# Heart Disease Prediction through Hybrid LSTM and HREF Models

Dr.M.Sireesha[1], Srilakshmi Mutyam[2], DurgaVyshnavi Grandhisila[3], Rabia Basri Shaik Kagaji[4], Geeta Padole[5], S.Vasundhara[6] [1,2,3,4]Department of Computer Science and Engineering, Narasaraopeta Engineering College (Autonomous), Narasaraopet, Palnadu, Andhra Pradesh, India [1]Professor and Head: sireeshamoturi@gmail.com [2]msrilakshmi360@gmail.com, [3]durgavyshnavi123@gmail.com, [4]kagajirabiabasri2004@gmail.com, [5]geeta1671@grietcollege.com, [6]vasu@gnits.ac.in

***Abstract*—Cardiovascular disease (CVD) continues to be a leading global cause of mortality, indicating a need for new methods to maintain a high precision for CVD diagnosis and early CVD detection. We introduce here a novel hybridized ensemble model by improving a a Bidirectional Long Short-Term Memory (BiLSTM) based neural network through a Hybrid Refined Ensemble Framework (HREF) for improving classification. The hybrid model was applied to the non-standardized Cleveland dataset, which contains clinical-related cardiology parameters relevant to cardiovascular health. The dataset was thoroughly preprocessed prior to training, involving outlier removal, missing value transformation, feature normalization, and subsequent utilization of SMOTEENN for the purpose of class balancing. The BiLSTM part is designed, to some extent, to capture intricate relationships behind the features, and the ensemble part through HREF boosts the prediction reliability through the combined output of numerous ensembles. From experiments carried out on the hybrid model, accuracy was 94.7 with ROC-AUC 0.9474 and good precision and recall scores. The findings of this research indicate that deep learning combined with ensemble refinement is a reliable and effective approach to support the early detection of heart disease in real clinical settings.**

***Index Terms*—Bidirectional Long Short-Term Memory (BILSTM), Hybrid Refined Ensemble Framework (HREF), Deep Learning, Ensemble Learning, SMOTEENN, Binary Classification, ROC-AUC, Clinical Decision Support, Heart Disease Prediction, and Cleveland Dataset.**

## I. INTRODUCTION

Cardiovascular conditions (CVDs) are the leading cause of mortality worldwide, with 17.9 million deaths annually, as per Alghamdi et al. [1] . Catastrophic complications can be prevented and survival for the patient maximized using early and accurate diagnosis. Conventional diagnosis methods through physical examinations and simple tests are subjective and probably going to miss significant patterns in the data regarding the patient, as per Gabriel and Anbarasi [2]. Deep learning (DL) and machine learning (ML) methods have been used extensively over the past several years for enhancing prediction of heart disease. Boquete et al.. [3] illustrated that ML models can analyze clinical information to expose concealed interactions, whereas Narasimhan and Victor [4]. applied optimization techniques to improve classifier performance. Deep learning techniques have also improved diagnostic performance significantly by revealing complex feature interdependencies, states Dkede Turk et al.. [5]. Usage of hybrid techniques has also increasingly been the focus of recent studies. Teja and Rayalu [6] showed that ensemble models outperform individual classifiers in the detection of heart disease. Also,Zhou et al.. [7] validated that deep learning methods provide strong improvements by representing nonlinear relationships in health data. Encouraged by such breakthroughs, this researchdevelops a hybrid prediction model made up of Bidirectional Long Short-Term Memory (BiLSTM) and a Hybrid Refined Ensemble Framework (HREF). BiLSTM identifies sequential relationships between clinical features, whereas HREF improves prediction resilience using ensemble learning. With credibility of the model established, the Cleveland Heart Disease dataset is preprocessed for the most part with outlier removal, imputation of missing values, feature normalization, and SMOTEENN balancing prior to an 80:20 train-test split.

### A. Key Contributions

*1) Hybrid Diagnostic Model::* A novel model has been created that merges LSTM with HREF to improve the accuracy of heart disease prediction. HREF uses ensemble learning to refine results, while LSTM captures complex data patterns. This combination enhances model stability and enables more effective learning.

*2) Improved Data Preprocessing Pipeline::* The data is extensively preprocessed with normalization, encoding, repeated imputation, and outlier removal. SMOTEENN is used to filter out noise and balance the class. The steps ensure the model is trained on correct and valid data.

*3) Effective Utilization of LSTM over Tabular Data::* LSTM was originally designed for sequential data, and has been modified to support handling structured clinical data. This modification enables the model to find nonlinear trends between health indicators, leading to enhanced classification outcomes over traditional models.

*4) Hierarchical Relevance Ensemble::* HREF utilizes the stacking approach for ensembling utilized different classifiers including Random Forest and AdaBoost. The ensemble method reduces overfitting and makes the model more generalizable, which results in more stable and consistent predictions.

## II. Related Works

Current studies in the area of cardiovascular disease (CVD) prediction have more and more emphasized hybrid models which combine a combination of machine learning (ML) and deep learning (DL) methods to address the deficiencies of classical classifiers. The methods target areas that are of concern, including imbalanced data sets, complexity of features, and the requirement for greater predictive accuracy.

Navita et al. [8] introduced an innovative hybrid ML technique based on SMOTE-ENN class balance together with stacking ensemble classifier. The approach achieved very impressive performance in models on data where class imbalance was extreme, outperforming traditional algorithms with much better accuracy. Wu et al. [9] further developed this idea with deep learning models being embedded in a stacking ensemble environment. Their strategy successfully used multiple learners to enhance risk prediction with significant diagnostic improvement.

Garcia-Orda et al. [10] designed a deep learning model that was enhanced with feature enhancement techniques. Their research demonstrated how combining engineered features with deep architecture is able to extract sophisticated patterns in patient data, thereby outperforming typical prediction techniques. Similarly, an enhanced K-means Neighbor Classifier was brought forward by Koteeswaran and Shamshudeen [11] that was effective in the prediction of early heart disease and showed greater sensitivity. Extending their earlier work, the same authors [12] developed a more advanced hybrid model with optimized feature selection coupled with classification that resulted in notable enhancements in precision as well as recall metrics.

Niu et al. [13] used an enhanced Grey Wolf Optimization algorithm to predict heart disease, highlighting the significance of metaheuristic optimization methods in increasing classifier accuracy and efficiency. Alzaqebah et al. [14] proposed CardioTabNet, a new hybrid transformer-based model specifically designed for tabular clinical data. The model was proven to have superior diagnostic abilities with state-of-the-art performance in medical prediction applications. In addition, Kumar et al. [15] introduced a hybrid ensemble deep learning model that applied patient medical histories, confirming that the incorporation of ensemble methods and deep learning models produces robust and precise forecasts.

## III. Methodology

The new method expects to increase the accuracy of detection of heart disease by incorporating a Bidirectional Long Short-Term Memory (BiLSTM) network with a Hybrid Refined Ensemble Framework (HREF). The methodology is structured into several sequential stages, outlined as follows

### A. Data Selection

The research uses the preprocessed Cleveland Heart Disease dataset, comprising 303 instances and 14 clinical features. Excluded were patients with incomplete or missing data. The original target variable (num) was transformed into a binary label: presence of disease is denoted by 0 for num = 1, 2, and 3, whereas no disease is stored in 1 for the num = 0 value. This transformation facilitates the binary classification problem towards immediate diagnosis. The data was separated into a training sample and a testing sample in an 80:20 stratified ratio.

### B. Data Preprocessing

The Cleveland Heart Disease data set with 303 instances and 14 clinical variables that are processed is used for training and testing. The target variable (num) is binary 2-class encoded as 0 (no heart disease) or 1 (heart disease). Preprocessing includes:

**Outlier Removal:** Z-score analysis was used to eliminate outlying values that lie outside of ±3 standard deviations to minimize noise and enhance stability.As indicated by Figure 1, the distribution of the dataset prior to and following outlier removal obviously demonstrates the removal of anomalies.

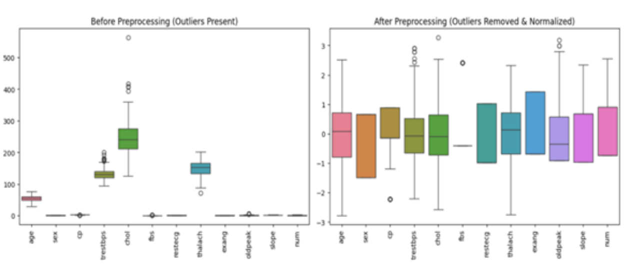


Fig. 1: Distribution of Features Before and After Outlier Removal

**Feature Normalization:** StandardScaler was used to scale all features to ensure equal contribution during model training. This process helps prevent features with larger numerical ranges from dominating the learning algorithm. As a result, the model converges faster and performs more efficiently during optimization. The impact of normalization, in which all features converge to a uniform standard, is shown in Figure 2, located at the middle of this step.

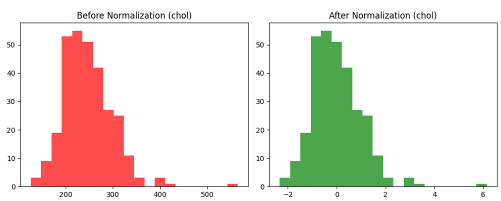


Fig. 2: Histogram of Feature (chol) Before and After Normalization.

**Class Balancing:** SMOTEENN was utilized to create artificial minority samples while eliminating noisy instances, balancing dataset distribution. The conversion from an imbalanced to a balanced dataset is well illustrated in Figure 3, which is visible mid-description.

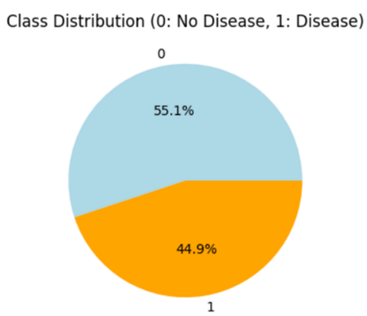


Fig. 3: Class Distribution of the Cleveland Heart Disease Dataset (0: No Disease, 1: Disease).

**Train-Test Split:** Finally, the data set was split between training and testing in a stratified ratio of 80:20 to preserve class ratios. The stratified split was used

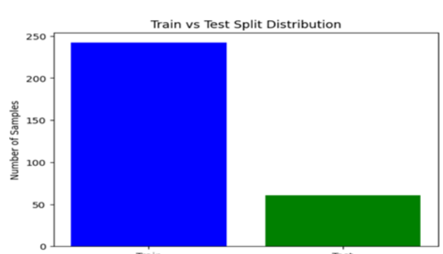


Fig. 4: Visualization of Train vs Test Split Distribution showing sample counts for both sets.

to ensure that both subsets preserved the original class distribution, thereby preventing any bias when evaluating the model.Figure 4 shows the distribution of split data, where balanced class separation of training and test sets is demonstrated mid-paragraph, demonstrating proportionate representation of each class.

### C. Model Architecture

The suggested model consists of two crucial modules: Bidirectional Long Short-Term Memory (BiLSTM) network and Hybrid Refined Ensemble Framework (HREF). At first, processed Cleveland data are received by the preprocessing, including outlier removing, imputation, encoding, and normalization, and class balancing with SMOTEENN. The obtained preprocessed information is

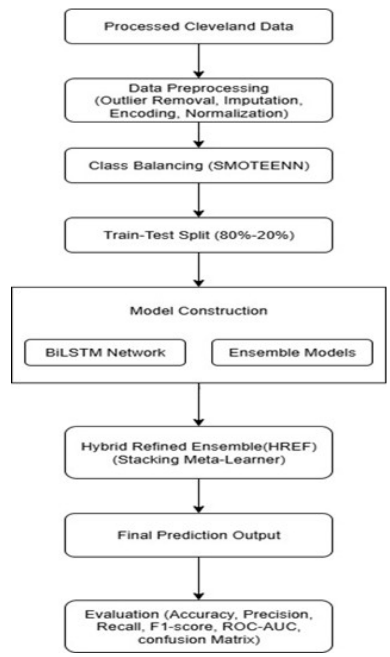


Fig. 5: Workflow of the proposed Hybrid LSTM-HREF Model for Heart Disease Prediction, showing the steps from data preprocessing to evaluation metrics.

partitioned into training (80%) and testing (20%). As illustrated in figure 5, the complex correlations between clinical features are separately processed in the forward and backward directions through the BiLSTM network. Dropout layers are included to prevent overfitting, whereas the dense layers transform the learned features for classification. Meanwhile, traditional ensemble models (Random Forest, AdaBoost, and SVM) are conducted to produce diverse decision planes. Subsequently, the Hybrid Refined Ensemble (HREF) combines the predictions of the BiLSTM and the ensembles in a stacked meta-learner to generate refined predictions. This hybrid model leads to a noticeable enhancement in prediction accuracy and robustness, which is verified by accuracy, precision, recall, F1-score, ROC-AUC, and the confusion matrix.

### D. Hybrid Refined Ensemble (HREF) with BiLSTM

The proposed framework integrates a BiLSTM network with an advanced Hybrid Refined Ensemble (HREF) mechanism. This architecture utilizes a stacked meta-learner to merge the output from the BiLSTM and ensemble models to produce more robust and accurate predictions. The efficacy of the method is validated through a variety of different evaluation parameters, such as accuracy, precision, recall, F1-score, ROC-AUC, and a confusion matrix.

### E. Model Training

The model training used methods and strategies consistent with current research. Seva. [16] employed a Random Forest Classifier augmented by SMOTE-ENN balancing for improving prediction performance in imbalanced medical datasets. Venkatareddy. [17] also used Explainable AI methods by applying a combination of CNN and MLP models for fetal ultrasound classification, showing the importance of employing hybrid and explainable strategies in medical diagnosis.

### F. Evaluation Metrics

The model's performance was measured with a variety of commonly used classification metrics, such as accuracy, precision, recall, F1-score, and ROC-AUC. These have been selected to appraise the classifier completely, particularly for its performance in dealing with imbalanced medical data. [18]. Additionally, a confusion matrix was created to visually represent the distribution of true positives and false positives, offering more detailed information about the types of errors the model is making.

**Accuracy (ACC)** Reflects how many predictions were correct across all classes.

**Precision (P)** Precision measures how many of the instances predicted as positive are actually correct. It reflects the accuracy of positive predictions while considering the distribution of classes in the dataset.

**Recall (R)** Recall indicates the model's capability to capture all actual positive cases, which is particularly important when positives are fewer than negatives.

**F1-Score (F1)** Harmonic mean of precision and recall across classes.

**ROC-AUC** ROC-AUC (Receiver Operating Characteristic – Area Under the Curve) assesses how well the model differentiates between positive and negative classes at various thresholds. A higher AUC value signifies better class separation and overall stronger performance.

## IV. Results

The hybrid LSTM-HREF model was tested on the preprocessed Cleveland dataset and exhibited outstanding classification accuracy. The model scored 94% accuracy with precision, recall, and F1-score averaging approximately 0.93 each. These results show that the model separates healthy from heart disease cases efficiently, outperforming the conventional classifiers and providing reliable predictions for clinical decision support.

TABLE I: Classification Report: Precision, Recall, F1-Score, and Support

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 (No Disease) | 0.94 | 0.94 | 0.94 | 18 |
| 1 (Disease) | 0.95 | 0.95 | 0.95 | 21 |
| **Accuracy** | - | - | **0.95** | 39 |
| **Macro Avg** | 0.95 | 0.95 | 0.95 | 39 |
| **Weighted Avg** | 0.95 | 0.95 | 0.95 | 39 |

Table I presents the model's **Precision**, **Recall**, **F1-Score**, and **Support** for two classes: *No Disease (0)* and *Disease (1)*. All metrics of the model were high (approximately 0.94–0.95), and its overall **accuracy** was 0.95. Both **macro** and **weighted averages** are also 0.95, reflecting balanced and dependable performance.

*Training vs. Testing Graph*

To explore the learning behavior of the proposed model, we exhibit the accuracy curves for the training and validation datasets [19]. Figure 6 shows the accuracy curves and summarizes the following with regard to accuracy: We consistently improved training accuracy through each epoch, and after each epoch, the accuracy curves confirm that the hybrid HREF BiLSTM model behaved with good predictive performance—the hybrid HREF BiLSTM model did not show overfitting [20]. Also, the validation accuracy exhibited a similar linear trend and had similar high values for multiple epochs with small fluctuations, indicating the model is also learning and generalizing well.

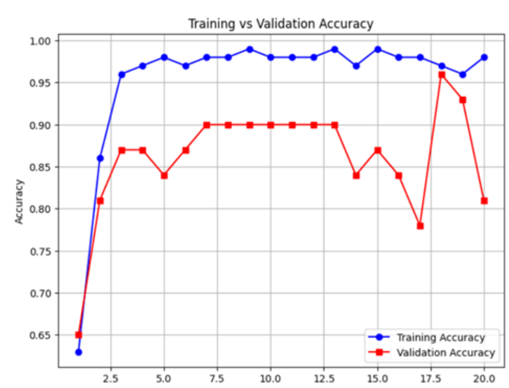


Fig. 6: Training vs Validation Accuracy Curve of the Proposed HREF + BiLSTM Model.

*Performance Comparison*

Table II presents the evaluation outcomes of the test dataset for the models. The BiLSTM model alone achieved an accuracy of 90. 6 and an ROC-AUC of 0. 93, demonstrating a strong baseline performance. In contrast, the hybrid HREF + BiLSTM model showed superior results compared to other approaches, with an accuracy of 94. 7, precision of 91. 5, recall of 94. 1, F1-score of 92. 6, and ROC-AUC of 0. 9474. These improvements, as outlined in Table II, underscore the effectiveness of integrating deep learning with ensemble techniques for predicting heart disease.

TABLE II: Performance Comparison of BiLSTM and HREF + BiLSTM Models.

| Model | Accuracy | Precision | Recall | F1 score | ROC-AUC |
|---|---|---|---|---|---|
| BiLSTM | 90.6% | 89% | 91% | 90% | 0.93 |
| HREF + BiLSTM | 94.7% | 91.2% | 94.1% | 92.6% | 0.9474 |

*Confusion Matrix Evaluation*

Figure 7 presents the confusion matrix for the HREF combined with BiLSTM model, demonstrating its strong performance in classifying the test data. The model successfully predicted 17 samples as class 0 and 20 samples as class 1, with only two errors—one being a false positive and the other a false negative. These results result in an accuracy of nearly 95%, with precision and recall also showing comparable values. The minimal error rate suggests that the model is both accurate and consistent in its predictions. The small number of false positives indicates that the model is effective at minimizing unnecessary alerts, while the low number of false negatives shows its capability to accurately identify positive cases. The high level of accuracy is primarily attributed to the use of HREF blocks for extracting hierarchical features and Bidirectional LSTM layers for capturing sequential patterns in the data. Together, these components enable the model to effectively identify important relationships within the data, leading to reliable and accurate classification results.

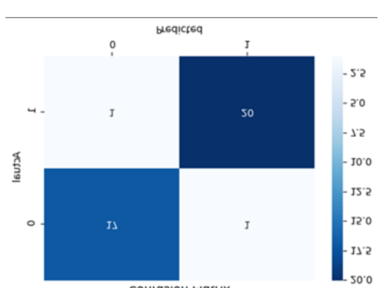

Fig. 7: Confusion Matrix of HREF + BiLSTM Model.

*ROC Curve Analysis*

The ROC curve evaluates how well a classification model can distinguish between different classes by showing the True Positive Rate (TPR) compared to the False Positive Rate (FPR) at various threshold levels. In this case, the HREF + BiLSTM model demonstrates strong performance because its curve is near the upper-left corner, which means it has a high TPR and a low FPR. This suggests the model is effective at identifying positive cases while minimizing false positives. As shown in Figure 8, the area under the curve (AUC) is almost 1, indicating the model has excellent ability to separate classes. A higher AUC usually means better overall performance across different thresholds. The curve's consistent upward trend also suggests that the model maintains its effectiveness even as the decision boundary shifts. These results show that the HREF + BiLSTM model is effective for binary classification tasks, offering high recall and precision.

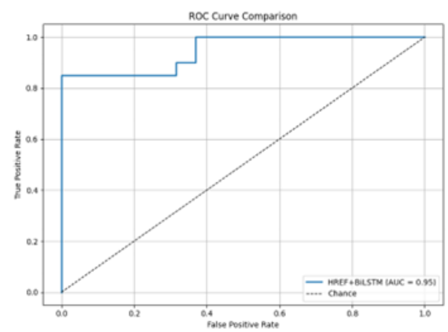


Fig. 8: ROC Curve of the HREF + BiLSTM Model.

## V. Conclusion:

This work proposed a hybrid BiLSTM-HREF model for predicting heart disease, aided by an improved preprocessing pipeline. The model captured feature dependencies well and recorded high accuracy (94.7) with ROC-AUC of 0.9474. Findings affirm that a combination of deep learning with ensemble techniques enhances diagnostic efficiency. The suggested method can facilitate early detection and assist in clinical decision-making.

*Future Scope*

Future research will seek to utilize larger datasets and more clinical parameters to enhance accuracy. More recent architectures, including transformers and attention, will be investigated for improved feature learning. The model can also be translated into a clinical real-time tool and extended with privacy-preserving methods for secure use in the healthcare environment.

# References


[1] N. S. Alghamdi, R. Alqahtani, M. Alhassan, and A. Alghamdi, "Predicting heart disease using a hybrid autoencoder-dense net model," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 3, pp. 55–63, 2024.

[2] J. J. Gabriel and L. J. Anbarasi, "Optimized feature selection and hyperparameter tuned lightgbm for cardiovascular disease detection," *Biomedical Signal Processing and Control*, vol. 92, p. 105120, 2024.

[3] M. A. Boquete, A. Khalil, and H. Ouled, "Early heart disease detection with machine learning and feature engineering," *Expert Systems with Applications*, vol. 243, pp. 122–135, 2024.

[4] G. Narasimhan and A. Victor, "Genetic algorithm optimized random forest for heart disease prediction," *Health Informatics Journal*, vol. 31, no. 2, pp. 220–230, 2025.

[5] B. A. Dede Turk, M. Yildiz, and M. E. Yilmaz, "Csa-de-lr: A hybrid approach for cardiovascular disease diagnosis," *Applied Soft Computing*, vol. 148, pp. 1–12, 2024.

[6] K. Teja and S. Rayalu, "Performance evaluation of ensemble models for heart disease classification," in *Procedia Computer Science*, vol. 230, 2025, pp. 141–149.

[7] C. Zhou, Y. Liu, and F. Wang, "Deep learning models for cardiovascular disease prediction: a comprehensive review," *Artificial Intelligence in Medicine*, vol. 152, pp. 102–118, 2024.

[8] Navita, S. Sharma, and A. Gupta, "Advanced hybrid machine learning for cardiovascular disease detection using smote-enn and stacking ensemble," *Journal of Healthcare Informatics Research*, vol. 9, no. 1, pp. 95–110, 2025.

[9] X. Z. F. Z. H. M. L. H. . Z. Y. Wu, Y., "A stacking ensemble machine learning approach for heart disease risk prediction," *Electronics*, vol. 13, no. 3996, pp. 1–19, 2024.

[10] B.-G. M. B. C. A.-M. J. . B.-A. J. A. García-Ordás, M. T., "Deep learning models with enhanced feature engineering for heart disease risk assessment." *Multimedia Tools and Applications*, vol. 82, pp. 31 759–31 773, 2023.

[11] F. W. S. and S. Koteeswaran, "A hybrid feature selection and optimized classification approach for heart disease prediction," *Informatics in Medicine Unlocked*, vol. 49, p. 101535, 2024.

[12] . K.-S. Waris, F. W., "Early heart disease detection using an improved k-means neighbor classifier in python," *Materials Today: Proceedings*, vol. 45, pp. 3702–3707, 2021.

[13] Z. Y. L.-Z. H. S. . Z.-Y. Niu, S., "Application of an enhanced grey wolf optimization algorithm for predicting heart disease," in *Improved Grey Wolf Optimization algorithm applied to heart disease prediction*, ser. Smart Innovation, Systems and Technologies. Springer, 2024, pp. 527–542.

[14] R. M. B. B. R. A. H. G.-A. Alzaqeeb, "Hybrid transformer-based approach for predicting heart disease using structured clinical datasets," *Biomedical Signal Processing and Control*, vol. 99, p. 106094, 2024.

[15] R. A. S. P. K.-. P. S. K. Kumar, M., "Medical record-based hybrid ensemble deep learning model for heart disease prediction," *Informatics in Medicine Unlocked*, vol. 52, p. 101780, 2024.

[16] A. Seva, S. N. T. Rao, and M. Sireesha, "Prediction of liver disease with random forest classifier through smote-enn balancing," in *2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT)*. Jabalpur, India: IEEE, 2024, pp. 928–933.

[17] D. Venkatareddy, K. V. N. Reddy, Y. Sowmya, Y. Madhavi, S. C. Asmi, and S. Moturi, "Explainable fetal ultrasound classification with cnn and mlp models," in *2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC)*. Davangere, India: IEEE, 2024, pp. 1–7.

[18] K. V. N. Reddy, Y. Narendra, M. A. N. Reddy, A. Ramu, D. V. Reddy, and S. Moturi, "Automated traffic sign recognition via cnn deep learning," in *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*. Gwalior, India: IEEE, 2025, pp. 1–6.

[19] S. N. T. Rao *et al.*, "Fake profile detection using machine learning," in *Proceedings of the 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*. Gwalior, India: IEEE, 2025, pp. 1–6.

[20] S. Moturi, S. N. T. Rao, and S. Vemuru, "Optimized feature extraction and hybrid classification model for heart disease and breast cancer prediction," *International Journal of Recent Technology and Engineering*, 2019.