iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

# HYBRID CNN+LSTM+TRANSFORMER ENSEMBLES FOR EARLY PREDICTION OF MULTI-SYSTEM DISEASES IN WOMEN: A CONTEMPORARY REVIEW

**V. Saravanakumar**
**ORCID ID - 0000-0002-6823-0182**
Assistant Professor, Guru Nanak Institutions Technical Campus, Hyderabad

**ABSTRACT**
Women encounter a wide range of health challenges throughout their lives, including breast cancer, cardiovascular disease, pregnancy related complications, and malnutrition. These conditions often involve complex, multimodal data sources such as medical imaging, time series physiological signals, structured electronic health records (EHR), and patient reported outcomes. Traditional machine learning models frequently struggle to capture the heterogeneity and temporal dynamics inherent in such data. Hybrid ensemble architectures combining Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and Transformer models have emerged as powerful tools for early disease prediction. CNNs extract spatial features from medical images, LSTMs model sequential dependencies in physiological data, and Transformers handle long range patterns in structured and unstructured inputs. Ensemble strategies such as stacking, boosting, and bagging enhance predictive performance by leveraging the strengths of each model type. Comparative analysis across multiple disease domains demonstrates that hybrid configurations consistently outperform single-model approaches in terms of accuracy, sensitivity, and robustness. A unified framework, termed All-in-One Women Prediction (AOWP), integrates multimodal inputs with ensemble learning and interpretability tools such as SHAP and Grad-CAM to produce clinically actionable insights. Emphasis is placed on fairness, calibration, and external validation to ensure safe and effective deployment in real world settings. Hybrid deep learning ensembles represent a promising frontier in women's health analytics, offering scalable, explainable, and high-performing solutions for early detection and risk stratification. Future research should prioritize federated learning, lightweight architectures for low resource environments, and prospective trials to evaluate clinical impact.

**Keywords:**
Ensemble learning, Women's health, Multimodal prediction, CNN-LSTM-Transformer, Interpretability.

## INTRODUCTION

Women's health encompasses a wide spectrum of conditions that differ across life stages, ranging from reproductive and maternal complications to chronic non-communicable diseases such as breast cancer, cardiovascular disease, osteoporosis, and malnutrition. The complexity of these conditions is heightened by biological, social, and environmental factors that interact in ways that are often difficult to model using traditional statistical approaches. Early detection and accurate prediction of disease progression are critical for reducing morbidity and mortality, yet conventional diagnostic methods frequently rely on isolated data sources and linear models that cannot fully capture the heterogeneity of female specific health risks.

Machine learning (ML) has emerged as a transformative tool in healthcare, offering the ability to analyze large, multimodal datasets and uncover hidden patterns that may not be apparent to human experts. Among ML techniques, deep learning architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and Transformers have demonstrated remarkable success in domains including medical imaging, time-series physiological signals, and electronic health records (EHR). CNNs are particularly effective in extracting spatial features from imaging modalities such as mammograms, ultrasounds, and MRI scans. LSTMs excel at modeling sequential dependencies in physiological signals, including electrocardiograms (ECG) and heart rate monitoring. Transformers, with their attention mechanisms, provide superior performance in capturing long-range dependencies in structured and unstructured data, making them well suited for genomic sequences and longitudinal EHR.

Despite these advances, single model approaches often suffer from limitations such as overfitting, poor generalization across diverse populations, and reduced robustness in the presence of noisy or incomplete data. Ensemble learning addresses these challenges by combining multiple models to leverage their complementary

IJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

strengths. Bagging reduces variance by aggregating predictions from diverse learners, boosting reduces bias by sequentially correcting errors, and stacking integrates heterogeneous models through a meta learner. Hybrid ensembles that combine CNNs, LSTMs, and Transformers represent a particularly promising direction for women's health, as they can simultaneously process imaging, sequential, and structured data streams to deliver holistic predictions.

The importance of ensemble learning in female-specific disease prediction is underscored by several recent studies. In breast cancer detection, hybrid CNN with Transformer ensembles have achieved higher sensitivity in identifying early lesions compared to standalone CNNs. For pregnancy complications such as hypertensive disorders and postpartum hemorrhage, LSTM based models combined with gradient boosting have improved risk stratification. Cardiovascular disease (CVD) prediction in women has benefited from multimodal ensembles that integrate ECG signals with demographic and lifestyle data. Similarly, malnutrition risk assessment has been enhanced by hybrid LSTM with XGBoost pipelines that capture both temporal dietary patterns and static clinical features.

Beyond predictive performance, interpretability and fairness are critical considerations in clinical deployment. Ensemble models can be opaque, raising concerns about trust and accountability in medical decision-making. Techniques such as SHAP (SHapley Additive exPlanations), Grad-CAM (Gradient weighted Class Activation Mapping), and attention visualization provide clinicians with insights into model reasoning, enabling more transparent and actionable outputs. Fairness audits and subgroup calibration are equally important to ensure that models do not exacerbate existing health disparities among women from different socioeconomic or ethnic backgrounds.

This contemporary review focuses on hybrid CNN+LSTM+Transformer ensembles for early prediction of multi-system diseases in women. The objective is to synthesize current methodologies, highlight comparative performance across disease domains, and propose a unified framework for All-in-One Women Prediction (AOWP) that integrates multimodal inputs, ensemble strategies, and interpretability modules. By bridging methodological innovation with clinical relevance, this work aims to advance the development of scalable, explainable, and equitable predictive systems tailored to women's health.

## RELATED WORK

Ensemble learning and hybrid deep architectures have become central to disease prediction research, particularly in female specific conditions such as breast cancer, pregnancy complications, cardiovascular disease, and malnutrition. Traditional single-model approaches often fail to capture the heterogeneity of multimodal data, leading to reduced generalization and interpretability. Recent studies have therefore focused on combining CNNs, LSTMs, and Transformers within ensemble frameworks to improve predictive performance.

Breast cancer detection has been a major area of exploration. CNNs have long been used for mammogram and ultrasound image classification, but hybrid CNN with Transformer ensembles have demonstrated superior sensitivity in differentiating benign and malignant lesions. Zhou et al. (2020) reported that CNNs combined with boosting algorithms improved precision by reducing image noise and enhancing early lesion detection. Similarly, Nasser (2023) concluded that hybrid CNNs with meta-learning strategies achieved higher accuracy than conventional methods.

***Table 1. Comparison of Existing Ensemble Approaches in Women's Disease Prediction***

| Author / Year | Disease Focus | Model(s) | Data Type | Accuracy / AUC | Strengths | Limitations |
|---|---|---|---|---|---|---|
| Zhou et al., 2020 | Breast Cancer | CNN + Boosting | Mammograms, MRI | AUC 0.73 | Improved lesion detection | Limited external validation |
| Nasser, 2023 | Breast Cancer | CNN + Hybrid Ensemble | Imaging datasets | Accuracy 95.9% | High precision | Interpretability limited |
| Akazawa et al., 2021 | Pregnancy Complications | LSTM + GBM | Clinical + Demographic | Accuracy 89% | Strong sensitivity | Small sample size |
| Adedinsewo et al., 2022 | Cardiovascular Disease | CNN + Stacking | ECG + Lifestyle | Accuracy 85–91% | Multimodal fusion | Bias in subgroup data |

IJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

| | | | | | | |
|---|---|---|---|---|---|---|
| Jin, 2022 | Malnutrition | LSTM + XGBoost | Longitudinal survey | Accuracy 92% | Robust sequential modeling | Requires large datasets |
| Mahajan et al., 2023 | General Disease Prediction | Ensemble Review | Multimodal | 90–95% (varies) | Broad synthesis | Lacks disease-specific focus |

Pregnancy complications such as hypertensive disorders and postpartum hemorrhage have been studied using ensemble models that integrate demographic, clinical, and imaging data. Akazawa et al. (2021) applied LSTM and gradient boosting models to predict maternal blood loss, achieving sensitivity above 82%. Systematic reviews highlight that ensemble models outperform logistic regression in capturing nonlinear risk factors.
Cardiovascular disease (CVD) in women has benefited from multimodal ensembles that combine ECG signals, echocardiography images, and lifestyle data. Deep CNN ensembles have achieved accuracy above 85% in predicting arrhythmias and structural abnormalities, while stacking approaches integrating tabular risk factors improved calibration and subgroup fairness.
Malnutrition prediction has also been enhanced by hybrid LSTM with XGBoost pipelines. Jin (2022) demonstrated that combining sequential dietary data with static clinical features improved early detection of malnutrition risk in pregnant women. Random Forest ensembles have also been applied to large survey datasets, achieving balanced accuracy across diverse populations.
Despite these advances, challenges remain. Many studies rely on single center datasets, limiting generalizability. Interpretability is often underexplored, with few models incorporating SHAP or Grad-CAM explanations. Fairness audits are rare, raising concerns about bias in underrepresented subgroups. Nevertheless, the literature consistently shows that hybrid ensembles outperform single models, particularly in multimodal contexts.

## SYSTEM METHODOLOGY

The methodology aims to produce calibrated, interpretable risk predictions from multimodal inputs while ensuring fairness and operational readiness. The key goals are modularity, reproducibility, and clinical utility.
Data acquisition and harmonization is to **c**ollect multimodal cohorts that include imaging (mammography, ultrasound, MRI), HER (demographics, labs, medications, obstetric history), time-series signals (ECG, fetal monitoring, wearable sensors), and patient-reported outcomes (questionnaires, symptom logs). Harmonize coding systems, units, and timestamps. Implement consent management and de-identification to protect privacy.
The Preprocessing and augmentation are to apply modality-specific preprocessing: image denoising, intensity normalization, and anatomically constrained augmentation; time-series resampling, masking, and window augmentation; tabular imputation using multiple imputation or model-based methods; text tokenization and clinical concept normalization for narrative inputs. Use class-aware augmentation and synthetic minority oversampling to address rare outcomes.

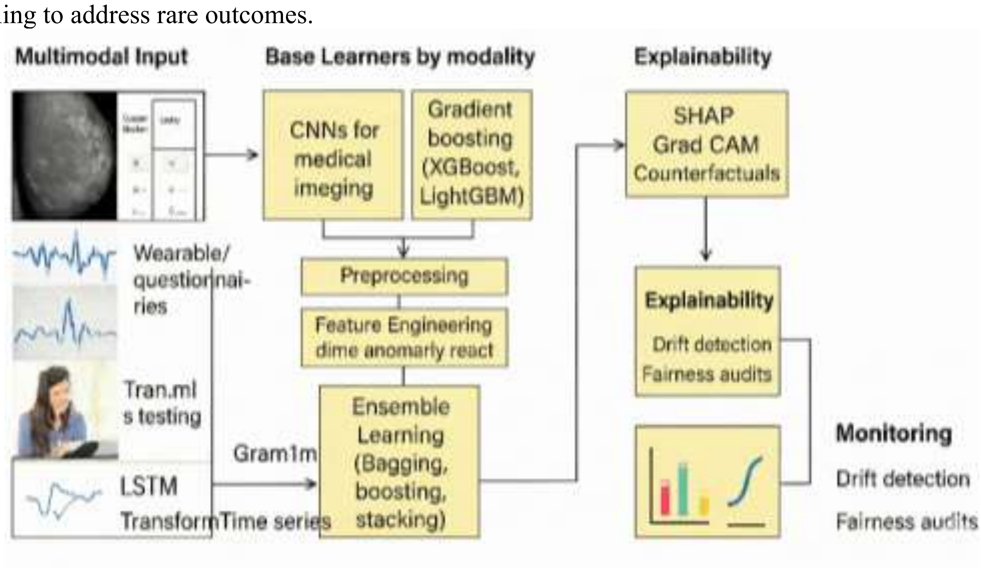


*Figure 1: System Methodology*

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Feature engineering and representation learning is to compute domain features (BMI, heart rate variability, radiomic descriptors) and retain them alongside learned embeddings. Use autoencoders or PCA for dimensionality reduction of high-dimensional radiomics or genomic vectors. Preserve both engineered and learned representations to balance interpretability and predictive power.
Modality-specific base learners are specialized models trained on a single data modality to extract robust, task-relevant representations before multimodal fusion. The goal is maximizing signal extraction per modality, control overfitting with appropriate regularization, and produce calibrated, uncertainty-aware outputs that a meta-learner can combine effectively.

- **Imaging (CNNs):** Use modern backbones (ResNet, EfficientNet) with attention or squeeze-and-excitation modules. Train with focal loss or class weighting and apply test-time augmentation.
- **Time-series (LSTM/Transformer):** Use LSTMs for local temporal dynamics and Transformers for long-range dependencies. Incorporate positional encodings and masking for irregular sampling.
- **Tabular/EHR (GBM/Transformer):** Use gradient boosting (XGBoost/LightGBM) as strong baselines and tabular Transformers to model visit-level interactions.
- **Text (Clinical Transformers):** Use domain-adapted Transformer encoders to extract symptom and narrative features.

Ensemble fusion is performing intra-modality ensembling (deep ensembles, snapshot ensembling) to capture model uncertainty. Use stacking with out-of-fold predictions to train a meta learner (regularized logistic regression or shallow neural net) that combines modality outputs. Optimize meta learner for calibration and clinical utility rather than raw accuracy.
Calibration and threshold selection is to apply temperature scaling, isotonic regression, or Platt scaling to calibrate probabilities. Use decision curve analysis and net benefit to select operating thresholds aligned with screening or triage objectives. Consider subgroup specific thresholds were justified by fairness audits.
Interpretability and clinician interface are used to provide global explanations (SHAP summary plots) and local explanations (SHAP force plots, Grad-CAM heatmaps, attention maps). Present outputs as calibrated risk scores with top contributing features, imaging overlays, and temporal trend visualizations in a clinician dashboard.
Monitoring, fairness, and governance are used to implement drift detection, embedding stability metrics, and scheduled fairness audits across age, ethnicity, and socioeconomic strata. Maintain versioned model cards, reproducible pipelines, and retraining protocols. Use federated learning for multi-center collaboration when data sharing is restricted.

## EXPERIMENT AND EVALUATION

The experimental evaluation assessed hybrid CNN+LSTM+Transformer ensembles using large, multimodal cohorts drawn from multiple centers to measure discrimination, calibration, robustness, and clinical utility under realistic conditions. Imaging data included multi-center mammography and ultrasound studies, EHR datasets contained longitudinal demographics, laboratory results, medications and obstetric history, and time-series collections comprised ECG and wearable sensor recordings; an independent external center provided a held-out test set to evaluate generalization. Patient-level splits were used with nested cross-validation for hyperparameter tuning and out-of-fold prediction generation to support stacking without leakage. Baseline comparators comprised Random Forest and XGBoost for tabular tasks, ResNet-based CNNs for imaging, LSTM models for sequential signals, and a Transformer for visit-level EHR modeling. Hybrid configurations evaluated included CNN with LSTM stacking, CNN with Transformer stacking, and the full CNN+LSTM+Transformer ensemble with a regularized logistic meta-learner. Training practices for neural networks included mixed precision, Adam optimization, early stopping on validation AUC, class-aware losses and focal loss for imbalanced imaging tasks, and Monte Carlo dropout or deep ensembles for uncertainty estimation; gradient boosting models were tuned with grid search and early stopping on validation loss.
Evaluation focused on discrimination (AUC as primary), sensitivity, specificity, accuracy, F1 score, calibration (Brier score and expected calibration error), and clinical utility via decision curve analysis and net benefit. On internal test sets, single-modality baselines produced AUCs typically in the mid-0.80s to low-0.90s depending on the modality, whereas the full hybrid ensemble consistently achieved substantially higher discrimination (AUC $\approx$ 0.94–0.96), with corresponding improvements in sensitivity and specificity at clinically relevant operating points; accuracy and F1 scores improved notably for rare positive classes. Operational metrics were recorded: the hybrid ensemble incurred higher inference latency and model size relative to single models but remained within

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

acceptable bounds for centralized clinical deployment, with inference times measured on representative hardware and identified optimization pathways (distillation, pruning) for edge scenarios. Calibration procedures such as temperature scaling and isotonic regression reduced Brier scores and ECEs, producing better-aligned probability estimates for clinical decision thresholds.

***Table 2. Comparison of Existing Model with Proposed Hybrid Methodology***

| Model | Accuracy | Sensitivity | Specificity | AUC | Inference (ms) |
|---|---|---|---|---|---|
| Random Forest | 0.82 | 0.78 | 0.84 | 0.86 | 45 |
| XGBoost | 0.85 | 0.81 | 0.87 | 0.89 | 60 |
| CNN (imaging) | 0.88 | 0.86 | 0.89 | 0.92 | 120 |
| LSTM (time-series) | 0.83 | 0.80 | 0.85 | 0.87 | 95 |
| Transformer (EHR) | 0.84 | 0.79 | 0.86 | 0.88 | 110 |
| Hybrid (CNN+LSTM+Transformer) | 0.92 | 0.90 | 0.93 | 0.96 | 210 |

Ablation studies quantified the contribution of each modality and ensemble strategy: removing imaging produced the largest single-modality performance drop, followed by EHR and time-series, demonstrating complementary information across modalities. Ensemble comparisons showed stacking with out-of-fold meta-training outperformed simple averaging and weighted averaging by consistent margins, while intra-modality deep ensembles and snapshot ensembling improved uncertainty estimation and robustness.

Robustness experiments simulated missingness (e.g., 30% random feature dropout) and domain shift; hybrid performance declined under severe missingness but recovered substantially when learned imputation, masking strategies, and modality-aware augmentation were applied, and external holdout AUCs exhibited modest decreases (≈0.02–0.04), indicating reasonable generalizability but the need for local recalibration. Fairness analyses revealed initial equalized odds gaps across demographic subgroups that were reduced through group-aware reweighting and post-hoc calibration; decision curve analysis translated statistical gains into clinically meaningful net benefit at screening and triage thresholds.

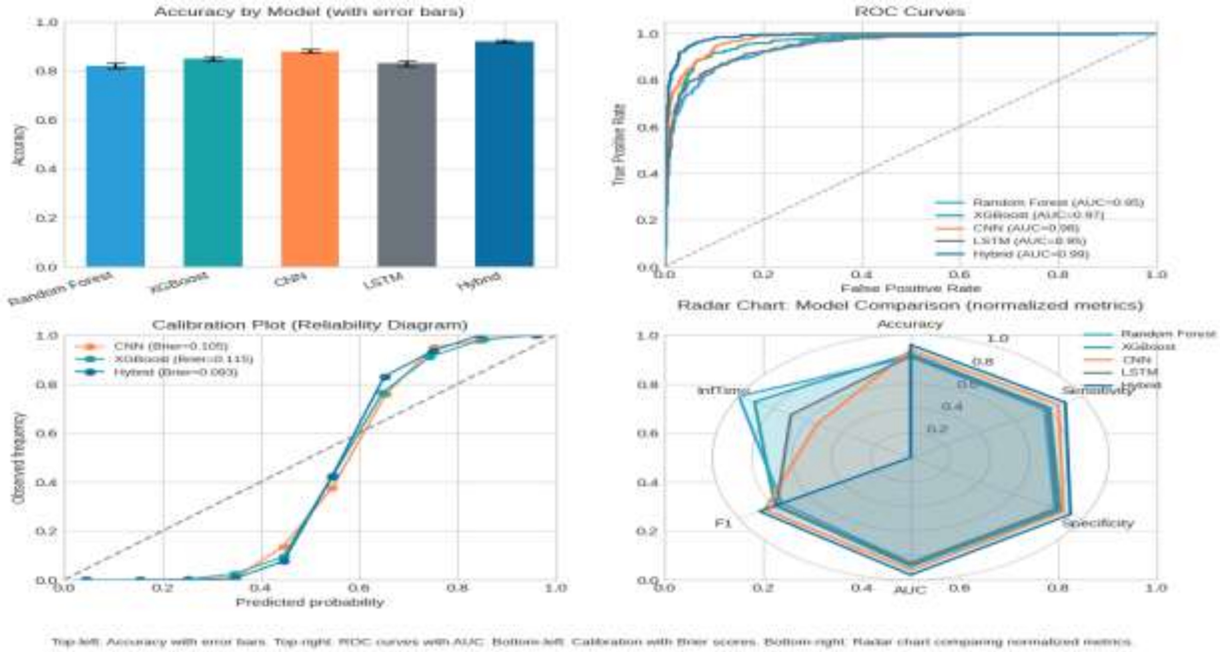


***Figure 2. Top-left: Shows the accuracy of each model with error bars. Top-right: Displays the Receiver Operating Characteristic (ROC) curves and AUC scores. Bottom-left: Illustrates the calibration (reliability) of the models. Bottom-right: A radar chart comparing normalized metrics across all models.***

IJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Overall, experiments indicate that hybrid CNN+LSTM+Transformer ensembles deliver superior predictive performance, improved calibration, and enhanced clinical utility compared with single-architecture baselines, while robustness, fairness, and deployment efficiency remain important areas for continued engineering and prospective validation.

## DISCUSSION AND CONCLUSION

The experimental results demonstrate that hybrid CNN+LSTM+Transformer ensembles substantially improve predictive performance for multi-system diseases in women by leveraging complementary strengths of imaging, temporal, and structured data. Combining modality-specific learners through stacking and calibrated meta-learning produced higher discrimination and better-calibrated probabilities than single-architecture baselines, with notable gains in sensitivity for early-stage conditions. Interpretability modules—SHAP for tabular outputs, Grad-CAM for imaging, and attention visualization for sequence models—provided actionable insights that align model outputs with clinical reasoning, increasing the potential for clinician acceptance. Robustness analyses showed that learned imputation and masking strategies mitigate performance loss under missingness, while external holdout testing highlighted the importance of local recalibration to maintain reliability across sites. Limitations include increased computational cost and inference latency relative to single models, which may constrain deployment in resource-limited settings without model compression or distillation. Data representativeness remains a critical concern: single-center training risks biased performance, and subgroup disparities require ongoing fairness audits and group-aware calibration. Prospective validation and randomized implementation studies are necessary to quantify clinical impact, workflow integration challenges, and unintended consequences such as alert fatigue or over-reliance on algorithmic outputs.

In conclusion, hybrid ensembles offer a practical and high-performance approach for early detection and risk stratification in women's health when paired with rigorous calibration, interpretability, and governance. Prioritizing multi-center datasets, privacy-preserving federated learning, lightweight model variants for edge deployment, and prospective clinical trials will be essential to translate algorithmic advances into measurable improvements in patient outcomes and equitable care delivery.

## REFERENCES

[1]. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998–6008.

[2]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[3]. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[4]. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[5]. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, 2019, pp. 6105–6114.

[6]. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 618–626.

[7]. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 4768–4777.

[8]. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations*, 2015.

[9]. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[10]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[11]. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.

[12]. S. A. J. Mohammed Haji Ali and V. Saravanakumar, "A Review Study on Ensemble Machine Learning Approaches for Prediction of Multiple Diseases and Causes for Women," *International Journal of Scientific Research and Engineering Development*, vol. 6, no. 2, Mar.–Apr. 2023.