential DS-Layer Classifier took 35 epochs to give a standard accuracy. The

Table 1
Classification report of each classifier after image enhanced with fuzzy logic.

Category	Models	Metrics	Maternal cervix	Fetal femur	Fetal abdomen	Fetal brain	Fetal thorax	Other
		Precision	0.97	0.79	**0.85**	**0.94**	0.60	**0.90**
	PreLUNet	Recall	**0.99**	0.56	0.80	0.92	**0.95**	0.78
FLCE		F1 Score	**0.99**	0.66	**0.84**	0.93	0.73	**0.83**
		Precision	0.97	**0.81**	0.78	0.91	**0.89**	0.80
	Squeeze	Recall	0.99	0.59	**0.81**	**0.96**	0.79	0.84
FLCE	NET	F1 Score	0.98	0.68	0.80	**0.94**	0.83	0.82
		Precision	0.97	**0.81**	0.78	0.91	**0.89**	0.80
	CNN+DS	Recall	0.96	**0.78**	0.78	0.92	0.56	**0.89**
FLCE	Evidential Classifier	F1 Score	0.98	**0.71**	0.49	0.93	**0.95**	0.78
		Precision	**0.99**	0.70	0.83	0.70	0.68	0.71
	Swin	Recall	0.92	0.49	0.30	0.86	0.75	0.80
FLCE	Transformer	F1 Score	0.95	0.58	0.44	0.77	0.71	0.75
		Support	421	271	397	754	436	1058

Table 2
Classification report of each classifier after image enhanced with histogram equalization.

Category	Models [testing images	Metrics	Maternal cervix	Fetal femur	Fetal abdomen	Fetal brain	Fetal thorax	Other
		Precision	0.98	**0.78**	0.83	0.94	0.59	**0.89**
HE	PreLUNet	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
		Precision	**0.98**	0.77	**0.93**	0.98	**0.90**	0.80
HE	SqueezeNET	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**
		Precision	0.88	0.77	0.75	**0.98**	0.81	0.68
HE	CNN+DS	Recall	0.83	0.61	0.74	0.69	0.78	**0.88**
	Evidential Classifier	F1 Score	0.89	0.73	0.82	0.79	0.79	0.77
		Precision	0.99	0.73	0.67	0.55	0.77	0.81
HE	Swin	Recall	0.98	0.71	0.49	**0.93**	0.95	0.78
	Transformer	F1 Score	0.98	0.47	0.50	0.69	0.74	0.74
		Support	421	271	397	754	436	1058

confusion matrix in Fig. 7 was created based on the results of testing the model using testing data. The results showed most of the images in each class were correctly categorized.

For fuzzy logic contrast enhancement and histogram equalization, the SqueezeNET training and validation accuracy curve is depicted in Fig. 8. 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 is shown in Fig. 9 was created based on the results of testing the model using testing data after 25 epochs of model training. The results showed that while there were a few misclassifications, most of the images in each class were correctly categorized. The PreLUNet training and validation accuracy for fuzzy logic contrast enhancement and histogram equalization is shown in Fig. 10. After training for 25 epochs, the training accuracy increased progressively from above 0.6654 to 0.9653 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 0.8759. 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. The confusion matrix for PreLUNet is shown in Fig. 11. Fig. 12 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 accuracy reached 0.7979 and validation accuracy 0.7471 in the case of Fuzzy Logic Contrast Enhancement.

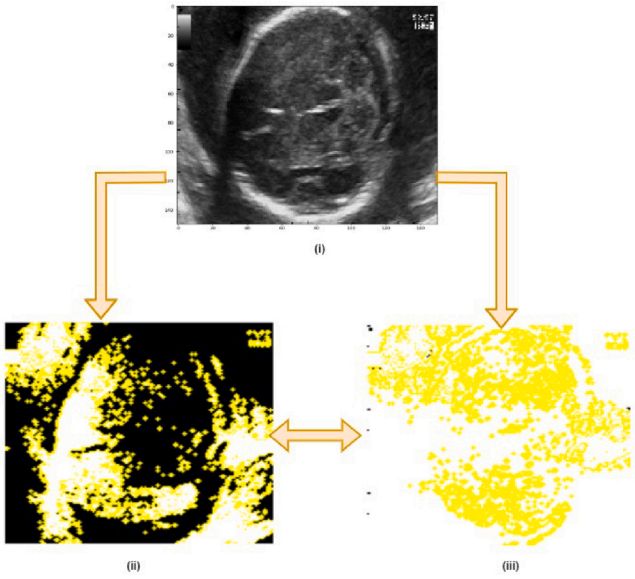

Fig. 14. Mask representing Super Pixels in lime algorithm.

The classification reports in Table 1 and able 2 again proved how remarkably the architectures performed when tested with testing data. A total of 3337 testing images were used. A weighted average of recall and precision is the F1 score. False positive and false negative results can occur in accuracy and recall, as is well known, thus both are taken into account. The recall is the classification model's capacity to recognize each data point in a pertinent class. Precision: a classification

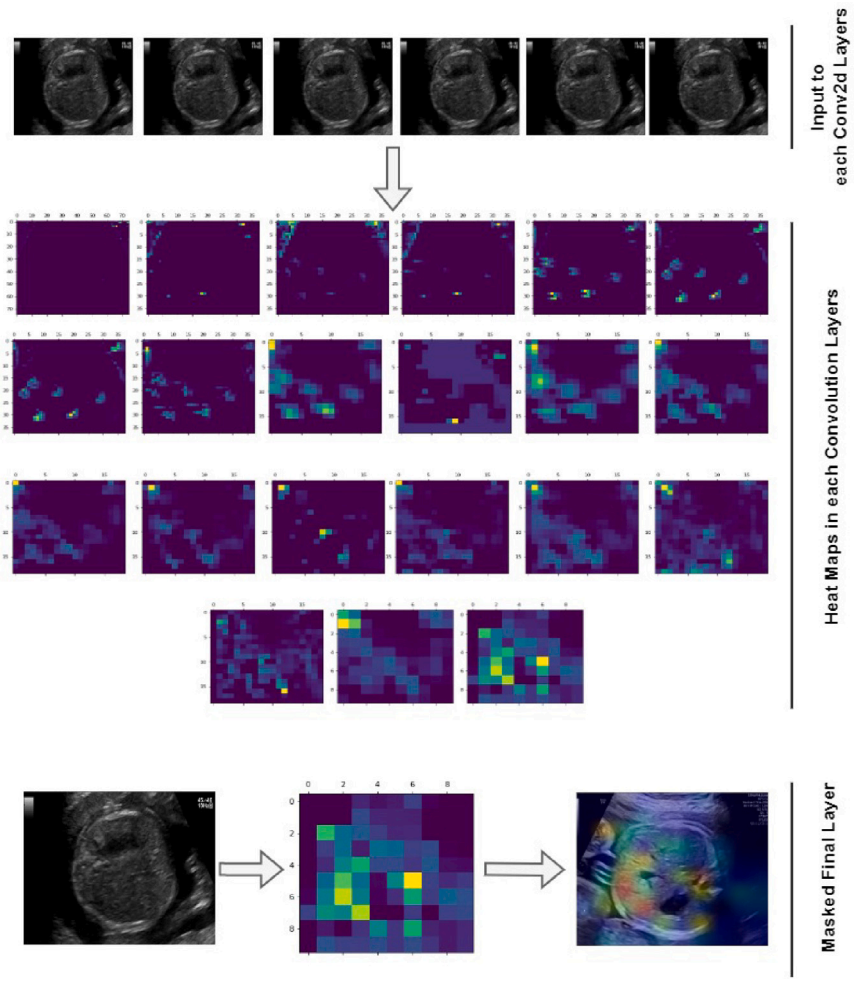

Fig. 15. Heat maps generated from each convolution layers of SqueezeNET and the masked output.

model's capacity to only return data points that belong to a certain class (see Table 2).

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. The number of test images from each class was used as indicated on the support. With precision values of 0.56 for CNN+DSLayer for Fuzzy Logic Contrast Enhancement and 0.81 for Histogram Equalizer, the results for the Fetal Thorax class are a little less impressive.

The 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 curve in Fig. 13 resembled that the Swin Transformer took the shortest time i.e. 296.516 s. CNN+DS Layer took 334.369 s to get trained for 25 epochs. The highest time is taken by PreLUNet is 1404.51 s.

5. 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. Thanks to XAI, developers, decision-makers, customers and other stakeholders can better understand an AI forecast [35]. Understanding the behavior of our machine learning model is more important.

Local Interpretable and Model-independent [LIME] is referred to as explanations [36]. Python module to better comprehend the actions of the proposed black-box classifier model. As of today, LIME is capable of detecting text, images, and tabular data using classifiers. The introduction of LIME is a ground-breaking explanation technique that accurately 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 LimeImageExplainer (), 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 arguments passed as parameters are — *images* that 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* — to set the number of fictitious data points that LIME will generate that are similar to our input [37].

The explanation object provides a get_image_and_mask() method that accepts predicted labels for the previously parsed 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

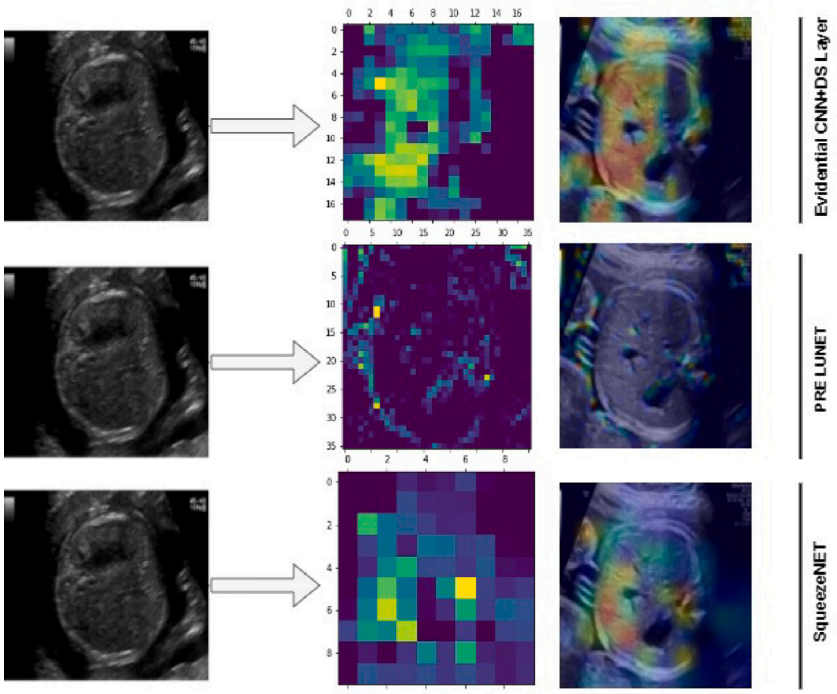

Fig. 16. Heat maps generated from the last conv2d layer of each classifiers.

were used to make the prediction are represented in the returned image using the associated mask. The CNN+DS Layer was tested with a sample of images from the Fetal Brain. The Super pixels in Fig. 14 show the pixels that were responsible for the classifier to correctly classify that test image as the Fetal Brain.

The output is also explained using GradCam. Using GradCAM activation maps and heatmaps displayed at various SqueezeNET levels in Fig. 15, we qualitatively evaluated network-identified regions of interest. The original image and heatmap are layered to identify the key areas in the picture in order to foresee model interpretability and to do further analysis. We used Conv 2d_1 to Conv 2d_23 of SqueezeNET for visualizing the mask output. In this study, we used the GradCAM approach to: (1) view and compare a model's various layers to determine the decision of the model; and (2) pinpoint the network segments that have the biggest impact. (3) to evaluate the 23 convolution layers of the models' activation maps. The prediction of the SqueezeNET architecture is explained using the GradCam by generating heat maps of all the features from the convolution layers of SqueezeNET. The SqueezeNET is given the image of the fetal abdomen as input. Fig. 15 shows the heat maps of the outcomes from each convolution layer in the SqueezeNET architecture. Conv2d_1 to Conv2d_23 outputs are displayed. As can be seen, the heat map is more concentrated in the last output, highlighting the relevant regions resulted in its correct classification by SqueezeNET. Fig. 16. demonstrates the differences in the heat maps and the final masked output when the same input is given to SqueezeNET, DNN+DS Evidential classifier and the PreLUNet.

6. Conclusion and future work

This study demonstrates how to identify fetal planes in ultrasound pictures using a Dempster Shafer-based evidential CNN classifier. This paper also illustrates the application of Swin Transformer in the analysis of medical images. Once the inputs were improved with Histogram Equalization and Fuzzy Logic Contrast, the classifiers functioned with much better precision. In the results section, the outputs from each enhancement are compared. All four of the classifiers, including CNN+DS, Swin Transformer, SqueezeNET, and PreLUNet, performed admirably. 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 future purpose of this study is to continue classifying images using U-NET architecture segmentation to group the brain images into the aforementioned classes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

The authors are grateful to King Saud University, Riyadh, Saudi Arabia for funding this work through Researchers Supporting Project Number- RSP2023R18.

References

[1] John Hopkins Medicine Health, 2022, <https://www.hopkinsmedicine.org/health/treatment-tests-andtherapies/fetal-ultrasound>. (Accessed 22 December 2022).

[2] Ultrasound Mayo Clinic, 2008, 2008<https://www.mayoclinic.org/tests-procedures/fetal-ultrasound/about/pac-20394149>. (Accessed 22 December 2022).

[3] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[4] T. Gabruseva, D. Poplavskiy, 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.

[5] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, 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.

[6] P. Gupta, Pneumonia detection using convolutional neural networks, Sci. Technol. 7 (01) (2021) 77–80.
[7] S. Sajid, S. Hussain, A. Sarwar, Brain tumor detection and segmentation in MR images using deep learning, Arab. J. Sci. Eng. 44 (11) (2019) 9249–9261.
[8] M. Siar, 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.
[9] T. Saba, A.S. Mohamed, M. El-Affendi, J. Amin, M. Sharif, Brain tumor detection using fusion of hand crafted and deep learning features, Cogn. Syst. Res. 59 (2020) 221–230.
[10] M. Dildar, S. Akram, M. Irfan, H.U. Khan, M. Ramzan, A.R. Mahmood, S.A. Alsaiari, A.H.M. Saeed, M.O. Alraddadi, M.H. Mahnashi, Skin cancer detection: a review using deep learning techniques, Int. J. Environ. Res. Public Health 18 (10) (2021) 5479.
[11] J. Daghrir, L. Tlig, M. Bouchouicha, 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.
[12] K.M. Hosny, M.A. Kassem, 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.
[13] S. Joshi, S. Kumar, Image contrast enhancement using fuzzy logic, 2018, arXiv preprint arXiv:1809.04529.
[14] G. Raju, M.S. Nair, A fast and efficient color image enhancement method based on fuzzy-logic and histogram, AEU-Int. J. Electron. Commun. 68 (3) (2014) 237–243.
[15] Z. Tong, P. Xu, T. Denoeux, An evidential classifier based on Dempster–Shafer theory and deep learning, Neurocomputing 450 (2021) 275–293.
[16] P. Sridar, A. Kumar, A. Quinton, R. Nanan, J. Kim, R. Krishnakumar, Decision fusion-based fetal ultrasound image plane classification using convolutional neural networks, Ultrasound Med Biol 45 (5) (2019) 1259–1273.
[17] M. Komatsu, A. Sakai, R. Komatsu, R. Matsuoka, S. Yasutomi, K. Shozu, A. Dozen, H. Machino, H. Hidaka, T. Arakaki, et al., Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning, Appl. Sci. 11 (1) (2021) 371.
[18] H. Dou, D. Karimi, C.K. Rollins, C.M. Ortinau, L. Vasung, C. Velasco-Annis, A. Ouaalam, X. Yang, D. Ni, A. Gholipour, A deep attentive convolutional neural network for automatic cortical plate segmentation in fetal MRI, IEEE Trans. Med. Imaging 40 (4) (2020) 1123–1133.
[19] P. Liu, H. Zhao, P. Li, F. Cao, Automated classification and measurement of fetal ultrasound images with attention feature pyramid network, in: Second Target Recognition and Artificial Intelligence Summit Forum, Vol. 11427, SPIE, 2020, pp. 661–666.
[20] K. Sentz, S. Ferson, Combination of Evidence in Dempster–Shafer Theory, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States); Sandia ..., 2002.
[21] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size, 2016, arXiv preprint arXiv:1602.07360.
[22] S.-H. Tsang, 2022, https://towardsdatascience.com/review-squeezenet-image-classification-e7414825581a. (Accessed 22 December 2022).
[23] S. Parvathavarthini, K. Sharvanthika, N. Bohra, S. Sindhu, Performance analysis of squeezenet and densenet on fetal brain MRI dataset, in: 2022 6th International Conference on Computing Methodologies and Communication, ICCMC, 2022, pp. 1340–1344, http://dx.doi.org/10.1109/ICCMC53470.2022.9753874.
[24] M. Odusami, R. Maskeliunas, R. Damaševičius, S. Misra, Comparable study of pre-trained model on alzheimer disease classification, in: Computational Science and Its Applications–ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13–16, 2021, Proceedings, Part V 21, Springer, 2021, pp. 63–74.
[25] M. Rasool, N.A. Ismail, A. Al-Dhaqm, W. Yafooz, A. Alsaeedi, A novel approach for classifying brain tumours combining a SqueezeNet model with SVM and fine-tuning, Electronics 12 (1) (2023) 149.
[26] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.
[27] X.P. Burgos-Artizzu, D. Coronado-Gutiérrez, B. Valenzuela-Alcaraz, E. Bonet-Carne, E. Eixarch, F. Crispi, E. Gratacós, Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes, Sci. Rep. 10 (1) (2020) 1–12.
[28] NGUYEN VUONG, Fuzzy logic - Image contrast enhancement, 2022, <https://www.kaggle.com/code/nguyenvlm/fuzzy-logic-image-contrast-enhancement>. (Accessed 22 December 2022).
[29] S. Sudhakar, Histogram equalization image contrast enhancement, 2016, https://towardsdatascience.com/histogram-equalization-5d1013626e64. (Accessed 06 December 2022).
[30] Kyaw Saw Htoon, A tutorial to histogram equalization, 2020, https://medium.com/@kyawsawhtoon/a-tutorial-to-histogram-equalization-497600f270e2. (Accessed 06 December 2022).
[31] P.V. Oak, R. Kamathe, Contrast enhancement of brain MRI images using histogram based techniques, Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng. 1 (3) (2013) 90–94.
[32] N. Senthilkumaran, J. Thimmiaraja, Histogram equalization for image enhancement using MRI brain images, in: 2014 World Congress on Computing and Communication Technologies, 2014, pp. 80–83, http://dx.doi.org/10.1109/WCCCT.2014.45.
[33] M. Ravikumar, B. Shivaprasad, D.S. Guru, Enhancement of MRI brain images using fuzzy logic approach, in: Recent Trends in Image Processing and Pattern Recognition: Third International Conference, RTIP2R 2020, Aurangabad, India, January 3–4, 2020, Revised Selected Papers, Part II 3, Springer, 2021, pp. 131–137.
[34] 2022, https://github.com/tongzheng1992/E-CNN-classifier/tree/main/libs. (Accessed 22 December 2022).
[35] T. Leers, The explainable AI boom: Why is XAI important? And why now? - Tim Leers — dataroots.io, 2022, https://dataroots.io/research/contributions/why-xai-and-why-now. (Accessed 05 December 2022).
[36] M.T. Ribeiro, S. Singh, 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.
[37] R. Winastwan, Interpreting image classification model with lime, in: Medium, Towards Data Science, 2021, URL https://towardsdatascience.com/interpreting-image-classification-model-with-lime-1e7064a2f2e5.