Chapter 1: Introduction

1.1 Background of the Study

Timely detection of fetal brain anomalies plays a critical role in enhancing prenatal counseling and improving postnatal neurological outcomes. Ultrasonography remains a widely used and effective clinical tool for identifying such anomalies. However, the interpretation of ultrasound images is subject to variability due to factors such as operator experience, fetal positioning, and image quality, which can limit diagnostic accuracy. In recent years, the development of artificial intelligence (AI)-based image processing techniques has offered a novel perspective in the evaluation of fetal brain ultrasound images. AI algorithms have the potential to reduce subjectivity and human-related variability, thereby promoting more standardized and reliable assessment processes. This research aims to investigate the diagnostic contribution and potential clinical benefits of analyzing archived fetal brain ultrasound images using AI-driven methods.

1.2 Statement of the Problem

Variability in the interpretation of fetal brain ultrasound images—arising from factors such as operator dependence, image quality, and fetal positioning—poses significant challenges to diagnostic accuracy and standardization. This variability can hinder the timely and accurate identification of fetal brain anomalies in clinical practice. Artificial intelligence (AI)-based analytic approaches have the potential to minimize human-related interpretative differences and enhance consistency in diagnostic processes. However, there is a notable lack of research evaluating the reliability and clinical applicability of AI models trained on expert-labeled, archived fetal brain ultrasound images. This knowledge gap hinders both researchers and clinicians seeking to clarify the role of AI in the detection of fetal brain anomalies and in clinical decision-making. Therefore, this study aims to assess how the analysis of archived 2D fetal brain ultrasound images—labeled by experts and processed with AI models —can impact diagnostic accuracy and the standardization of assessments.

1.3 Purpose and the Significance of Study

The primary purpose of this research is to evaluate the diagnostic accuracy and clinical applicability of artificial intelligence (AI)-based analysis of archived 2D fetal brain ultrasound images, focusing specifically on the distinction between normal and abnormal findings as determined by expert assessment. By systematically investigating the potential of AI in standardizing and supporting diagnostic processes, this study seeks to contribute valuable evidence to the fields of medical imaging and prenatal diagnostics.

This research holds significant implications for both academic circles and clinical practice. By examining the performance and practical use of AI-assisted image analysis in the context of fetal brain assessment, the study aims to:

1. Advance Scientific Knowledge:

   The study will provide new insights into the capabilities and limitations of AI algorithms for evaluating fetal brain ultrasound images. It will determine how successful AI algorithms are in compared to experienced clinicians in detecting and classifying structural and developmental abnormalities. This evidence will enrich the current body of literature on medical imaging, machine learning applications, and fetal anomaly detection. It will also explore scenarios in which AI performance may be suboptimal, including rare conditions, poor-quality imaging, or demographic variations. This work will highlight areas for technological refinement and inspire further interdisciplinary research in fetal neuroimaging.

2. Inform Clinical Practice:

   The findings will equip clinicians with evidence-based information on the integration of AI tools into prenatal diagnostics. By clarifying the strengths and weaknesses of AI support, the research will empower healthcare professionals to make more informed decisions in routine fetal anomaly screening. The study will quantify the time savings and workflow improvements that AI tools can bring to routine and complex cases.

3. Guide Technology Adoption and Policy:

   The results of this study may guide hospital administrators and policymakers in decisions regarding the adoption of AI-based diagnostic support systems. Data-driven insights will highlight the potential benefits and considerations for implementing such technologies in clinical workflows.

4. Enhance Patient Care:

Ultimately, this research aims to improve prenatal care by supporting early and accurate detection of fetal brain anomalies. By identifying effective diagnostic strategies, the study has the potential to positively impact clinical outcomes for expectant mothers and their babies.

1.4 Research Questions

1. How accurately can a deep learning model, trained on expert-labeled 2D fetal brain ultrasound images, distinguish between normal and abnormal intracranial findings?
2. Does the diagnostic performance of the model differ between the lateral ventricle and other intracranial regions?
3. What types of anomalies are most frequently misclassified by the model, and what factors contribute to these misclassifications?
4. To what extent can AI-assisted image analysis reduce inter-observer variability in the assessment of fetal brain ultrasound images?
5. What practical and ethical considerations arise when integrating AI-based decision support systems into routine prenatal screening?

1.5 Definitions of Terms

- **Artificial Intelligence (AI):**
  Computer systems or algorithms capable of performing tasks that typically require human intelligence, such as pattern recognition, decision-making, and problem-solving. In this study, AI refers specifically to deep learning models used for analyzing medical images.
- **Deep Learning:**
  A subset of machine learning based on artificial neural networks with multiple layers, designed to automatically learn complex patterns from large amounts of data.
- **2D Fetal Brain Ultrasound:**
  A medical imaging technique that uses sound waves to produce two-dimensional images of the developing fetal brain during pregnancy.

- **Lateral Ventricle:**
  A pair of structures within the fetal brain that contain cerebrospinal fluid; abnormalities in size or shape can be indicative of certain brain anomalies.
- **Other Intracranial Regions:**
  Areas within the fetal brain apart from the lateral ventricles, including various anatomical structures such as the midline, thalamus, basal ganglia, cerebral hemispheres, and cortical surfaces. Evaluation of these regions is essential for the detection of a broad spectrum of brain anomalies beyond those limited to the ventricles.
- **Expert Labeling:**
  The process of classifying medical images (as normal or abnormal) by highly experienced physicians based on standardized diagnostic criteria.
- **Diagnostic Accuracy:**
  The ability of a test or model to correctly identify the presence or absence of disease or abnormality, often quantified by sensitivity, specificity, and overall accuracy

Chapter 2: Literature Review

Prenatal ultrasonography is the gold standard imaging method for the detection and assessment of fetal structural anomalies. Its widespread adoption is largely due to its non-invasive nature, real-time imaging capability, and cost-effectiveness. However, the diagnostic accuracy of ultrasound often depends on the operator’s experience, fetal position, and equipment quality. Drukker et al. (2022) demonstrated that there is substantial inter-observer variability in identifying and interpreting fetal structures, which can directly influence clinical outcomes. These human-factor limitations, particularly in the evaluation of the fetal brain and other intracranial regions, have highlighted the need for new technologies to improve standardization and diagnostic accuracy.

In recent years, artificial intelligence (AI)—especially deep learning-based approaches—has revolutionized the field of medical image analysis. These methods, based on artificial neural networks, have shown remarkable success in tasks such as pattern recognition, segmentation, and classification. Wu et al. (2023) reported that AI algorithms have been utilized for identifying standard anatomical planes (biparietal head diameter, abdomianal circumference, femur length, etc…), detecting structural anomalies, and guiding sonographers during image acquisition in fetal anomaly screening. The integration of AI into ultrasound workflows not only increases accuracy and reproducibility but also provides a solution to one of the most critical challenges in prenatal imaging: operator-dependent variability.

Applications of AI in the analysis of fetal brain and other intracranial ultrasound images have been the subject of an increasing number of studies in the literature. Deep learning models have been used for segmenting brain structures, classifying normal and abnormal findings, and automatically recognizing key anatomical landmarks. Especially when trained on large, expertly labeled datasets, these models have demonstrated high performance in terms of diagnostic accuracy, sensitivity, and specificity—comparable to or even exceeding that of experienced clinicians in certain tasks (Belciug et al., 2024; Wu et al., 2023). AI-based systems enable the systematic assessment of intracranial regions beyond the lateral ventricles, such as the midline, thalamus, basal ganglia, and cerebral hemispheres, thereby offering an advantage in detecting complex anomalies that may be missed due to human oversight or fatigue.

Nevertheless, there are various methodological and practical limitations in the current literature regarding the application of AI models. Many studies have focused only on a single anatomical region or used a single AI approach, limiting their generalizability and diagnostic power. Belciug et al. (2024) emphasize that more recent approaches involve the combination of multiple algorithms and testing on large, diverse datasets. Validation studies have reported accuracy rates above 90% for certain tasks; however, model performance may vary depending on image quality, gestational age, and the type of anomaly assessed. Moreover, the need for large, expertly labeled datasets, the risk of overfitting, and the lack of standardized evaluation protocols remain significant barriers to the widespread clinical use of AI-based decision support systems.

In conclusion, the future of AI in prenatal ultrasound imaging depends on the development of robust, interpretable, and clinically validated systems. Belciug et al. (2024) and Wu et al. (2023) note that current research aims not only to detect a broad range of congenital anomalies, but also to provide clinicians with actionable real-time feedback through intelligent decision support systems. As the technology matures, such systems are expected to reduce the gap between centers with varying expertise, decrease diagnostic errors, and improve perinatal outcomes. However, attention to ethical issues, data privacy, and the need for multidisciplinary collaboration will remain indispensable as AI continues to transform prenatal diagnostic practices.

Chapter 3: Methodology

3.1 Research Design

This study employs an observational research design to evaluate the diagnostic accuracy of artificial intelligence (AI)-based analysis of 2D fetal brain ultrasound images. The research is based on archived ultrasound images that were previously acquired during routine clinical practice. No new imaging procedures, interventions, or biological sample collections will be performed as part of this study. Instead, existing anonymized ultrasound data will be systematically re-evaluated by expert clinicians to establish reference classifications (normal or abnormal), which serve as ground truth labels for subsequent AI model training and testing.

The study will be conducted in several phases. Initially, expert clinicians will review and label all eligible ultrasound images, focusing on key intracranial regions such as the lateral ventricles and other relevant brain structures. These expert-labeled images will then be used to train a deep learning-based model, which is subsequently tested on a separate dataset to assess its ability to accurately distinguish between normal and abnormal findings. The model's performance will be evaluated using standard statistical metrics such as sensitivity, specificity, accuracy, and F1 score. Throughout the research process, all patient identifiers will be removed to ensure data anonymity and participant confidentiality. This design allows for the systematic assessment of AI-assisted diagnosis using real-world clinical data, while minimizing risk to participants and adhering to all relevant ethical guidelines.

3.2 Setting and Participants/Sampling

The research will be conducted using archived 2D fetal brain ultrasound images obtained from routine clinical examinations. The dataset comprises images from a diverse population of pregnant individuals who underwent standard prenatal ultrasound screening between the 18th and 24th weeks of gestation. By utilizing data collected over an extended period and from a broad patient base, the study aims to ensure comprehensive representation of various fetal brain presentations encountered in real-world clinical practice.

Eligible ultrasound images will be selected based on strict inclusion criteria, such as maternal age between 18 and 45 years, clear visualization of the lateral ventricles and other intracranial regions, and the absence of any personally identifiable information. Images that do not meet diagnostic quality standards, have ambiguous gestational age, or contain duplicate or incomplete records will be excluded from the analysis. This approach is designed to maximize the reliability and validity of the findings while minimizing potential sources of bias.

The use of anonymized clinical data ensures that no direct contact with patients is required, and no interventions or additional procedures are performed. By systematically analyzing a large and representative sample of archived images, this research seeks to provide robust evidence on the diagnostic performance of artificial intelligence models in detecting fetal brain anomalies across a broad clinical spectrum. The working dataset currently comprises 645 images (413 normal, 232 anomalous); 116 anomalous samples are expert-vetted synthetic images used for training only, while validation/test use real images only.

3.3 Data Collection and Data Analysis

3.3.1 Data Collection

The data collection phase will be carried out in two main steps, involving the systematic evaluation of 2D fetal brain ultrasound images from the clinical digital archive.

1. Clinical Archive Screening and Image Selection:

The study will utilize images obtained from cases that underwent routine prenatal ultrasound screening between the 18th and 24th weeks of gestation and were stored in the institutional archive. Selection will be performed according to well-defined inclusion and exclusion criteria; only images that are diagnostically adequate, anonymized, and complete will be included in the study. The process of image selection will be conducted meticulously by the research team in accordance with the study protocol.

2. Expert Labeling:

The selected ultrasound images will be independently reviewed by clinicians with over 15 years of experience in the field, and each image will be classified as either normal or abnormal, focusing particularly on the lateral ventricles and other intracranial regions. In cases with abnormal findings, further sub-categorization will be performed according to the specific pathology. In instances of disagreement among experts, a consensus method will be used to finalize the label.

All data will be fully anonymized prior to analysis, and no personal or identifiable information will be present in the records. During data collection, strict adherence to ethical guidelines and patient confidentiality will be ensured.

3.3.2 Data Analysis

The analysis of the collected data will encompass the development and clinical validation of the artificial intelligence (AI) model.

Quantitative Analysis:

This study will employ a Python-based deep learning workflow to classify 2D fetal brain ultrasound images as normal or anomalous. In addition to the images stored in institutional datasets, some synthetic images will be generated to alleviate class imbalance and improve model generalization if needed. All synthetic images will be reviewed and judged by fetal sonography specialist Dr. Nazlı Yenigül and Dr. Bilge Çetinkaya Demir.

Training/Validation policy: Real (archival) images and synthetic anomalous images will be used for training the model and only real images will be used for validating and testing the model. Images will be standardized to 224×224 and normalized to the backbone's channel statistics. To address residual imbalance and improve minority-class sensitivity, the training pipeline will apply class-aware sampling/weighted loss and light data augmentation (small rotations, horizontal flips, mild brightness/contrast jitter). A modern CNN/Transformer classifier will be trained with early stopping. Exact architecture and hyper-parameters will be finalized in order to achive the best model performance.

Development credit. The AI model and the synthetic anomaly image generation pipeline will be designed and implemented by AI Engineer Erdeniz Ünvan.

Primary outcomes: Performance will be reported using accuracy, sensitivity (recall), specificity, precision, and F1-score metrics and a confusion matrix will be provided. All data will be fully anonymized and stored on secured, access-controlled institutional servers.

Qualitative Analysis:

Cases that are misclassified by the model and the possible contributing factors (e.g., image quality, anatomical variations) will be reviewed by experts, and emerging patterns and limitations will be thematically reported. By comparing the model outputs with expert assessments, a deeper understanding of the clinical applicability of AI will be gained.

All analyses will be performed using Python and relevant libraries for medical image processing and machine learning. The results will be discussed in the context of the research questions, providing practical implications for clinical implementation and future research. Data security and compliance with ethical standards will be prioritized throughout every stage of the analysis.

3.4 Validity and Reliability

Ensuring the validity and reliability of this research is essential for the credibility of its findings. To achieve this, every stage of the study has been carefully planned. Clear inclusion and exclusion criteria have been established for the selection of ultrasound images, ensuring that only diagnostically relevant and high-quality data are analyzed. All images will be independently reviewed and labeled by at least two experienced clinicians. In cases of disagreement, a consensus procedure will be employed to enhance the reliability of the reference labels.

To maintain data integrity, all image data will be anonymized and securely stored, with access strictly controlled throughout the study. Ethical guidelines for the use of clinical data will be strictly followed, ensuring respect for patient privacy and data protection.

By employing systematic and transparent procedures at every stage—from data selection to expert labeling, model validation, and statistical analysis—this research aims to produce reliable and valid results that can meaningfully contribute to clinical practice and future scientific investigations.

3.5 Ethical Issues and Limitations

This research is committed to upholding ethical standards at every stage. All ultrasound images used are fully anonymized and contain no personally identifiable information, with data security maintained at the highest level. The study’s design and procedures will be reviewed and approved by the relevant institutional ethics committee to ensure full compliance with ethical guidelines. Since no direct data is collected from participants, informed consent is not required; nevertheless, all obligations regarding confidentiality and patient privacy will be strictly fulfilled.

Synthetic images will be generated to enhance representation of anomalous patterns without reproducing any individually identifiable image. They will contain no personal identifiers and will not be one-to-one copies of any patient scan. Synthetic images will be used only for training; all evaluation will be performed on real archival images. All data (real and synthetic) will be stored in de-identified form on secure institutional infrastructure with role-based access control. No data will be shared outside the institution at the individual level; only aggregate statistics will be reported.

Certain limitations of the study should also be acknowledged. The analyzed dataset is limited to a single healthcare institution and includes only prenatal ultrasound images obtained between the 18th and 24th weeks of gestation, which may restrict the generalizability of the findings to broader populations. Additionally, as the study is confined to existing archived images, it is not possible to evaluate clinical variables or long-term outcomes. Potential issues such as diagnostic image quality and inter-expert variability in labeling may also influence the reliability of the findings.

In light of these ethical considerations and limitations, the study aims to contribute meaningfully to scientific knowledge and clinical practice by transparently and responsibly investigating the value of AI-assisted ultrasound analysis.

References

1. Belciug, S., Ivanescu, R. C., Serbanescu, M. S., Ispas, F., Nagy, R., Comanescu, C. M., ... & Iliescu, D. G. (2024). Pattern Recognition and Anomaly Detection in fetal morphology using Deep Learning and Statistical learning (PARADISE): protocol for the development of an intelligent decision support system using fetal morphology ultrasound scan to detect fetal congenital anomaly detection. *BMJ open*, *14*(2), e077366.
2. Wu, H., Wu, B., Lai, F., Liu, P., Lyu, G., He, S., & Dai, J. (2023). Application of artificial intelligence in anatomical structure recognition of standard section of fetal heart. *Computational and Mathematical Methods in Medicine*, *2023*(1), 5650378.
3. Drukker, L., Sharma, H., Karim, J. N., Droste, R., Noble, J. A., & Papageorghiou, A. T. (2022). Clinical workflow of sonographers performing fetal anomaly ultrasound scans: deep-learning-based analysis. *Ultrasound in Obstetrics & Gynecology*, *60*(6), 759-765.