

MINOR PROJECT SYNOPSIS

XAI IMPLEMENTATION FOR PCOS DETECTION



Submitted By:

Prateek Kumar	22103052
Vedant Singh Chauhan	22103058
Shivam Singh	22103031

Submitted To:

Dr. Keshav Ajmera
Ms. Kirti Jain

Mentor:

Dr. Kavita Pandey

INTRODUCTION

The integration of artificial intelligence (AI) and machine learning (ML) in healthcare has revolutionized the ability to diagnose, predict, and manage a wide range of medical conditions. These advanced technologies offer the potential for highly accurate disease detection, personalized treatment plans, and the automation of diagnostic processes, significantly improving patient care and reducing human error. However, as AI and ML models become more complex, their decision-making processes are often seen as opaque, leading to challenges in trust and acceptance within the medical community.

To address this, explainable AI (XAI) has emerged as a critical innovation, offering transparency by enabling healthcare professionals to understand, interpret, and trust AI-driven predictions. XAI provides insights into how models arrive at their decisions, fostering accountability, enhancing clinical decision-making, and ensuring that AI systems support medical experts rather than replace them. The combination of AI/ML with XAI is transforming healthcare by making advanced technologies more accessible, reliable, and aligned with ethical standards, ultimately improving patient outcomes and safety.

The combination of AI/ML with XAI is transforming healthcare by making advanced technologies more accessible, reliable, and aligned with ethical standards, ultimately improving patient outcomes and safety. As healthcare increasingly relies on AI-driven models for diagnostics, ensuring that these systems are interpretable and trustworthy becomes crucial.

In this context, our research focuses on leveraging Explainable AI for the detection of polycystic ovary syndrome (PCOS). After analyzing 12 research papers on various diseases such as diabetes, cardiovascular conditions, allergies, and asthma, I chose PCOS due to its growing prevalence and diagnostic complexity. By implementing XAI on machine learning and deep learning models for PCOS detection, the project aims to provide transparency in the decision-making process, ensuring that medical professionals can trust and effectively use these tools. This research highlights the importance of explainability in AI healthcare applications, helping bridge the gap between advanced technology and practical, trustworthy healthcare solutions.

LITERATURE REVIEW

Disease (Year)	Data Sets	ML Models Used	Evaluation Metrics	XAI Models Used
HIV (2021)	Case-control study on HIV patients	Clustering, Logistic Regression, Decision Trees, Neural Networks	Precision, Recall, Accuracy, Gini Importance, VI	LIME, Shapley, ICE
Cardiovascular (2022)	Dataset with 303 instances, 14 attributes	SVM, AdaBoost, KNN, Bagging, LR, Naive Bayes	AUC, ROC, Sensitivity, Specificity, F1-Score	SHAP, LIME
Allergy (2021)	Intradermal skin test of 878 patients (India)	Decision Tree, SVM, Random Forest	Accuracy, Sensitivity	Post-hoc Explainability
Asthma (2023)	Cohort from NCT04796844 trial	Logic Learning Machine	Accuracy, F1-Score, PPV, NPV, TPR, TNR	Rule-based if-then rules
Liver (2023)	Indian Patient Liver Dataset	Deep Learning (Keras-Tensorflow)	Accuracy, Precision, Recall, F-Measure	SHAP
Monkeypox (2023)	Images from Kaggle (Monkeypox, Measles, etc.)	ResNet-18, ResNet-50, ResNet-101, SqueezeNet	Accuracy, Precision, Recall, F1-Score	LIME
Breast Cancer (2023)	Public data of Indonesian women	XGBoost, Logistic Regression, Random Forest, SVM	Accuracy, Precision, Recall	SHAP
Diabetes (2023)	Diabetes dataset (768 rows, 9 columns)	KNN, Naive Bayes, SVM, Decision Tree, Random Forest, LR	F1 Score, Accuracy, Precision, Recall, ROC AUC, Time Taken	SHAP, LIME
PCOS (2023)	Kaggle PCOS dataset	RF, ADB, GB, XGB, CATB, PODBoost	Accuracy, Error-Rate, ROC-AUC, Recall, Precision, F1-Score	LIME
PCOS (2023)	PCOS dataset from Kaggle (541 instances)	Logistic Regression, RF, DT, Naive Bayes, SVM, KNN, XGBoost, AdaBoost, Stacking	Accuracy, Precision, Recall, F1-Score, AUC	Local and Global Explainability
Thyroid (2023)	UCI Machine Learning Repository	Decision Tree, Random Forest, Gradient Boosting, Naive Bayes, KNN, LR, SVM	Accuracy, Precision, Recall, F1-Score	SHAP, LIME

ABOUT PCOS

Polycystic Ovary Syndrome (PCOS) is a prevalent hormonal disorder affecting women, commonly diagnosed through ultrasound imaging of the ovaries. Machine learning techniques, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable accuracy in medical image analysis tasks. However, their “black-box” nature makes them unsuitable for critical medical diagnostics, where the interpretability of decisions is paramount. This project seeks to develop a CNN model for detecting PCOS from ultrasound images while leveraging Explainable AI (XAI) techniques like Grad-CAM to ensure transparency, trust, and clinical validation. By using XAI, clinicians can understand and trust the model’s decisions, making it an integral tool for PCOS diagnosis.

PCOS TABLE

Author	Disease (Year)	Data Sets	Models Used	Evaluation Metrics	XAI Models Used
Khanna et. al	PCOS (2023)	541 patients from Kerala, India	Logistic Regression, Random Forest, SVM, etc.	Accuracy, precision, recall, AUC-ROC, etc.	SHAP, LIME, ELI5, QLatice
Moral et. al	PCOS (2024)	PCOS dataset from Kaggle (541 records)	Logistic Regression, Random Forest, Gradient Boosting	Accuracy, Error-Rate, ROC-AUC, etc.	LIME
Elmannai et. al	PCOS (2023)	PCOS dataset from Kaggle (541 instances)	Logistic Regression, Random Forest, Naive Bayes	Accuracy, precision, recall, F1 Score, etc.	Local and global explainability techniques
Çiçek et. al	PCOS (2021)	“Polycystic ovary syndrome” dataset from Kaggle	Random Forest	Accuracy, sensitivity, specificity, etc.	LIME
Özmen et. al	PCOS (2023)	Ultrasound images of 54 patients	CNN, SqueezeNet	MSE, PSNR, accuracy, etc.	None
Umapathy et. al	PCOS (2024)	Two datasets with ultrasound images	Random Forest, MobileNet, ResNet152V2, etc.	Accuracy, sensitivity, specificity, AUC	None
Alamoudi et. al	PCOS (2023)	Ovary ultrasound images and clinical data (285 patients)	VGG-16, VGG-19, DenseNet121, DenseNet201, etc.	Accuracy, precision, recall, F1 score, specificity	None
Suha et. al	PCOS (2022)	594 ovary ultrasound (USG) images	CNN with transfer learning (VGGNet16), XGBoost	Accuracy, precision, sensitivity, F1 score, AUC-ROC	None

RESEARCH GAP

While deep learning and explainable AI (XAI) have gained significant traction in healthcare, there remains a substantial gap in applying these technologies specifically to **Polycystic Ovary Syndrome (PCOS)** prediction from ultrasound images. Most current research has focused on other diseases, such as diabetes, cardiovascular conditions, and cancers, with limited attention given to PCOS despite its growing prevalence. Although advanced AI models have demonstrated excellent performance in medical image analysis, their black-box nature poses a significant barrier to their adoption in sensitive clinical applications, where interpretability is crucial.

In the case of PCOS detection, accurate diagnosis can be particularly challenging, and clinicians often need tools that not only provide reliable predictions but also allow them to understand how those predictions are made. The current literature lacks robust work that combines **deep learning** with **XAI** to address these challenges, especially in the context of PCOS.

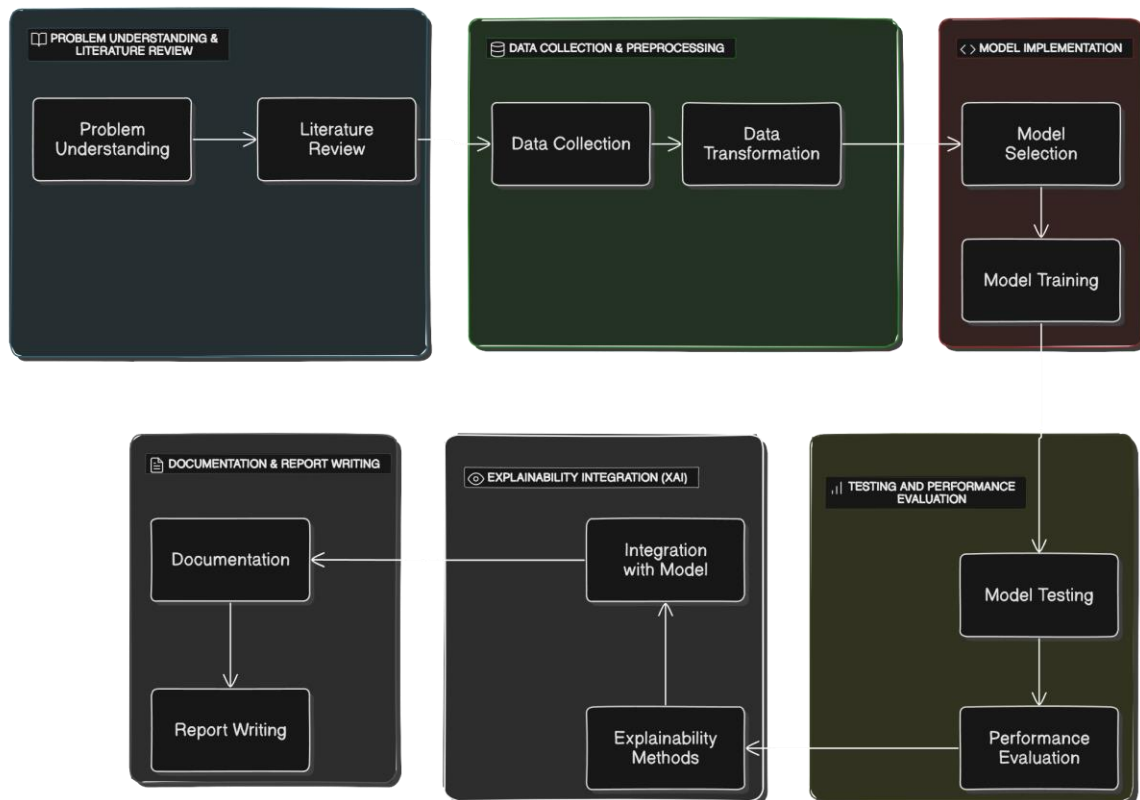
By focusing on the integration of deep learning models with **XAI techniques**, this project aims to close the research gap. It will enable more transparent and interpretable models for PCOS detection, making it possible for clinicians to visualize and trust the key factors that influence the AI's decision-making process. This gap highlights the importance of our work in making AI models more applicable and trustworthy in real-world clinical settings, particularly for diseases like PCOS that demand high levels of diagnostic accuracy and interpretability.

PROBLEM STATEMENT

The diagnosis of PCOS through ultrasound images presents a significant challenge in terms of accuracy and interpretability. While CNNs can achieve high performance in classifying medical images, their black-box nature raises concerns about the trustworthiness of their predictions in sensitive healthcare applications.

The challenge is to build a CNN-based system for detecting PCOS from ultrasound images that is not only accurate but also interpretable. By applying Explainable AI (XAI) techniques, the model's internal workings can be visualized, highlighting the key areas of the image that influenced the model's decision, making the predictions more understandable and trustworthy to clinicians.

FLOW OF ACTION



MODEL SELECTION AND SETUP

CNN ARCHITECTURE: Selected the CNN MobileNet, known for its high accuracy and efficiency in image classification tasks.

PRE-TRAINED MODEL: Using a pre-trained model and fine-tune the weights for the specific task of PCOS detection.

MODEL SETUP: Define the input layer and configure the output layer for binary classification (PCOS-positive or PCOS-negative).

MODEL TRAINING

TRAINING: We will be training the model on the augmented dataset, while monitoring performance on the validation set.

TESTING: Test the trained model on the unseen test dataset to evaluate its generalization capability.

PERFORMANCE METRICS: Compute key performance metrics like Accuracy to quantify the model's ability to detect PCOS.

GRAD-CAM INTEGRATION

IMPLEMENT GRAD-CAM TO GENERATE CLASS ACTIVATION HEATMAPS: Visualizing the regions in the ultrasound images that contributed the most to the model's diagnosis. These heatmaps will overlay on the original images, highlighting areas such as ovarian cysts or other relevant features that the model used for its decision-making.

VISUAL EXPLANATIONS: The generated heatmaps provide interpretable insights into the CNN model's decision, revealing whether the model is focusing on clinically relevant areas.

ANALYSIS OF GRAD-CAM HEATMAPS

Evaluate the heatmaps to ensure that the CNN model is focusing on meaningful regions in the images (e.g., cystic formations in the ovaries). If the model is focusing on irrelevant regions, adjustments can be made to the architecture or training data.

TRUST BUILDING: By demonstrating that the CNN model bases its predictions on medically relevant areas, trust in the model is established. This transparency bridges the gap between AI predictions and clinical decision-making, making the model a reliable tool for PCOS diagnosis.