Informatics Institute of Technology

In Collaboration With

University of Westminster, UK

University of Westminster, Coat of Arms

PcoExplaGen

Multi-Omics and Clinical Data Integration for Enhanced Early Detection of Polycystic Ovarian Syndrome (PCOS) Using Explainable Deep Learning

A Bioinformatics Approach

A dissertation by

Thushini Abeysuriya

W1867426/20210156

Supervised by

Dr.Sugandima Vidanagamachchi

Submitted in partial fulfilment of the requirements for the Beng Software Engineering degree at the University of Westminster.

April 2024

ABSTRACT

Polycystic Ovary Syndrome (PCOS) is a complex endocrine disorder affecting up to 10% of women of reproductive age, with early detection remaining challenging due to its heterogeneous presentation. Traditional diagnostic approaches relying solely on clinical markers often miss subtle molecular indicators, while genomic analyses alone fail to incorporate crucial clinical context, resulting in delayed diagnosis and suboptimal treatment outcomes.

This project addresses these limitations through a novel multi-modal approach integrating omics data (from GEO datasets including GSE5090, GSE54248, GSE54250) with clinical biomarkers using a sophisticated architecture combining Graph Neural Networks (GNNs) and feed-forward networks. The implementation features separate processing pathways: omics data flows through GNN layers (GATConv/GCNConv) modeling gene-gene interactions, while clinical markers pass through conventional neural networks. These pathways converge in a fusion mechanism with adaptive weighting based on clinical indicators like AMH levels. Domain adaptation techniques mitigate batch effects across datasets, while transfer learning leverages pre-trained encoders to improve generalizability.

Comprehensive evaluation demonstrates the system's effectiveness with a cross-validated mean AUC of 0.653 (±0.116). Refined threshold tuning yields a precision of 0.975, recall of 0.870, and classification accuracy up to 95%. The integration of explainable AI tools provides transparent visualization of feature importance, showing that AMH levels, β-HCG measurements, and their derived ratios consistently emerge as critical predictors. This approach not only enhances diagnostic accuracy but also provides actionable insights into the molecular mechanisms underlying PCOS.

Keywords: Polycystic Ovary Syndrome, Bioinformatics, Explainable AI, Machine Learning, Deep Learning, Transfer Learning, Domain Adaptation, Omics Data

Subject Descriptors : Applied Computing → Life and Medical Sciences → Bioinformatics
Applied Computing → Life and Medical Sciences → Computational Biology → Biological Networks
Computing Methodologies → Machine Learning → Machine Learning Algorithms → Feature Selection