Classification of Common Fetal Anatomical Planes from Ultrasound Imaging using Dempster Shafer Theory and Deep Learning

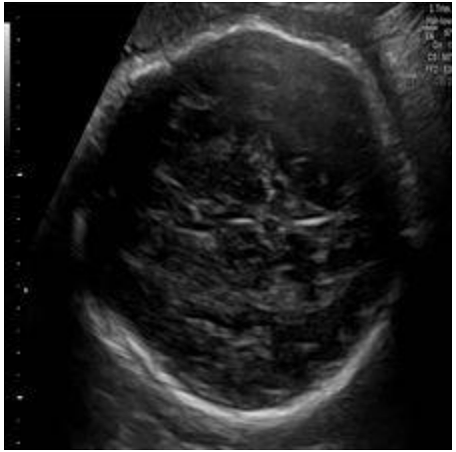
Members:

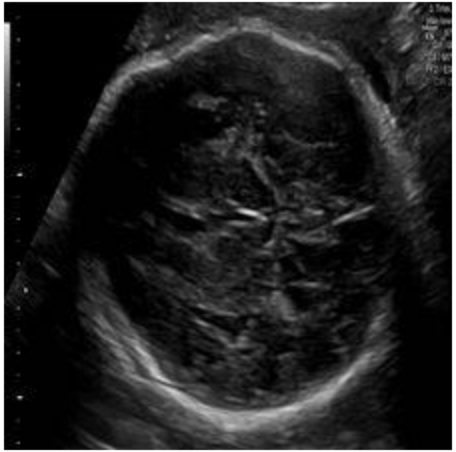
A.M. Tayeful Islam (19101107)
Marshia Nujhat (ID: 19101100)
Atanu Roy (ID: 19101267)
Ahmed Mayeesha Reza Agomoni (19101181)

Supervisor:

Dr. Md. Golam Rabiul Alam
Professor, Department of Computer Science and Engineering
School of Data and Sciences, Brac University

Fetal Anatomical Planes

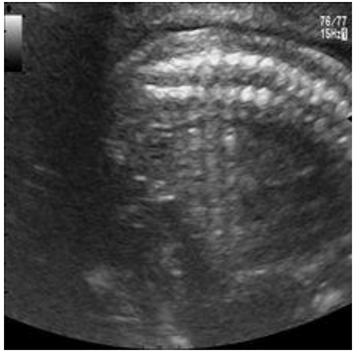




What will be the decision of the model?

Fetal Anatomical Planes

Uncertainty



Confusing Images

More than one
class in an image

Outliers

Motivation

- Accurately classify the images into eight widely used fetal anatomical planes (Abdomen, Femur,Maternal cervix ,Trans-thalamic, Trans-cerebellum, Trans-ventricular , Thorax and Other)
- Reduce time and manpower required to classify these images.
- Classify images into a set of possible results in the cases of high noise and uncertainty .

Research Problem

- US images are hard to classify even with expert eyes due to speckle noise speckle noise, low signal-to-noise ratio, varying intensities of acoustic shadows, motion blurring, missing boundaries and inter-operator errors.
- Inaccurate results in the case of brain planes and uncertain situation
- Requires three stages of man power to classify these images

Existing Work

- Common Fetal plane segmentation using deep CNN
- Evidential fusion model
- Evidential classifier using fitnet-classifier

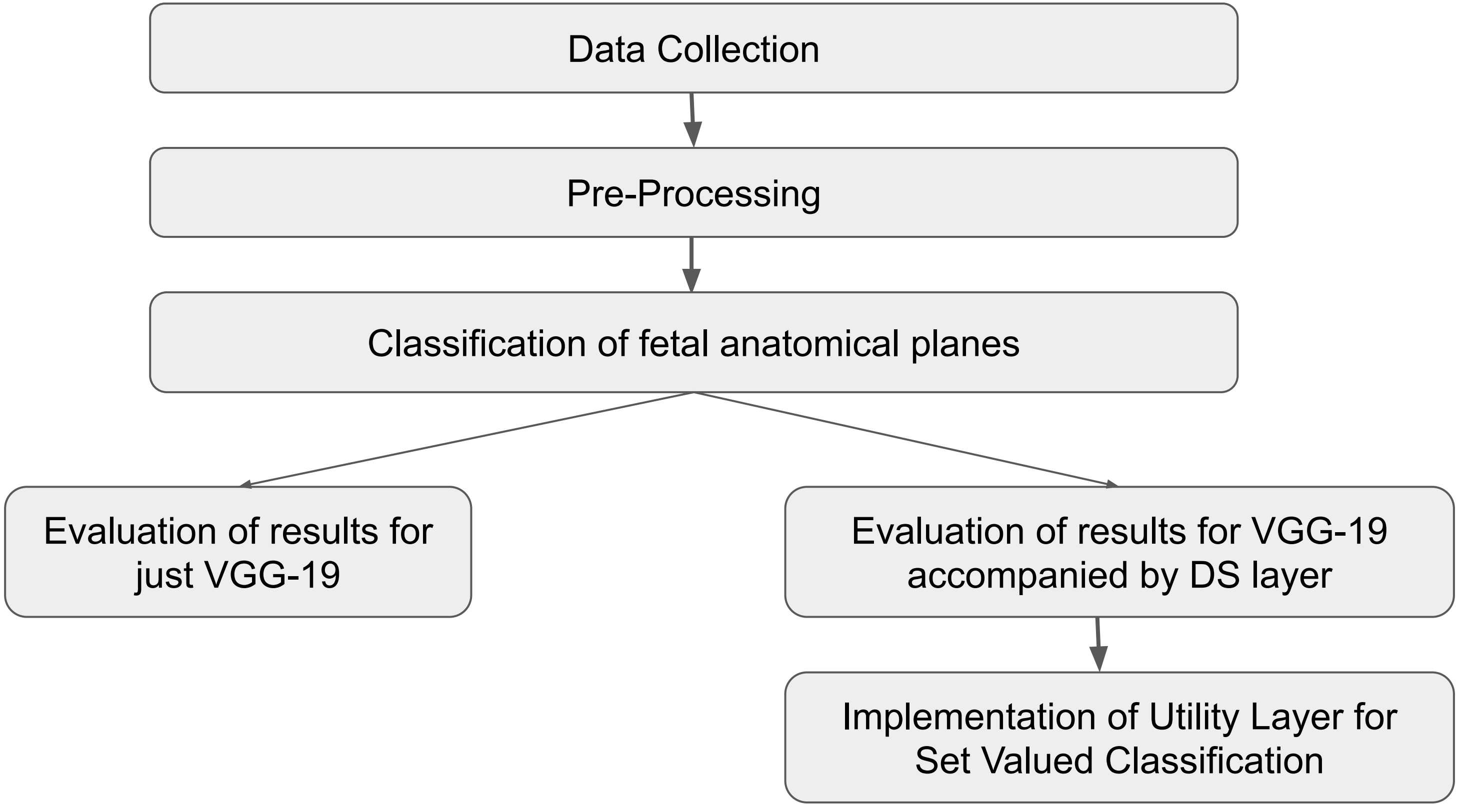
Limitations:

- No work done on fetal US images that combined CNN and DST to gain output
- Not much improvement in accuracy by using DST
- In DST based evidential classifier using FitNet-4, accuracy without DST was 85% and with DST, 85.7%.

Research Contribution

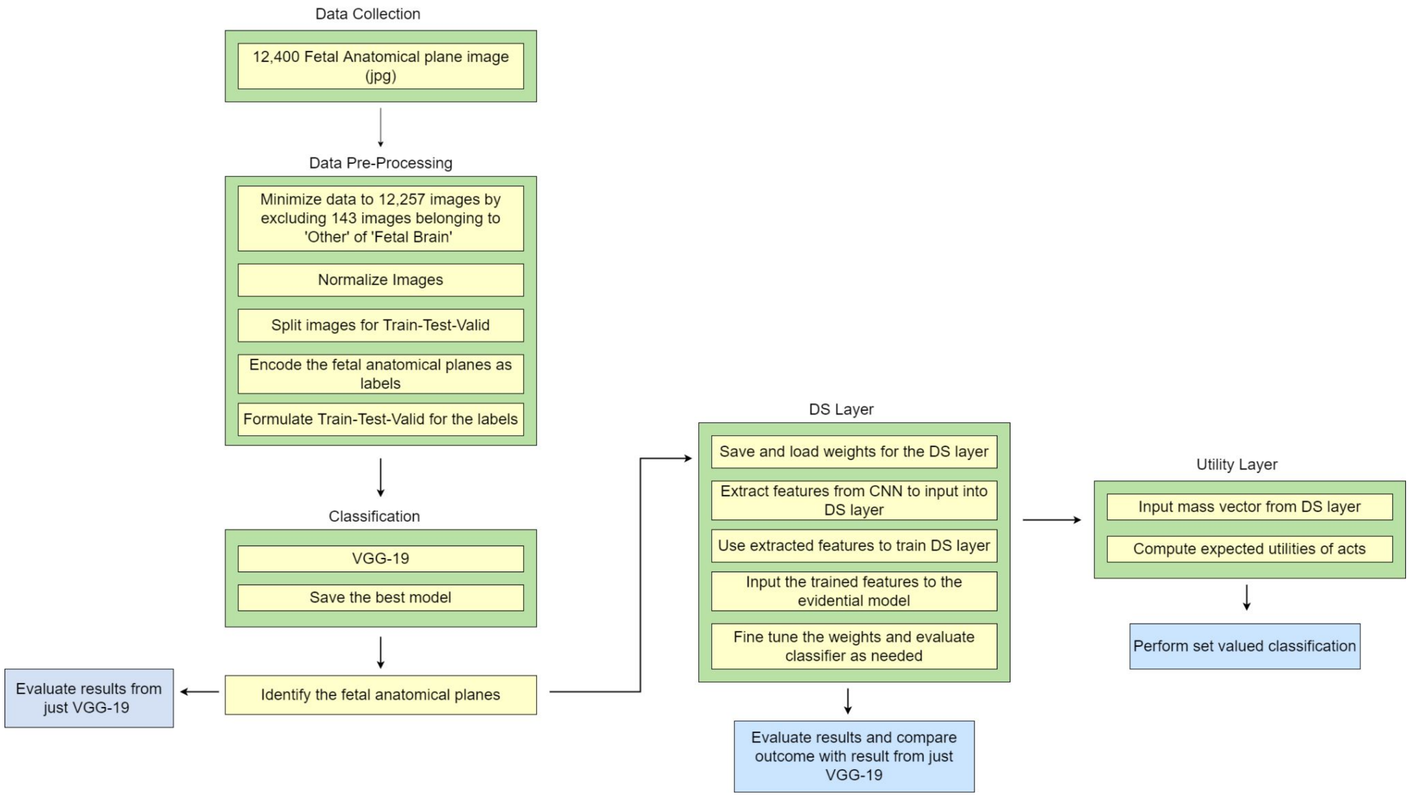
- We proposed an evidential classifier using an existing high performing CNN model, i.e. VGG-19 with DST to predict the different classes of common fetal anatomical planes as there are high possibilities of uncertainty in fetal ultrasound images
- Our proposed model gave us better accuracy and a significantly satisfactory result for cases of uncertainty .
- We noted the changes in accuracy and uncertainty in the data and model when VGG-19 was applied only and when a DS layer was applied over the CNN layer.
- Our proposed model gave us an utility layer which gives us more information about the prediction of the model in cases of high uncertainty and presence of outliers

Methodology



Methodology:

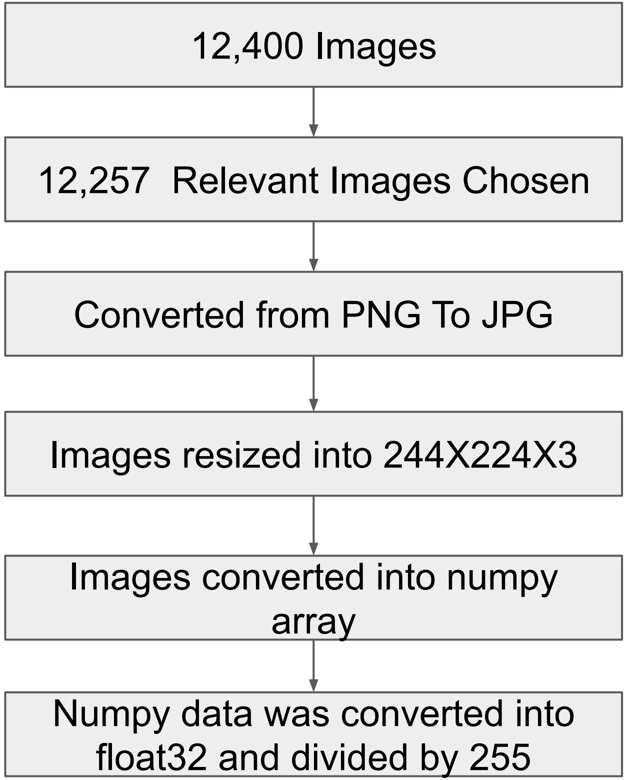
(Top Level Overview of Proposed Model)



Data Collection

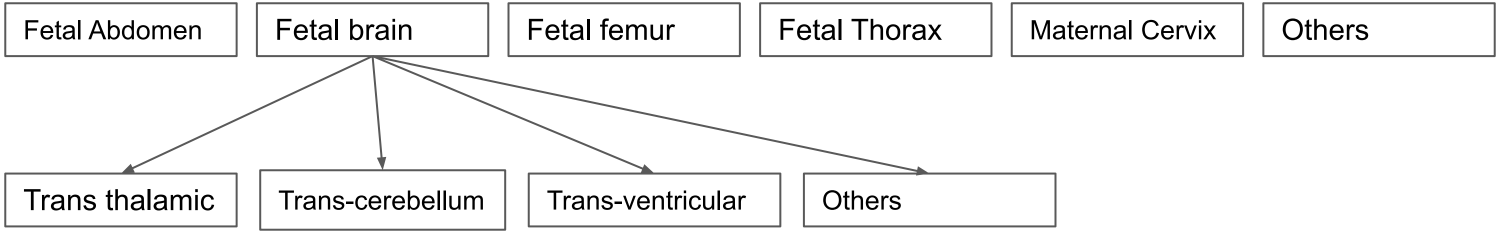
- The data was collected from a publicly available dataset consisting total 12400 images of 1792 patients from four hospitals in Spain.
- The machines that were used to collect these data are Voluson E6, Voluson S8, Voluson S10 and Aloka.
- The images were labeled according to its fetal plane
- Anonymized patient IDs were given for each image, where IDs are a consecutive 4 digit number ordering patients in ascending chronological order according to their first visit.

Image Pre-Processing



Data Pre-Processing

Initial classes of the fetal anatomical images:



For our analysis we have deleted the class “Fetal brain” and worked with its 3 subclasses. So in the end we ended up with the following 8 classes. This made the classification easier and more precise. We have also excluded the “Fetal Brain” and “Other” images

Fetal Abdomen | Trans thalamic | Trans-cerebellum | Trans-ventricular | Fetal femur | Fetal Thorax | Maternal Cervix | Others

We have used One hot encoding for these classes and considered them as labels.

Data Pre-Processing

(Number of images per class)

Anatomical Planes used for detection	Number of Images
Fetal abdomen	711
Trans-thalamic	1,638
Trans-cerebellum	714
Trans-ventricular	597
Fetal femur	1,040
Fetal thorax	1,718
Maternal cervix	1626
Other	4,213
Total	12,257

Model Specifications:

(VGG-19)

Layers of VGG-19

- 16 Convolution Layers, accompanied by ReLU activation functions
- 5 MaxPool Layers
- 3 Fully Connected Layers
- 1 Softmax Layers

- Number of filters/kernels, K
- Filter Size, F
- Stride, S
- Amount of Zero Padding applied, P

Parameters

Equations:

➢ After Convolution Layer, [1]

$$H_2 = \left(\frac{H_1 - F + 2P}{S}\right) + 1$$

where, W2 = H2

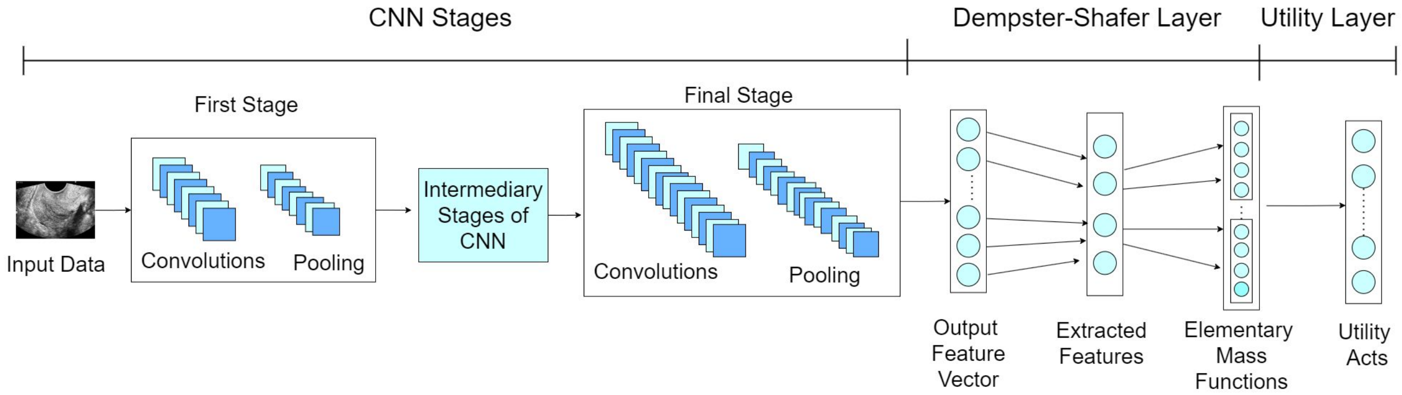
➢ After Pooling Layer,

$$W_2 = \left(\frac{W_1 - F}{S}\right) + 1$$

where, W2 = H2

Model Specifications:

(Dempster-Shafer based Evidential Classifier)



- Input data goes through multi stage CNN (VGG-19)
- Features extracted from CNN converted to elementary mass functions
- Output of DS layer: mass vector - measure of belief of a sample belonging to a class
- Utility layer computes expected utilities of acts and performs set valued classification

$$s^i = \alpha^i \exp(-\left(\eta^i d^i\right)^2) \quad i = 1, \ldots, n,$$

$$m^i(\{\omega_j\}) = h^i_j s^i, \quad j = 1, \ldots, M$$

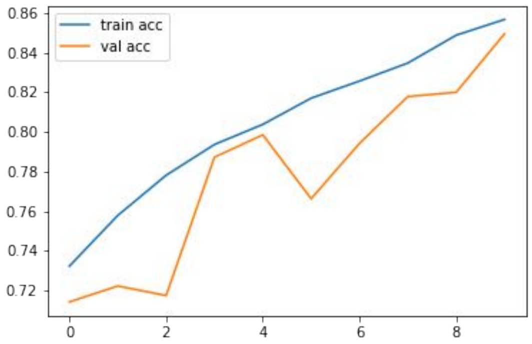
Results and Analysis:

Results from VGG-19:

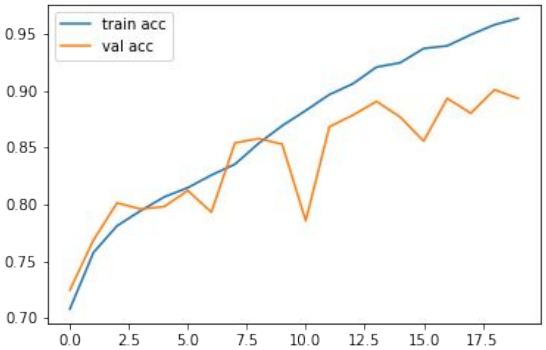
	precision	recall	f1-score	support
0	0.85	0.81	0.83	108
1	0.85	0.80	0.83	158
2	0.89	0.91	0.90	229
3	0.99	0.99	0.99	267
4	0.90	0.90	0.90	661
5	0.74	0.77	0.76	96
6	0.84	0.72	0.78	271
7	0.42	0.78	0.54	49
accuracy			0.87	1839
macro avg	0.81	0.84	0.82	1839
weighted avg	0.88	0.87	0.87	1839

Results and Analysis:

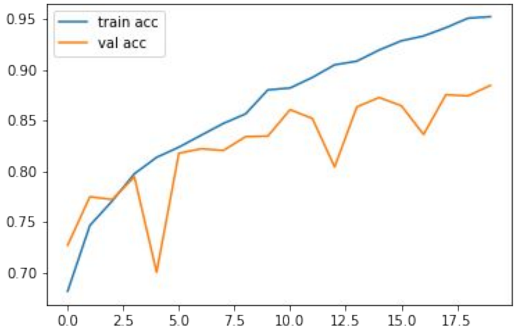
Accuracy Graphs Obtained for Different Prototypes:



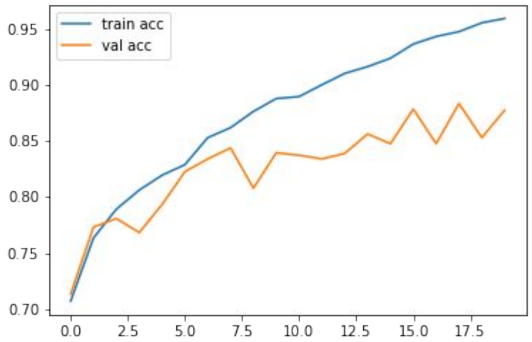
(20 Prototypes)



(30 Prototypes)



(50 Prototypes)



(100 Prototypes)

Results and Analysis:

Classification Reports Obtained for Different Prototypes:

	precision	recall	f1-score	support
0	0.57	0.88	0.69	67
1	0.74	0.86	0.79	138
2	0.94	0.80	0.87	319
3	0.99	0.98	0.99	240
4	0.90	0.89	0.89	632
5	0.05	0.60	0.10	10
6	0.89	0.61	0.72	363
7	0.60	0.71	0.65	70
accuracy			0.82	1839
macro avg	0.71	0.79	0.71	1839
weighted avg	0.88	0.82	0.84	1839

(20 Prototypes)

	precision	recall	f1-score	support
0	0.90	0.87	0.88	106
1	0.83	0.91	0.87	172
2	0.96	0.91	0.94	266
3	1.00	1.00	1.00	263
4	0.92	0.92	0.92	597
5	0.81	0.80	0.80	98
6	0.87	0.76	0.81	271
7	0.52	0.77	0.62	66
accuracy			0.89	1839
macro avg	0.85	0.87	0.86	1839
weighted avg	0.90	0.89	0.89	1839

(30 Prototypes)

	precision	recall	f1-score	support
0	0.92	0.82	0.87	119
1	0.75	0.92	0.83	134
2	0.94	0.94	0.94	247
3	1.00	0.98	0.99	241
4	0.92	0.90	0.91	636
5	0.59	0.84	0.69	80
6	0.86	0.72	0.78	299
7	0.61	0.72	0.66	83
accuracy			0.87	1839
macro avg	0.83	0.86	0.84	1839
weighted avg	0.88	0.87	0.88	1839

(50 Prototypes)

	precision	recall	f1-score	support
0	0.92	0.89	0.90	101
1	0.83	0.83	0.83	152
2	0.96	0.92	0.94	273
3	0.99	1.00	0.99	251
4	0.92	0.94	0.93	614
5	0.90	0.77	0.83	131
6	0.84	0.76	0.80	270
7	0.38	0.72	0.50	47
accuracy			0.89	1839
macro avg	0.84	0.85	0.84	1839
weighted avg	0.90	0.89	0.89	1839

(100 Prototypes)

Results and Analysis:

Performance Metrics:

- Accuracy
- Precision
- Recall
- F1 score
- Macro avg
- Weighted avg

Better Result:

Obtained with 30 prototypes

Results and Analysis:

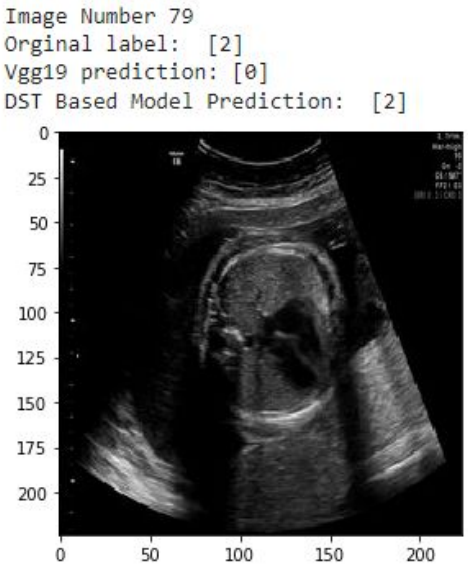
Performance Comparison:

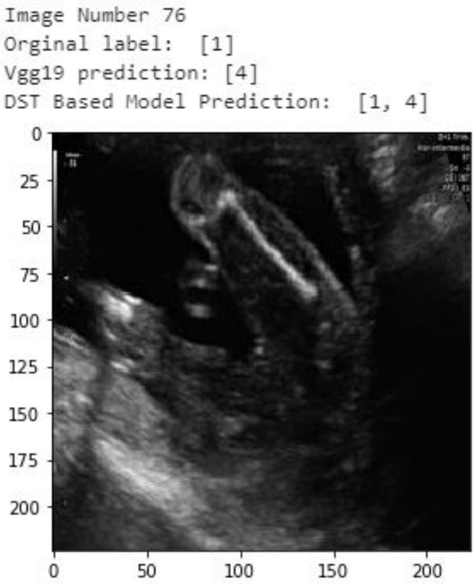
VGG-19 : 87% Accuracy

VGG-19 + DST (30 prototypes) : 89% Accuracy

Results and Analysis:

Comparative analysis between VGG-19 and VGG-19 + DST Prediction:





Results and Analysis:

Comparative analysis between VGG-19 and VGG-19 + DST Prediction:

Image Number 81
Orginal label: [6]
Vgg19 prediction: [7]
DST Based Model Prediction: [6, 7]

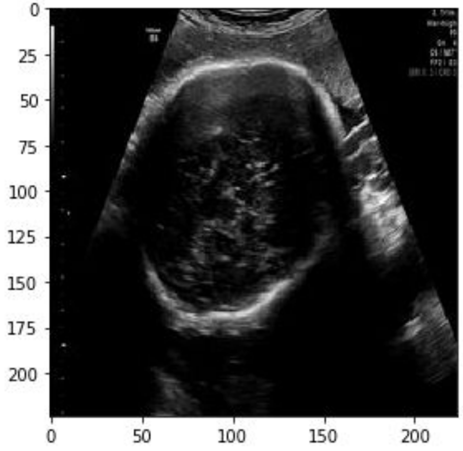
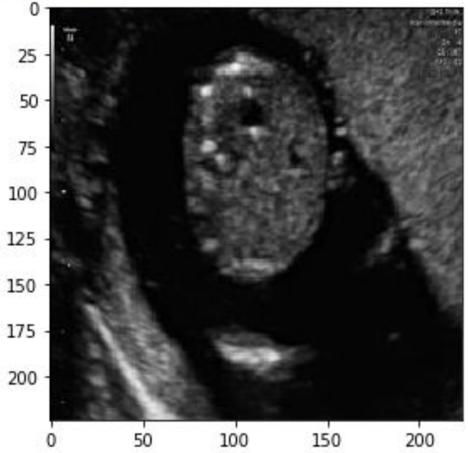


Image Number 22
Orginal label: [0]
Vgg19 prediction: [0]
DST Based Model Prediction: [0, 5]



Conclusion

From the above analysis, we came to the following conclusions:

- Comparative results of our proposed model, VGG-19 with DS layer, shows the uncertainty of predicting the label as opposed to just a prediction of VGG-19
- Better classification accuracy than just using VGG-19 solely.
- Set valued classification allowed us to not only demonstrate the uncertainty of both the result and the model, giving the attending physician better insight
- Aids in mitigating the misclassification of Ultrasound images

Based on the conclusions drawn, we want to work on the following things in the future:

- Use multiple models to create evidential fusion model with DST
- Implement XAI to better understand the predictions made by the model

Reference

1) B. Gao and L. Pavel, “On the properties of the softmax function with application in game theory and reinforcement learning,” arXiv preprint arXiv:1704.00805, 2017.

The better designed slides start from here, Sir.

Classification of Common Fetal Anatomical Planes from Ultrasound Imaging using Dempster Shafer Theory and Deep Learning

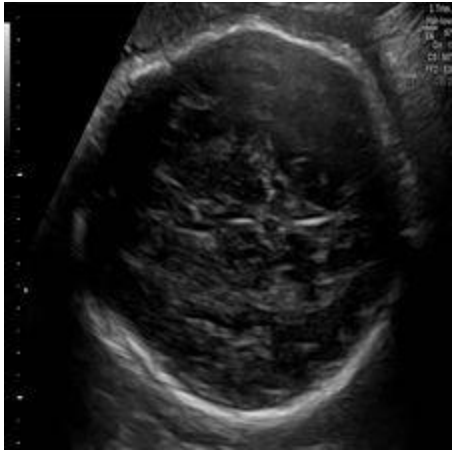
Members:

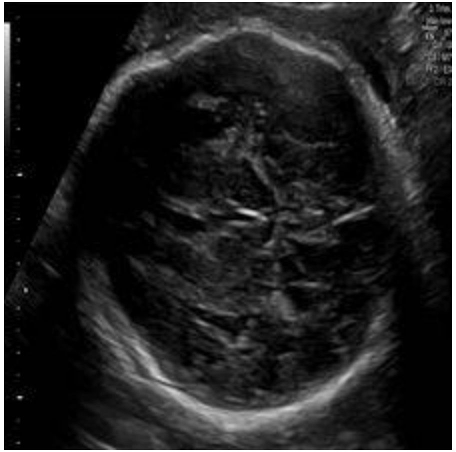
A.M. Tayeful Islam (19101107)
Marshia Nujhat (ID: 19101100)
Atanu Roy (ID: 19101267)
Ahmed Mayeesha Reza Agomoni (ID: 19101181)

Supervisor:

Dr. Md. Golam Rabiul Alam
Professor, CSE
School of Data and Sciences
Brac University

Fetal Anatomical Planes



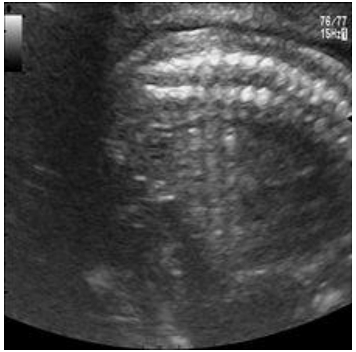


What will be the decision of the model?

Fetal Anatomical Planes

Uncertainty

Confusing



Multiple class

Outliers

Motivation

- Accurately classify the images into eight widely used fetal anatomical planes (Abdomen, Femur,Maternal cervix ,Trans-thalamic, Trans-cerebellum, Trans-ventricular , Thorax and Other)
- Reduce time and manpower required to classify these images.
- Classify images into a set of possible results in the cases of high noise and uncertainty.

Research Problem

- US images are hard to classify even with expert eyes due to speckle noise speckle noise, low signal-to-noise ratio, varying intensities of acoustic shadows, motion blurring, missing boundaries and inter-operator errors.
- Inaccurate results in the case of brain planes and uncertain situation
- Requires three stages of man power to classify these images

Existing Problem

- Common Fetal plane segmentation using deep CNN
- Evidential fusion model
- Evidential classifier using fitnet-classifier

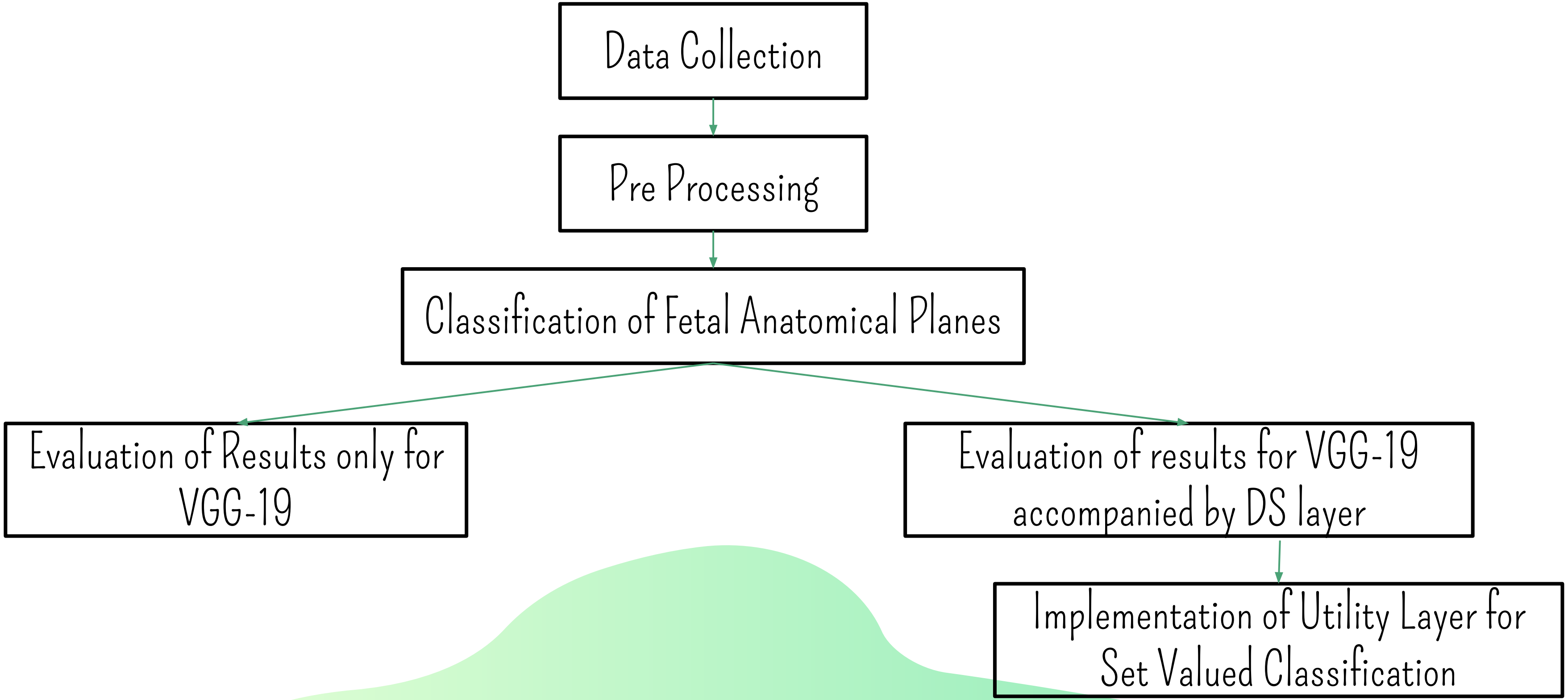
Limitations

- No work done on fetal US images that combined CNN and DST to gain output
- Not much improvement in accuracy by using DST
- In DST based evidential classifier using FitNet-4, accuracy without DST was 85% and with DST, 85.7%.

Research Contribution

- We proposed an evidential classifier using an existing high performing CNN model, i.e. VGG-19 with DST to predict the different classes of common fetal anatomical planes as there are high possibilities of uncertainty in fetal ultrasound images
- Our proposed model gave us better accuracy and a significantly satisfactory result for cases of uncertainty
- We noted the changes in accuracy and uncertainty in the data and model when VGG-19 was applied only and when a DS layer was applied over the CNN layer
- Our proposed model gave us an utility layer which gives us more information about the prediction of the model in cases of high uncertainty and presence of outliers

Methodologies



Data Collection

- The data was collected from a publicly available dataset consisting total 12400 images of 1792 patients from four hospitals in Spain.
- The machines that were used to collect these data are Voluson E6, Voluson S8, Voluson S10 and Aloka.
- The images were labeled according to its fetal plane
- Anonymized patient IDs were given for each image, where IDs are a consecutive 4 digit number ordering patients in ascending chronological order according to their first visit.

Image Pre-Processing

12,400 Images

↓

12,257 Relevant Image Chosen

↓

Converted Images to PNG format

↓

Images resized into 244x224x3

↓

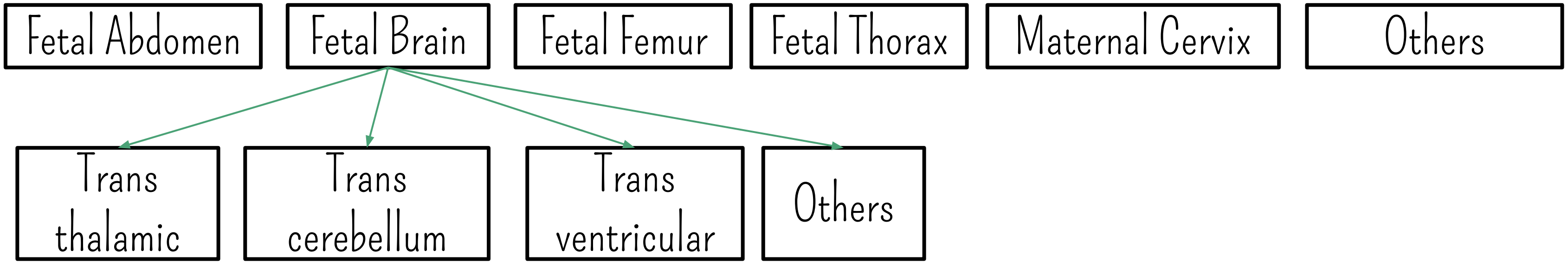
Images converted into numpy array

↓

Numpy data was converted into float32 and divided by 255

Data Pre-Processing

Initial classes of the fetal anatomical images:



For our analysis we have deleted the class "Fetal brain" and worked with its 3 subclasses. So in the end we ended up with the following 8 classes. This made the classification easier and more precise. We have also excluded the "Fetal Brain" and "Other" images

Data Pre-Processing

For our analysis we have deleted the class "Fetal brain" and worked with its 3 subclasses. So in the end we ended up with the following 8 classes. This made the classification easier and more precise. We have also excluded the "Fetal Brain" and "Other" images

Fetal Abdomen	Trans thalamic	Trans cerebellum	Trans ventricular	Fetal Femur	Fetal Thorax	Maternal Cervix	Others

We have used One hot encoding for these classes and considered them as labels.

Data Pre-Processing

Class distribution of the dataset

Anatomical Planes used for detection	Number of Images
Fetal abdomen	711
Trans-thalamic	1,638
Trans-cerebellum	714
Trans-ventricular	597
Fetal femur	1,040
Fetal thorax	1,718
Maternal cervix	1626
Other	4,213
Total	12,257

Model Specification (VGG-19)

Layers of VGG-19

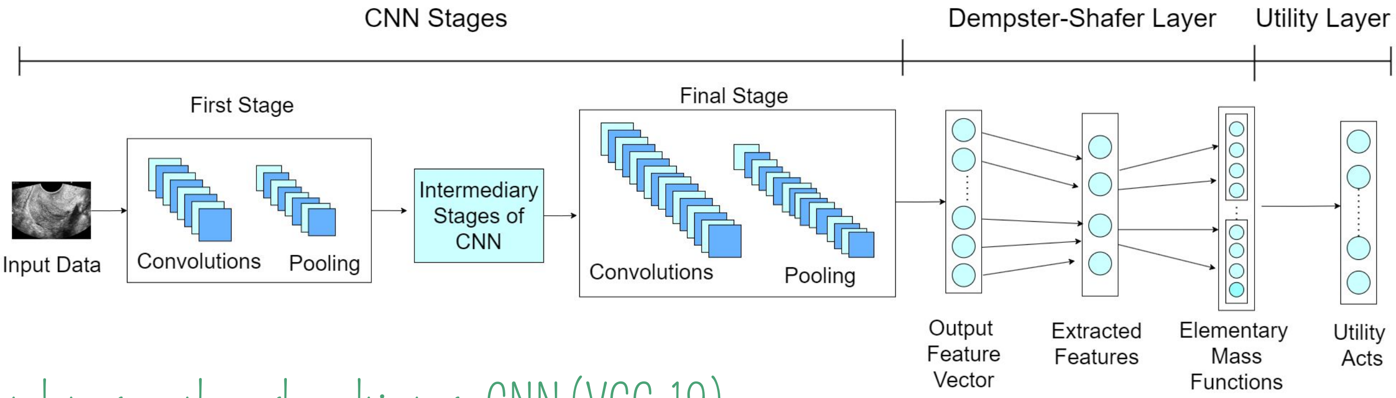
- 16 Convolution Layers, accompanied by ReLU activation functions
- 5 MaxPool Layers
- 3 Fully Connected Layers
- 1 Softmax Layers

Parameters

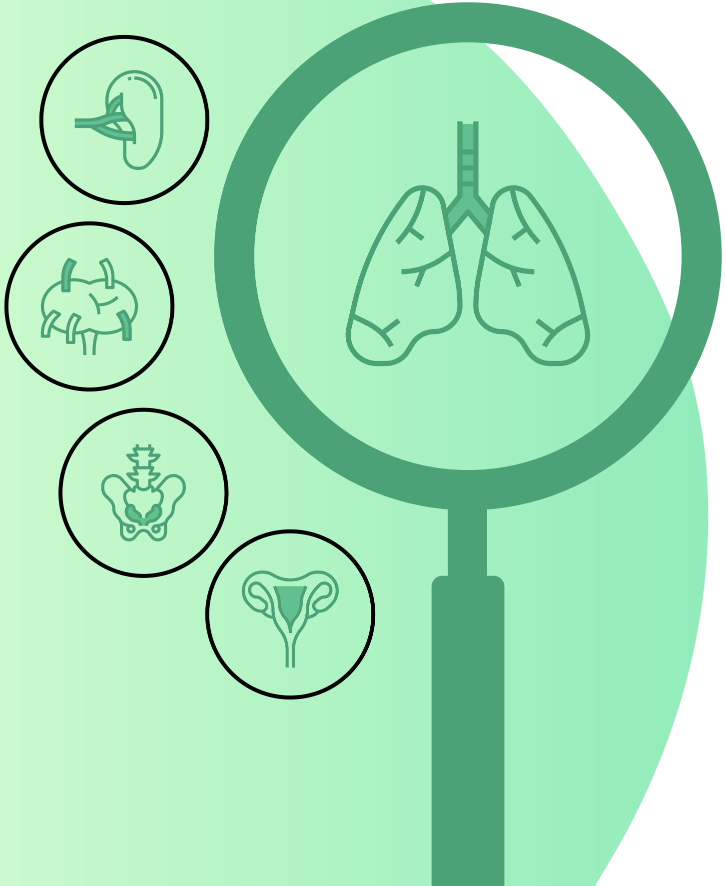
- Number of filters/kernels, K
- Filter Size, F
- Stride, S
- Amount of Zero Padding applied, P

Model Specification

(DST Based Evidential Classifier)



- Input data goes through multi stage CNN (VGG-19)
- Features extracted from CNN converted to elementary mass functions
- Output of DS layer: mass vector - measure of belief of a sample belonging to a class
- Utility layer computes expected utilities of acts and performs set valued classification

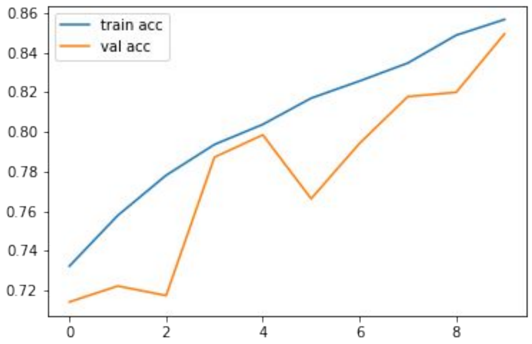


Results & Analysis

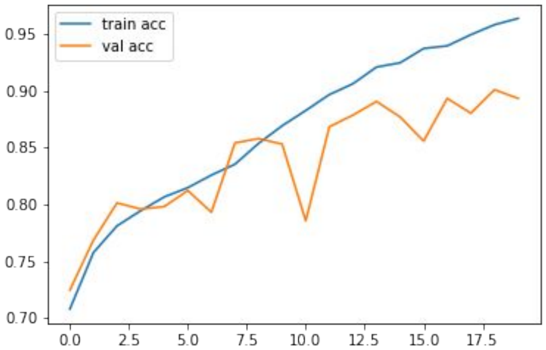
Results from VGG-19

	precision	recall	f1-score	support
0	0.85	0.81	0.83	108
1	0.85	0.80	0.83	158
2	0.89	0.91	0.90	229
3	0.99	0.99	0.99	267
4	0.90	0.90	0.90	661
5	0.74	0.77	0.76	96
6	0.84	0.72	0.78	271
7	0.42	0.78	0.54	49
accuracy			0.87	1839
macro avg	0.81	0.84	0.82	1839
weighted avg	0.88	0.87	0.87	1839

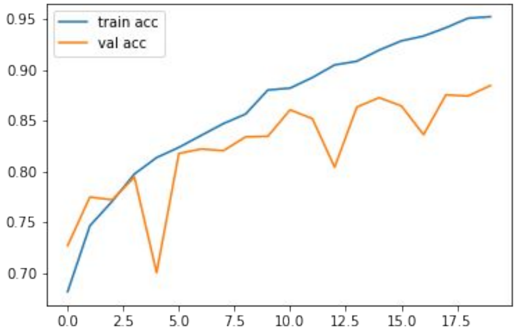
Accuracy Graphs Obtained for Different Prototypes



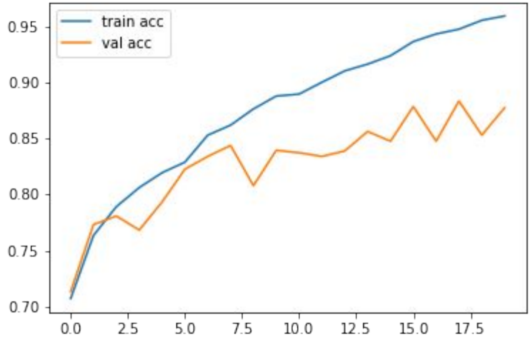
(20 Prototypes)



(30 Prototypes)



(50 Prototypes)



(100 Prototypes)

Classification Reports Obtained for Different Prototypes

	precision	recall	f1-score	support
0	0.57	0.88	0.69	67
1	0.74	0.86	0.79	138
2	0.94	0.80	0.87	319
3	0.99	0.98	0.99	240
4	0.90	0.89	0.89	632
5	0.05	0.60	0.10	10
6	0.89	0.61	0.72	363
7	0.60	0.71	0.65	70
accuracy			0.82	1839
macro avg	0.71	0.79	0.71	1839
weighted avg	0.88	0.82	0.84	1839

(20 Prototypes)

	precision	recall	f1-score	support
0	0.90	0.87	0.88	106
1	0.83	0.91	0.87	172
2	0.96	0.91	0.94	266
3	1.00	1.00	1.00	263
4	0.92	0.92	0.92	597
5	0.81	0.80	0.80	98
6	0.87	0.76	0.81	271
7	0.52	0.77	0.62	66
accuracy			0.89	1839
macro avg	0.85	0.87	0.86	1839
weighted avg	0.90	0.89	0.89	1839

(30 Prototypes)

	precision	recall	f1-score	support
0	0.92	0.82	0.87	119
1	0.75	0.92	0.83	134
2	0.94	0.94	0.94	247
3	1.00	0.98	0.99	241
4	0.92	0.90	0.91	636
5	0.59	0.84	0.69	80
6	0.86	0.72	0.78	299
7	0.61	0.72	0.66	83
accuracy			0.87	1839
macro avg	0.83	0.86	0.84	1839
weighted avg	0.88	0.87	0.88	1839

(50 Prototypes)

	precision	recall	f1-score	support
0	0.92	0.89	0.90	101
1	0.83	0.83	0.83	152
2	0.96	0.92	0.94	273
3	0.99	1.00	0.99	251
4	0.92	0.94	0.93	614
5	0.90	0.77	0.83	131
6	0.84	0.76	0.80	270
7	0.38	0.72	0.50	47
accuracy			0.89	1839
macro avg	0.84	0.85	0.84	1839
weighted avg	0.90	0.89	0.89	1839

(100 Prototypes)

Reason for working with 30 prototypes:

Performance Metrics:

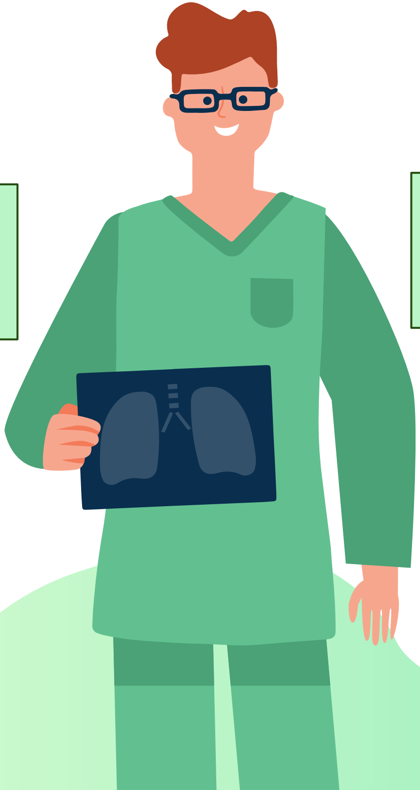
- Accuracy
- Precision
- Recall
- F1 score
- Macro avg
- Weighted avg

Better Result:

Obtained with 30 prototypes

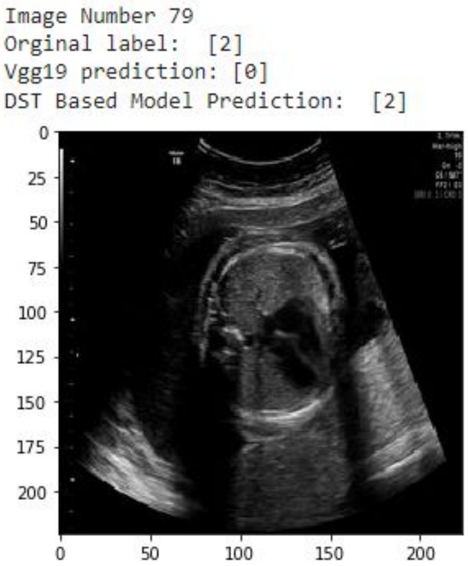
Performance Comparison

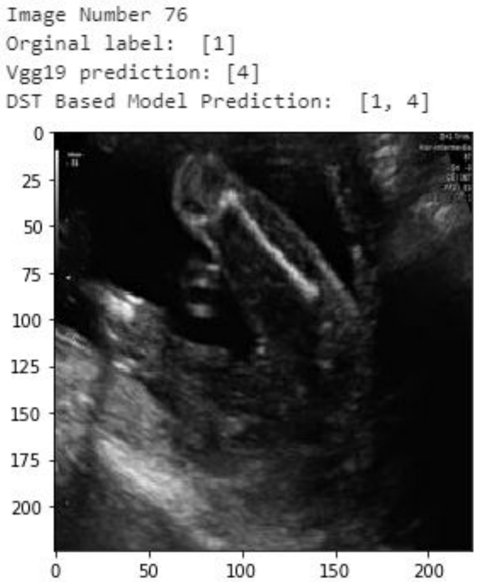
VGG-19:
87% Accuracy



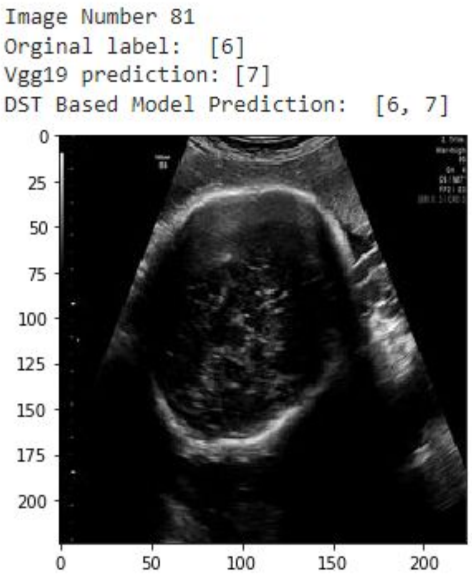
VGG-19 + DST (30 prototypes):
89% Accuracy

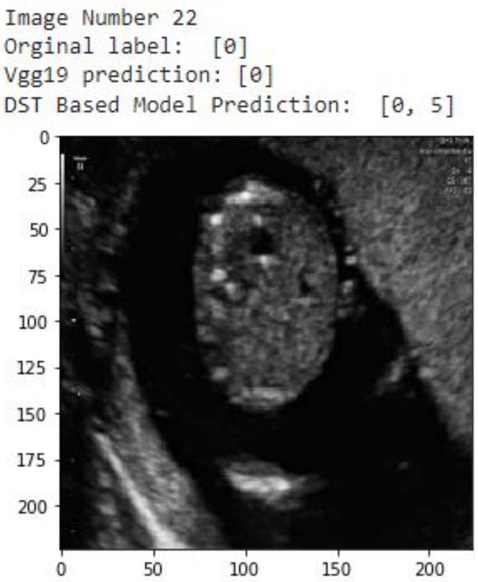
Comparative analysis between VGG-19 and VGG-19 + DST Prediction





Comparative analysis between VGG-19 and VGG-19 + DST Prediction





Conclusion

From the above analysis, we came to the following conclusions:

- Comparative results of our proposed model, VGG-19 with DS layer, shows the uncertainty of predicting the label as opposed to just a prediction of VGG-19
- Better classification accuracy than just using VGG-19 solely.
- Set valued classification allowed us to not only demonstrate the uncertainty of both the result and the model, giving the attending physician better insight
- Aids in mitigating the misclassification of Ultrasound images

Based on the conclusions drawn, we want to work on the following things in the future:

- Use multiple models to create evidential fusion model with DST
- Implement XAI to better understand the predictions made by the model

References

- B. Gao and L. Pavel, “On the properties of the softmax function with application in game theory and reinforcement learning,” arXiv preprint arXiv:1704.00805, 2017.


Any
Questions?