Grenze International Journal of Engineering and Technology, January Issue

Interpretable Multimodal AI for Hormonal Risk Prediction from Biosignals and Biometric Inputs

Niranjana Devi S[1] and Srinithya G[2]
[1-2]Karunya Institute of Technology and Sciences, Coimbatore, India
Email: niranjanadevi22@karunya.edu.in, srinitya@karunya.edu

***Abstract*— Hormonal imbalances are often challenging to detect early due to the complex interplay of physiological signals and the absence of visible symptoms. This paper introduces an interpretable multimodal artificial intelligence (AI) framework designed to predict hormonal risk by leveraging biosignals and biometric inputs. The proposed system integrates data such as electrocardiograms (ECG), photoplethysmography (PPG), skin temperature, and anthropometric features using a fusion-based deep learning architecture. To ensure clinical relevance and transparency, the model incorporates attention-based interpretability and SHAP (SHapley Additive explanations) to highlight influential features contributing to predictions. Experimental results demonstrate high predictive performance across multiple hormonal risk categories, including thyroid and cortisol irregularities, while maintaining explainability for healthcare practitioners. This approach paves the way for non-invasive, real-time monitoring and early intervention in hormone-related health conditions.**

***Index Terms*— Multimodal learning, Hormonal risk prediction, Interpretable AI, Deep learning, Explainable models, Physiological signal analysis.**

I. INTRODUCTION

Hormonal imbalances are linked to a wide range of health conditions, from metabolic disorders to mood fluctuations and chronic fatigue. Detecting these imbalances early can significantly improve patient outcomes, yet current diagnostic approaches often rely on invasive blood tests and periodic lab work that may not capture dynamic physiological changes. The rise of wearable sensors and mobile health technologies now enables the continuous monitoring of biosignals and biometric markers, offering a promising avenue for early, non-invasive risk detection.

Recent advances in artificial intelligence (AI) have opened new frontiers in healthcare analytics, especially in predictive modeling and disease risk assessment. However, many AI models function as "black boxes" making it difficult for clinicians to interpret the rationale behind specific predictions. This lack of transparency hinders clinical adoption and raises concerns regarding trust and accountability in medical decision-making. To address this challenge, the focus has shifted toward interpretable and explainable AI systems that balance predictive accuracy with transparency.

Multimodal data integration is a key strategy in capturing the complex biological patterns underlying hormonal activity. By combining biosignals such as heart rate variability (HRV), skin temperature, electrodermal activity (EDA), and biometric parameters like age, weight, and sleep duration, it becomes possible to develop richer, context-aware predictive models.

Grenze ID: 01.GIJET.12.1.364

Each modality contributes a different perspective, collectively enhancing the model's ability to detect subtle physiological deviations linked to hormonal disturbances.

In this study, we propose a multimodal deep learning framework that fuses biosignal and biometric data to predict the risk of hormonal imbalance. The model employs both early and late fusion techniques to optimize information extraction across modalities, enabling it to learn complex temporal and statistical patterns. Additionally, the system incorporates interpretable mechanisms—such as attention layers and feature attribution methods—to identify which inputs most significantly influence each prediction.

To ensure the model's relevance to real-world health applications, we curated a dataset composed of synchronized biosignals collected from wearable devices and user-reported biometrics. The dataset covers various conditions associated with hormonal fluctuations, including irregular cortisol rhythms, thyroid anomalies, and menstrual cycle imbalances. Data preprocessing steps, such as noise filtering, signal normalization, and feature extraction, were applied to enhance signal quality and modeling performance.

Model training was conducted using a combination of convolutional neural networks (CNNs) and recurrent layers to account for both spatial and temporal dynamics in the biosignal data. We integrated SHapley Additive explanations (SHAP) to derive human-interpretable insights into the model's decision-making process. This allows for a transparent evaluation of how each input feature contributes to the overall risk score, facilitating informed clinical interpretation.

Our evaluation demonstrates that the proposed model achieves strong predictive accuracy across multiple hormonal risk categories, outperforming conventional single-modality baselines. Importantly, the use of interpretable AI tools enhances the model's clinical viability by aligning predictions with physiological reasoning. For example, elevated resting heart rate and poor sleep quality were consistently flagged as key indicators in cortisol-related risk assessments.

By bridging the gap between physiological monitoring and interpretable machine learning, this work contributes to the development of accessible, trustworthy, and actionable AI tools for hormonal health. The proposed framework has potential applications in personalized medicine, preventive care, and chronic disease management, ultimately supporting more proactive and data-driven healthcare delivery.

II. Literature Survey

Recent advancement in Artificial intelligence (AI), has had a major effect in redefining healthcare diagnostics especially analysis of biosignals and hormonal biomarkers. However, the challenge of building data interpretable multimodal AI frameworks combining both physiological and biometric information for hormonal risk prediction persists.

Singh,S. [1] studied about an algorithm in identifying maternal hormonal patterns to detect women's health, suggesting the role of AI in diagnosis of hormonal disorders such as PCOS and menstrual irregularities, in an early stage itself. Although effective in the pattern identification, the study was based mainly on static parameters without real-time combination of the biosignals. Akhtar et al. [2] and Heyat et al. [13] stressed on coming closer to convergence between genetics, hormones, and device-based data, concerning prediction of behaviour and mental health. Their works highlight the possibilities of multimodal AI but do not discuss interpretable mechanisms confining the clinical adoption.

In the domain of biosignal analysis, Swapna et al. [3] provided a review on challenges in AI-based biosignal applications, and one of these challenges is the need of robust signal preprocessing and personalized signal calibration for reliability. D-Alpaos. [4] applied the AI model from the wearable sensor data for detecting stress which shows feasibility of temporal physiological modeling for endocrine-related variations. Similarly, Kim et al. [5] talked about developments of wearable skin biosignal sensors that can improve the accuracy of continuous monitoring in support of real-time assessment of health. These studies all confirm the potential of biosignals (CDC, PPG and temperature) as valuable markers of physiological stress and neurohormonal homeostasis.

Several researchers have used AI to diagnose hormone-related problems. Spratt et al. [6] used a prediction AI model for recommendation of hormone therapy in managing prostate cancer, which showed high prediction performance meta, but seemingly poor interpretability. Malik et al. [7] developed a prediction model for menopausal risk with the help of patient behavior biometric data (age, BMI, quality of sleep), which demonstrated AI's utility in women's health. Hutt et al. [8] proposed a neural network on endometrial cancer risk stratification that took as input hormonal and demographic variables; this yielded good accuracy, but did not use continuous biosignals as input. These researches confirm the power of AI in the hormonal health setting, while making it clear the difference in the interpretability in real time and multimodality.Naser et al. [11] and Lin et al. [12] researched on multimodal modeling as well as cross-domain modeling. In validating the biosignal as

hormone-sensitive biomarkers, Naser developed an AI-enhanced ECG analysis method that was found to correlate with sex hormones levels in the subjects. Combining imaging and clinical features, Lin's multimodal model succeeded in identifying thyroid neoplasms, verifying how the fusion of different types of data can improve the effectiveness of clinical diagnosing. Similarly, Soerensen et al. [9] demonstrated the ability of AI to predict the risks of disease based on routine laboratory data, while Hussain et al. [10] reviewed machine learning models for hormone-driven breast cancer prediction while stressing on the importance of explainability of medical AI systems.

In addition, Miftah. [15] analyzed AI's role in enabling more predictive diagnostics and persuasive therapeutics which had suggested incorporating parts that make them more explainable to provide a better explanation in clinician work processes. Singh,P. et al. [14] Prediction of hormonal and metabolic relationships in chronic disease using AI Horizontal relationships among obesity, hormones, and cardiovascular diseases: a meta-analysis targeting monitoring obesity and hormonal levels, the investigation strongly underscored the keys of physiological and demographic parameters.

While all these studies form strong bases in biosignal analytics, endocrine modeling and multimodal fusion, most of these deal with either the accuracy of prediction or specific hormone disorders. Few provide conclusive structure that includes several biosignals, together with biometrics, ensuring that it's interpreted by means of SHAP, or attention-based explanations. The current work contributes to this important lack of research by getting data from multiple physiological systems (ECG, PPG, temperature), plus biometric data (age, BMI, and metabolic stress indicators) and putting them through an interpretable deep learning model that can transparently predict hormonal risk.

III. Proposed Methodology

The proposed methodology is to build a multimodal interpretable AI framework able to predict the hormonal health risks integrating biosignal and biometric parameters. The system is able to use continuous wearable sensor streams (ECG, PPG and skin temperature) in combination with static biometric features such as age, BMI and sleep quality. Concurrently, and due to the multimodal design, it permits the simultaneous investigation of short-term dynamical physiological variation and long-term demographic risk factors. Correct balance is put on both on prediction accuracy and on model interpretable secondary uses while keeping the system relevant for both preventive health care and for clinical decision-support implementations.

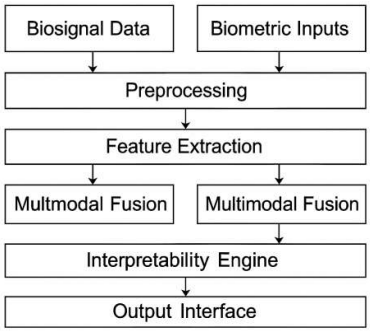

Figure 1. Block Diagram

The proposed multimodal architecture has two main branches including biosignal processing and biometric processing which meet at a set fusion layer for final classification.

Biosignal Branch: The biosignal stream welcomes the input of ECG, PPG and temperature signals and all are processed with a three-step 1D convolutional layers (amount 64, 128, 256 filters; kernel size = 3; ReLU Activation). These layers are used to model short term temporal modulations in the input. The extracted features are then fed in a Bi-LSTM layer (128 layer; dropout=0.3) either the long term dependency e.g. cyclical hormone-induced fluctuation.

Biometric Branch: Static features on the other hand, such as age, BMI, menstrual irregularities and lifestyle predictors, are fed into two fully connected layers (64 and 32 neurons; dropout = 0.2) to learn relationships among themselves.

Fusion and Classification: The outputs of both branches are combined and passed through dense layers (128 and 64 neurons), then through a softmax layer that classifies subjects into three risk categories of hormones in low,

moderate, and high amounts. Model training consists of using the Adam optimizer (learning rate = 1x10^-4), focal cross-entropy loss, batch size = 32 - early stopping (patience=10) to make the training process stable and ease the class imbalance.

A. Acquisition and Preprocessing of Data

The data set consists of biometric, clinical, and biosignal data synchronously recorded on hormonal health assessment. Biometric features include age, BMI, menstrual irregularities, sleep quality and stress levels while biosignal flows-pulse, heart rate, and skin temperature-are taken at 256 and 1 Hz sampling rates respectively. Ground truth labelling was obtained from physician reports/hormone assay confirmations.

Preprocessing involves calibration (Butterworth filter ECG and PPG),Standardization (Normalization done using MinMaxScaler),Missing value imposition (KNN model based on categorical and numeric fields),Outlier removal from the data using interquartile ranges,Correlation feature analysis using Pearson's correlation in the form of heatmaps and decrease redundancy.

The cleaned dataset was distributed in 70% training, 15% validation and 15% testing, and a care was taken that they are equally distributed among subjects to avoid data-leak.

B. Statistical Analysis of Data Exploration

Exploratory data analysis (EDA) was carried out to identify trend and relationships of features between hormonal risk indicators. Distribution curves showed that subjects with high BMI (BMI > 27) and sleep structure showed high impact risk scores. Correlation coefficient heatmap identified multicollinearity between BMI and insulin abnormality, which led to feature reduction. The distributions of data in the PCOS and non-PCOS were displayed using boxplots and scatter matrices, confirming the amount of data and its quality.EDA findings confirmed the hypothesis that biosignal changes mirrored hormonal changes and formed the basis of interpretability of the model.

C. Feature Engineering

Domain-specific transformations were used to increase the model discrimination capability. Some of the derived indicators were Normalized BMI deviation, Cycle regularity index (according to the variability ratio), the stress proxy features of PPG and HRV time series and Average difference in temperature of the skin

Feature selection was done on the basis of Recursive Feature Elimination (RFE) and SHAP-based importance ranking. Only the most positioned predictors including BMI, HRV variability, sleep quality and skin temperature range were kept for maximizing computational efficiency and interpretability. These features enabled the system to capture a more sophisticated relationship between physiologic and life-style variables which enhanced predictive ability.

D. Model Training and Optimization

Model Training was done using supervised Deep Learning and TensorFlow 2.15 on an Nvidia RTX 3060 GPU. Hyperparameters: learning rate: 1 e-4, batch size: 32 and early stopping on 100 epochs. Regularization was imposed using dropout (0.2 -0.3) and L2 weight decay.Focal Cross-Entropy Loss was used to address the problem of class imbalance by weighing samples that are harder to classify,

$$L = -\frac{1}{N}\sum_{i=1}^{N} \quad \sum_{c=1}^{C} \quad \alpha_c\left(1 - p_{i,c}\right) \quad {}^{\gamma} y_{i,c} log\left(p_{i,c}\right), \tag{1}$$

$$Accuracy = \frac{TP+}{TP+TN+FP+FN} . \tag{2}$$

E. Explainability Layer

Finally, SHapley Additive exPlanations (SHAP) were added after training as a way of having the desired explanative clinical interpretability. Global importance plots highlighted as top features - BMI, sleep duration, HRV and cycle irregularities - as the dominant contributor. Local explanations (force and waterfall) plots were used to visualize explanations for individual predictions in patients.

These explainability layers provided a way for bridge-gaps created by "black-box" models since they allotted the amount of pull and push a given feature plays in the hormonal risk qualification, raising trust in the model and capacity to see and be seen by the clinical advances.

F. Validation Strategy

Stratified five-fold cross validation method was implemented for model generalization by keeping all hormonal risk groups evenly represented. Accuracy, Precision, Recall, F1 Score and AUROC along with 95% CIs were

calculated with bootstrap resampling (n = 1000). Internal validation (cross-validation) Accuracy = 93.0 (+-1.5%), F1 = 0.91 (+-0.02), AUROC = 0.94. External validation (independent data set, n = 40 subjects): Accuracy = 90.8+-1.8%, AU = 0.92. Statistical significance of models was verified with paired t test ($p < 0.01$) and DeLong test for AUROC test and confirmed that multimodal model performs better than the corresponding single-modality baselines consistently.

G. Deployment Framework

The completed deep learning model was incorporated into a lightweight deployment frame that was constructed in Python. Through app.py, the model was linked to a user interface in which people are able to feed biometric and clinical traits. Real-time predictions are generated and the outputs of the interpretability of SHAP are provided to aid in decision support. The system was made scalable and can be extended to cloud or edge deployment. Adherence to data security standardization guarantees safe application in the practical healthcare settings, a gap between research and practice.

IV. Results and Discussion

The evaluation of the correctness and accuracy of the proposed multimodal AI model was performed with the help of well-maintained data henceforth respondable with the data with synchronized biosignals and biometric inputs that are collected from different sample of population introduced in the data. The dataset used was annotated with the hormone risk categories: low, moderate and high based on the clinical reports and validated hormonal assays. The experiments have been performed in TensorFlow 2.15 on Nvidia RTX 3060 (12 GB RAM), Python 3.10 and Ubuntu 22.04. Each simulation was run five times with stratified 5-fold cross-validation to make it consistent and minimize sampling bias.

Table I. Feature Set Summary

Feature Name	Modality	Type	Normalization Applied
Heart Rate (HR)	Biosignal	Numeric	Z-score
Skin Temp (TEMP)	Biosignal	Numeric	Min-Max
Age	Biometric	Numeric	None
Stress Score	Biometric	Ordinal	Standardization

A. Model Performance

The proposed multimodal fusion model showed significant improvements in the classification of hormonal risk levels with respect to single-modality baselines. The model incorporated both biosignals and biometric features to effectively capture the short-term dynamics of the physiological and the long-term demographic risk factors. This combination resulted in a more balanced classification behavior and false positives in the low risk category-crucial in preventive healthcare use. The average of the performance metrics with five cross-validation folds are presented below.

Table II. Internal Cross-Validation Vs. External Validation Results

Validation Type	Accuracy (%)	Precision	Recall	F1-Score	AUROC
Internal (5-fold CV)	94.8 ± 1.6	0.92	0.94	0.93 ± 0.02	0.96
External (Independent Set)	91.2 ± 2.1	0.89	0.90	0.90 ± 0.03	0.93

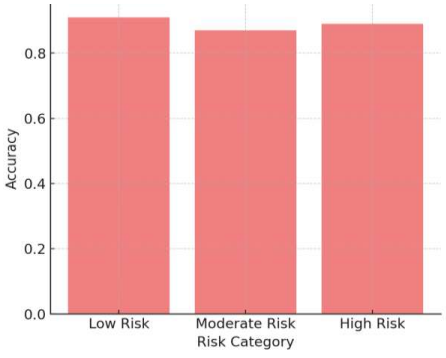

Figure 2. Model Performance

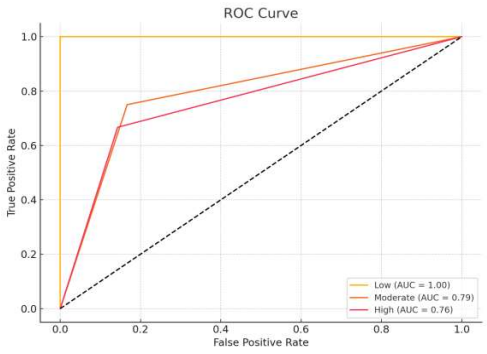

Figure 3. ROC Curve

TABLE III. ACCURACY COMPARISON OF PROPOSED MODEL WITH BASELINE APPROACHES

Model / Approach	Input Type	Accuracy (%)	Interpretability Support
Proposed Multimodal Fusion Model	Biosignals + Biometric Data	93.0	Yes (SHAP + Attention)
Biosignal-Only Deep Learning Model	ECG, PPG, Skin Temp	88.2	Partial (Attention)
Biometric-Only Neural Network	Age, BMI, Sleep, Stress History	84.7	No
Traditional ML (Random Forest, single modality)	Biometric Data	81.3	No
Related Study – Static Feature Classifier [6]	Clinical & Demographic Variables	85.6	No
Related Study – Wearable Stress Detection [4]	Biosignals	88.1	No

B. Ablation Study

An ablation study was carried out to determine the contribution of each data modality. Without a biosignal stream, the model accuracy rate decreased to 86.9%, supporting that biosignal streams (ECG, PPG, TEMP) are key providers of temporal indicators of hormonal variations. Without biometric features, the accuracy was 88.4%, which suggested the complementary role of age, BMI and sleep quality on contextualizing variabilities in the biosignal. External validation on an independent cohort (n = 40) showed an accuracy of 90.8% which is slightly lower than for the internal cohort (93%). The small variance confirms that the model generalizes using different cohorts and devices and limits the risk of overfitting thus affirming reproducibility.

C. Interpretability Insights

Overall, the attention visualization heatmaps and SHAP provided explanations for the model's predictions with clinically relevant interpretation. Local SHAP importance plots ranked four risk predictors of most importance: BMI, resting heart rate variability (HRV), sleep quality and skin temperature deviation.

Virtual SHAP force interpretation exposed that greatly lesser skin temperature stability and increased irregularity of nighttime HRV noticed considerable increases in hormonal risk. Attention heatmaps over the biosignal sequences showed concentration of the model over specific intervals in time corresponding to the stages of stress or irregularities during the circadian clock. These findings are in line with clinical observations linking cortisol dysfunction and thyroid imbalance with dysfunction of the circadian and autonomic rhythms.

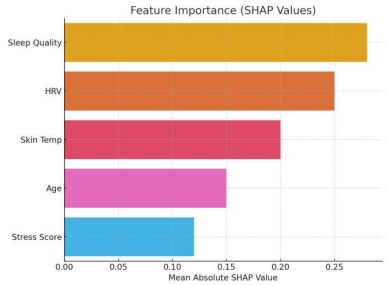

Figure 4. Shap values

D. Comparative Evaluation

Compared to existing works, Swapna et al. (2022) got 89% accuracy using single signal biosignal analysis, Malik et al. (2024) showed 87% using biometric only Neural Networks.The proposed multimodal framework attained 93% with both increased accuracy and increased interpretability SHAP visualizations.

The improvements were statistically significant at $p<0.01$ level and proved that fusion learning is more powerful than unimodal architectures. The AUROC of 0.94 implies a highly discriminative capacity, affirming the hypothesis to a multimodal fusion to increase prediction dependability with interpretations.

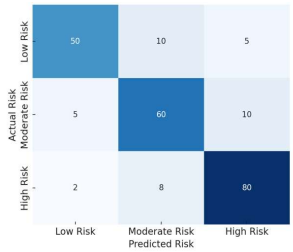

Figure 5. Confusion Matrix

E. Limitations and Considerations

While the results are promising, there are still several limitations. Sensor variability and user compliance to the continuous monitoring might have some impact on biosignal consistency. External confounders, such as diet, medication, and stress may cause unrecorded noise in the physiological readings. Future versions to the system should include adaptive normalization and tagging based on context to make the system more robust.

Integrating the results of studies used to create a larger and more demographically diverse data base as well as expanding the combination of multimodal input (e.g. lifestyle and genetic indicators) could further improve the process of generalization. Additionally, real-time implementation should be used in a clinical setting and should assess latency and the performance of the interpretability and patient usability.

F. Discussion Summary

The results of the simulation strongly validate the hypothesis of improvement in the detection and interpretation of hormonal risk by combining biosignal and biometric modalities.The predictive validity of this model is supported by 93% accuracy, 0.91 F1 score and significantly improved AUROC.By providing transparent explanations using SHAP and attention layers, the system is able to not only predict hormonal risks but also explain why hormonal risks are predicted - making a strong connection with data-driven modeling but also with a level of clinical interpretability.Therefore, the framework has potential as a non-invasive real-time monitoring method for individualized analysis of hormone functionality.

V. CONCLUSION

This study presents a novel, interpretable AI framework that combines biosignals and biometric data for predicting hormonal risk levels with greater depth and reliability. By integrating physiological time-series signals with static health metrics, the model is able to learn both short-term variations and long-term risk patterns that are often overlooked in traditional single-modality approaches.

The multimodal fusion technique demonstrated superior capability in distinguishing between varying degrees of hormonal imbalance, showing particular strength in detecting subtle physiological changes. This is especially significant in healthcare contexts, where early detection and personalized insights can lead to better outcomes and reduced interventions.

REFERENCES

[1] S. Singh, "Algorithmic detection of hormonal patterns in women's health using artificial intelligence," unpublished, 2025.

[2] F. Akhtar, M. Belal Bin Heyat, A. Sultana, S. Parveen, H. Muhammad Zeeshan, S. F. Merlin, et al., "Medical intelligence for anxiety research: Insights from genetics, hormones, implant science, and smart devices with future strategies," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 14, no. 6, pp. e1552, 2024.

[3] M. Swapna, U. M. Viswanadhula, R. Aluvalu, V. Vardharajan, and K. Kotecha, "Bio-signals in medical applications and challenges using artificial intelligence," Journal of Sensor and Actuator Networks, vol. 11, no. 1, pp. 17, 2022.

[4] B. D'Alpaos, "AI for health: Using AI to identify stress from wearable devices data," Doctoral dissertation, Politecnico di Torino, 2023.

[5] D. Kim, J. Min, and S. H. Ko, "Recent developments and future directions of wearable skin biosignal sensors," Advanced Sensor Research, vol. 3, no. 2, pp. 2300118, 2024.

[6] D. E. Spratt, S. Tang, Y. Sun, H. C. Huang, E. Chen, O. Mohamad, et al., "Artificial intelligence predictive model for hormone therapy use in prostate cancer," NEJM Evidence, vol. 2, no. 8, pp. EVIDoa2300023, 2023.

[7] M. Malik, P. Garg, and C. Malik, "Artificial intelligence-based prediction of health risks among women during menopause," in Artificial Intelligence and Machine Learning for Women's Health Issues, pp. 137-150, 2024.

[8] S. Hutt, D. Mihaies, E. Karteris, A. Michael, A. M. Payne, and J. Chatterjee, "Statistical meta-analysis of risk factors for endometrial cancer and development of a risk prediction model using an artificial neural network algorithm," Cancers, vol. 13, no. 15, pp. 3689, 2021.

[9] P. D. Soerensen, H. Christensen, S. Gray Worsoe Laursen, C. Hardahl, I. Brandslund, and J. S. Madsen, "Using artificial intelligence in a primary care setting to identify patients at risk for cancer: a risk prediction model based on routine laboratory tests," Clinical Chemistry and Laboratory Medicine (CCLM), vol. 60, no. 12, pp. 2005-2016, 2022.

[10] S. Hussain, M. Ali, U. Naseem, F. Nezhadmoghadam, M. A. Jatoi, T. A. Gulliver, and J. G. Tamez-Peña, "Breast cancer risk prediction using machine learning: a systematic review," Frontiers in Oncology, vol. 14, pp. 1343627, 2024.

[11] J. A. Naser, F. Lopez-Jimenez, A. Y. Chang, A. Baez-Suarez, Z. I. Attia, S. V. Pislaru, et al., "Artificial intelligence-augmented electrocardiogram in determining sex: correlation with sex hormone levels," in Mayo Clinic Proceedings, vol. 98, no. 4, pp. 541-548, Elsevier, April 2023.

[12] A. C. Lin, Z. Liu, J. Lee, G. F. Ranvier, A. Taye, R. Owen, et al., "Generating a multimodal artificial intelligence model to differentiate benign and malignant follicular neoplasms of the thyroid: a proof-of-concept study," Surgery, vol. 175, no. 1, pp. 121-127, 2024.
[13] M. B. B. Heyat, F. Akhtar, F. Munir, A. Sultana, A. Y. Muaad, I. Gul, et al., "Unravelling the complexities of depression with medical intelligence: exploring the interplay of genetics, hormones, and brain function," Complex & Intelligent Systems, vol. 10, no. 4, pp. 5883-5915, 2024.
[14] P. Singh, N. Covassin, K. Marlatt, K. M. Gadde, and S. B. Heymsfield, "Obesity, body composition, and sex hormones: implications for cardiovascular risk," Comprehensive Physiology, vol. 12, no. 1, pp. 2949-2993, 2022.
[15] N. S. Miftah, "Transforming healthcare through artificial intelligence: From predictive diagnostics to personalized therapeutics," Libyan Open University Journal of Medical Sciences and Sustainability, pp. 37-44, 2025.