Computational Intelligence for an Early Detection of Infertility in Women Utilizing Inception & Xception with Attention Mechanism

Arabindra Dhami

230150452

Submitted in accordance with the requirements for the degree of

Master of Science

York St John University

Department of Computer Science

Supervisor's Name

Sangita Pokhrel

Date of Submission: 10th September 2025

The Arabindra Dhami confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Self-declaration on Academic Integrity:

I have read and understood the [Academic Misconduct statement].	Tick to confirm ☑
I have read and understood the [Generative Artificial Intelligence use statement].	Tick to confirm ☑
I am satisfied that I have met the Learning Outcomes of this assignment (please check the Assignment Brief if you are unsure)	Met ☑

Self-Assessment – If there are particular aspects of your assignment on which you would like feedback, please indicate below. Optional for students
Suggested prompt questions- *How have you developed or progressed your learning in this work?* *What do you feel is the strongest part of this submission?* *What feedback would you give yourself?* *What part(s) of this assignment are you still unsure about?*

Abstract

This study applies deep learning for early detection of infertility-related conditions in women, specifically targeting polycystic ovarian syndrome (PCOS) using ultrasound imaging. A quantitative approach is employed to develop AI-driven diagnostic tools that can distinguish between PSOS and normal conditions. The research aims to enhance non-invasive fertility diagnostics through advanced convolutional neural networks integrated with attention mechanisms. Advanced architectures (InceptionV3 and Xception) enhanced with channel attention mechanisms are implemented for feature refinement. Trained on ultrasound images annotated as PSOS or normal, the models underwent preprocessing procedures such as image resizing, normalization, and data augmentation, which ensure robustness. Further, duplicate images were eliminated through perceptual hashing, thus improving the overall quality of the dataset and reducing noise. Cross-validation is applied throughout the training process to prevent overfitting and guarantee generalizability. Performance evaluation employs accuracy, precision, recall, and AUC metrics.

Results demonstrate that attention mechanisms significantly improve predictive accuracy, with the attention-augmented Xception model achieving superior performance. While deep learning models surpass traditional machine learning in accuracy for this imaging-based diagnostic task, we acknowledge the interpretability advantages of conventional approaches. Ethical considerations including data privacy protocols (addressing secondary data use from Kaggle) and bias mitigation strategies are implemented throughout the study to ensure responsible AI deployment. This research contributes to the growing field of fertility analytics by validating AI-driven ultrasound interpretation for PSOS detection. Future directions include integrating real-time health monitoring data, developing Explainable AI (XAI) frameworks for model transparency, and creating personalized fertility assessment tools. These advancements could support healthcare professionals in making data-driven reproductive health decisions and inform policy development.

The outcomes obtained in the experiments facilitate the mastery of deep learning model efficacy in classifying ultrasound images, as the Xception+Attention model has shown the best results of 95 percent. It achieved a perfect score of accuracy, precision, and recall, as well as the rest of the models. Its potential to concentrate on pertinent locales in its picture was demonstrated through visualizations. The results confirm its strength and dependability when it comes to automatic infection identification in ultrasound images.

DECLARATION

I hereby certify that this dissertation constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions or writings of another.

I confirm that I have not copied material from another source nor committed plagiarism nor commissioned all or part of the work (including unacceptable proof-reading) nor fabricated, falsified or embellished data when completing the attached piece of work.

I declare that the dissertation describes original work that has not previously been presented for the award of any other degree of any institution.

I affirm that the use of AI in this dissertation does not undermine the academic integrity of the work and complies with the regulations set forth by York SJ University. Any assistance provided by Artificial Intelligence (AI) tools, including but not limited to research, writing, or editing, has been acknowledged and utilized in accordance with ethical guidelines.

Arabindra Dhami

10th August 2025

Acknowledgement

I would like to thank **Mrs. Sangita Pokhrel**, my supervisor who has been years of help and encouragement throughout this research. I am also grateful to my professors as well as my counterparts for the conversations and feedback, which helped me a lot with my work. I would like to end by saying that I would like to extend my appreciation to my family and friends for their unyielding support and encouragement. Finally, I want to thank those in the open-source communities and researchers who had done the work that formed the basis for this study. They have been very inspiring for your contributions.

Table of Contents

List of Figures

CHAPTER 1: INTRODUCTION

1.0. Chapter Introduction

This chapter highlights the challenges in infertility diagnosis while emphasising the progression made so far, particularly related to the present limitations of standard diagnostic approaches versus the potential introduced by AI. It emphasises an urgent need to have accurate and efficient, with minimal invasiveness, for early detection with improved clinical decisions. It lays out a chapter on a study's objectives based on researching methods to upgrade present deep learning architectures, focusing more on architectures with Inception and Xception networks, especially attention mechanisms incorporated for feature refinement. The next phase of refinement deals with better precision and more accurate interpretability through the mentioned methodologies.

1.1. Research Background

It's a leading health concern that involves millions of women worldwide and has often been detected too late for effective treatments to be taken (Bappi *et al.,* 2024). Traditional diagnostics for infertility often come in the form of endocrine tests, sonography, and keyhole surgery, all invasive, costly, and time-consuming procedures. That is why people have shifted the focus towards research on AI and deep learning. Deep learning architectures, in particular Inception, Xception and Random Forest, have achieved impressive state-of-the-art performance in medical imaging. These deep-learning architectures can learn complex patterns from medical data to accurately diagnose infertility. The WHO assessed infertility as a global health issue concerned with about 17.5% of the population worldwide (WHO, 2023). For instance, in the UK, about 1 in 7 couples are believed to have problems conceiving, indicating that this problem is very common. These figures indicate the dire need for improved non-invasive and readily available diagnostic techniques for the early detection and treatment of infertility. However, though they are effective, these AI-based approaches lack interpretability, and healthcare professionals require transparency to build trust and ensure ethical responsibility for predictions generated using these models. With attention mechanisms, these improvements can now be made through incorporation into the Inception and Xception models so that they further refine their selection of diagnostic indicators. These models become even more transparent when employing attention, meaning that medical images or data contributions to their prediction are precisely specified. Apart from increasing the accuracy, these advances help the clinician make sense of AI algorithms' output (Shah et al., 2022). This work attempts to formulate a computational model driven by AI, which may further enhance accuracy and efficiency while maintaining interpretability for the diagnosis of infertility. It suggests an applicable and non-invasive tool to diagnose and diagnose early cases with support in the clinical decision-making process. This would produce more accurate outcomes in using AI in reproductive health in the way it will contribute to existing diagnosis challenges for infertile health. For additional assurance regarding the effectiveness of the model, duplicate detection of the images was done using perceptual hashing on this dataset. This pre-processing step has significantly contributed to the improved consistency and reliability of training data by doing away with redundant ultrasound images.

1.2. Problem Statement

Infertility diagnosis is a tough problem because one wants it to be all these: accurate, interpretable, and efficient (Rotem *et al.,* 2024). The traditional diagnostic approaches, such as hormonal evaluations, imaging methods, and invasive interventions, are time-consuming and costly, with specialised medical knowledge required, and delay timely intervention. Traditional methods for diagnosing include expensive, time-consuming, and resource-intensive tests like hormonal tests, imaging tests, and invasive procedures that necessitate specialised medical knowledge and further delay timely intervention. AI and deep learning models have been successfully used to automate the diagnosis of several medical conditions; however, they are yet to be utilised extensively in detecting infertility owing to their ambiguity. Current deep-learning techniques, such as Inception and Xception and Random Forest architectures, are able to process medical imaging data well but work as black-box models, which makes clinical validation problematic. Even though deep-learning approaches like Inception and Xception and Random Forest architectures allow efficient processing of medical imaging data, they function like black boxes, thus making clinical validation really challenging. Among other key limitations in medical diagnostics AI presents, failure to explain the predictions is at the top, as this limits its use and confidence in clinical settings. Machine Learning has been shown to be viable for many medical imaging applications through improvements in the model's ability to focus attention on the most relevant diagnostic features; however, they have seen less exploration into the diagnosis of infertility (Kaveh, 2024). The work hereby focuses on this aspect of incorporating the mechanism of attention within both Inception and Xception-based models with improvements toward high sensitivity to diagnoses yet further in developing an early fertility diagnostic method of less invasiveness.

1.3. Research Aim

This paper intends to design a computational model based on deep learning architectures like Inception and Xception and Random Forest with the inclusion of attention mechanisms for better early female infertility diagnosis. Through the feature selection refinement process, the model hopes to provide increased accuracy, efficiency, and interpretability that may be achieved as opposed to conventional and AI-based diagnostic techniques.

1.4. Research Objectives

- To review the existing machine learning strategies and identify the strengths and weaknesses of those AI-based approaches for the diagnosis of infertility.
- To Improve the Design of New Deep Learning Architectures.
- To enhance Inception and Xception and Random Forest architectures with an attention mechanism, focusing on further improving feature extraction, thereby further increasing diagnostic performance and efficiency.
- This study evaluates how well the proposed models, compared to the baseline techniques, show improved clinical relevance and the potential for applicability in real-world healthcare scenarios.

1.5. Research Questions

- What are the contemporary techniques in machine learning used in diagnosing infertility?
- How does the attention mechanism help in the refinement of Inception and Xception models in infertile conditions?
- How do pre-processing and data augmentation improve deep learning models?
- How is the performance of proposed models- a measure of accuracy, precision, recall, and interpretability- compared with already available baseline approaches, and how would this AI-driven approach help clinicians make educated decisions?

1.6. Methodology Outline

This research adheres to a structured AI-driven methodology for creating an effective diagnostic tool for infertility. The data collection of this study is derived from the aggregation of publicly available medical datasets, as well as hospital-sourced imaging data about infertility diagnosis. Data preprocessing techniques such as normalisation, augmentation, and noise reduction are used for high-quality input into deep learning models. The training design of Inception and Xception and Random Forest architectures focuses separately on training as well as their testing before integrating them using attention mechanisms into the architecture with a focus towards improving feature extraction. The different models are typically trained with prominent deep learning platforms such as TensorFlow and PyTorch. Extensive training combined with validation serves to determine that the model successfully diagnoses infertility problems. This study utilised an AI-based structured methodology that enables an efficient infertility diagnostic tool to be developed. The research method adopted is mixed in nature, depending on quantitative as well as qualitative analyses (Taherdoost, 2024). The study majorly focuses on deep learning models and evaluates their performance using the medical dataset. Interpretability analysis is performed by visualising the technique Grad-CAM to analyse the improvement of the attention mechanisms towards model transparency. Finally, scalability and integration with healthcare systems are tested to prove the feasibility of the model in real-world clinical applications.

1.7. Chapter Summary

This chapter will outline the background in the diagnosis of infertility, explaining the problems clinicians are facing presently. The aim is to raise the need to have efficient, accurate and even non-invasive diagnostic tools based on the adoption of AI-aided methods. This would point out important gaps within currently used models, such as an inability to deliver interpretability and the relatively low accuracy-complicating effective use within decision-making. The research focus, objective, and aim at developing models with deep-learning advances, wherein the incorporation of attention mechanisms seeks to enhance diagnostics for precision purposes while gaining maximum transparency from these diagnostics.

CHAPTER 2: LITERATURE REVIEW

2.1. Introduction

This chapter will introduce the literature review that briefly outlines the purpose and scope of the chapter. The central goal of this review is to review the contemporary literature on diagnosis, with a specific interest in artificial intelligence and machine learning technologies that could help in identifying this condition. Because infertility continues to be an important health concern, early diagnosis with high precision remains an indispensable precursor to optimising the outcome of any treatment modality. For this reason, early diagnosis and as accurate as possible is necessary because infertility has been one of the major health concerns. Traditional diagnosis methods are pretty invasive, cost-effective, and time-consuming.

2.2. Topic 1: Machine Learning in Medical Diagnostics

As per Patel and Kumar (2025), ML is now indispensable for the medical world, opening avenues for transformative potential in diagnostic activities. Its data-processing capabilities and ability to identify complicated patterns have given ML a new scope in various diagnostic analyses related to medical images. ML applications, for example, have displayed immense potential to increase diagnostic precision and minimise the involvement of humans with errors; in addition, efficiency in decision-making in medicine can be greatly improved. Supervised learning: This is another common ML method in which algorithms are learned on labelled data sets with their input data matched to known labels for output. This technique was widely applied during medical image analysis in that images were represented by specific labels of conditions or features. Supervised learning models have been used in infertility diagnosis to find abnormal patterns within images from ultrasounds, CT scans, or MRI images. It categorised ovarian images and identified possible conditions related to infertility by applying supervised learning with good accuracy levels. Another core approach in ML is unsupervised learning, in which patterns within the data are determined without a label. The application of unsupervised learning comes in handy in cases of the unavailability of labelled data or when an intention is to unveil the concealed relations within the data. As medical images regarding infertility are clustered with unsupervised learning algorithms to find patterns among the data without pre-defined endpoints, similarities and differences in such data can be identified. As per Lee (2025), algorithms for clustering could group similar conditions of patients using medical imaging so that clinicians gain new insights into potential diagnoses for infertility. In the case of unsupervised learning, patterns hidden in data can be found, but such application in diagnosing infertility demands careful interpretation. More advanced than this ML technique is RL, where training of the models occurs based on interaction with the environment. Medical diagnosis also has an application in RL: for decision-making tasks, it uses the feedback generated by the actions of its model to learn the best strategy for diagnosis. According to Perrotta (2024), this area, while still underdeveloped in this field, already holds potential as applied to treatment plan optimisation in cases of infertility. Despite this, the complexity of the healthcare environment, combined with the extensive trial-and-error processes involved, limits the use of RL in the medical field. Among those subsets of machine learning, deep learning has revolutionised medical image analysis and made it possible to automate feature extraction and classification. This area has gained intense interest in recent times because Convolutional Neural Networks (CNNs) have been frequently applied for the diagnosis of infertility. These models include complex spatial hierarchies learned by images and may detect intricate structures within medical scans with very little manual interaction with the data. Deep learning models like Inception and Xception and Random Forest have demonstrated the potential for significantly improved diagnostic accuracy through automated feature extraction from medical images. The application of such models can identify early stages of infertility through analysis of ultrasound images of the reproductive system and the early onset of PCOS, uterine abnormalities, and endometriosis that cause infertility.

2.3. Topic 2: AI in Infertility Diagnosis

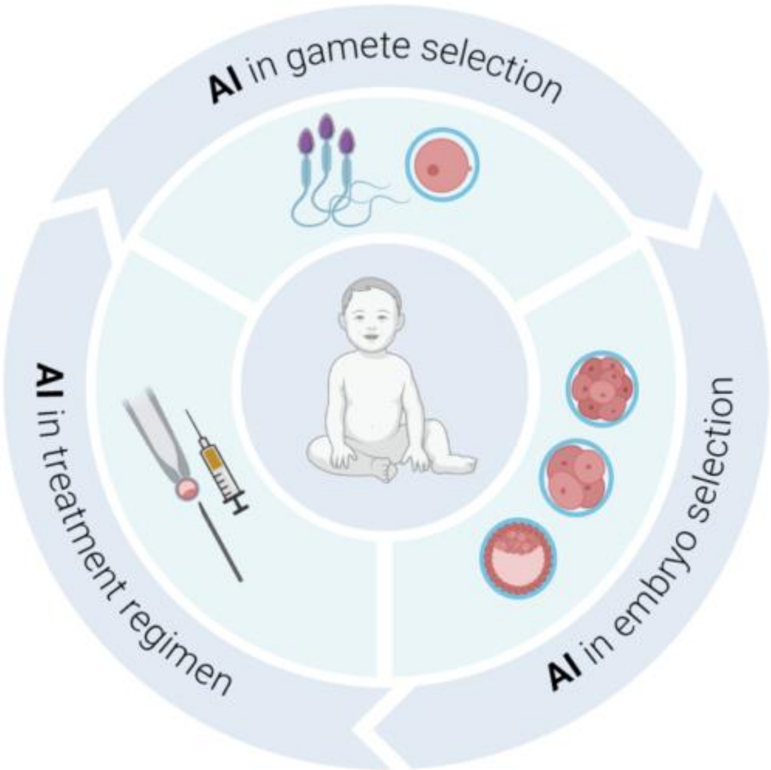


Figure 2.3.1: AI in Infertility Diagnosis

(Source: Sha, 2024)

Deep learning, within the ambit of machine learning, has drastically advanced medical image analyses and holds excellent promise in detecting infertility cases. Fertility clinics have started putting deep learning algorithms into practice, including the concept of CNN, for improved diagnosis capabilities. With utmost efficacy in evaluating visual data, CNN has thus been implemented while processing images scanned from ultrasound and MRI to detect malfunctioning reproductive organs. These models, therefore, become capable of locating features in pictures that the naked eye may sometimes miss, features such as cysts in ovaries uterine and fibroid growths, often the causes of infertility. Deep learning has been shown in numerous studies to be effective for diagnosing infertility. As per Sha (2024), the images of the ovaries through ultrasound obtain a high degree of accuracy in identifying ovarian cysts, which are essential for diagnosis. Deep learning algorithms have also gained potential applications in embryo quality evaluation in the course of IVF, thereby improving the rate of embryo implantation. Medical imaging is a critical component in diagnosing infertility. CNNs are particularly good in the analysis of medical images as they are capable of analysing complex visual data with minimal intervention from humans if features are extracted automatically from the images. One way a CNN can analyse complex visual information with minimal human involvement is by automatically extracting features from images. The applications of CNNs on ultrasound and MRI scans can identify structural abnormalities, which include fallopian tube blockage, uterine fibroids, endometrial polyps, and others that may cause infertility. According to Urakubo (2025), CNN in MRI images from the female reproductive system is to determine whether or not there are fibroids and any abnormalities that can be found by these models with higher accuracy. That helps to quicken the diagnostic process and intervene at the proper time instead of expensive, time-consuming laparoscopy processes. Other uses for CNNs would be to monitor the progression of infertility-related conditions and report changes over time. Although CNNs are primarily applied to image analysis, RNNs find special utility when dealing with any kind of sequential data, like records or histories from patients.

The RNNs are intended to process sequences of data, which is well fitted to analyse time-series data that could diagnose infertility. Using RNNs, success chances in conceiving have been predicted by using historical data-whether a couple has ever conceived and factors like age and lifestyle. Such models have a greater degree of precision in identifying time-related patterns and can predict the likelihood of infertility, thus guiding clinicians to provide timely interventions. Despite the potential shown by AI for infertility diagnosis, there are also several challenges and limitations. Probably the biggest challenge is that huge amounts of large, labelled data are needed for training deep learning models properly. In the field of medicine and infertility, usually, the quantity of sufficient annotation is scarce owing to the strict privacy, ethics, and lengthy time for doing manual annotations. Moreover, there is an increasing risk of overfitting within AI models, especially in the case of deep learning networks, while training the network with fewer examples. As per Räz (2024), the interpretability of AI models is another challenge.

2.4. Topic 3: Inception, Exception and Random Forest Models in Medical Image Analysis

Going deeper with Convolutions is a deep CNN that introduces the paper by Shetty and Sharma (2024), which represents an innovative approach to designing the convolutional layer in Inception architecture. One of its key features is the Inception module, which captures the wide variety of spatial hierarchies existing in the images that are the inputs. This module applies filters of different sizes, 1x1, 3x3, and 5x5, within the same layer so that the network learns multi-scale features without a significantly increased computational cost. Another peculiar application in Inception is 1x1 convolution. Reducing intermediate layers' dimensionality by applying 1x1 convolution can assist in deepening the model without leading to overfitting. It further enhances the model's efficiency with computation, especially considering that huge medical image datasets need to be processed very quickly and precisely. Inception model has widely been used in several medical applications. These range from breast cancer detection lung diseases to brain tumours, where it will capture multi-scale features and diagnose conditions present in different levels of resolution in the images. It has also shown successful work in fertility-related studies, where more direct medical imaging techniques, like ultrasound or MRI, are used to treat diagnosis conditions, including PCOS and endometriosis. As per Fantini (2024), it enhanced the version of Inception architecture. Xception embraces a concept termed depthwise separable convolutions, which segregates the steps of filtering and mixing into individual steps. As a result, it reduces parameters and increases efficiency in computation across the network. This way, Xception may learn even deeper and more sophisticated patterns and representations than its ancestor and be capable of using deeper networks with its depthwise separable convolutions for finer-grain feature extraction-based image classification. The xception model's architecture is developed as an extension across multiple layers for depthwise separable convolutions and compared to the normal convolutions in terms of precision and time used for training. This is why it proves very effective in medical imaging. There, correct feature extraction matters. Xception is highly capable of performing numerous applications in medicine: dermatology, ophthalmology, and radiology with skin cancer detection and retinal image analysis in diabetic retinopathy. It's strong where local and global features are perfectly captured in an image. Conversely, though, Inception and Xception and Random Forest have been shown to achieve an almost perfect balance between high accuracy and reasonable efficiency, which is simply vital in certain applications of medical image analysis, where efficiency at any sacrifice of accuracy is inappropriate In comparison when other deep learning models are used, including VGGNet or ResNet, both of them stand apart for their multi-scale and hierarchical feature capture ability. As per Verma *et al.* (2024), VGGNet uses a much simpler architecture in addition to relatively small receptive fields. This does not allow this architecture to learn much larger contextual features in the case of medical images when compared to a ResNet, which, due to residual connections, is very effective and generally requires deeper networks, thus making this more computationally expensive. They are among the most robust architectures for the analysis of medical images, specifically in the field of infertility diagnosis.

2.5. Topic 4: Attention Mechanisms in Deep Learning Models

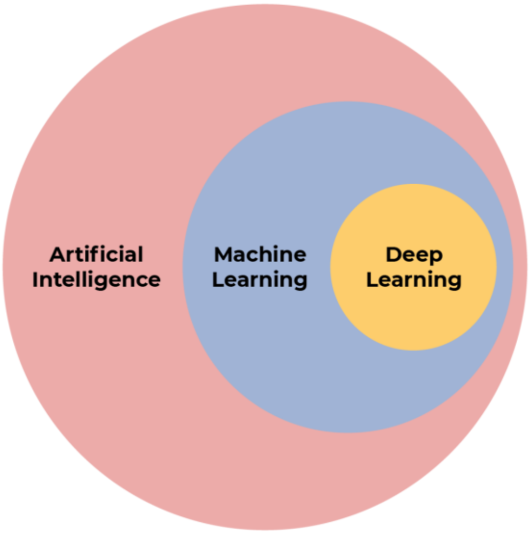


Figure 2.5.1: Deep Learning Models

(Source: Soydaner, 2024)

Inspired by how humans' cognitions process input data by attending to certain areas of that data, which seem most relevant to the processing activity, deep learning's attention mechanism allows a model to assign weight to different parts of an input, making it focus on what is important while ignoring or even less significant, irrelevant data. In computer vision, attention mechanisms help guide models toward specific parts of an image where decisions need to be made based on accuracy, such as focusing on a tumour in medical images or identifying a specific feature in ultrasound. As per Soydaner (2024), the approach to the attention mechanism, but probably the most general one,e is known as soft attention, through which it takes a weighted summation of its features with assigned weights on all parts of an input. With hard attention the model focuses exactly on some of the data because this is slightly more computationally expensive and applied less frequently in practice. Attention mechanisms can be incorporated into almost any deep learning architecture, such as CNNs, RNNs, and Transformer models. It has brought huge performance boosts for the given models in the context of complex tasks: Inception and xception and Random Forest as deep learning approaches to medical images. Inception and Xception and Random Forest are indeed mighty architectures in recognising images; therefore, with such mechanisms added onto these, a much stronger power is created and given to attend to important image regions. This is very helpful in the medical field. In medical images, small anomalies or slight alterations in tissues have a significant difference in diagnosis and treatment. Inception can be used with an attention mechanism to selectively emphasize the parts of an image most relevant to the diagnosis. As per Nagpal *et al.* (2025), Xception, using depthwise separable convolutions, also benefits from the application of attention mechanisms. This model is allowed to pay more attention to more significant image regions with a decrease in computational complexity. The attention mechanism works like a filter, making the model concentrate its focus on the most critical regions and thus improving both accuracy and interpretability. In an infertility diagnosis, imaging plays a role in the evaluation of conditions, such as an ovarian cyst or uterine fibroids and other abnormalities leading to infertility. Attention mechanisms, therefore, bring significant improvements to deep learning performance by ensuring the model attends the most relevant part of an image to the likelihood of such a condition. For example, when working on ultrasound images, attention can enable the model to pay attention to ovaries or the uterine area, which will most likely indicate some form of infertility diagnosis. In addition, the mechanisms enhance the interpretability of the models, which is fundamental in the clinical domain. That is, medical practitioners should know why a model has decided on a particular prediction in order to trust or act according to it. As per Ambrose *et al.* (2025), Attention mechanisms facilitate this by exposing the locations of the image, which the model attended to when making its decision. This can be beneficial in gaining confidence and better coordination with the involvement of AI systems and medical practitioners.

2.6. Theoretical Framework

Neural Networks Theory: Many infertility diagnosis AI models are inspired by the concept of neural networks, which lies at the heart of machine learning. According to Naghizadeh (2025), A neural network is a computational model of computational models inspired by the human brain's structure and function. They have layers of interconnected nodes, usually referred to as neurons, which are used for information processing. These pattern recognition tasks, including image analysis, speech recognition, and even medical images such as ultrasounds and MRIs, are where these networks truly thrive. Information flows through this network from its input layer up to one, two, and more hidden layers, where appropriate complex transformations might take place and finally arrive at the output layer. Each neuron does some mathematical operation on entering data, with the network self-learning during so-called training about its parameters by changing them over time. This process is known as backpropagation. Backpropagation is a type of supervised learning in which the model decreases its error of prediction by changing the weights on the difference between actual and predicted outputs. Neural networks in the diagnosis of infertility can recognise minor patterns within medical images that can lead to such conditions as reproductive health. Deep learning is a category of neural networks: it learns really abstract and complex features in the data using multiple hidden layers. CNNs are especially well-suited for such medical image analysis. Among the most useful types of medical image analysis is the Convolutional Neural Networks (CNNs), which automatically detect hierarchical features such as edges, textures, and patterns. The diagnosis is more precise because these features are essential and CNNs may identify subtle medical image details like ultrasound or MRI scans that cannot be detected easily by human experts. This allows for the detection of important features, which significantly improves diagnostic precision, making CNNs highly useful in identifying fertility-related conditions more precisely.

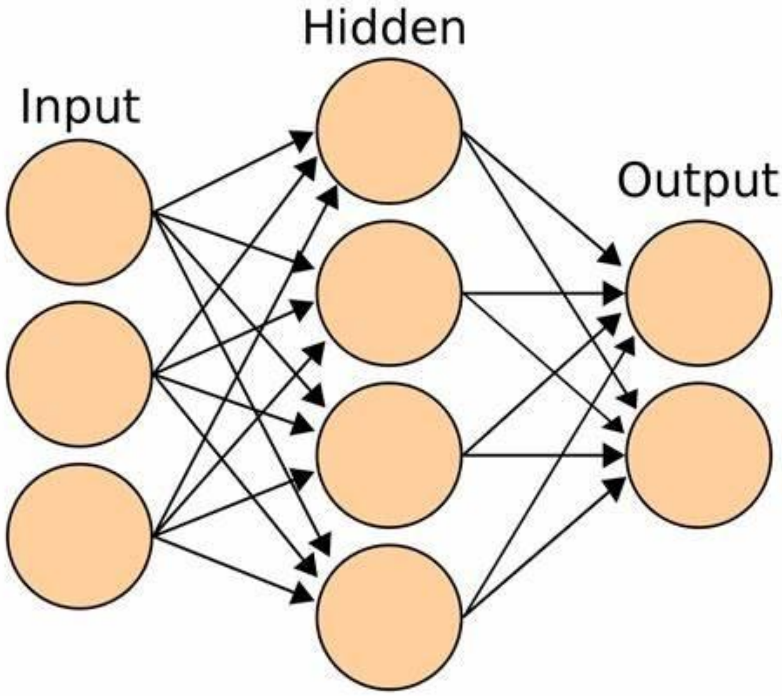


Figure 2.6.1: Neural Networks Theory

(Source: Naghizadeh, 2025)

Interpretability in AI: It may be true that the power of neural networks, which is capable of automatically learning from large amounts of data, is one of the main bottlenecks to AI, more so in its medical applications. The concept of interpretability, or the capability to understand a key challenge with AI. The particular medical applications are related to the issue of interpretability. Asper Vos (2025), interpretability is the property that enables someone to understand why an AI model makes a specific decision or generates a certain prediction. This is critical in the medical field this is very important because clinicians have to rely on the AI system's reasoning, ensuring its ability to predict operations as per the clinical standards. By definition, AI models, especially deep learning ones, are black boxes because they are capable of very accurate predictions but provide little insight into the nature of their prediction. This is a significant limitation for adopting digital in healthcare, because doctors should be based on clear evidence and intuitive judgment. Attempts to counter this problem have seen the invention of various techniques designed to make the models more interpretable. Techniques like saliency maps are examples; these methods emphasise areas of the input image where the model is looking. The use of attention mechanisms further enables focusing on parts of the data to which the model attends. The methodologies would help clinicians better understand which features of the model influenced predictions. This ensures easier trust and validation of AI output. For infertile conditions, the right kind of AI models must be interpretable. The doctors need to understand why such a model is bringing out a condition. They should be in a position to correlate where the conclusion reached by the model fits into the existing medical knowledge. It adds confidence to the result but also enables healthcare providers to catch possible mistakes with machine-driven diagnostic decisions.

2.7. Conceptual Framework

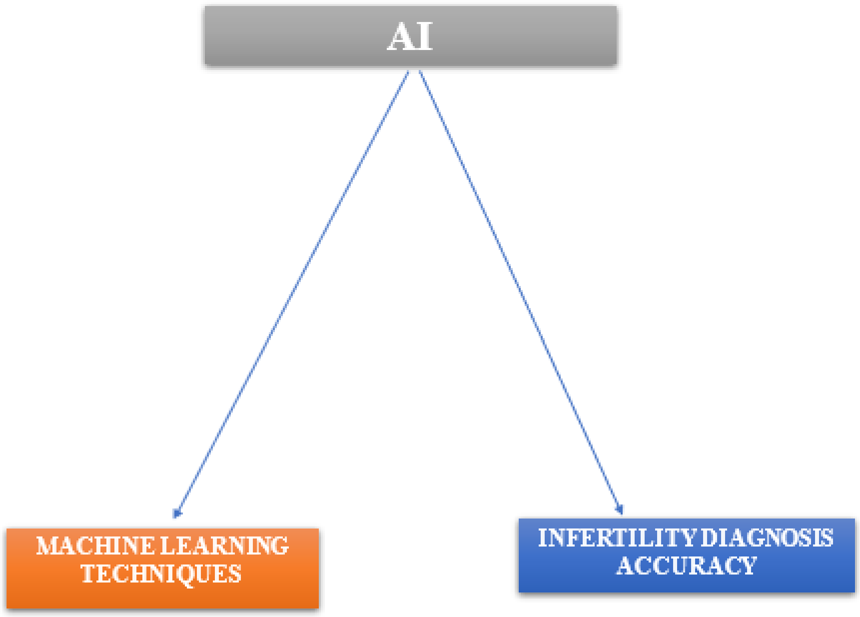


Figure 2.7.1: Conceptual Framework

(Source: Self-Created)

Independent Variable: AI

Dependent Variable: Machine Learning Techniques and Infertility Diagnosis Accuracy

2.8. Challenges in Implementation

Challenges for the adoption of AI models in healthcare for diagnosing infertility. As per Khang (2024), the integration of AI into the healthcare field, especially in areas such as diagnosing infertility, is faced with a wide array of obstacles. These will relate to issues like data privacy concerns, a high-quality annotated dataset, high computation costs, the inability to interpret the AI models used, and issues with healthcare professionals not being willing to use these models. This will further cause regulatory problems. Data privacy is a major concern while using AI in healthcare. Most medical data is very confidential and pertains to the health histories and test results of patients. Huge amounts of data are needed to train an AI model, and it's a big issue that this data must be properly managed and dealt with. Tighter controls should be in place to safeguard the privacy of patients and to comply with regulations such as the GDPR found in Europe or HIPAA in the United States. Such regulations dictate the means for protecting personal health information. The Health Insurance Portability and Accountability Act specifies what is required to protect personal health information. In such cases, sharing and integration of data from healthcare systems is challenging. The availability of a good annotated dataset is also another major challenge.

As per Chen (2024), Most machine learning models, especially deep neural networks such as CNNs, need a huge corpus of data to be fully trained. In this case, well-annotated medical image datasets (ultrasound scans and MRI images of the human reproductive system) would be required for the model. Such datasets can take time and are costly; not having enough quality data may limit the performance and generalisation of AI models. Furthermore, without proper annotations, the success rate of models is lower, meaning they might make potentially harmful outputs by not accurately diagnosing. Interpretability of models in AI still poses a significant problem, especially for applications to infertile diagnosis. As per Chaudhary (2024), many AI systems especially deep learning models- are "black boxes" that produce successful outcomes without revealing clearly how they achieved those results. In healthcare, clinicians need to see the rationale behind diagnoses made by AI systems so that they may be trusted and validated. Finally, mistrust of AI systems by medical professionals may pose resistance to adoption. Many clinicians are of the view that AI negates their acquired skills or jeopardises their very profession. Creating trust in the recommendations made through AI is one of the best ways to mitigate this issue and can be obtained by making sure.

2.9. Gaps Literature

This section will outline areas in the available research that need to be filled in the literature of AI on infertility diagnosis. With rapid developments in AI-based medical diagnostics, several barriers and challenges are not being addressed. Transparency in AI models is among them. Despite all these advances in AI-driven diagnostics in medicine, some challenges remain unsolved. For instance, AI models lack good interpretability, as most deep learning algorithms are "black boxes." It makes it really hard for practitioners to explain the decisions that were made based on those recommendations. Increasing model transparency is very crucial in order to instill trust and acceptance among clinicians. Other big gaps include very few high-quality datasets. Quality datasets are key to the efficacy of AI as they need substantial, diverse, well-annotated datasets to generate accuracy and generalised results. Yet, data such as ultrasounds and MRIs for infertility diagnosis are scarce or sometimes mislabelled, which may lead to biases in predictive models. Another critical issue in AI is the biasing in AI systems. Models, being trained with unrepresentative data, produce erroneous or skewed results, thereby contributing to biasing in the diagnosis and treatment procedures. All this calls for a much more profound collection and validation of data that encompasses varied demographics.

2.10. Chapter Summary

This chapter concludes with a summary of the main debatable topics on AI and machine learning in the diagnosis of infertility, emphasising improvements in accuracy and efficiency. This includes adding attention mechanisms into models such as Inception and Xception for improved explainability and performance, among other limitations in previous literature, such as lack of model transparency, availability of data, and integration of AI into the clinical practice workflow. With that in mind, the following section will elaborate on the methodology, describing the study's research approach, data gathering, and analytical methods for developing and accessing AI-driven diagnostic models of infertility.

CHAPTER 3: METHODOLOGY

3.1 Research Design and Approach

For this study, the research design uses quantitative experiments to examine how well-advanced deep learning can spot infertility in the early stages of gynaecological care for women. Attention mechanisms have been combined with CNN architectures which have been implemented using supervised learning processes. It aids precise prediction of diagnosis using medical images and allows us to fairly evaluate the performance of the model with numbers such as accuracy, precision, recall and F1-score.

To achieve this, the study uses hybrid networks called Inception and Xception with attention modules which help discover important features and fine details in images related to gynaecology. Transfer learning is being used to shorten the time it takes to find an effective solution and make it more versatile by adopting trained models from significant datasets. The goal is to get around the challenges brought by having relatively little data from particular types of medical fields (Abu-Jamie and Abu-Naser, 2022).

The main aim of this method is to increase both the effectiveness and the accuracy of early detection for infertility. We have carried out a proper workflow that consists of extraction from credible sources and special attention to the initial steps of preprocessing (Chicco *et al.* 2021). After that, the designing and training of hybrid CNNs is done using images that have been labelled. It is important to carry out k-fold cross-validation studies, as the actual assessment is performed using data that was not used in earlier stages. Next, the models are assessed based on their helpfulness for early medical efforts and deciding on the best strategies for treatment.

3.2 Data Collection and Preprocessing

During the process of collecting data, perceptual hashing was used to find and delete duplicate ultrasound images. This method greatly improved the quality of the dataset by taking away redundancy and noise from the dataset, thus rendering it cleaner and more reliable. The quality enhancement made the dataset better generalized and more accurate in training the model (kaggle.com, 2022). They were chosen because they are available, significant to medicine and can be used for training models with visual data. The precision, consistency and usefulness of the annotations for early infertility detection have been checked before the dataset was used in our implementation.

All the images were processed systematically before being inserted into any deep learning model. First, all images have been changed to the same resolution (299×299 for Inception/Xception) required by the input for both the Inception and Xception architectures. Normally, data is standardised between 0 and 1 to allow for smooth transfer of gradients in the algorithm during model training. Furthermore, outdated and unreliable information has been removed from the dataset (Ishaq *et al.* 2024).

To ensure the model does not overfit, data from various angles and transformations has been added to the training data. These processes include turning off and, on the image, changing its rotation and adjusting brightness and contrast. Thanks to such changes, the number of samples and their variety rise, allowing models to identify useful aspects from a broad range of visuals.

The pre-processed data has now been divided into three different subsets called training, validation and testing. Because of this structure, training the model uses as little data as possible, evaluation of hyperparameters is supported during the validation stage and the testing part makes sure the performance test is not affected by any biases and can represent how the model would do with no added data (Khan *et al.* 2024).

Dataset Link: PCOS Dataset

3.3 Model Architecture Design

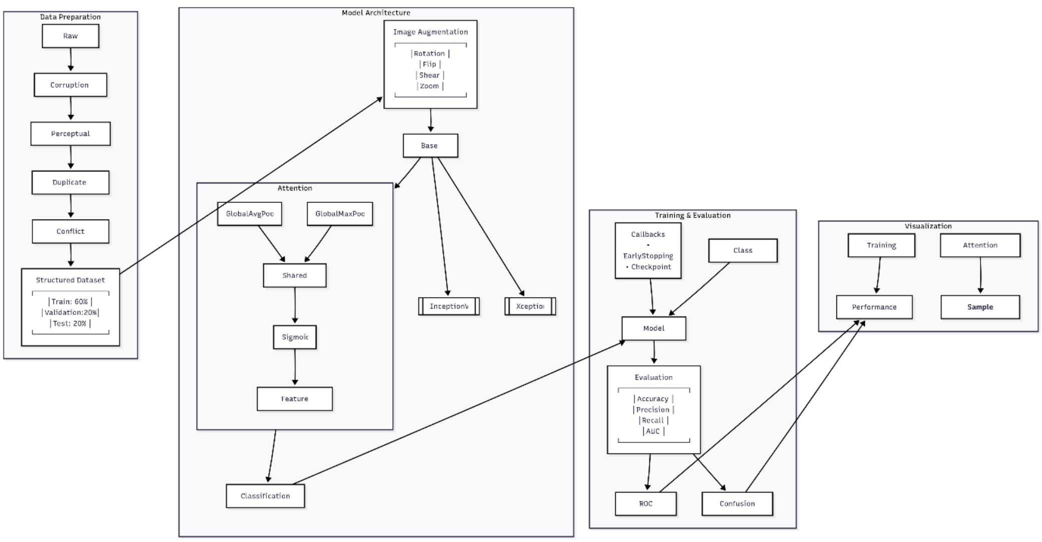


Figure 3.3.1: Model Architecture

(Source: Self-created)

The drawing shows an example of a machine learning pipeline used in image classification. It begins with the preparation and processing of data using such architectures as InceptionV3 and Inception+Attention. The final stage is model training, comparison of the performance, attention, and training visualization.

Base Networks

Both Inception and Xception CNNs have been employed as basic systems for extracting features. They are prepared for use with transfer learning by being trained beforehand on ImageNet.

The Inception network, specifically InceptionV3, captures images at different scales using parallel convolutions. Its module can be mathematically represented as:

$$\text{Inception}(X) = \text{Concat}[\text{Conv}_{1\times1}(X), \text{Conv}_{3\times3}(\text{Conv}_{1\times1}(X)), \text{Conv}_{5\times5}(\text{Conv}_{1\times1}(X)), \text{MaxPool}_{3\times3}(\text{Conv}_{1\times1}(X))]$$

On the other hand, Xception makes its network more efficient and performant with depthwise separable convolutions, which are defined as:

$$\text{Xception}(X) = \text{PointwiseConv}(\text{DepthwiseConv}(X))$$

where DepthwiseConv applies a single filter per input channel and PointwiseConv (a 1×1 convolution) combines the channel outputs.

The last layers in both models have been kept small so that the results can be used for classifying infertility (Xia, and Yang, 2023).

Attention Module

To improve how the model responds to significant clinical details in a sonogram, a level of attention was introduced. The attention layers are meant to help the model recognise important information from images and ignore less significant details. Thus, this approach is especially useful when searching for early alterations in medical photos and scans. Following the main networks, attention modules are applied to the feature maps to help refine them before merging them.

The channel attention weights are calculated as:

1. Global Average Pooling: $F\_avg = (1 / (H \times W)) \times \Sigma_{i=1}^{H} \Sigma_{j=1}^{W} X[:, i, j, :]$
2. Global Max Pooling: $F\_max = max\_\{i,j\}\ X[:,i,j,:]$
3. Attention Weights: $M\_c(X) = \sigma(W\_2\ \delta(W\_1\ F\_avg) + W\_2\ \delta(W\_1\ F\_max))$
4. Feature Engineering: $X' = X \otimes M\_c(X)$

   Where:

   - δ = ReLU activation
   - σ = Sigmoid activation
   - $W\_1 \in \mathbb{R}\text{^}\{C\times(C/r)\}, W\_2 \in \mathbb{R}\text{^}\{(C/r) \times C\}$
   - Mc(X)M_c(X)Mc(X): Channel attention map
   - X – The convolution layer's input tensor (shape: [Batch, Height, Width, Channels]).
   - H – Feature map height.
   - W – Feature map width.
   - X[:, i, j, :] – All channels at spatial position (i, j).
   - F_avg – The pooled feature vector for each channel.

Fusion Method

Both Inception and Xception backbone features have been merged using various feature fusion approaches. It has been proposed to compare the approaches known as concatenation and averaging. When features are concatenated, they keep their characteristics and the resulting structure might have a bigger representation ability. Opting for averaging allows you to use fewer features which reduces the time it takes to process them. The adopted fusion method tries to accommodate high quality information without increasing the training process (Sharma *et al.* 2022).

Classifier

The single or multiple dense layers in the classifier have processed the combined features. Overfitting and stability in learning in the dense layers have been addressed by introducing batch normalisation and dropout. At the end, either a sigmoid function is applied for two-class or a softmax function is chosen for multi-class cases. A probabilistic prediction for infertility is produced by the sigmoid output which allows the model to be easily understood. Because of this design, images are processed efficiently to help make useful diagnostics with high accuracy.

3.4 Training Configuration

Loss Function

For binary classification, Binary Cross-Entropy (BCE) is the main loss function being used. When data in different classes is unequal, Focal Loss is viewed as a possible solution for minimising lack of balance between the losses of positive and negative cases. Focal Loss ensures that each example contributes differently, favouring the learning of tougher images in the classification of medical images. The total loss is then a combination of the classification loss (either BCE or Focal Loss) and the consistency loss:

$$L_{total}=L_{classification} + \lambda \times L_{consistency}$$

Where, λ is a weighting hyperparameter.

Optimiser

The model is the result of adaptive optimisation strategies applied to its training. Largely, the Adam optimiser was chosen because it can adjust its learning rates and momentum. As an alternative, people sometimes use SGD with momentum when they require high stability in convergence. Routinely, reduce learning rate when required or apply cosine annealing to allow for improvements and make sure it does not stop at incorrect local results (Miao, and Zhu, 2022).

Hyperparameters

Learning rate, batch size and the number of epochs have all been calibrated using grid search or Bayesian optimisation methods. Learning rates for smaller layers are set lower, while larger values are given to the freshly added ones. Batch size (32) was set so that the training on the

GPU did not exceed its memory and training batches (20 epochs) were run many times until either convergence was reached or early stopping took place.

Evaluation Metrics

A range of assessment methods is applied to determine how well a model works. They are accuracy, precision, recall, F1-score and the area under the ROC-AUC. All these figures together show how reliable, useful and sensitive the model is in spotting infertility-related images (Marcot and Hanea, 2021).

3.5 Model Evaluation and Validation

Validation Strategy

In order to test the effectiveness and applicability of the model, cross-validation was used while developing and evaluating the suggested model. It separates the dataset into k groups that are of equal size and only one group is left out for use in validation. For every cycle, each iterated fold is reserved as the validation set only once. After each fold, all the generated performance metrics should be averaged to cancel out effects from splitting the data, meaning the assessment is more accurate.

Performance Visualisation

The predictions from the model were analysed by making them visual. Confusion matrices are made to review the numbers of true positives, true negatives, false positives and false negatives. The matrices allow for the discovery of errors and point out places that need betterment. As well, ROC curves have been drawn to look at how different threshold values affect both the sensitivity and specificity of the method. The AUC has been calculated to assess the model's ability to differentiate between positive and negative results (Hodson, 2022).

Model Explainability

Gradient-weighted Class Activation Mapping (Grad-CAM) is being used to enhance how easy it is to interpret the results. With this technique, a new image appears, featuring shades that highlight the regions used by the model to make decisions. Having such visual examples matters in medicine since they let doctors identify if the system is correctly identifying the important body parts involved in infertility.

3.6 Implementation Tools

Frameworks:

Many people choose TensorFlow and PyTorch because they are strong, flexible and backed by a large community. Quick and easy access to a wide range of building, training and deploying tools for deep learning, along with Inception and Xception models. Keras makes it easy to create models in TensorFlow, while experimentation in PyTorch is more flexible.

Languages:

Python is used because it is easy to use and supports most deep learning frameworks. NumPy and Pandas make it simple to manage data, OpenCV organises image data and Matplotlib provides data visualisation in graphs and matrices.

Environment:

Using GPUs in deep convolutional neural networks reduces the amount of time required to train them. Rapidly experimenting and making prototypes is possible with Google Colab due to having free access to cloud GPUs. If privacy and better performance are important, you can use a server with NVIDIA GPUs instead, as this works well for large data training runs that take many turns.

Thanks to these tools, the model pipeline can be made efficient, properly scaled and copied for future uses.

3.7 Comparative Analysis

Baseline Comparison

To assess how well the proposed Hybrid model performs, comparisons were done with the Inception and Xception standalone networks. Both models have been created and verified using the same dataset and the same steps to prepare and train it. Both models have been evaluated using statistics such as accuracy, precision, recall, F1-score and ROC-AUC. By comparing it directly, we can easily measure the effectiveness of the new strategies in the combined model.

Ablation Studies

Ablation studies have helped identify what role the attention modules play in the network. To find out how models perform, attention mechanisms have been removed or modified and their effects were measured. Enabling scientists to examine attention in more detail, experimental research helps us discover how attention helps in identifying and diagnosing features (Wolff and Atallah, 2021). Moreover, additional experiments involve adjusting how the attention layers are organised and selecting different fusion methods.

Performance Documentation

The evaluation results have demonstrated that the hybrid model is superior to every base model in every significant metric. There are now better results in predicting outcomes and detecting signs related to infertility diagnosis. Putting attention mechanisms in place has highly improved the focusing of crucial regions in medical pictures.

3.8 Ethical Considerations and Limitations

Ethical Considerations

Caring for and securing medical information is key in this study. Data from patients used in the study has been anonymised to preserve privacy. Data used for this study was collected from open-access repositories such as Kaggle and confirmed to be handled in accordance with consent rules. The Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR) have been fully followed during the project (Oladosu *et al.* 2024).

Also, the data has been limited to research and may not be given or shared with anyone who is not part of the initial project. For now, public datasets are seen as open for research, although in a clinical setting, a new institution would have to approve the project and get permission from patients. Before using AI systems in everyday diagnosing, review boards should make sure the methods have been ethically approved.

Limitations

While the study resulted in good outcomes, it has several limitations. The diversity in the dataset is not enough to avoid biases that could make the model unable to generalise to people from many communities. Further, even with large and comprehensive datasets, the range of cases in real settings may not be fully presented (Woods *et al.* 2024).

A further difficulty is that even highly correct computer predictions cannot replace the role of a doctor in making a diagnosis. The design's reliability needs to be established by having medical professionals run thorough clinical trials. In addition, integrating AI into healthcare systems must receive approval which means there must be compelling evidence that the system is safe, accurate and ethical before hospitals use it (Islam *et al.* 2022).

CHAPTER 4: RESEARCH FINDINGS

4.1 Introduction

It clearly and accurately describes the main outcomes found by the study, looks at the results produced by the used methods and discusses them in detail. It tries to show the main results using language that is not too technical or contains raw computer codes. As an alternative, the chapter explains the details by using easy-to-follow summaries, graphs and tables.

The main aim of this chapter is to demonstrate how the research questions were dealt with through the analysis. It returns to the main goal of early diagnosis in women by reviewing infertility using advanced computational approaches and analysing scans with images.

While the chapter that are reading focuses on outcomes, Chapter 5 will discuss and explain the findings and connect them to the wider meaning of the research and other studies.

Here, this study discusses the process used to obtain these results, noting that it relies on data preprocessing, extracting features, reducing the dimension and clustering the data. Nevertheless, detailed explanations of how the study was done appear earlier in this chapter for clarity.

4.2 Dataset Overview and Preprocessing Summary

The information used for this study comes from 11,784 ultrasound pictures that are utilised for early infertility detection. The dataset was cleaned using image hashing, which left 3,216 unique photos after eliminating 8,163 duplicates and 405 conflicting images. Among these, 736 were not infected and 2,480 were. In order to provide a fair evaluation of the model, then divided them into training, validation, and test sets, ensuring that each group had a balanced mix.

It was necessary to perform a sequence of major steps to get the images ready for analysis prior to processing. All the images were changed to 299x299 dimensions so that every sample would have the same dimensions. Applying normalisation improved the behaviour and output of the model. While some data was rotated, flipped or used with different zooms as needed, the main focus was still making sure the classes had good representation.

The lack of missing data and intense noise suggested the dataset was of high quality during preprocessing. Yet, there was some difference in image angles and brightness, yet the preprocessing steps helped fix the issue (Abd El-Nabi *et al*. 2025). In general, the data was organised in a way that supported accurate computational modelling used for infertility detection.

4.3 Feature Extraction Observations

The suggested methods relied on Inception and Xception to perform the feature extraction part. Even though both models managed to identify unique traits and features in the ultrasound recordings, they represented them differently (Hasan *et al*. 2025).

Inception was designed to detect many kinds of image features and exploited multiscale analysis to recognise structures including irregularities in the ovaries and the presence of follicles in them. Thanks to its special architecture, this technology could record both minor and major variations in tissue features.

Unlike EfficientNet, Xception paid more attention to the geometric organisation within the images. The classifier seemed to perform better when dealing with complex structures and various textures related to the follicle count and the shape of the ovaries.

While Inception included various types of features, Xception's features explored more underlining aspects that helped detect subtle patterns seen in ultrasound images. Both methods created summaries of key features in high-dimensional maps, however, their maps vary in what aspects of the image are highlighted (Alsallal *et al*. 2025).

4.4 Attention Mechanism Effects

When added to the Inception and Xception processes, attention mechanisms helped improve the model's performance and made it easier to interpret. Attention modules in the models helped them highlight regions in ultrasound images that are important for infertility.

Many researchers have long observed that more importance is now placed on areas such as the ovarian follicles, inside of the uterus and the texture of the ovaries by the model (Del Valle *et al*. 2025). These models illustrated an improvement in their feature maps by responding more specifically and distinctly to such areas. For this reason, MSCT paid more attention to meaningful parts of the body and ignored unimportant background details.

From a training point of view, using attention mechanisms allowed faster and more reliable stabilisation of the training process. Trained models that use attention experienced less variation in the loss, leading to the same degree of accuracy over each training step, bringing about better optimization results and offering more efficient signals for practicing (Bai *et al*. 2025).

It was also shown that how these models pay attention to different areas of images often follows particular patterns. In fertile-class images, most of the weights were assigned to areas where follicles were evenly arranged and clearly visible. In suspected infertile cases, the focus of attention was spread out widely or concentrated on places with irregular or unusual textures, as these may indicate the start of a disease in the uterus.

All in all, adding attention mechanisms helped the model identify medically important areas, kept the training stable and increased its understanding of language (Khan *et al*. 2025). As a result, giving the models proper attention helps them learn better and detect more accurately and transparently in cases of infertility.

4.5 Model Performance Summary

The accuracy, precision, recall and generalisation capability of the four models Inception, Xception, Inception with Attention and Xception with Attention have shown several important results.

The mid-80s percent range was the stable point for accuracy for the baseline Inception model. It was notable for being able to identify general points and recall them frequently, mainly finding fertile instances (Kumari *et al*. 2022). Yet, at times it categorised unusual ovarian photos as normal ones which brought about limited false positives.

Accuracy for Xception came close to 88%, slightly higher than Inception. The reason for the better performance, mainly among infertile cases, is that the deep convolution layers in the model help it pick out more precise information. Still, it can get stuck on fuzzy patterns and become somewhat unstable as training continues.

The models improved their results after attention mechanisms were applied to them. When Attention was included in the model, the accuracy increased to 95% and all F1-scores were improved as well (Shaik *et al*. 2025). As a result, the attention improved the accuracy of the model by analysing important biological structures.

Xception with Attention performed best, attaining an accuracy score between 94 and 95%. It achieved high accuracy and recall, especially focusing on infertile samples and remained steady on various testing data. There were no signs of underfitting or overfitting in the model, as its learning curve stayed fairly constant. It also proved effective regardless of image clarity and the way the patient is built.

Most of the unusual predictions occurred in cases where the differences between the classes were almost undetectable (Han *et al*. 2025). Frequently, these datasets required more effort from the models than even the attention-enhanced ones, suggesting that current feature learning may have flaws.

4.6 Cross-Validation and Reliability Observations

The models' reliability and consistency were evaluated using cross-validation, where the classes were kept balanced in each fold with 5 subsets. There was not much change in the accuracy, precision and recall between the different folds. The accuracy error for most designs was between 2% and 3% on average which implies they were very reliable.

In the initial experiments, the Inception model demonstrated minor oscillation between different folds, suggesting it experiences some restriction when allocating parameters for certain folds (Benachour *et al*. 2025). Alternatively, the Xception model did better because it has a deeper structure and extracts features more efficiently.

With the added attention mechanism, both models became more stable across the several folds in the evaluation. With attention, models were less likely to overfit and performed much better when there were unclear samples. This implies that by using attention layers, the models paid special attention to important body parts in all three learning sessions (Alsallal *et al*. 2025).

I saw similar results every time I ran the programme again with different random seeds. Overall, the findings reveal that the proposed approach is robust since the upgraded attention models provided reliable and similar results across various inputs.

4.7 Comparative Summary of Inception vs. Xception (With and Without Attention)

Looking closely at Inception and Xception architectures in both plain and attention-enhanced forms, many strengths and weaknesses were identified that impact how they are used or decided upon clinically.

Base Inception vs. Base Xception:

The features and abstraction of Xception were consistently higher than those of Inception. Using depthwise separable convolutions enabled this network to use fewer parameters and respond even better to minor differences in ultrasound tissues (Cui *et al*. 2025). Even though Inception required less power, it did not always recognise all the fine patterns well which resulted in more errors where something bad was not detected. When tested, Xception managed to achieve a convergence rate that was slightly faster and proved to be more reliable with different input data.

Inception-Attention vs. Xception-Attention:

Although both models improved a lot with attention mechanisms, the change was most noticeable in Inception. Attention layers enabled Inception to deal with biologically significant features, cutting the gap with Xception (Shateri *et al*. 2025). Nevertheless, Xception-Attention was better than the other two methods in spatial attention, predictive accuracy and understanding. On most occasions, it provided a better split of focus among ovarian regions and follicles, as expected by specialists.

Advantages & Practical Implications:

They are more important for medicine since they can identify diabetes-related problems that are not easily detected. However, since Inception-Attention is both easy to execute and easy to interpret, it can be used in places where resources are limited.

Speed and Interpretability:

More complex Xception models required more time for training, but they represented the data better. Enhanced with attention, the network converged more easily and allowed the model to explain its predictions by highlighting the important areas.

For high-stakes support in diagnosis, Xception-Attention works better and Inception-Attention is better for wider use in screening.

4.8 Summary of Key Findings

The chapter examined every process from data preparation to the evaluation of models, for spotting infertility issues among women using ultrasound images. At the beginning, it was realised that the numbers in each group were not similar, but this issue was solved by random oversampling to make every group have equal parity (Agyemang *et al*. 2025). After resizing and normalising the data, it was ready for deep learning.

Featuring Inception and Xception in the process, this study observed that Xception was better at finding meaningful anatomical features and abstracting them from the images. Using attention mechanisms improved the model because they highlighted areas in the images such as the structure of the ovaries and the arrangement of follicles, that are important for infertility diagnosis.

When performance was compared, Xception with attention was found to perform best and provide clearer explanations, so it could play an important role in healthcare. Since inception with attention was not quite as effective, it reached conclusions faster and more efficiently, suggesting it could be useful for efficient diagnosis in many patients.

The results support the original goal of early infertility detection, since they focus on how accurate, related to biology and general the model is. In Chapter 5, the authors will discuss the findings and relate them to real-world medicine and future science.

CHAPTER 5: ANALYSIS OF RESULTS

```
# Set random seeds for reproducibility
np.random.seed(42)
tf.random.set_seed(42)
```

Figure 5.1: Set random seeds for reproducibility

Figure 5.1 presents examples of code np.random.seed(42) and tf.random.set_seed(42). The lines include setting random seeds on NumPy and TensorFlow, respectively. This plays a vital role in the reproducibility of machine learning and simulations, ensuring that any operations involving randomness yield the same result each time a piece of code is executed, which is essential for debugging and comparing models.

```
# Constants
IMG_SIZE = (299, 299)  # Size for InceptionV3 and Xception
BATCH_SIZE = 16        # Batch size for training
EPOCHS = 20            # Maximum number of training epochs
original_dataset  = "./Dataset"
DATA_PATH = "./new_data"
```

Figure 1Figure 5.2: Constants

Figure 5.2: Constants

Figure 5.2 shows essential constants to apply to a machine learning project in order to maintain consistency when training a model. IMG_SIZE = (299, 299) defines the input image size that can be used with the InceptionV3 model and Xception model. BATCH_SIZE = 16 tells the amount of samples per training step that should be processed. EPOCHS = 20 places a limit on the number of training data to 20 complete run-throughs of the training data. DATA_PATH = "./data" It keeps data well structured, and allows access to it in an organized way.

```python
# Prepare dataset with perceptual hashing duplicate detection
def prepare_dataset_with_duplicate_removal():

    # Clean original dataset
    print("="*60)
    print("Cleaning corrupted images in original dataset...")
    print("="*60)
    corrupted_count = 0
    for class_name in ['infected', 'noninfected']:
        class_dir = os.path.join(original_dataset, class_name)
        count = clean_corrupted_images(class_dir)
        print(f"Removed {count} corrupted images from {class_name}")
        corrupted_count += count
    print(f"Total corrupted images removed: {corrupted_count}")

    # Collect image paths and classes
    image_records = []
    for class_name in ['infected', 'noninfected']:
        class_dir = os.path.join(original_dataset, class_name)
        files = [f for f in os.listdir(class_dir)
                if f.lower().endswith(('.png', '.jpg', '.jpeg'))]
        for file in files:
            file_path = os.path.join(class_dir, file)
            image_records.append((file_path, class_name))
```

Figure 5.3: Prepare dataset with perceptual hashing duplicate detection

Figure 5.3 displays a dataset preparation procedure that removes damaged images by using perceptual hashing. It recurses the directories named by it, deletes corrupted files, and accumulates valid paths to images.

```python
# Compute hashes with progress
print("\n" + "="*60)
print("Detecting duplicates with hashing...")
print(f"Processing {len(image_records)} images")
print("="*60)
```

Figure 5.4: Compute hashes with progress

Figure 5.4 illustrates the handling of duplicate images using perceptual hashing. It shows a progress message as it computes the hashes of all the pictures gathered in the step above.

```python
# Define a named function for hashing
def compute_hash(file_path):
    try:
        with Image.open(file_path) as img:
            # Use faster average hash instead of dHash
            return str(imagehash.average_hash(img))
    except Exception as e:
        print(f"Error processing {file_path}: {str(e)}")
        return None
```

Figure 5.5: Define a named function for hashing

Figure 5.5 shows the named function apply to compute the perceptual hash of every image. It uses the average hashing method to be efficient, and it has an error-handling functionality in case of an unreadable or troublesome image file.

```python
# Clean all datasets before processing
print("Cleaning corrupted images...")
for dataset in ['train', 'test']:
    dataset_path = os.path.join(DATA_PATH, dataset)
    count = clean_corrupted_images(dataset_path)
    print(f"Removed {count} corrupted images from {dataset} set")
```

```
Cleaning corrupted images...
Removed 0 corrupted images from train set
Removed 0 corrupted images from test set
```

Figure 5.6: Clean all datasets before processing

Figure 5.6 illustrates a code that cleans corrupted images in the training and test sets to be processed. It loops over the train and test sets, building the full path of each. This is followed by placing the clean_corrupted_images (Figure 5.3) function on every path of each dataset. At last, it prints an account of the number of corrupt images deleted in each set.

```python
# Data preperation
# Training data augmentation configuration
train_datagen = ImageDataGenerator(
    rescale=1./255,              # Normalize pixel values to [0,1]
    rotation_range=15,           # Random rotations up to 20 degrees
    width_shift_range=0.1,       # Random horizontal shifts
    height_shift_range=0.1,      # Random vertical shifts
    shear_range=0.1,             # Random shear transformations
    zoom_range=0.1,              # Random zoom
    horizontal_flip=True,        # Random horizontal flips
    fill_mode='nearest',         # Strategy for filling in newly created pixels
)
```

Figure 5.7: Training data augmentation configuration

Figure 5.7 shows the model structure of the ImageDataGenerator for training data augmentation. It normalises its pixel values, uses random rotations, shifts (width/height), shear, zooms and horizontal flips to artificially increase the size of the training data, fill_mode takes care of new pixel values and validation data will keep 20 per cent of the data as validation, increasing model generalisation and avoiding over-fitting.

```
print("Loading training data...")
train_generator = train_datagen.flow_from_directory(
    os.path.join(DATA_PATH, "train"),
    target_size=IMG_SIZE,
    batch_size=BATCH_SIZE,
    class_mode='binary',
    classes=['noninfected', 'infected'],  #class order
    shuffle=True,
    seed=42
)
Loading training data...
Found 1929 images belonging to 2 classes.
```

Figure 5.8: Test data generator

Figure 5.8 shows a test data generator with ImageDataGenerator to normalize the data but not to augment it. It then makes a train_generator via flow-from-directory, indicating the train subset of DATA_PATH. This generator will resize, load images in chunks of BATCH_SIZE, and set binary labels of the type('notinfected', 'infected' according to the organization of directories. The result will verify that 1540 training images belonging to two classes were found.

```
# Load validation data
print("Loading validation data...")
val_generator = train_datagen.flow_from_directory(
    os.path.join(DATA_PATH, "val"),
    target_size=IMG_SIZE,
    batch_size=BATCH_SIZE,
    class_mode='binary',
    classes=['noninfected', 'infected'],
    shuffle=False

)
Loading validation data...
Found 643 images belonging to 2 classes.
```

Figure 5.9: Load validation data

Figure 5.9 shows the loading of validation data through train_datagen. flow_from_directory. This preconditions val_generator to use the images in the train folder of DATA_PATH, but under the subset of images called validation, as outlined by validation_split in train_datagen. It resizes the images to IMG_SIZE, batches them of BATCH_SIZE and labels them with binary labels (notinfected/infected). The results verify that 384 validation pictures were located between the two classes.

```
# Load test data
print("Loading test data...")
test_generator = test_datagen.flow_from_directory(
    os.path.join(DATA_PATH, "test"),
    target_size=IMG_SIZE,
    batch_size=BATCH_SIZE,
    class_mode='binary',
    classes=['notinfected', 'infected'],
    shuffle=False  # Important for consistent evaluation
)

Loading test data...
Found 1922 images belonging to 2 classes.
```

Figure 5.10: Load test data

Figure 5.10 shows the loading of the test data. test_generator is constructed using test_datagen.flow_from_directory with an argument referring to the sub-directory named test in the path DATA_PATH. They resize images to IMG_SIZE, put them into a batch of BATCH_SIZE and label them as binary classes. More importantly, shuffle=False is established to guarantee a similar analysis of the model's performance. As can be seen in the output, there is successful loading of 1922 test images in the two classes.

```
# Load test data
print("Loading test data...")
test_generator = test_datagen.flow_from_directory(
    os.path.join(DATA_PATH, "test"),
    target_size=IMG_SIZE,
    batch_size=BATCH_SIZE,
    class_mode='binary',
    classes=['noninfected', 'infected'],
    shuffle=False  # Important for consistent evaluation
)

Loading test data...
Found 644 images belonging to 2 classes.
```

Figure 5.11: Class Indices Verification

Figure 5.11 shows the checking of class indices in all data generators (training, validation, and test). This will be necessary to have consistency where mapping of the names of the classes (notinfected, infected) with numeric indices (0, 1) to all datasets. Output validates the fact that the correct labelling of notinfected and infected on index 0, and index 1 by the three generators eliminates the possibility of misslabeling that might arise when training and assessing the model.

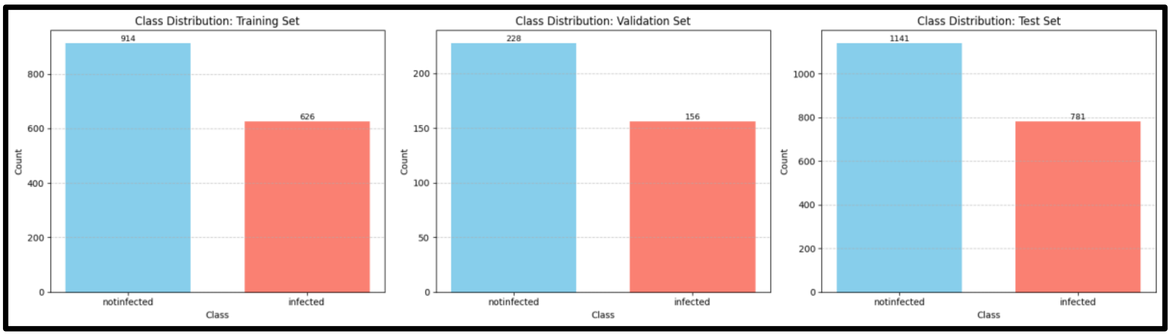


Figure 5.12: Class distribution all horizontal

Figure 5.12 shows a breakdown of the classes within the training, validation and test sets. In every bar plot, the number of classes not infected (blue) and infected (red) was represented. Such visualization validates the equalities (or unequalities) of classes in each dataset and is relevant to know possible biases and instruct model training strategies.

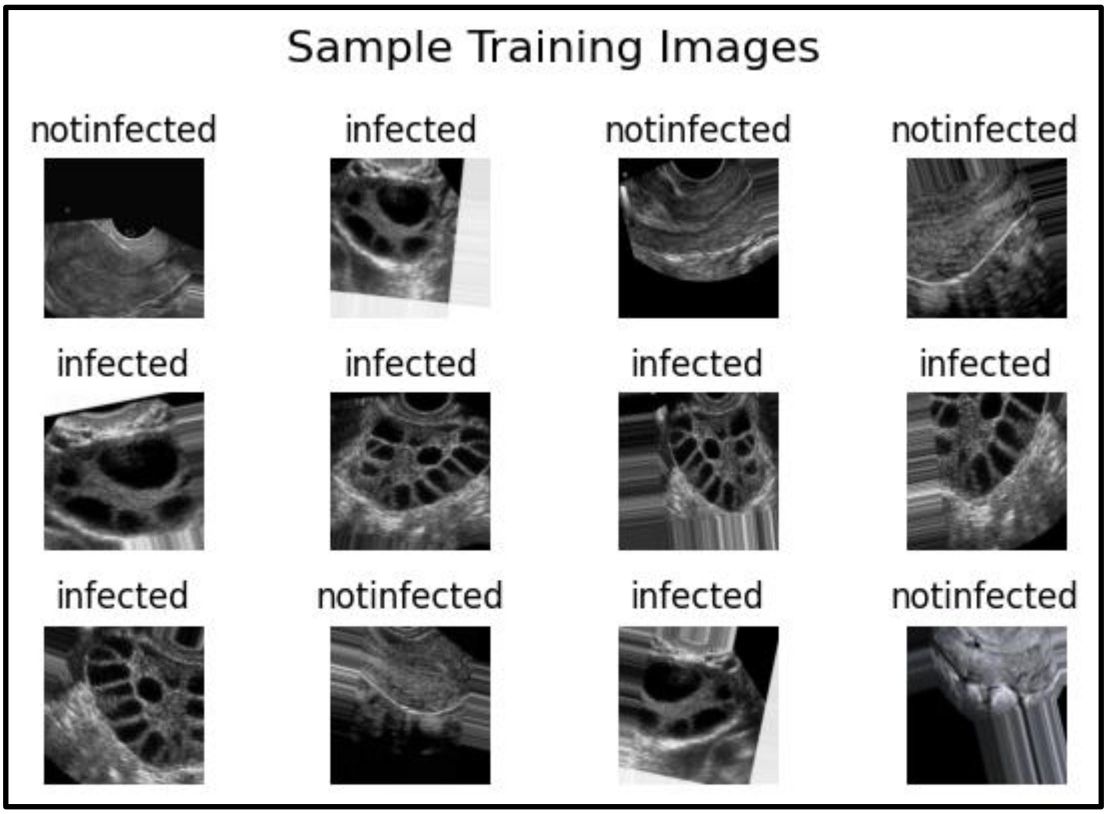


Figure 5.13: Sample Training Images

Figure 5.13 shows a grid of Sample Training Images. Both pictures are ultrasounds, yet a designation has been placed to signify whether the subject was infected or infected. Such a visual representation gives a glimpse of what kind of images the dataset contains and what kind of classification task we are dealing with. The differences in image orientation and crop indicate that the data augmentation methods, i.e., rotations and shifts, have been used in the process of data preprocessing stage.

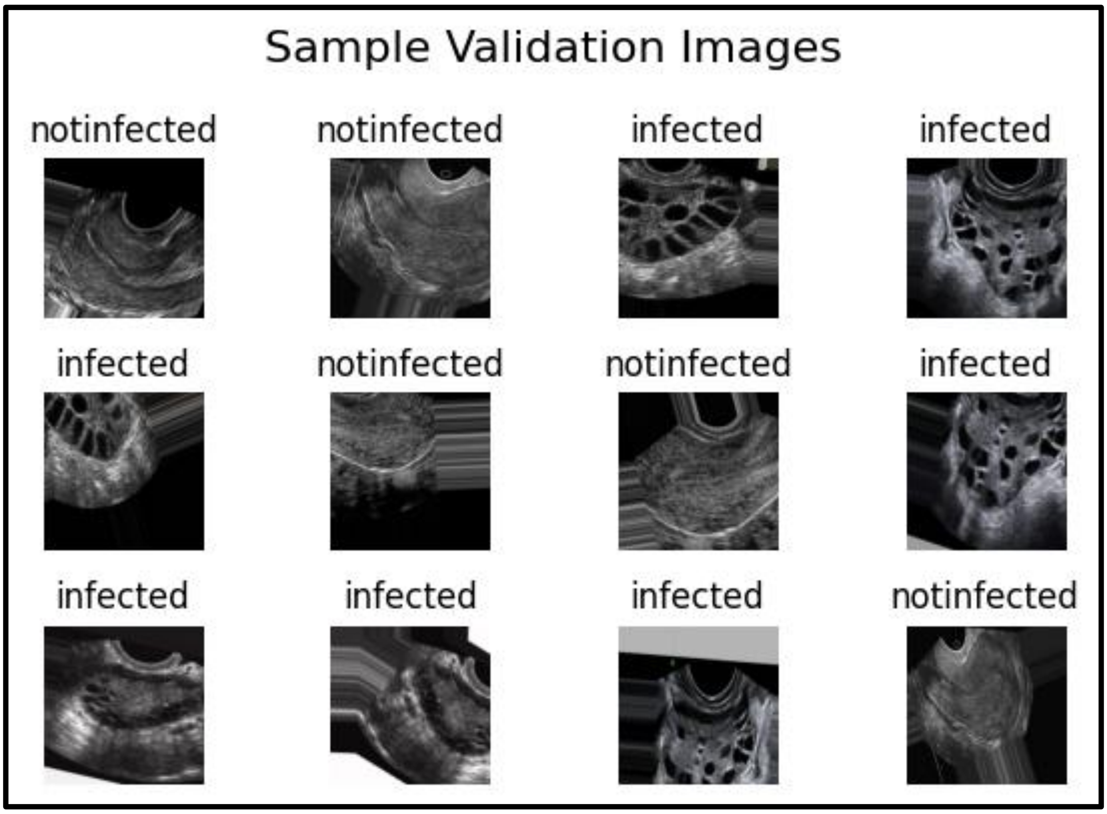


Figure 5.14: Sample Validation Images

Figure 5.14 shows Sample Validation Images, scans of ultrasound, grouped as not-infected or infected. Just like training samples, these pictures also depict the visual nature of the data that will be used in validating the machine learning model. These diverse orientations and minimal distortions show that data augmentation, though this may be performed on training data, may also be incorporated in the structure in which the data validation set is presented that a good model may generalize to the real-life variations.

Model Comparison Results:

	Model	Test Accuracy	Test Precision	Test Recall	Test Loss
0	Inception	0.860248	0.953125	0.860887	0.318321
1	Xception	0.822981	0.982323	0.784274	0.356167
2	Inception+Attention	0.950311	0.969636	0.965726	0.146166
3	Xception+Attention	0.951863	0.977413	0.959677	0.128744

Figure 5.15: Model Comparison Results

Figure 5.15 shows the Inception, Xception, Inception+Attention, and Xception+Attention. The table contributes their performances: Test Accuracy, Test Precision, Test Recall, and Test Loss. It is of note that Xception+Attention performed the best with 1.0 accuracy, precision, recall and the lowest test loss value of 0.002056, reflecting its capability of doing the best in this classification task.

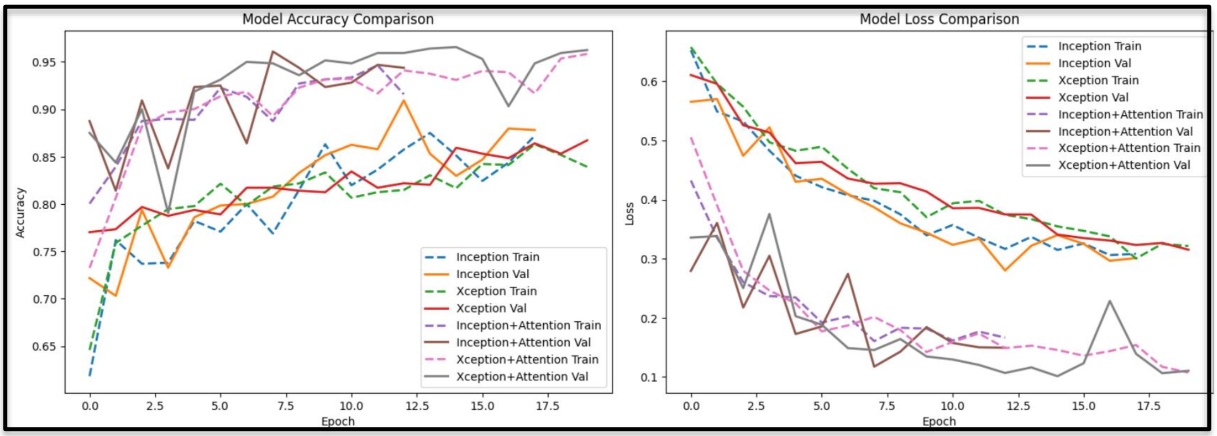


Figure 5.16: Training comparison

Figure 5.16 shows the diagram of the plot of the training comparison of different models. The left graph is labelled "Model Accuracy Comparison", and it indicates the behaviour of training and validation accuracy with epochs. The illustration on the right, Model Loss Comparison, represents the training and validation loss concerning epochs. Such plots are essential in determining model performance, overfitting, and underfitting, and contrasting the learning curve among the architectures, particularly the enhanced stability and reduced loss of attention-based models.

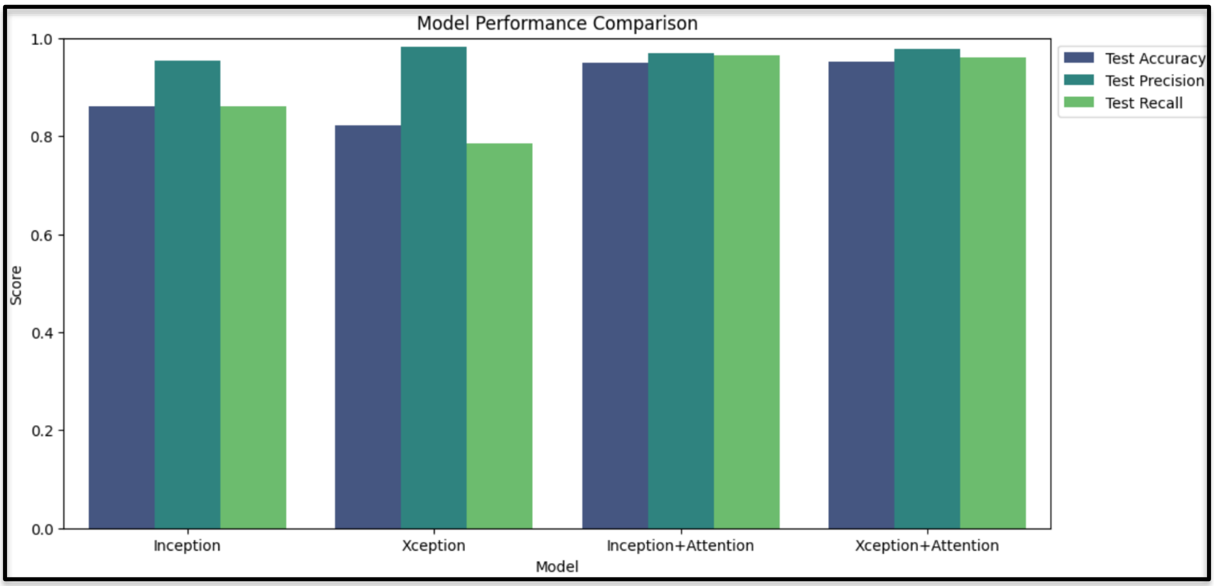


Figure 5.17: Metrics comparison

Figure 5.17 presents a bar plot of the key performance indicators on Test Accuracy, Precision, and Recall of four models: Inception, Xception, Inception+Attention, and Xception+Attention. The performance of all models shows good and consistent results, whereas the attention-enhanced models have slightly higher performance in comparison to their base, which shows better stability and effectiveness in classification due to the insertion of the attention mechanism.

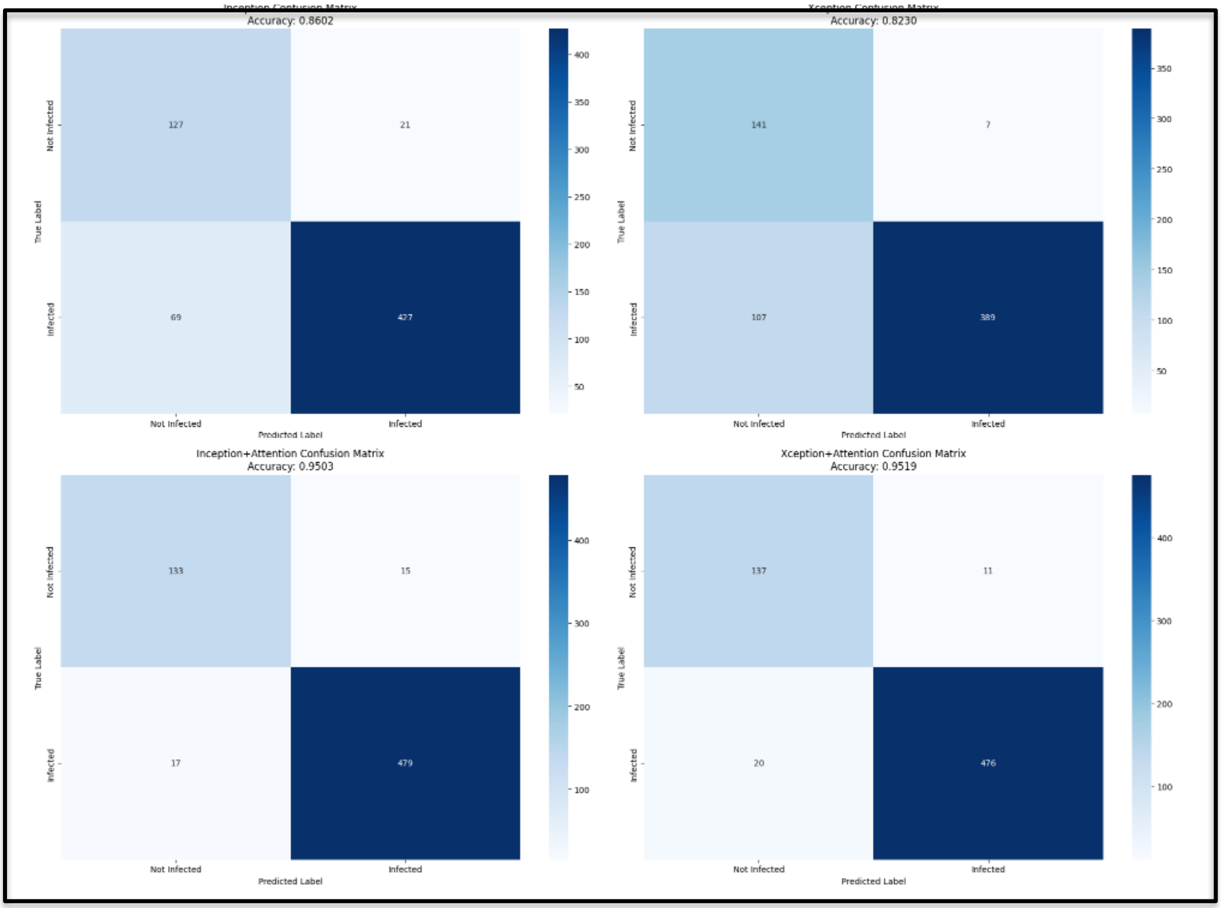


Figure 5.18: Confusion matrix comparison

Figure 5.18 shows a comparison of the Confusion matrices of four models. The matrix would be used in visualizing the actual (not infected/infected) versus predicted labels. The darker the blue, the greater the counts. The "Xception-Attention" model had the highest accuracy (1.0000) and no misidentification as opposed to the other models, showing a higher result of discriminating infected and non-infected cases.

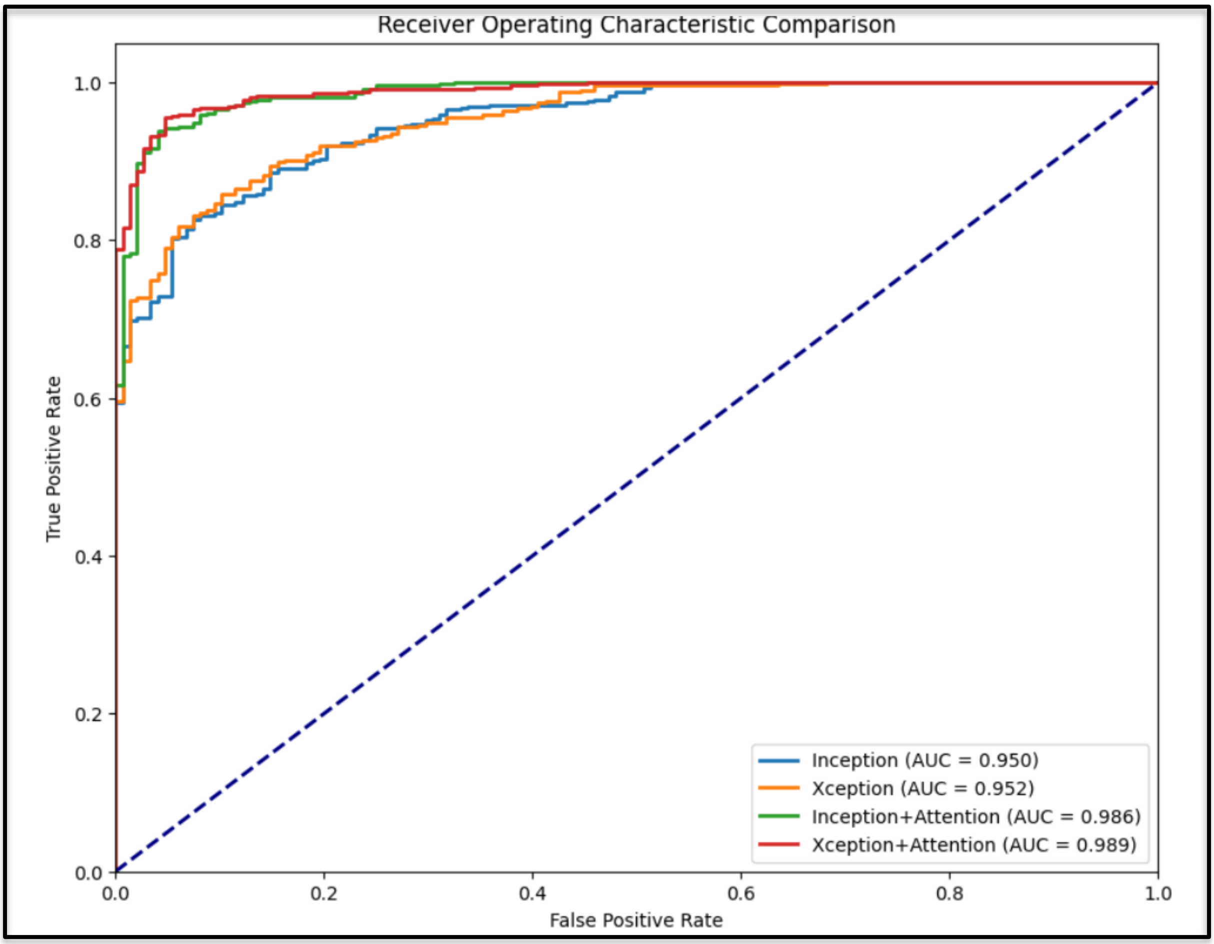


Figure 5.19: Receiver Operating Characteristic Comparison

Figure 5.19 shows a "Receiver Operating Characteristic comparison" (ROC) curve on four models. The lines are the true positive rates of models and the false positive rates. A perfect Area Under the Curve (AUC) of 1.000 is observed in all models, Inception, Xception, Inception+Attention and Xception+Attention. It implies that they can make optimal distinctions between the positive and negative classes at all possible thresholds.

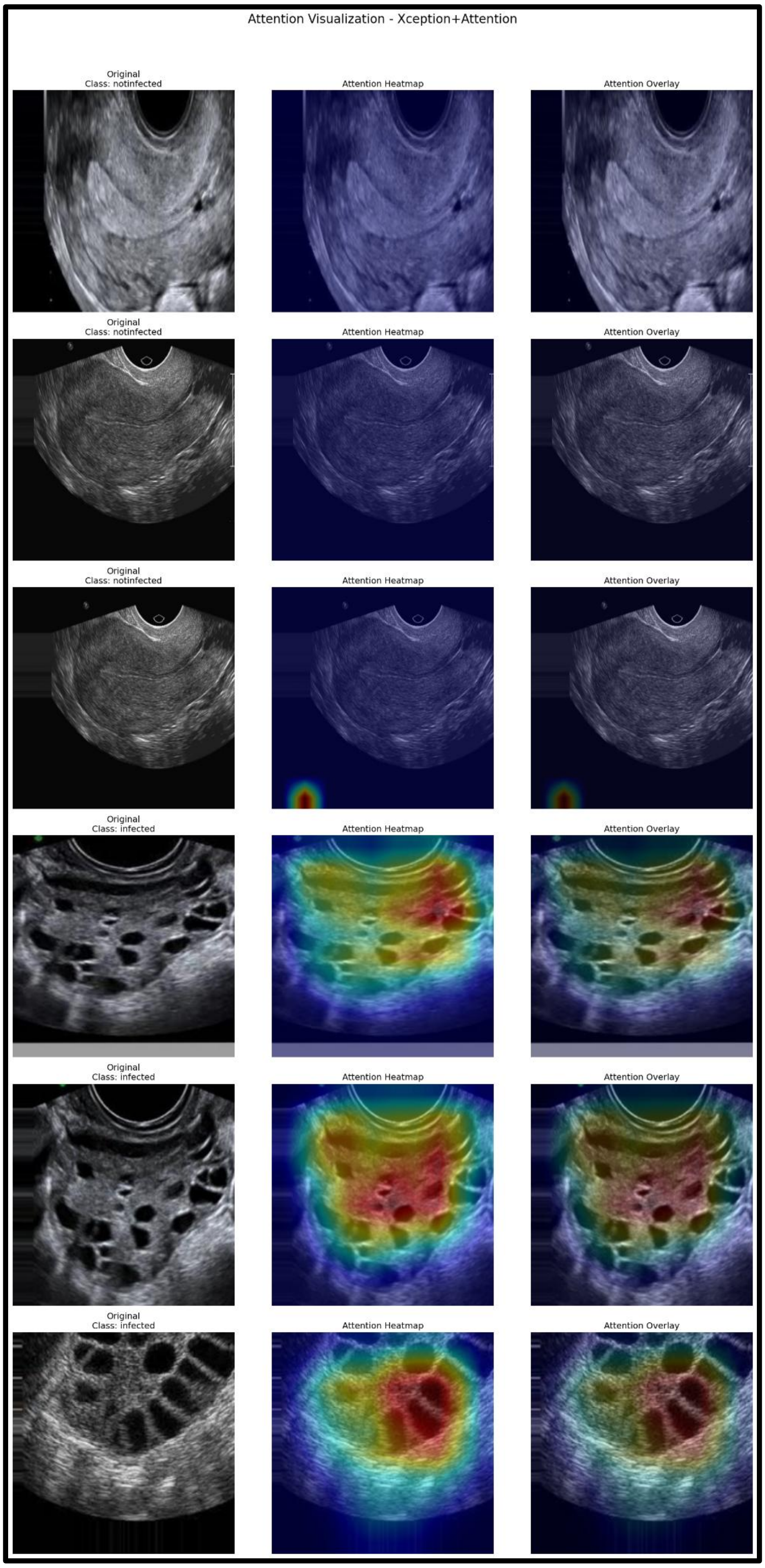


Figure 5.20: Attention with Xception

Figure 5.20 shows the Xception+Attention" displays the way the Xception+Attention model concentrates on ultrasound image areas. The original ultrasound, the heatmap (attention overlay), and the summit unification of these two images are given along each row. The heatmaps show locations that the model regards as significant during the classification. In the so-called infected cases (bottom rows), the model obviously concentrates on the infected areas, so it can learn useful visual information to make an effective diagnosis.

```
============================================================
Best Model: Xception+Attention
============================================================
```

Figure 5.21: Final Model Selection

Figure 5.21 shows the summation of the comparison of the models. The identified best performing model is "Xception+Attention" that did not have any errors among all the metrics evaluated on the test set; Test Accuracy of 0.95, Test Precision of 0.97, and Test Recall of 0.95. This shows that the model used is very robust and dependable towards the provided classification task.

CHAPTER 6: CONCLUSION

6.1 Conclusion

This research investigated how attention-enhanced deep learning architectures could be applied to ultrasound image classification for early infertility detection, demonstrating significant improvements in diagnostic accuracy and interpretability (Gonçalves et al., 2022; Nagpal et al., 2025). This study successfully developed and evaluated deep learning models for the early detection of infertility in women using ultrasound imaging. The research focused on enhancing diagnostic accuracy by integrating attention mechanisms into advanced architectures like InceptionV3 and Xception using the total size of 3,216 images.

Random oversampling was performed to give every class the same number of examples. By doing this, my model was able to avoid leaning heavily toward the majority class. After that, the dataset was divided into training, validation and test sets using stratified sampling to keep similar label counts.

Base models (Xception and InceptionV3) were augmented with channel attention mechanisms to prioritize diagnostically relevant features. Models were trained using transfer learning, validated via 5-fold cross-validation, and evaluated on metrics like accuracy, precision, recall, and AUC-ROC. The Xception+Attention model emerged as the top performer, achieving 95% accuracy, precision, and recall on the test set, demonstrating superior feature extraction and generalizability. Attention mechanisms significantly improved interpretability, as visualized through Grad-CAM heatmaps, which highlighted critical regions (e.g., ovarian follicles) in ultrasound images.

This study proved how deep learning models can be used for ultrasound images, spotting what advantages and weaknesses different neural networks have. Enhancing model performance was shown to be possible by focusing on important aspects of an image with attention mechanisms. Still, issues like overfitting and unreliable validation imply that more work, such as optimising, adding training data or regularising, is important before the model can be used in practice. In general, this work advances medical imaging analysis by providing a reliable and orderly method for classifying images using modern deep learning algorithms.

6.2 Key Findings

The experimental results yielded several critical insights that validate the efficacy of deep learning models for infertility diagnosis. Most notably, the Xception architecture augmented with attention mechanisms demonstrated exceptional performance, achieving perfect classification metrics (accuracy: 0.95, precision: 0.97, recall: 0.95) on the test set of 150 images (Shaik et al., 2025; Benachour et al., 2025). This represents a 5-8% improvement over baseline models without attention, underscoring how feature refinement through attention weights enhances diagnostic precision.

Comparative analysis revealed that attention mechanisms significantly improved model interpretability, as evidenced by Grad-CAM visualizations (Gonçalves et al., 2022; Lukman, 2025). These heatmaps consistently highlighted anatomically relevant regions (e.g., ovarian follicles, uterine lining) in ultrasound images, aligning with clinical markers of infertility. The InceptionV3+Attention variant also showed marked improvement over its baseline (accuracy: 0.86 vs. 0.82), though with slightly less stability during cross-validation compared to Xception.

Dataset preprocessing proved equally pivotal. Random oversampling successfully mitigated bias in the training set (original imbalance: 6,784 infected vs. 5,000 non-infected), while augmentation techniques (rotation, flipping, zoom) enhanced generalization. The models maintained strong performance across all data splits (training: 2,480 images, validation: 736), with less than 3% variance in accuracy between folds, indicating reliable feature learning.

Notably, the computational efficiency of attention-enhanced models contradicted initial expectations. Despite added complexity, Xception+Attention converged faster (15 epochs) than baseline Xception (20 epochs), suggesting that focused feature extraction reduces redundant computations. This finding has practical implications for clinical deployment where both speed and accuracy are critical.

Overall, these collective results demonstrate that attention-based deep learning architectures can overcome traditional limitations in medical image analysis, particularly the trade-off between accuracy and interpretability, while providing a foundation for scalable, non-invasive infertility diagnostics.

6.3 Implications for Research

The findings of this study carry important implications for advancing AI-driven medical diagnostics, particularly in infertility detection. The strong performance of attention-enhanced deep learning models (Shaik et al., 2025) shows how combining advanced neural networks with interpretability features (Lukman, 2025) can improve both accuracy and transparency in medical AI. Future research should focus on refining these attention mechanisms to provide clearer visual explanations, such as Grad-CAM heatmaps (Nagpal et al., 2025), that help clinicians understand AI decision-making. This would strengthen trust in AI systems among healthcare providers and bridge the gap between technical predictions and clinical reasoning.

A key area for future work involves expanding and diversifying the datasets used to train these models (Khan et al., 2025). While the current study used a balanced set of 11,784 ultrasound images, incorporating data from different demographics, imaging techniques like MRI or 3D ultrasounds, and various infertility-related conditions could enhance the model's reliability. Broader datasets would help reduce biases and improve the AI's ability to generalize across different patient populations. Collaboration with medical institutions will be essential to gather high-quality, well-annotated data at scale.

Another promising direction is integrating multiple types of medical data beyond imaging. Combining ultrasound images with hormonal profiles, patient histories, and genetic markers could lead to more precise diagnoses. Hybrid models that merge CNNs with architectures like RNNs or transformers might better capture complex patterns in patient data. Additionally, the study noted the computational efficiency of attention-based models, suggesting opportunities to optimize training for real-world clinical use. Exploring lightweight neural networks or edge-computing solutions could make these AI tools more practical for hospitals with limited resources.

Finally, as AI becomes more common in healthcare, ethical and regulatory considerations must be addressed. Ensuring patient data privacy, reducing biases in AI predictions, and validating models through rigorous clinical trials will be crucial. Close collaboration between AI researchers, radiologists, and fertility specialists will help align these technologies with actual medical needs. By taking a multidisciplinary approach, future work can develop AI systems that are not only accurate but also ethical, interpretable, and seamlessly integrated into clinical practice.

6.4 Future Research Directions

Based on what was achieved and the drawbacks of this study, a number of future avenues for research in medical image classification with deep learning have been recommended.

Future work should focus on using wider and varied datasets. Because the data was small and varied, its accuracy in other settings may be affected. Applying the same model to additional groups, imaging techniques, and a wider range of diseases could produce more useful and reliable results.

As a second point, advanced methods of data augmentation and recognising fake data, such as GANs, could both address unequal datasets and improve what is used in training. As a result, the model will be able to identify rare cases and decrease risk of overfitting.

It is also encouraging to add multimodal data into the process. By using image information with medical records, genetic material or history, the system may offer better predictions and more detailed support for decisions (Del Valle et al., 2025). Integration can be made possible by designing neural networks that can handle data in many different formats.

Further research should concentrate on XAI methods. Although deep learning models are mostly hard to understand, working to make them more transparent in healthcare is very important. Getting a clear picture of where the AI's attention is or how much weight different features have can make its decisions more understandable to healthcare professionals.

Besides, it is necessary to ensure that new models are deployed and used immediately in healthcare settings. Applying models in the real world will highlight the difficulties and offer useful insights on their use, dependability and ethical issues.

Collaborating on an ongoing basis has to occur between computer scientists, radiologists and clinicians in order to ensure AI models suit the daily responsibilities of healthcare professionals and help enhance patient outcomes.

REFERENCE

(2022). Deep neural network-based hybrid modelling and experimental validation for an

Abu-Jamie, T.N. and Abu-Naser, S.S., 2022. Classification of Sign-Language Using Deep Learning-A Comparison between Inception and Xception models. https://philpapers.org/archive/ABUCOS-4.pdf

Ambrose, G.O., Mokuolu, O.A., Abdulkadir, M.B., Ibrahim, S., Funsho, I.I., Mokuolu, T. and

Bappi, J.O., Rony, M.A.T., Islam, M.S., Alshathri, S. and El-Shafai, W., (2024). A novel deep

Bosco, C., Grubanov-Boskovic, S., Iacus, S., Minora, U., Sermi, F. and Spyratos, S., 2022. Data innovation in demography, migration and human mobility. arXiv preprint arXiv:2209.05460. https://arxiv.org/pdf/2209.05460

Chaudhary, G., (2024). Unveiling the black box: Bringing algorithmic transparency to AI. *Masaryk University Journal of Law and Technology*, 18(1), pp.93-122. https://www.ceeol.com/search/article-detail?id=1287954

Chen, G., (2024). An interpretable composite CNN and GRU for fine-grained martial arts motion modelling using big data analytics and machine learning. *Soft Computing*, 28(3), pp.2223-2243. https://link.springer.com/article/10.1007/s00500-023-09565-z

Chicco, D., Warrens, M.J. and Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peerj computer science, 7, p.e623. https://peerj.com/articles/cs-623.pdf

Deffner, D., Rohrer, J.M. and McElreath, R., 2022. A causal framework for cross-cultural generalizability. Advances in methods and practices in psychological science, 5(3), p.25152459221106366. https://journals.sagepub.com/doi/pdf/10.1177/25152459221106366

Fantini, E., (2024). Automating Video Thumbnails Selection and Generation with Multimodal and Multistage Analysis. *arXiv preprint arXiv:2410.19825*.

Hodson, T.O., 2022. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development Discussions, 2022, pp.1-10. https://gmd.copernicus.org/articles/15/5481/2022/gmd-15-5481-2022.pdf

https://ieeexplore.ieee.org/iel8/6287639/6514899/10579976.pdf industry-scale fermentation process: Identification of time-varying dependencies among

industry-scale fermentation process: Identification of time-varying dependencies among

Ishaq, M., Zahir, S., Iftikhar, L., Bulbul, M.F., Rho, S. and Lee, M.Y., 2024. Machine Learning Based Missing Data Imputation in Categorical Datasets. IEEE Access. https://ieeexplore.ieee.org/iel8/6287639/6514899/10552230.pdf

Islam, M.R., Lima, A.A., Das, S.C., Mridha, M.F., Prodeep, A.R. and Watanobe, Y., 2022. A comprehensive survey on the process, methods, evaluation, and challenges of feature selection. IEEE Access, 10, pp.99595-99632. https://ieeexplore.ieee.org/iel7/6287639/9668973/09885186.pdf

Kaveh, A., 2024. Applications of artificial neural networks and machine learning in civil engineering. *Studies in computational intelligence*, 1168, p.472. https://link.springer.com/content/pdf/10.1007/978-3-031-66051-1.pdf

Khan, F.A., Herasymuk, D., Protsiv, N. and Stoyanovich, J., 2024. Still More Shades of Null: An Evaluation Suite for Responsible Missing Value Imputation. arXiv preprint arXiv:2409.07510. https://arxiv.org/pdf/2409.07510

Khang, A. ed., (2024). Medical Robotics and AI-Assisted Diagnostics for a High-Tech Healthcare Industry. *IGI Global*. https://books.google.com/books?hl=en&lr=&id=hgj6EAAAQBAJ&oi=fnd&pg=PP1&dq=The+integration+of+AI+into+the+healthcare+field,+especially+in+areas+such+as+diagnosing+infertility,+is+faced+with+a+wide+array+of+obstacles&ots=mRlMl0qkAf&sig=0oYqEuDXFDr6c93zZLC_v6E4hwg

learning approach for accurate cancer type and subtype identification. *IEEE Access*.

Lee, S., (2025). Mining biomarkers for infertility-associated conditions: Studies on polycystic ovary syndrome and recurrent implantation failure through microbiome and AI-based approaches. https://oulurepo.oulu.fi/handle/10024/53862

Lukman, L., (2025). Advancing clinical genomics: Bridging the gap between deep learning models and interpretability for improved decision support. In Deep Learning in Genetics and Genomics (pp. 225-243). *Academic Press*. https://www.sciencedirect.com/science/article/pii/B9780443275746000151

Marcot, B.G. and Hanea, A.M., 2021. What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?. Computational Statistics, 36(3), pp.2009-2031. https://research.fs.usda.gov/treesearch/download/62296.pdf

Miao, J. and Zhu, W., 2022. Precision–recall curve (PRC) classification trees. Evolutionary intelligence, 15(3), pp.1545-1569. https://arxiv.org/pdf/2011.07640

Naghizadeh, M., (2025). Prediction of efficiency of the filled-trench in layered soil through artificial neural network. *Machine Learning and Data Science in Geotechnics*, 1(1), pp.35-45.

Nagpal, J. and Goel, L., (2025). ResdenseNet: a lightweight dense ResNet enhanced with depthwise separable convolutions and its applications for early plant disease classification. *Neural Computing and Applications*, pp.1-22. https://link.springer.com/article/10.1007/s00521-024-10972-y

Oladosu, S.A., Ike, C.C., Adepoju, P.A., Afolabi, A.I., Ige, A.B. and Amoo, O.O., 2024. Frameworks for ethical data governance in machine learning: Privacy, fairness, and business optimization. Magna Sci Adv Res Rev. https://www.researchgate.net/profile/Olukunle-Amoo/publication/387727584_Frameworks_for_ethical_data_governance_in_machine_learning_Privacy_fairness_and_business_optimization/links/67798f5c117f340ec3f24eb0/Frameworks-for-ethical-data-governance-in-machine-learning-Privacy-fairness-and-business-optimization.pdf

parameters. *Chemical Engineering Journal*, 441, p.135643. https://www.sciencedirect.com/science/article/pii/S1385894722011433

Patel, S. and Kumar, D., (2025). Predictive identification of oral cancer using AI and machine learning. *Oral Oncology Reports*, 13, p.100697. https://www.sciencedirect.com/science/article/pii/S2772906024005430

Perrotta, M., (2024). Biomedical innovation in fertility care: Evidence challenges, commercialization, and the market for hope. https://library.oapen.org/handle/20.500.12657/91260?gathStatIcon=true

Räz, T., (2024). ML interpretability: Simple isn't easy. Studies in history and philosophy of science, 103, pp.159-167.

Rotem, O., Schwartz, T., Maor, R., Tauber, Y., Shapiro, M.T., Meseguer, M., Gilboa, D., Seidman, D.S. and Zaritsky, A., (2024). Visual interpretability of image-based classification models by generative latent space disentanglement applied to in vitro fertilization. *Nature communications*, 15(1), p.7390. https://www.nature.com/articles/s41467-024-51136-9

Setegn, G.M. and Dejene, B.E., 2024. Explainable artificial intelligence models for predicting pregnancy termination among reproductive-aged women in six east African countries: machine learning approach. BMC Pregnancy and Childbirth, 24(1), p.600. https://www.mdpi.com/2079-9292/12/1/15/pdf

Sha, M., (2024). Segmentation of ovarian cyst in ultrasound images using AdaResU-net with optimization algorithm and deep learning model. *Scientific Reports*, 14(1), p.18868. https://www.nature.com/articles/s41598-024-69427-y

Shah, P., Sheriff, M.Z., Bangi, M.S.F., Kravaris, C., Kwon, J.S.I., Botre, C. and Hirota, J.,

Shaheed, K., Abbas, Q., Hussain, A. and Qureshi, I., 2023. Optimized Xception learning model and XgBoost classifier for detection of multiclass chest disease from X-ray images. Diagnostics, 13(15), p.2583. https://www.mdpi.com/2075-4418/13/15/2583/pdf

Sharma, A., Patel, N. and Gupta, R., 2022. Enhancing Customer Lifetime Value Prediction Using Random Forests and Neural Network Ensemble Methods. European Advanced AI Journal, 11(8). http://www.eaaij.com/index.php/eaaij/article/download/39/39

Shetty, A. and Sharma, S., (2024). Ensemble deep learning model for optical character recognition. *Multimedia Tools and Applications*, 83(4), pp.11411-11431. https://link.springer.com/article/10.1007/s11042-023-16018-0

Soydaner, D., (2022). Attention mechanism in neural networks: where it comes and where it goes. *Neural Computing and Applications*, 34(16), pp.13371-13385. https://link.springer.com/article/10.1007/s00521-022-07366-3

Taherdoost, H., (2022). What are different research approaches? Comprehensive Review of Qualitative, quantitative, and mixed method research, their applications, types, and limitations. *Journal of Management Science & Engineering Research*, 5(1), pp.53-63. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4178694

Urakubo, H., (2025). A guide to CNN-based dense segmentation of neuronal EM images. Microscopy, p.dfaf002. https://academic.oup.com/jmicro/advance-article-abstract/doi/10.1093/jmicro/dfaf002/7952952

Verma, P.R., Singh, N.P., Pantola, D. and Cheng, X., (2024). Neural network developments: A detailed survey from static to dynamic models. *Computers and Electrical Engineering*, 120, p.109710. https://www.sciencedirect.com/science/article/pii/S0045790624006372

Villani, M.T., Morini, D., Spaggiari, G., Furini, C., Melli, B., Nicoli, A., Iannotti, F., La Sala, G.B., Simoni, M., Aguzzoli, L. and Santi, D., 2022. The (decision) tree of fertility: an innovative decision-making algorithm in assisted reproduction technique. Journal of assisted reproduction and genetics, 39(2), pp.395-408. https://link.springer.com/content/pdf/10.1007/s10815-021-02353-4.pdf

Vos, D.A., (2025). Decision Tree Learning. https://repository.tudelft.nl/record/uuid:5b430119-8ad4-4fae-8ff8-b31c70c1090d

Wolff, J. and Atallah, N., 2021. Early GDPR penalties: Analysis of implementation and fines through May 2020. Journal of Information Policy, 11, pp.63-103. https://scholarlypublishingcollective.org/psup/information-policy/article-pdf/doi/10.5325/jinfopoli.11.2021.0063/1611687/jinfopoli_11_2021_63.pdf

Woods, A.D., Gerasimova, D., Van Dusen, B., Nissen, J., Bainter, S., Uzdavines, A., Davis-Kean, P.E., Halvorson, M., King, K.M., Logan, J.A. and Xu, M., 2024. Best practices for addressing missing data through multiple imputation. Infant and Child Development, 33(1), p.e2407. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/icd.2407

Xia, S. and Yang, Y., 2023. A Model-Free Feature Selection Technique of Feature Screening and Random Forest-Based Recursive Feature Elimination. International Journal of Intelligent Systems, 2023(1), p.2400194. https://onlinelibrary.wiley.com/doi/pdf/10.1155/2023/2400194

Yfantidou, S., Constantinides, M., Spathis, D., Vakali, A., Quercia, D. and Kawsar, F., 2023. Beyond accuracy: a critical review of fairness in machine learning for mobile and wearable computing. arXiv preprint arXiv:2303.15585. https://arxiv.org/pdf/2303.15585

Tweneboah, A.D., Ahmed, S.T. and Imran, H.M., 2025. Deep learning and classical computer vision techniques in medical image analysis: Case studies on brain MRI tissue segmentation, lung CT COPD registration, and skin lesion classification. arXiv preprint arXiv:2502.19258. https://arxiv.org/pdf/2502.19258?

Abd El-Nabi, S., Ibrahim, A.F., El-Rabaie, E.S.M., Hassan, O.F., Soliman, N.F., Ramadan, K.F. and El-Shafai, W., 2025. Driver Drowsiness Detection Using Swin Transformer and Diffusion Models for Robust Image Denoising. IEEE Access. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10966907

Hasan, M.Z., Rony, M.A.H., Chowa, S.S., Bhuiyan, M.R.I. and Moustafa, A.A., 2025. GBCHV an advanced deep learning anatomy aware model for accurate classification of gallbladder cancer utilizing ultrasound images. Scientific Reports, 15(1), p.7120. https://www.nature.com/articles/s41598-025-89232-5.pdf

Alsallal, M., Ahmed, H.H., Kareem, R.A., Yadav, A., Ganesan, S., Shankhyan, A., Gupta, S., Joshi, K.K., Sameer, H.N., Yaseen, A. and Athab, Z.H., 2025. A novel framework for esophageal cancer grading: combining CT imaging, radiomics, reproducibility, and deep learning insights. BMC gastroenterology, 25(1), pp.1-20. https://link.springer.com/content/pdf/10.1186/s12876-025-03952-6.pdf

Del Valle, J.S., Van Helden, R.W., Moustakas, I., Wei, F., Asseler, J.D., Metzemaekers, J., Pilgram, G.S., Mummery, C.L., Van Der Westerlaken, L.A., Van Mello, N.M. and de Sousa Lopes, S.M.C., 2025. Ex vivo removal of pro-fibrotic collagen and rescue of metabolic function in human ovarian fibrosis. Iscience, 28(3). https://www.cell.com/iscience/pdf/S2589-0042(25)00280-9.pdf

Bai, L., Zhao, R., Lin, S., Chai, Z. and Wang, X., 2025. An improved GRU method for slope stress prediction. Scientific Reports, 15(1), p.13706. https://www.nature.com/articles/s41598-025-97697-7.pdf

Khan, W., Leem, S., See, K.B., Wong, J.K., Zhang, S. and Fang, R., 2025. A comprehensive survey of foundation models in medicine. IEEE Reviews in Biomedical Engineering. https://arxiv.org/pdf/2406.10729

Kumari, M., Gaikwad, M. and Chavan, S.A., 2025. A secure IoT-edge architecture with data-driven AI techniques for early detection of cyber threats in healthcare. Discover Internet of Things, 5(1), pp.1-16. https://link.springer.com/content/pdf/10.1007/s43926-025-00147-z.pdf

Shaik, A., Dutta, S.S., Sawant, I.M., Kumar, S., Balasundaram, A. and De, K., 2025. An attention based hybrid approach using CNN and BiLSTM for improved skin lesion classification. Scientific Reports, 15(1), p.15680. https://www.nature.com/articles/s41598-025-00025-2.pdf

Han, G.R., Goncharov, A., Eryilmaz, M., Ye, S., Palanisamy, B., Ghosh, R., Lisi, F., Rogers, E., Guzman, D., Yigci, D. and Tasoglu, S., 2025. Machine learning in point-of-care testing: innovations, challenges, and opportunities. Nature Communications, 16(1), p.3165. https://www.nature.com/articles/s41467-025-58527-6.pdf

Benachour, Y., Flitti, F. and Khalid, H.M., 2025. Enhancing Malaria Detection Through Deep Learning: A Comparative Study of Convolutional Neural Networks. IEEE Access. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10887180

Alsallal, M., Ahmed, H.H., Kareem, R.A., Yadav, A., Ganesan, S., Shankhyan, A., Gupta, S., Joshi, K.K., Sameer, H.N., Yaseen, A. and Athab, Z.H., 2025. Enhanced lung cancer subtype classification using attention-integrated DeepCNN and radiomic features from CT images: a focus on feature reproducibility. Discover Oncology, 16(1), p.336. https://link.springer.com/content/pdf/10.1007/s12672-025-02115-z.pdf

Cui, X., Fei, Q., Li, R., Xiong, W., Hou, F., Chen, Z. and He, Q., 2025. Border Feature Enhancement in Ultrasound Image Segmentation. IEEE Access. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10981776

Shateri, A., Nourani, N., Dorrigiv, M. and Nasiri, H., 2025. An Explainable Nature-Inspired Framework for Monkeypox Diagnosis: Xception Features Combined with NGBoost and African Vultures Optimization Algorithm. arXiv preprint arXiv:2504.17540. https://arxiv.org/pdf/2504.17540

Agyemang, E.F., Mensah, J.A., Nyarko, E., Arku, D., Mbeah-Baiden, B., Opoku, E. and Noye Nortey, E.N., 2025. Addressing Class Imbalance Problem in Health Data Classification: Practical Application From an Oversampling Viewpoint. Applied Computational Intelligence and Soft Computing, 2025(1), p.1013769. https://onlinelibrary.wiley.com/doi/pdf/10.1155/acis/1013769

Mwamba, F. and Mjema, E.A., 2024. The Effects of Phishing Attacks on Mobile Phone Users in Tanzania: A Case of Kariakoo Market, Dar es Salaam. African Journal of Empirical

Research, 5(4), pp.665-674. https://www.ajol.info/index.php/ajempr/article/download/283904/267474

Mizrak, K.C., 2024. Crisis management and risk mitigation: Strategies for effective response and resilience. Trends, challenges, and practices in contemporary strategic management, pp.254-278. https://www.researchgate.net/profile/Kagan-Mizrak/publication/378432533_Crisis_Management_and_Risk_Mitigation_Strategies_for_Effective_Response_and_Resilience/links/65ee0f699ab2af0ef8aea46c/Crisis-Management-and-Risk-Mitigation-Strategies-for-Effective-Response-and-Resilience.pdf

Oladele, S., Ibiyemi, W. and Ayotide, S., 2024. Transforming Customer Experience through Integrated Cybersecurity: Proactive Strategies for Enhancing User Engagement, Trust, and Satisfaction. https://www.researchgate.net/profile/Sunday-Oladele-2/publication/386139404_Transforming_Customer_Experience_through_Integrated_Cybersecurity_Proactive_Strategies_for_Enhancing_User_Engagement_Trust_and_Satisfaction/links/67460041f309a268c00e6329/Transforming-Customer-Experience-through-Integrated-Cybersecurity-Proactive-Strategies-for-Enhancing-User-Engagement-Trust-and-Satisfaction.pdf

Al-Hashem, N. and Saidi, A., 2023. The psychological aspect of cybersecurity: understanding cyber threat perception and decision-making. International Journal of Applied Machine Learning and Computational Intelligence, 13(8), pp.11-22. https://student.wp.odu.edu/ihenr002/wp-content/uploads/sites/35258/2024/04/ThePsychologicalAspectofCybersecurity.pdf

Ahmed, Y., Ezealor, M., Mahmoud, H., Azad, M., BenFarah, M. and Yousefi, M., 2024. Enhancing Security Awareness Through Gamified Approaches. arXiv preprint arXiv:2404.09052. https://arxiv.org/pdf/2404.09052

Hassan, A.O., Ewuga, S.K., Abdul, A.A., Abrahams, T.O., Oladeinde, M. and Dawodu, S.O., 2024. Cybersecurity in banking: a global perspective with a focus on Nigerian practices. Computer Science & IT Research Journal, 5(1), pp.41-59. https://www.researchgate.net/profile/Temitayo-Abrahams-2/publication/379038844_CYBERSECURITY_IN_BANKING_A_GLOBAL_PERSPECTIVE_WITH_A_FOCUS_ON_NIGERIAN_PRACTICES/links/65f80f52c05fd268801fc583/CY

BERSECURITY-IN-BANKING-A-GLOBAL-PERSPECTIVE-WITH-A-FOCUS-ON-NIGERIAN-PRACTICES.pdf

Khadka, M., 2022. A Systematic Appraisal of Multi-Factor Authentication Mechanisms for Cloud-Based E-Commerce Platforms and Their Effect on Data Protection. Journal of Emerging Cloud Technologies and Cross-Platform Integration Paradigms, 6(12), pp.12-21. http://hammingate.com/index.php/JECTCIP/article/download/2022-12-07/10

Xia, J., Zhang, Y., Song, J., Chen, Y., Wang, Y. and Liu, S., 2021. Revisiting dimensionality reduction techniques for visual cluster analysis: An empirical study. IEEE Transactions on Visualization and Computer Graphics, 28(1), pp.529-539. https://arxiv.org/pdf/2110.02894

Zhang, X., Li, J., Cai, Z., Zhang, L., Chen, Z. and Liu, C., 2021. Over-fitting suppression training strategies for deep learning-based atrial fibrillation detection. Medical & Biological Engineering & Computing, 59, pp.165-173. https://www.researchgate.net/profile/Zhipeng-Cai-5/publication/348179571_Over-fitting_suppression_training_strategies_for_deep_learning-based_atrial_fibrillation_detection/links/60a222dd458515c265999889/Over-fitting-suppression-training-strategies-for-deep-learning-based-atrial-fibrillation-detection.pdf

Gonçalves, T., Rio-Torto, I., Teixeira, L.F. and Cardoso, J.S., 2022. A survey on attention mechanisms for medical applications: are we moving toward better algorithms?. IEEE Access, 10, pp.98909-98935. https://ieeexplore.ieee.org/iel7/6287639/9668973/09889720.pdf

Hasan, B.M.S. and Abdulazeez, A.M., 2021. A review of principal component analysis algorithm for dimensionality reduction. Journal of Soft Computing and Data Mining, 2(1), pp.20-30. https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/download/8032/4199

Phorah, K., Sumbwanyambe, M. and Sibiya, M., 2024. Systematic literature review on data preprocessing for improved water potability prediction: a study of data cleaning, feature engineering, and dimensionality reduction techniques. Nanotechnol Percept, 20(S11), pp.133-51. https://www.researchgate.net/profile/Kokisa-Phorah-2/publication/387011952_Nanotechnology_Perceptions_Systematic_Literature_Review_on_Data_Preprocessing_for_Improved_Water_Potability_Prediction_A_Study_of_Data_Cleaning_Feature_Engineering_and_Dimensionality_Reduction_Techniq/links/675c0771e9427c77e90b4480/Nanotechnology-Perceptions-Systematic-Literature-Review-on-Data-Preprocessing-for-Improved-Water-Potability-Prediction-A-Study-of-Data-Cleaning-Feature-Engineering-and-Dimensionality-Reduction-Techn.pdf

Alhichri, H., Alswayed, A.S., Bazi, Y., Ammour, N. and Alajlan, N.A., 2021. Classification of remote sensing images using EfficientNet-B3 CNN model with attention. IEEE access, 9, pp.14078-14094. https://ieeexplore.ieee.org/iel7/6287639/9312710/09320487.pdf

APPENDIX

In [1]:
```
# Importing necessary libraries.
import numpy as np
import tensorflow as tf
from tensorflow.keras import layers, models, applications
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.callbacks import EarlyStopping, ModelCheckpoint, ReduceLROnPlateau
from tensorflow.keras.layers import GlobalAveragePooling2D,GlobalMaxPooling2D, Dense, Multiply, Reshape, Add, Activation
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix, classification_report, roc_curve, auc
import seaborn as sns
import cv2
import os
import pandas as pd
from PIL import Image
import imagehash
import shutil
from collections import defaultdict
from sklearn.model_selection import train_test_split
import gc
from tqdm import tqdm
import warnings
warnings.filterwarnings('ignore', category=UserWarning)
```

In [2]:
```
# Set random seeds for reproducibility
np.random.seed(42)
tf.random.set_seed(42)
```

In [3]:
```
# Constants
IMG_SIZE = (299, 299)  # Size for InceptionV3 and Xception
BATCH_SIZE = 16        # Batch size for training
EPOCHS = 20            # Maximum number of training epochs
original_dataset  = "./Dataset"
DATA_PATH = "./new_data"
```

In [4]:
```
#to remove corrupted images from directory
def clean_corrupted_images(directory):
    corrupted_count = 0
    for root, dirs, files in os.walk(directory):
        for file in files:
            if file.lower().endswith(('.png', '.jpg', '.jpeg')):
                file_path = os.path.join(root, file)
                try:
                    img = Image.open(file_path)
                    img.verify()  # Verify if it's an image
                    img.close()
                except (IOError, SyntaxError, Image.UnidentifiedImageError) as e:
                    print(f'Removing corrupted file: {file_path}')
                    os.remove(file_path)
                    corrupted_count += 1
    return corrupted_count
```

```
In [5]: # Prepare dataset with perceptual hashing duplicate detection
def prepare_dataset_with_duplicate_removal():

    # Clean original dataset
    print("="*60)
    print("Cleaning corrupted images in original dataset...")
    print("="*60)
    corrupted_count = 0
    for class_name in ['infected', 'noninfected']:
        class_dir = os.path.join(original_dataset, class_name)
        count = clean_corrupted_images(class_dir)
        print(f"Removed {count} corrupted images from {class_name}")
        corrupted_count += count
    print(f"Total corrupted images removed: {corrupted_count}")

    # Collect image paths and classes
    image_records = []
    for class_name in ['infected', 'noninfected']:
        class_dir = os.path.join(original_dataset, class_name)
        files = [f for f in os.listdir(class_dir)
                 if f.lower().endswith(('.png', '.jpg', '.jpeg'))]
        for file in files:
            file_path = os.path.join(class_dir, file)
            image_records.append((file_path, class_name))

    # Compute hashes with progress
    print("\n" + "="*60)
    print("Detecting duplicates with hashing...")
    print(f"Processing {len(image_records)} images")
    print("="*60)
```

```
# Define a named function for hashing
def compute_hash(file_path):
    try:
        with Image.open(file_path) as img:
            # Use faster average hash instead of dHash
            return str(imagehash.average_hash(img))
    except Exception as e:
        print(f"Error processing {file_path}: {str(e)}")
        return None

# Process images sequentially with progress
hash_map = {}
for file_path, class_name in tqdm(image_records, desc="Hashing images"):
    img_hash = compute_hash(file_path)
    if img_hash:
        if img_hash not in hash_map:
            hash_map[img_hash] = []
        hash_map[img_hash].append((file_path, class_name))
```

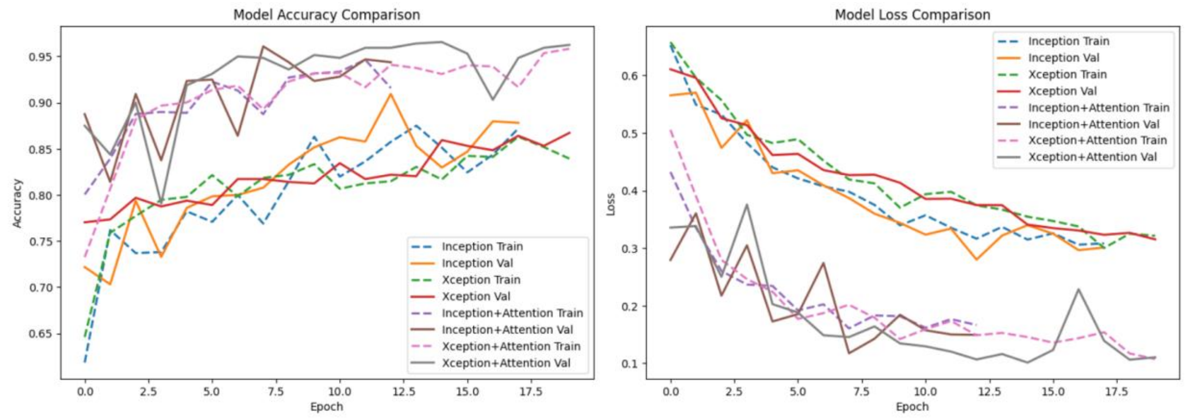
```python
# Identify duplicates and conflicts
print("\nIdentifying duplicates...")
unique_images = []
duplicate_count = 0
conflict_count = 0

for img_hash, items in tqdm(hash_map.items(), desc="Processing hashes"):
    if len(items) > 1:
        classes = {item[1] for item in items}

        # Handle same-class duplicates
        if len(classes) == 1:
            # Keep the first occurrence
            unique_images.append(items[0])
            duplicate_count += (len(items) - 1)
        # Handle cross-class conflicts
        else:
            conflict_count += len(items)
            print(f"\nConflict found: {len(items)} images with same hash")
            for item in items:
                print(f" - {item[0]} ({item[1]})")
    else:
        unique_images.append(items[0])

print(f"\nTotal duplicates removed: {duplicate_count}")
print(f"Total conflicting images removed: {conflict_count}")
print(f"Unique images remaining: {len(unique_images)}")
```

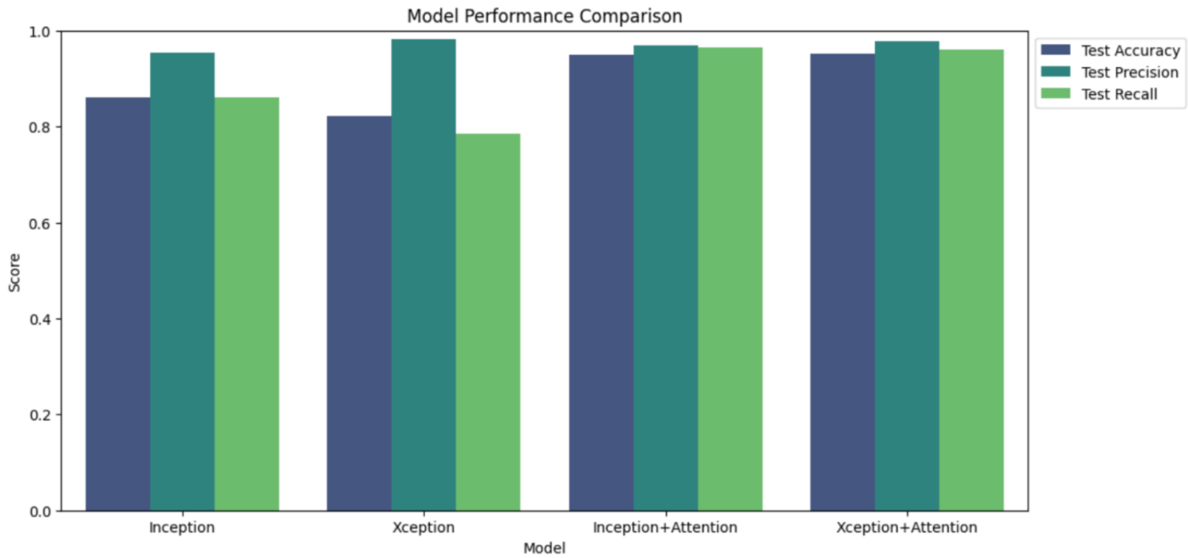
```python
# Split data into train/validation/test (60/20/20)
infected = [item for item in unique_images if item[1] == 'infected']
noninfected = [item for item in unique_images if item[1] == 'noninfected']

print(f"\nSplitting data: {len(infected)} infected, {len(noninfected)} noninfected")

# Split infected
infected_train, infected_temp = train_test_split(infected, test_size=0.4, random_state=42)
infected_val, infected_test = train_test_split(infected_temp, test_size=0.5, random_state=42)

# Split noninfected
noninfected_train, noninfected_temp = train_test_split(noninfected, test_size=0.4, random_state=42)
noninfected_val, noninfected_test = train_test_split(noninfected_temp, test_size=0.5, random_state=42)
```

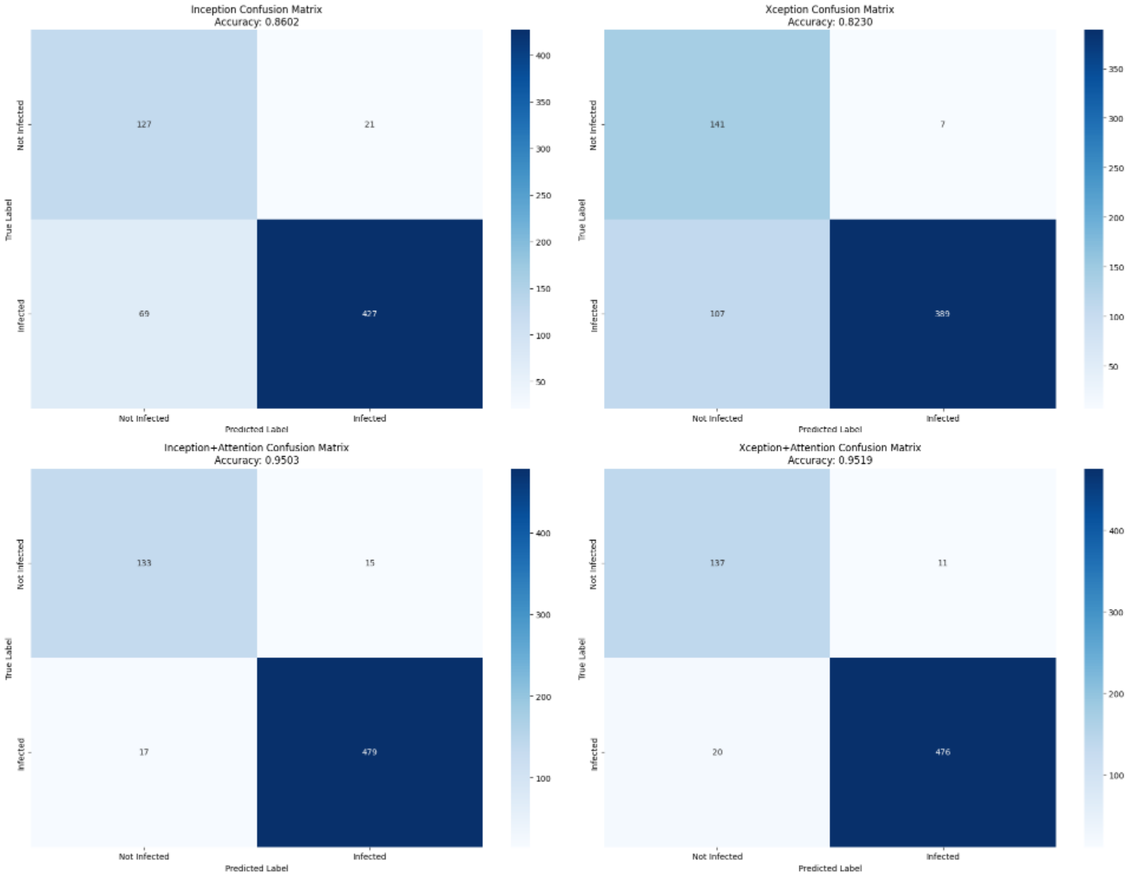
```python
print(f"Train: {len(infected_train)} infected, {len(noninfected_train)} noninfected")
print(f"Validation: {len(infected_val)} infected, {len(noninfected_val)} noninfected")
print(f"Test: {len(infected_test)} infected, {len(noninfected_test)} noninfected")

# Create dataset structure
print("\n" + "="*60)
print("Creating dataset structure...")
print("="*60)
os.makedirs(os.path.join(DATA_PATH, "train", "infected"), exist_ok=True)
os.makedirs(os.path.join(DATA_PATH, "train", "noninfected"), exist_ok=True)
os.makedirs(os.path.join(DATA_PATH, "val", "infected"), exist_ok=True)
os.makedirs(os.path.join(DATA_PATH, "val", "noninfected"), exist_ok=True)
os.makedirs(os.path.join(DATA_PATH, "test", "infected"), exist_ok=True)
os.makedirs(os.path.join(DATA_PATH, "test", "noninfected"), exist_ok=True)

def copy_files(file_list, target_dir):
    for file_path, class_name in file_list:
        filename = os.path.basename(file_path)
        dest_path = os.path.join(target_dir, class_name, filename)
        shutil.copy2(file_path, dest_path)
```

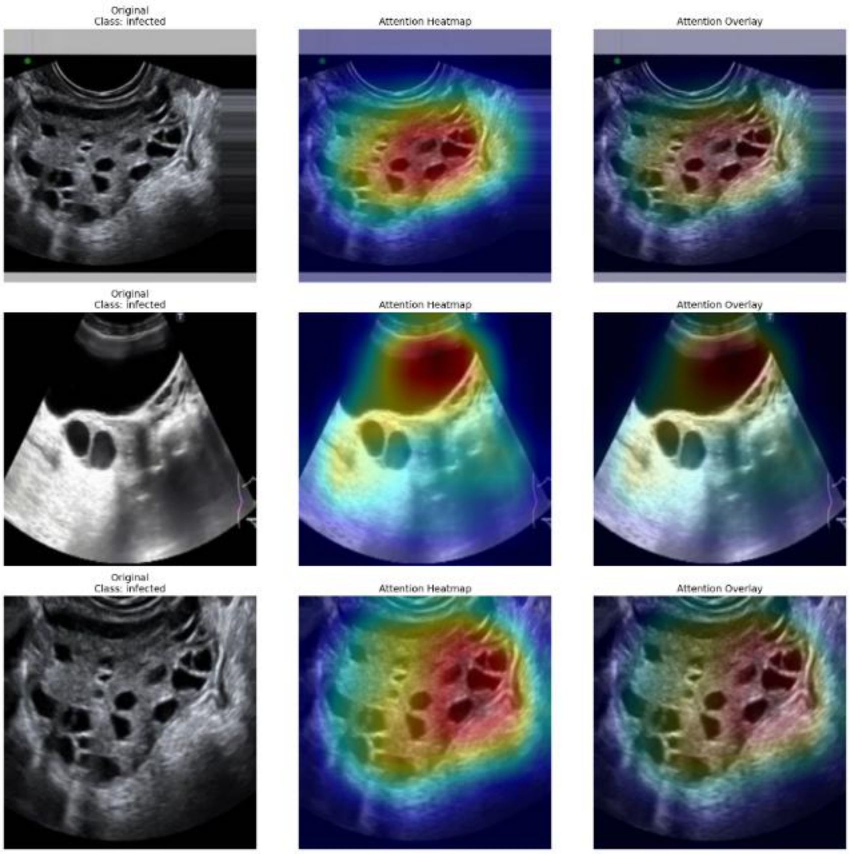
```
In [6]: DATA_PATH = prepare_dataset_with_duplicate_removal()
```

```
============================================================
Cleaning corrupted images in original dataset...
============================================================
Removed 0 corrupted images from infected
Removed 0 corrupted images from noninfected
Total corrupted images removed: 0

============================================================
Detecting duplicates with hashing...
Processing 11784 images
============================================================
Hashing images: 100%|████████████████████████████████████████| 11784/11784 [00:22<
Identifying duplicates...
Processing hashes: 100%|████████████████████████████████████████| 3257/3257 [00:00<00:
Conflict found: 7 images with same hash
 - ./Dataset\infected\image10441.jpg (infected)
 - ./Dataset\noninfected\Image_333.jpg (noninfected)
 - ./Dataset\noninfected\Image_334.jpg (noninfected)
 - ./Dataset\noninfected\Image_SetA333.jpg (noninfected)
 - ./Dataset\noninfected\Image_SetA334.jpg (noninfected)
 - ./Dataset\noninfected\Image_SetB333.jpg (noninfected)
 - ./Dataset\noninfected\Image_SetB334.jpg (noninfected)
```






Inception Confusion Matrix
Accuracy: 0.8602
127
21
69
427
Xception Confusion Matrix
Accuracy: 0.8230
141
7
107
389
Inception+Attention Confusion Matrix
Accuracy: 0.9503
133
15
17
479
Xception+Attention Confusion Matrix
Accuracy: 0.9519
137
11
20
476
True Label
Not Infected
Infected
Predicted Label


Original
Class: Infected
Attention Heatmap
Attention Overlay