Active Learning and CNNs for Reliable Detection of Fetal Intracranial Structures

Ana Carolina Morais[1] , **Fernanda Fernandes**[1]
[1]Department of Informatics Engineering, University of Coimbra
{anamorais,mrfernandes}@student.dei.uc.pt

Abstract

Misdiagnosis is a concern, in the field of healthcare as it can have impacts on patient outcomes and delay treatment procedures significantly. This project is focused on improving decision making by creating an Active Learning system that supports healthcare providers in identifying intracranial structures between weeks 11 to 14 of gestation. By utilizing a dataset consisting of 1,528 ultrasound images with a view the model makes use of a neural network (CNN) classifier to precisely recognize nine vital anatomical features such, as thalamus, midbrain and nasal bone. By incorporating a learning framework, into the systems operations in a step by step manner and with human involvement required to enhance its performance gradually over time compared to conventional passive learning methods will reduce the amount of data needed for training purposes.The findings reveal that the suggested system does not uphold a level of accuracy but also simplifies the diagnostic procedure by offering immediate assistance, to medical professionals and potentially lessening the risks linked to delayed diagnoses. This study emphasizes the impact that AI can have on diagnoses and underscores the necessity of creating efficient machine learning models to improve patient care effectively.

Keywords: Active Learning, CNNs, Fetal Health, Intracranial Structures, Ultrasound Imaging, Neural Networks, Healthcare Diagnostics, Artificial Intelligence, Deep Learning, modAL

1 Introduction

Advanced diagnostics in fetal healthcare face significant challenges, as errors in prenatal ultrasound interpretation can lead to delayed treatments and adverse outcomes for both mother and fetus [Leiserowitz and Herding, 2020]. With technology playing an increasingly vital role in obstetrics, precise identification of fetal structures has become critical [Yousefpour Shahrivar *et al.*, 2023].

This study introduces an Active Learning system [Pinto *et al.*, 2022] to aid healthcare professionals in identifying intracranial structures during the 11-to-14-week gestational period. Building on advances in ultrasound imaging and neural network-based systems, the framework leverages human expertise to efficiently use existing data and enhance diagnostic accuracy while addressing the challenges of large-scale neural network training [SciDev.Net, 2024].

Using a dataset of 1,528 ultrasound images, the system focuses on recognizing nine key anatomical features, such as the thalamus, midbrain, and nasal bone, which are critical for fetal development and improving prenatal care. By combining cutting-edge machine learning techniques with clinical expertise, the framework simplifies fetal structure identification, advancing the application of AI in healthcare diagnostics.

This paper's remaining sections are organized as follows: Materials, Methods, Results, Discussion, and Conclusion. The dataset, the preprocessing steps, and the tool utilized will all be covered in section 3. We will outline the plan to construct the model we suggest in section 4. The section 5 will present the outcomes derived from our model, and Section 6 will assess the results' level of significance. Section 7 will conclude with a summary of the work completed and some remarks.

2 Related Work

The application of artificial intelligence in medical diagnostics has grown significantly in recent years, demonstrating its potential to reduce diagnostic errors and improve patient outcomes. For example, [Jian *et al.*, 2021] proposes the use of multi scale feature concatenation networks and a simplified framework for detection and localization of myocardial infarction using ECG data, addressing challenges such as overfitting and underfitting, demonstrating improved performance compared to state of the art methods. [Safdar *et al.*, 2018] conducted a comprehensive review of machine learning based decision support systems for heart disease diagnosis, highlighting the effectiveness of various algorithms such as artificial neural networks and support vector machines in clinical applications, as always, noted the critical need for real time clinical data to enhance the training and accuracy of these systems in practical healthcare settings. These works emphasize the utility of machine learning in clinical settings, although they rely heavily on extensive labeled datasets, a challenge that limits scalability in resource constrained environments.

Active Learning has emerged as a promising solution to reduce the need for large labeled datasets while maintaining high model accuracy.[Mahapatra *et al.*, 2019] proposes an in novative medical image super resolution method using progressive generative adversarial networks, highlighting the effectiveness of the approach in improving the segmentation of vascular structures and micro aneurysms in retinal background images, which complement advances in image processing techniques in the area of medical analysis [Zhang *et al.*, 2021] developed a system Deep learning based image classification specifically designed for detecting retinal lesions and skin abnormalities. Their approach not only reduced the labeling requirements by 35% through the implementation of advanced convolution al neural network techniques but also achieved an accuracy that is comparable to traditional diagnostic methods, highlighting the potential of artificial intelligence to enhance diagnostic efficiency and accuracy in ophthalmology and dermatology .These findings highlight the effectiveness of AL in medical imaging, making it a viable alternative to traditional learning frameworks.

Despite its success in various medical applications, the use of Active Learning in fetal imaging remains sparsely explored. Most studies in this domain rely on passive learning methods. [Garcia-Canadilla *et al.*, 2020] highlights the potential of machine learning in fetal cardiology, demonstrating how these technologies can optimize image acquisition and improve the diagnosis of cardiac anomalies in fetuses, which complement our research on the application of artificial intelligence techniques in assessing neonatal conditions.In the context of prenatal assessment, [Phung *et al.*, 2023] developed a convolution neural network model that uses a feature rearrangement approach to predict Down syndrome, achieving superior results in terms of precision and recall compared to traditional methods, which highlights the effectiveness of artificial intelligence in improving of prenatal diagnosis. The need for more data efficient techniques is evident, especially for tasks requiring precise identification of anatomical features, such as intracranial structures, during early gestation.

Very recently [Sun *et al.*, 2024] developed a new model to measure fetal intracranial markers during the trimester, demonstrating strong consistency and correlation with manual measurements. Their work established normal reference ranges for key intracranial markers and highlighted the potential of AI to streamline sonographer tasks, achieving a measurement time significantly faster than traditional methods.

3 Materials

Database

The dataset[1] for this project consists of 1528 2D sagittal-view ultrasound images collected from Shenzhen People's Hospital. However, for training and validation purposes, we selected a curated subset of 810 images stored in the folder Set1-Training&Validation Sets CNN\Standard. This subset represents a standard plane classification set critical for our model's initial development. Additionally, we leveraged images from the folder Set2-Training&Validation Sets ANN Scoring system\Standard, which, despite the folder name not specifically indicating CNN usage, were identified as relevant through the provided ObjectDetection.xlsx file. These annotations specify the bounding box coordinates $(h_{\min}, w_{\min}, h_{\max}, w_{\max})$ and the structure label for each image. During preprocessing, each image in the selected subsets is cropped using the corresponding bounding box, resized to 224×224 pixels, normalized to $[0, 1]$ scale, and rearranged to match the expected format $[C, H, W]$ required by the PyTorch deep learning framework.

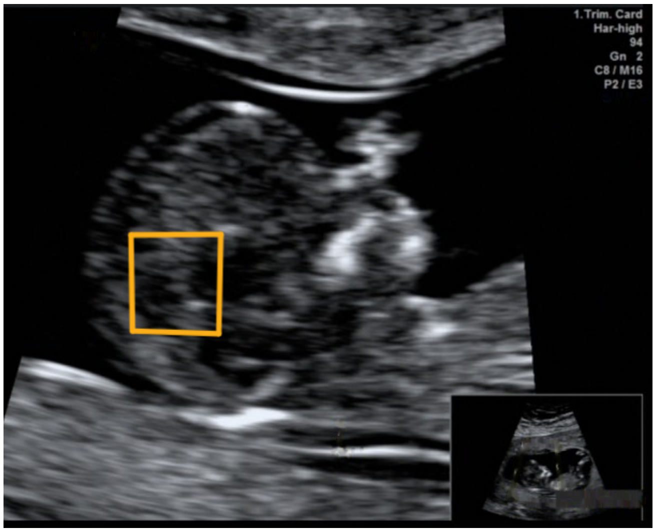

Figure 1: Example of a dataset image with the cutout

To test the model performance, additional images from the folders Internal Test Set and External Test Set are used. These datasets provide a separate evaluation set to validate the robustness and generalization of the trained model.

4 Methodos

This section outlines the approach used to address the problem of accurately identifying intracranial fetal structures in ultrasound images. The proposed system integrates a Convolutional Neural Network (CNN) for feature extraction and classification with an Active Learning framework to iteratively improve model performance while reducing the labeling effort required by domain experts.

4.1 Preprocessing

The preprocessing phase is essential to prepare the ultrasound images before feeding them into the Convolutional Neural Network. The process includes several critical steps to ensure that the data is in the correct format and ready for training.

[1] https://data.mendeley.com/datasets/n2rbrb9t4f/1

First, the images are loaded from the specified directory and each image is cropped according to predefined coordinates. These coordinates are retrieved from the metadata, which indicates the region of interest (ROI) of the image. Cropping is necessary to focus the model's attention on the relevant structures in the image, such as the thalamus or midbrain, while removing any unnecessary parts of the image. This step ensures that CNN can learn to identify only the most critical features.

Next, each image is resized to a fixed dimension of 224×224 pixels. This resizing is crucial because the model requires a consistent input size for all images to process them efficiently. Without resizing, the model would struggle to handle images of varying sizes, potentially leading to errors during training.

After resizing, the images are normalized by scaling the pixel values to the range [0, 1]. This step is essential because neural networks work better when input values are on a similar scale, which helps the optimization process and accelerates convergence during training. By dividing the pixel values by 255, we ensure that the images are scaled appropriately for the network.

Following normalization, the channels of each image are reordered from the default [height, width, channels] format to [channels, height, width], as required by most CNN frameworks. This reordering is necessary because the network expects input in a specific format, and changing the order ensures that the model can correctly process the image.

Finally, each image is assigned a label corresponding to the anatomical structure it contains. The labels are retrieved from the metadata and are used to train the model to classify the different structures in the fetal ultrasound images. Label assignment is a crucial step for supervised learning, as it provides the model with the correct target for each input image.

These preprocessing steps ensure that the dataset is properly formatted and ready for input into the CNN, enabling efficient and accurate model training.

4.2 Convolutional Neural Network (CNN)

The Convolutional Neural Network (CNN) serves as the backbone of the model for identifying the fetal anatomical structures from ultrasound images [Sarvamangala and Kulkarni, 2022]. CNNs are particularly effective in image classification tasks due to their ability to automatically learn hierarchical features from data, such as edges, textures, and complex patterns.

In our implementation, the CNN consists of several layers of convolutional filters, each followed by ReLU activation functions to introduce non-linearity and enable the network to learn more complex representations. After each convolutional layer, a max-pooling operation is applied to reduce the spatial dimensions of the feature maps, making the network more computationally efficient while preserving essential information. The model also includes dropout layers to prevent overfitting by randomly deactivating a fraction of neurons during training.

The final part of the network consists of fully connected layers, which take the extracted features and map them to the target classes (nine fetal anatomical structures). The network is trained using the cross-entropy loss function, and optimization is performed using the Adam optimizer, which adapts the learning rate based on the gradients during training.

4.3 Active Learning

Active Learning is an advanced machine learning paradigm that optimizes the model's learning process by selectively querying the most informative data points for labeling. Instead of relying on a large, fully labeled dataset, Active Learning seeks to minimize the amount of labeled data required while maintaining or even improving performance. The framework used in this study is modAL, a Python-based library built on top of scikit-learn, which provides flexibility for creating customized Active Learning workflows. modAL allows the seamless integration of Active Learning strategies with existing models, making it ideal for applications in medical image analysis where labeling can be time-consuming and costly. More information about Active Learning in [Wang *et al.*, 2024].

To implement Active Learning, we used modAL's ActiveLearner class with a Convolutional Neural Network classifier. The classifier was optimized using the Adam optimizer and trained with the CrossEntropyLoss function, suitable for multi-class classification tasks. Several hyperparameters, such as the query strategy, number of epochs for both the learner and the Active Learning loop, and the number of instances and queries, were tuned through experimentation to achieve the best performance.

Active Learning allows the model to select the most informative data points for training, which is particularly useful when there is limited labeled data. The system starts with a small labeled set and then queries the most uncertain instances from an unlabeled pool. These uncertain instances are considered the most valuable for improving the model's performance. In our implementation, two query strategies were tested: uncertainty sampling, where the model selects the most uncertain examples, and margin sampling, which focuses on instances with the smallest margin between the two most probable class predictions. Once labeled by an expert, these instances are added to the training set, and the model is retrained, refining its predictions with fewer labeled examples.

4.4 Metrics

As mentioned in earlier sections, a number of measures will be used to assess the effectiveness of the suggested model, including accuracy, sensitivity, specificity, positive and negative predictive value, and the area under the Receiver Operating Characteristic Curve. These indicators are crucial for evaluating the model's performance, especially when it comes to medical diagnostics, where misclassification might have serious consequences.

Accuracy The percentage of accurate forecasts both true positives and true negatives among all predictions is known as accuracy. It offers a broad indicator of the model's performance across all classes. However, because accuracy might be distorted by the dominant classes, it could not accurately represent the model's performance in imbalanced datasets.

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

Sensitivity Sensitivity, or Recall, also known as sensitivity, gauges how well the model can detect positive examples. Sensitivity is important in medical applications since it shows how well the model can detect diseases or abnormalities while reducing false negatives. Maximizing sensitivity is crucial in healthcare since a missed or delayed diagnosis could have serious repercussions.

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

Specificity Specificity, or the true negative rate, measures the model's ability to correctly identify negative instances. In the medical context, specificity ensures that healthy individuals are not incorrectly diagnosed, avoiding unnecessary and potentially harmful treatments. While specificity may not be as crucial as sensitivity in some situations, it still plays an important role in preventing false positives.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Area Under the ROC Curve The AUC measures the model's ability to distinguish between classes across all thresholds. It is particularly useful for evaluating models in multi-class problems, as it provides an overall assessment of performance. AUC values range from 0 to 1, with higher values indicating better discrimination between classes.

Since this is a multi-class problem with an imbalanced dataset, the metrics will be calculated using micro-averaging. Micro-averaging combines the contributions of all classes to compute a single averaged performance value. This approach is particularly useful when classes are not equally represented, ensuring that the performance of all classes is taken into account.

4.5 Experimental Setup

The experiments conducted in this study aimed to evaluate the performance of a Convolutional Neural Network (CNN) model integrated with Active Learning techniques. The goal was to assess how different configurations of the model, including variations in hyperparameters and Active Learning strategies, impact the model's performance in classifying fetal anatomical structures from ultrasound images.

Experimental Design A series of experiments were designed to evaluate the effect of different configurations on the model's performance. Each experiment involved varying key parameters, such as the number of layers in the CNN, the number of epochs for both the learner and Active Learning loop, the query strategy used, and the number of instances to be labeled during each iteration. The experiments were conducted using two primary Active Learning strategies: *margin sampling* and *uncertainty sampling*, with an additional test on *entropy sampling*.

5 Results

Table 1 summarizes the key outcomes of our experiments, including variations in model architecture, training epochs, sampling strategies, and resulting accuracies. After some test we decided to use a dropout rate of 0.6, a learning rate of 1e-4, and a weight decay of 1e-4 to mitigate the risk of overfitting, stabilize the optimization process, and enhance model generalization, particularly given the simple single-layer architecture and the limited size of the dataset. Among the tested configurations, ID 6 achieved the highest test accuracy of 84.91%, demonstrating a strong balance between model simplicity and performance. This configuration utilized a single-layer neural network trained for 40 epochs within the learner and 20 epochs in the Active Learning loop, employing uncertainty sampling as its sampling strategy.

Other notable configurations include ID 2 (test accuracy: 78.40%) and ID 10 (test accuracy: 79.56%), both of which employed a two-layer neural network with comparable sampling strategies but did not achieve the same level of generalization as ID 6. Configurations with deeper architectures (e.g., IDs 3 and 7) exhibited diminished performance, highlighting the risk of overfitting with excessive model complexity and limited training instances.

Additionally, the results from experiments shown in Table 2, further validate the findings from Table 1. Despite adjustments to these hyperparameters to stabilize training and reduce overfitting, the performance improvements were marginal compared to the results from Table 1. This suggests that the ID 6 configuration remains the most effective, even with additional tuning.

Further validating the effectiveness of ID 6, Table 3 highlights diagnostic performance metrics across various anatomical regions. Notably, the model demonstrated strong overall reliability with micro-averaged Sensitivity and Specificity of 81.13% and 97.64%, respectively, and an overall accuracy of 95.81%. These metrics confirm that ID 6 is not only effective in generalizing across different structures, but also robust in distinguishing positive and negative cases. Specific highlights include:

The Cisterna Magna achieving an impressive accuracy of 97.80%, supported by high Sensitivity (0.92) and Specificity (0.983). The Nuchal Translucency (NT) achieving an accuracy of 98.43%, reflecting its ability to reliably detect this feature while minimizing false positives and negatives. Near-perfect performance in regions like the Nasal Tip (97.48% accuracy) and Nasal Bone (96.85% accuracy), indicating strong adaptability of ID 6 for features with diverse challenges. The Palate, while achieving perfect Specificity and PPV, had a lower Sensitivity (0.537), highlighting opportunities to enhance the model's recall for certain conditions.

As shown in Figure 2, the incremental classification accuracy improves steadily with each query iteration, reflecting the benefits of uncertainty sampling in active learning. The training accuracy converges slightly faster than test accuracy, suggesting effective learning from queried instances without overfitting. This trend further validates the robustness of the ID 6 configuration across the active learning process.

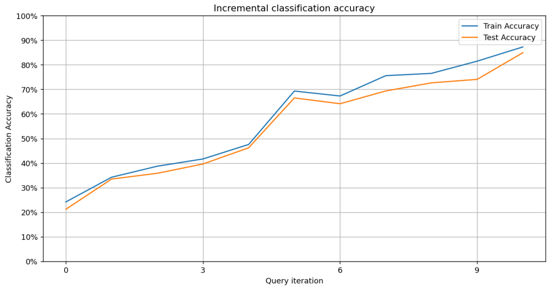

Figure 2: Incremental classification accuracy over query iterations.

6 Discussion

The superior performance of ID 6 can be attributed to its optimal trade-off between architectural simplicity, training epochs, and sampling strategy. Specifically: The single-layer network in ID 6 minimized the risk of overfitting, which was evident in deeper models (e.g., IDs 3 and 7). Simpler architectures proved effective for the relatively small dataset used in our study, avoiding unnecessary complexity that could lead to poor generalization. The choice of 40 epochs for the learner and 20 epochs in the AL loop ensured sufficient training without overextending the process. Configurations with fewer epochs (e.g., IDs 3 and 11) struggled to generalize effectively, while those with excessive epochs (e.g., ID 4) displayed signs of overfitting. These findings indicate the importance of balancing training duration to optimize performance. The use of uncertainty sampling in ID 6 allowed the model to prioritize instances that maximized learning potential, reducing data redundancy. This strategy outperformed alternatives such as entropy sampling (e.g., IDs 1 and 4) and margin sampling (e.g., IDs 7 and 10), which demonstrated lower efficiencies in leveraging the training data. The experiments conducted with a learning rate of 1e-4,dropout rate of 0.6, and weight decay of 1e-4, were aimed at stabilizing the optimization process and reducing overfitting. These parameters ensure that the model remains robust during training. However, the limited impact on performance, as observed in Table 2, indicates that the architectural simplicity and sampling strategy in ID 6 were the key contributors to its success. While the experiments summarized in Table 2 explored configurations with modified training epochs and optimized hyperparameters, the results still support the conclusion that the ID 6 configuration from Table 1 was the most effective. This configuration balanced performance, efficiency, and practicality, making it a superior choice compared to configurations with more complex architectures or adjusted hyperparameters.

The high micro-averaged Sensitivity (81.13%) and Specificity (97.64%) reported in Table 3 underline the diagnostic strength of ID 6 across diverse anatomical regions. The near-perfect Specificity and Accuracy in detecting features like Cisterna Magna (97.80%) and Nuchal Translucency (98.43%) demonstrate its reliability. However, the lower Sensitivity for conditions like the Palate (53.66%) indicates areas for future improvement.

The combination of dropout and weight decay in the experiments aimed to regulate the model's capacity to memorize the data. These techniques are known to improve generalization by mitigating overfitting, especially when datasets are small. However, the marginal improvement observed in Table 2 suggests that the structural simplicity of ID 6 already aligned well with the dataset characteristics, reducing the need for aggressive regularization techniques.

Despite extensive experimentation with various configurations and hyperparameter optimizations, the ID 6 setup from Table 1 consistently outperformed others. Its simplicity, balanced training strategy, and effective sampling approach made it the best choice for the task. The findings from Table 3 further reinforce its effectiveness in a practical diagnostic context. Future studies should investigate the scalability of this configuration to larger datasets and explore whether additional hyperparameter tuning could further enhance its performance.

7 Conclusion

This study demonstrated the potential of integrating Convolutional Neural Networks (CNNs) with an Active Learning framework to address the challenge of identifying fetal intracranial structures in ultrasound images. By utilizing Active Learning, the system was able to incrementally refine its performance while minimizing the labeling effort, offering a practical solution in scenarios where annotated data is scarce.

Despite the promising results, several limitations were observed. The relatively small size of the dataset constrained the model's capacity to generalize, leading to suboptimal sensitivity and specificity in some cases. Additionally, the model's reliance on a limited number of training samples increased the risk of overfitting, particularly in a complex domain such as fetal anatomy. These challenges underscore the need for larger and more diverse datasets to improve the model's robustness and reliability.

Future work should focus on addressing these limitations by exploring more sophisticated techniques, such as data augmentation and semi-supervised learning, to enhance the dataset's diversity and reduce the dependency on labeled data. Furthermore, integrating additional medical imaging modalities and domain-specific knowledge could further improve the system's diagnostic accuracy. Ultimately, this study highlights the importance of leveraging machine learning in medical diagnostics and provides a foundation for developing more efficient and reliable systems to support healthcare professionals.

Acknowledgments

We would like to acknowledge the assistance of several tools and resources that contributed to the development of this paper. ChatGPT was utilized to help refine and correct the writing throughout the manuscript. Additionally, GitHub Copilot provided valuable support in the execution of the code. We also wish to express our gratitude to the authors of [Pinto *et al.*, 2022] for their foundational work, which served as a basis for our code implementation.

References

[Garcia-Canadilla *et al.*, 2020] Patricia Garcia-Canadilla, Sergio Sanchez-Martinez, Fatima Crispi, and Bart Bijnens. Machine learning in fetal cardiology: What to expect. *Fetal Diagnosis and Therapy*, 47(5):363–372, 01 2020.

[Jian *et al.*, 2021] Jia-Zheng Jian, Tzong-Rong Ger, Han-Hua Lai, Chi-Ming Ku, Chiung-An Chen, Patricia Angela R. Abu, and Shih-Lun Chen. Detection of myocardial infarction using ecg and multi-scale feature concatenate. *Sensors*, 21(5), 2021.

[Leiserowitz and Herding, 2020] Gary S. Leiserowitz and Herman Herding. Misdiagnosis of a pelvic mass versus pregnancy. *PSNet [internet]*, 2020.

[Mahapatra *et al.*, 2019] Dwarikanath Mahapatra, Behzad Bozorgtabar, and Rahil Garnavi. Image super-resolution using progressive generative adversarial networks for medical image analysis. *Computerized Medical Imaging and Graphics*, 71:30–39, 2019.

[Phung *et al.*, 2023] Nhu Hai Phung, Chi Thanh Nguyen, Trung Kien Tran, Thi Thu Hang Truong, Danh Cuong Tran, Thi Trang Nguyen, and Duc Huy Do. A combination of multi-branch cnn and feature rearrangement for down syndrome prediction. In *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 001–006, 2023.

[Pinto *et al.*, 2022] Catarina Pinto, Juliana Faria, and Luis Macedo. An active learning-based medical diagnosis system. In *Progress in Artificial Intelligence: 21st EPIA Conference on Artificial Intelligence, EPIA 2022, Lisbon, Portugal, August 31–September 2, 2022, Proceedings*, page 207–218, Berlin, Heidelberg, 2022. Springer-Verlag.

[Safdar *et al.*, 2018] S. Safdar, S. Zafar, N. Zafar, et al. Machine learning based decision support systems (dss) for heart disease diagnosis: a review. *Artificial Intelligence Review*, 50:597–623, 2018.

[Sarvamangala and Kulkarni, 2022] D R Sarvamangala and Raghavendra V Kulkarni. Convolutional neural networks in medical image understanding: a survey. *Evol. Intell.*, 15(1):1–22, 2022.

[SciDev.Net, 2024] SciDev.Net. AI risks in healthcare: Misdiagnosis, inequality, and ethical concerns. https://www.news-medical.net/news/20240123/AI-risks-in-healthcare-Misdiagnosis-inequality-and-ethical-concerns.aspx, January 2024.

[Sun *et al.*, 2024] Lingling Sun, Junxuan Yu, Jiezhi Yao, Yan Cao, Naimin Sun, Keqi Chen, Yujia Lin, Chunya Ji, Jun Zhang, Chen Ling, Zhong Yang, Qi Pan, Ronghao Yang, Xin Yang, Dong Ni, Linliang Yin, and Xuedong Deng. A novel artificial intelligence model for measuring fetal intracranial markers during the first trimester based on two-dimensional ultrasound image. *International Journal of Gynecology & Obstetrics*, 167(3):1090–1100, 2024.

[Wang *et al.*, 2024] Haoran Wang, Qiuye Jin, Shiman Li, Siyu Liu, Manning Wang, and Zhijian Song. A comprehensive survey on deep active learning in medical image analysis. *Medical Image Analysis*, 95:103201, 2024.

[Yousefpour Shahrivar *et al.*, 2023] Ramin Yousefpour Shahrivar, Fatemeh Karami, and Ebrahim Karami. Enhancing fetal anomaly detection in ultrasonography images: A review of machine learning-based approaches. *Biomimetics (Basel)*, 8(7):519, November 2023.

[Zhang *et al.*, 2021] C. Zhang, F. He, B. Li, et al. Development of a deep-learning system for detection of lattice degeneration, retinal breaks, and retinal detachment in tessellated eyes using ultra-wide-field fundus images: a pilot study. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 259:2225–2234, 2021.

ID	Layers	Epochs (Learner)	Epochs (AL Loop)	Sampling Strategy	Instances to train	Queries	Train Accuracy	Test Accuracy
1	2	40	30	entropy_sampling	15	10	84.57%	72.84%
2	2	40	20	uncertainty_sampling	15	10	87.19%	78.40%
3	3	20	5	entropy_sampling	15	10	64.51%	59.88%
4	2	30	40	entropy_sampling	20	10	80.86%	74.07%
5	2	30	10	uncertainty_sampling	10	10	79.32%	72.84%
6	1	40	20	uncertainty_sampling	15	10	87.25%	84.91%
7	3	40	20	margin_sampling	15	10	77.93%	72.84%
8	2	25	5	margin_sampling	5	15	60.49%	56.17%
9	2	20	5	margin_sampling	15	10	79.01%	75.31%
10	2	40	20	margin_sampling	15	10	82.59%	79.56%
11	1	20	5	entropy_sampling	15	10	76.70%	73.46%

Table 1: Table of the best results.

ID	Epochs (Learner)	Epochs (AL Loop)	Train Accuracy	Test Accuracy
1	50	30	87.45%	81.76%
2	50	20	87.45%	78.93%
3	40	30	89.51%	79.63%
4	30	30	89.34%	83.33%
5	30	40	89.07%	83.96%

Table 2: Table of the best combinations taking into account the best combination from the previous table.

Scructure	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Accuracy
Thalami	0.761905	0.913043	0.571429	0.961832	0.893082
Midbrain	0.772727	0.956204	0.73913	0.963235	0.930818
Palate	0.536585	1	1	0.935811	0.940252
4th Ventricle	0.64	0.979522	0.727273	0.969595	0.95283
Cisterna Magna	0.92	0.982935	0.821429	0.993103	0.977987
Nuchal Translucency	0.875	0.993197	0.913043	0.989831	0.984277
Nasal Tip	0.90566	0.988679	0.941176	0.981273	0.974843
Nasal Skin	1	1	1	1	1
Nasal Bone	0.947368	0.971429	0.818182	0.992701	0.968553
Micro-Averaging Measurements	0.811321	0.976415	0.811321	0.976415	0.958071

Table 3: Table of experiment 6 results