High Performance Hybrid CNN CBAM Framework for High Sensitivity Heart Disease Classification

Giant Prakoso Amukti Wibowo[1], Slamet Riyadi[2]✉, Arlina Dewi[3], Mahendro Prasetyo Kusumo[4], Muhammad Abdul Haq[5], Imam Riadi[6]
[1-5]Universitas Muhammadiyah Yogyakarta, Indonesia
[6]Universitas Ahmad Dahlan, Indonesia

✉Corresponding Author: riyadi@mail.umy.ac.id

ABSTRACT

Cardiovascular diseases remain a paramount global health crisis, necessitating early and precise diagnostic interventions. While medical imaging is the clinical standard, manual interpretation is highly susceptible to visual fatigue and inter-observer variability. This study proposes a novel, highly robust Computer-Aided Diagnosis (CAD) framework that overcomes the spatial and textural limitations of standalone Convolutional Neural Networks (CNN) in heart disease image classification. A Feature- Level Ensemble (Hybrid) architecture was created by putting together the deep semantic features of ResNet50V2, the spatial boundaries of VGG16, and the parameter efficiency of EfficientNetV2B3. To directly deal with the loss of features caused by anatomical background noise, a Convolutional Block Attention Module (CBAM) was added to the EfficientNet pathway. This gave the network two-dimensional (channel and spatial) visual attention. To guarantee a thorough and impartial assessment, a complete restructuring of a dataset comprising 5,977 images was undertaken using an 80:10:10 stratified split, thereby eliminating the accuracy paradox resulting from class imbalance. The proposed Hybrid CBAM model significantly outperforms standalone baselines, with a peak accuracy of 94.00%. For clinical use, it was very important that the attention-guided ensemble had a Recall (sensitivity) of 0.94 for finding pathological cases and a Negative Precision of 0.96. This study definitively demonstrates that the integration of multi-model feature extraction with focused visual attention mechanisms yields a highly sensitive, reliable, and non-invasive automated screening instrument for the early detection of cardiovascular disease.

Keywords : CBAM, CNN, Feature-Level Ensemble, Computer Aided Diagnosis, Heart Disease, Hybrid Model.

A. Introduction

Cardiovascular diseases (CVDs) continue to be the primary cause of death globally, constituting a significant global health challenge that necessitates swift and precise diagnostic methods. Medical imaging techniques, including echocardiography, computed tomography (CT), and cardiac Magnetic Resonance Imaging (MRI), are essential non- invasive tools for evaluating cardiac function and detecting structural anomalies [1], [2], [3]. Nevertheless, the manual interpretation of these intricate visual data heavily depends on the subjective judgment of cardiologists. The conventional diagnostic process necessitates that clinicians carefully examine high-dimensional images, a procedure that is both time-intensive and prone to inter-observer variability and visual fatigue [4]. As a result, subtle pathological signs may be missed, especially in the early stages of cardiovascular diseases. This operational constraint underscores the need for the creation of robust Computer-Aided Diagnosis (CAD) systems to standardize assessments and substantially improve diagnostic efficiency[2], [5], [6].

In recent years, Convolutional Neural Networks (CNNs) have significantly changed medical image analysis, showing remarkable abilities in automatically extracting features

without needing manual design. Deep learning models have been successfully used in various diagnostic tasks, fundamentally changing how clinical data is processed and understood [7], [8], [9], [10]. However, despite these advancements, standalone CNN architectures often have important limitations when dealing with complex cardiovascular images. Standard models usually treat all pixels in a spatial grid equally, which makes single- architecture networks vulnerable to anatomical noise and unrelated background artifacts. Moreover, research suggests that aggressive down-sampling operations (pooling) in traditional networks often lead to the loss of crucial micro- structural details needed to distinguish subtle cardiac abnormalities from healthy tissues.

To overcome these intrinsic constraints, this investigation introduces a thorough and resilient diagnostic framework employing a Feature-Level Ensemble methodology. Rather than depending on a singular network, this research concurrently utilizes three separate, cutting-edge architectures: EfficientNetV2B3, VGG16, and ResNet50V2 [1], [5], [11]. Through the concatenation of these networks' outputs, the proposed model generates a multi-dimensional feature representation, thereby encapsulating both profound morphological patterns and nuanced spatial boundaries, and thus effectively mitigating the individual shortcomings of each foundational architecture.

The primary contribution of this research lies in the creation of a resilient Feature Level Hybrid Ensemble architecture, engineered to mitigate feature dilution stemming from anatomical noise. This hybrid structure intelligently integrates the multi-scale feature extraction proficiencies of three separate networks: VGG16, ResNet50V2, and an EfficientNetV2B3 pathway, which is specifically enhanced with a Convolutional Block Attention Module (CBAM). By incorporating dual-dimensional (channel and spatial) visual attention within the EfficientNetV2B3 branch, the hybrid model emphasizes crucial pathological coordinates while diminishing the influence of extraneous regions before feature concatenation. Supported by a rigorous 80:10:10 stratified data partitioning strategy to mitigate the accuracy paradox induced by extreme class imbalance, this multi-network synthesis delivers a highly reliable, precise, and automated screening tool for cardiovascular diseases.

B. Related Works

Cardiovascular diseases consistently represent a paramount challenge in global healthcare, driving the urgent need for rapid, non-invasive, and highly accurate diagnostic solutions. While medical imaging modalities such as echocardiography and Computed Tomography (CT) have become the clinical gold standard for screening structural and ischemic anomalies, the manual interpretation of these complex visual datasets remains deeply flawed. Traditional diagnostic workflows require clinicians to meticulously analyze high-dimensional images, a process that is not only extensively time-consuming but also highly susceptible to human error, visual fatigue, and significant inter-observer variability [10] [12]. Subtle pathological indicators, particularly in the early stages of cardiovascular deterioration, can easily be overlooked amidst the dense anatomical structures. Consequently, there is a critical necessity to transition from subjective human evaluation toward robust Computer-Aided Diagnosis (CAD) systems capable of standardizing evaluations and significantly enhancing diagnostic efficiency.

To address the limitations of manual clinical interpretation, Convolutional Neural Networks (CNNs) have become a significant technological advancement in medical image analysis. CNNs, by learning hierarchical spatial patterns directly from pixel data, remove the need for manual feature engineering, allowing for the quick and objective classification of medical images. Foundational architectures have historically been the standard in this area due to their specific structural advantages. For example, VGG16 has shown significant

effectiveness in capturing basic morphological boundaries and structural patterns through its deep, sequential arrangement of convolutional filters. At the same time, the ResNet architecture changed deep learning by introducing residual learning through skip connections, a method that allows the network to keep deep, abstract semantic features without suffering from the negative effects of gradient degradation [13]. Building on these established frameworks, modern compound-scaled models like EfficientNetV2 have improved computational efficiency by optimizing network depth, width, and resolution at the same time, thus achieving state-of-the-art accuracy on complex visual tasks while minimizing parameter overhead.

Despite the widespread adoption and documented successes of these standalone CNN architectures, recent literature consistently highlights their inherent limitations when processing highly nuanced medical images. Individual models frequently struggle to capture the full, complex spectrum of pathological indicators because their fixed architectural biases may cause them to miss either fine- grained textures or broader global anatomical contexts. Recognizing this vulnerability, researchers have increasingly transitioned toward ensemble learning paradigms to bolster diagnostic robustness. Recent research indicates that combining several CNN architectures, a method called Feature-Level Ensemble, significantly improves how well models generalize. By merging the output feature vectors from different networks – for example, combining ResNet's strong structural focus with EfficientNet's excellent texture processing – the resulting hybrid models show better diagnostic performance. This multi-dimensional feature synthesis has proven to be far more capable of navigating the severe intra-class variability and inter-class visual similarity commonly found in cardiovascular datasets [14], [15].

However, while ensemble models successfully provide a rich and diverse feature space, they introduce a new computational challenge: they often treat all extracted spatial information with equal weight, rendering the network vulnerable to anatomical background noise and irrelevant tissue structures. To resolve this feature dilution, the integration of visual attention mechanisms has become a critical strategy in contemporary deep learning research. The Convolutional Block Attention Module (CBAM) represents a significant methodological breakthrough in this specific area. Unlike standard pooling techniques that aggressively compress and discard spatial information, CBAM sequentially infers precise attention maps along two distinct dimensions [16]. The channel attention sub-module recalibrates the feature maps to prioritize "what" is clinically meaningful, such as specific pathological textures or lesions. Subsequently, the spatial attention sub-module directs the network to "where" these features are geographically located within the image matrix, effectively suppressing dark, irrelevant anatomical structures and forcing the model to scrutinize only the most critical diagnostic regions.

C. Method

This study employs a systematic computational approach to develop an automated system for classifying cardiovascular disease images. The methodology includes several sequential steps: data collection and organization, image pre-processing, the creation of a basic model, the design of a Feature-Level Ensemble (Hybrid) architecture, the integration of an attention mechanism, and a thorough evaluation.

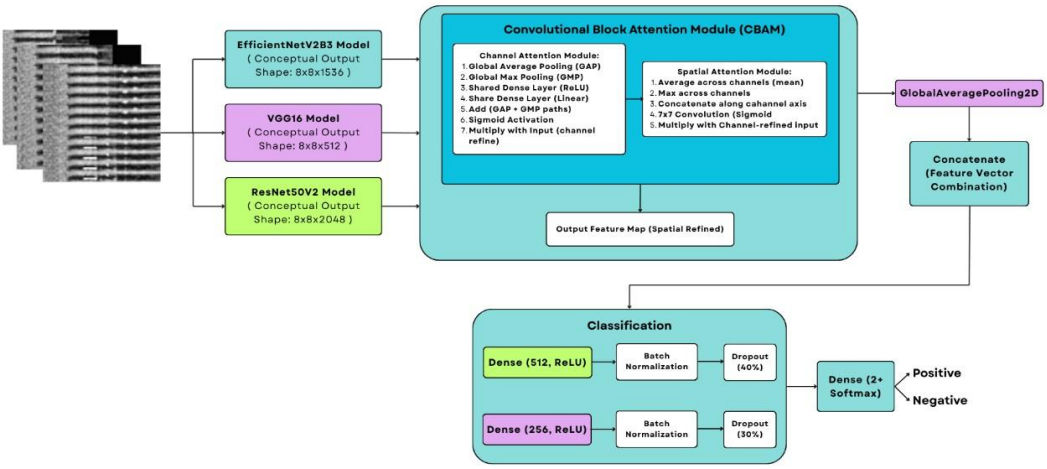

Figure 1. Proposed Model Architecture

1. Dataset Aggregation and Restructuring

The integrity of a deep learning model relies heavily on the quality and distribution of its training data. The primary dataset utilized in this study consists of medical images categorized into two diagnostic classes: Positive (indicating the presence of heart disease) and Negative (indicating healthy cardiac conditions). Initial exploratory data analysis revealed a severe class imbalance within the original testing subset (125 Positive vs. 1,066 Negative cases), a condition that frequently leads to the accuracy paradox in machine learning evaluations [1]. To rectify this and ensure a fair evaluation environment, a comprehensive data restructuring strategy was implemented. All images from the original predefined splits (train, validation, test) were aggregated into a single, unified repository comprising a total of 5,977 images (2,557 Positive and 3,420 Negative).

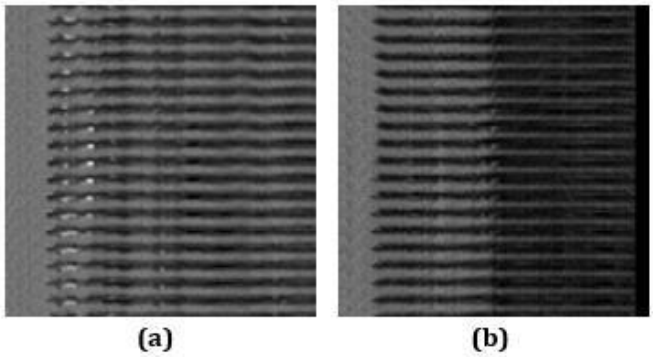

Figure 2. Sample CCTA scans illustrating the diagnostic classes: (a) positive for Coronary Artery Disease, (b) negative.

Following aggregation, the dataset was redistributed using a stratified splitting technique via the split-folders algorithm. The data was partitioned into three new subsets with an 80:10:10 ratio: 80% for training, 10% for validation, and 10% for testing. This stratification guarantees that the proportion of Positive and Negative cases remains consistent across all phases of model development, thereby providing a clinically realistic evaluation metric. The final distribution of the restructured dataset is detailed in Table 1.

Table 1. Dataset Splitting Summary

Class	Training (80%)	Validation (10%)	Testing (10%)	Total
Positive	2.054	256	256	2.577
Negative	2.735	342	343	3.420
Total	4.780	598	599	5.977

2. Image Pre-processing

Before being fed into the network, the raw images underwent standardized pre-processing to ensure they had the same dimensions and to speed up the gradient convergence during optimization. The images were dynamically loaded using Keras ImageDataGenerator. Each image was resized to a consistent size of 256 x 256 pixels to meet the input tensor requirements of the chosen Convolutional Neural Network (CNN) architectures. After that, pixel intensity normalization was applied using a rescaling factor of 1/255. The conversion process transforms the initial 8-bit pixel values, which span from 0 to 255, into a normalized continuous float range of [0,1]; this is essential for ensuring the stability of activation functions in deep neural networks. Moreover, to mitigate the potential for spatial biases to influence the evaluation phase, data shuffling was deliberately disabled (shuffle=False) for both the validation and testing generators.

3. Convolutional Neural Network Architectures

To establish a robust comparative baseline, this study implemented three state-of-the-art CNN architectures using a transfer learning paradigm. Specifically, EfficientNetV2B3, VGG16, and ResNet50V2 were instantiated using pre-trained weights from the ImageNet database. To function strictly as feature extractors, the foundational convolutional blocks of these models were frozen. EfficientNetV2B3, in particular, leverages Fused-MBConv layers alongside a progressive scaling approach to optimize parameter utilization. In contrast, VGG16 adopts a deep, sequential configuration of 3x3 convolutional filters, a design celebrated for its ability to discern consistent spatial hierarchies. Moreover, ResNet50V2 incorporates skip connections, or residual blocks, to successfully address the vanishing gradient issue, thus facilitating the extraction of highly abstract semantic features.

4. Proposed Model: Feature-Level Ensemble (Hybrid)

Relying on a single architecture often limits the diversity of extracted features. To address this, a novel Feature-Level Ensemble (Hybrid) model was engineered. In this architecture, a 256 x 256 x 3 input image is passed simultaneously through the frozen layers of EfficientNetV2B3, VGG16, and ResNet50V2. The resulting three-dimensional feature maps from each base model are flattened into one-dimensional vectors using GlobalAveragePooling2D. These three distinct vectors are then merged using a Concatenate layer, synthesizing a singular, highly dense feature representation.This fused vector is subsequently fed into a newly initialized classification head consisting of a 512-neuron Dense layer and a 256-neuron Dense layer. To prevent overfitting and covariate shift, BatchNormalization and 30-40% Dropout layers were strategically interleaved. The final output is generated via a 2- neuron Dense layer with a Softmax activation function, producing categorical probabilities for the Positive and Negative classes.

5. Integration of Attention Mechanism (CBAM)

CBAM serves as a lightweight, dual-attention mechanism that sequentially applies Channel and Spatial Attention to pinpoint critical pathological features ("what") and their exact locations ("where"), effectively suppressing background noise. Subsequently, the models were

compiled using the Adam optimizer (learning rate = 0.001) and categorical crossentropy, regulated by EarlyStopping and ReduceLROnPlateau callbacks. Final performance on the

599-image testing subset was rigorously quantified using standard confusion matrix metrics: Accuracy for overall correctness, Precision for diagnostic confidence, Recall (Sensitivity) to critically minimize False Negatives, and F1- Score to ensure robust, balanced performance across classes.

D. Result and Discussion

This chapter details the empirical results derived from the experimental scenarios, offering a comprehensive comparative analysis between thevaseline architectures and the proposed Feature-Level Ensemble model. Furthermore, it provides an in-depth clinical evaluation of the diagnostinc outputs generated by the heart disesase classification system.

1. Experimental Setup and Trainif Dynamics

To ensure rigorous and reproducible results, the training phase was meticulously standardized across all models. The networks were optimized using the Adam optimizer with an initial learning rate of 0.001, minimizing a categorical crossentropy loss function.To achieve optimal convergence and mitigate the risk of overfitting, a dynamic training strategy utilizing two primary callbacks was implemented:

a) Learning Rate Modulation (ReduceLROnPlateau): To ensure fine-grained convergence and prevent oscillation around the global minimum, the learning rate was dynamically decayed by a factor of 0.2 whenever validation loss stagnated for three consecutive epochs.
b) Early Stopping: To mitigate overfitting, training was halted and optimal weights were restored if validation accuracy stagnated for five consecutive epochs. This strategy ensured stable convergence across all architectures, efficiently concluding within 20–30 epochs.

2. Comparative Performance Analysis of Baseline Architectures

The initial phase of evaluation assessed three standalone Convolutional Neural Network (CNN) architectures using the rigorously structured testing dataset comprising 599 images. The results revealed a significant variance in feature extraction capabilities when pre-trained ImageNet weights were applied to complex cardiovascular imagery. As a standalone baseline, EfficientNetV2B3 exhibited the lowest diagnostic viability, recording an overall accuracy of 68.78%, alongside a Positive Precision of 0.63 and a Recall of 0.67. This phenomenon highlights a critical limitation of zero-shot transfer: without extensive fine-tuning by unfreezing the upper layers, pre-trained weights optimized for natural RGB images fundamentally struggle to map onto the highly specific, narrow-variance textural domains of grayscale medical imaging. The specific misclassification distribution of this model, which heavily leans towards False Negatives, is visually detailed in Figure 3.

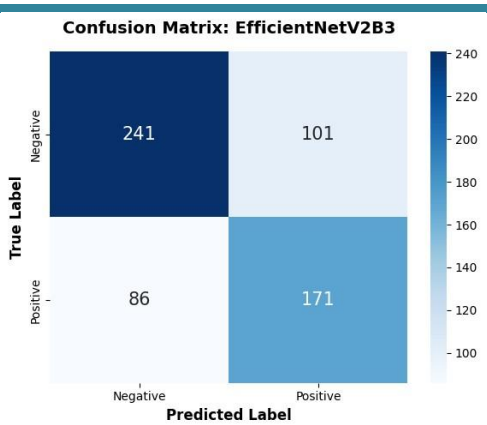

Figure 3. Confusion Matrix of EfficientNetV2B3

Table 2. Comparison Results

Model	Class	Accuracy	Precision	Recall	F1-Score
EfficientNetV2 B3 (Baseline)	Positive	68.78%	0.63	0.67	0.65
	Negative		0.74	0.70	0.72
VGG16 (Baseline)	Positive	86.48%	0.84	0.85	0.84
	Negative		0.89	0.87	0.88
ResNet50V2 (Baseline)	Positive	88.98%	0.86	0.89	0.87
	Negative		0.92	0.89	0.90
Hybrid Model (Standard)	Positive	92.82%	0.89	0.95	0.92
	Negative		0.96	0.91	0.94
Proposed Model (Hybrid CBAM)	Positive	94.00%	0.93	0.94	0.94
	Negative		0.96	0.95	0.95

In stark contrast to the limitations observed in EfficientNetV2B3, VGG16 demonstrated remarkable robustness in extracting spatial features, achieving a significantly higher accuracy of 86.48%. Despite being the oldest architecture among the evaluated baselines, its deep, sequential arrangement of 3x3 convolutional blocks consistently captured the morphological boundaries and structural patterns of cardiac tissues far more effectively than the frozen EfficientNet model. Figure 4 illustrates the confusion matrix for VGG16, showcasing its notably improved true positive detection rate.

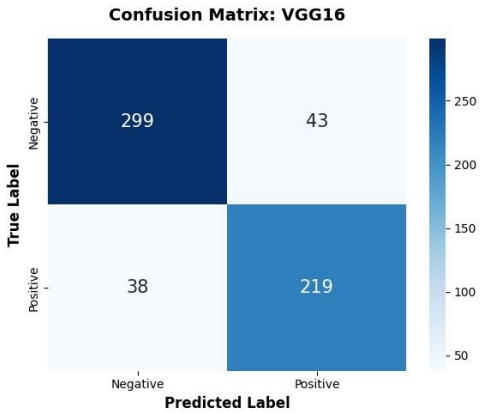

Figure 4. Confusion Matrix of VGG16

Building upon this capacity for spatial extraction, ResNet50V2 emerged as the superior standalone architecture in this study, peaking at an accuracy of 88.98%. The integration of residual blocks, or skip connections, facilitated unimpeded gradient flow throughout the exceptionally deep network. This distinct structural advantage enabled the model to preserve and recognize complex, high-level anatomical features without suffering from degradation, ultimately yielding a robust Positive Recall of 0.89 and a Negative Precision of 0.92. The comprehensive diagnostic behavior and class-wise predictions of this deep architecture are presented in Figure 5.

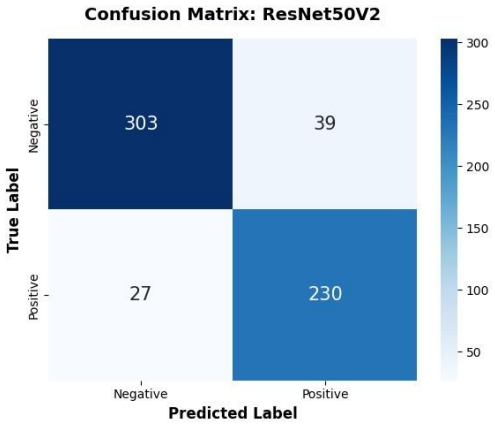

Figure 5. Confusion Matrix of ResNet50V2

3. The Efficacy of Feature-Level Ensemble and Attention Mechanisms

Operating on the premise that different CNN architectures capture distinct hierarchical features, the proposed Hybrid model concatenated the flattened feature vectors from all three baseline networks. This fusion strategy resulted in a significant accuracy surge to 92.82%. The performance leap proves that feature concatenation successfully creates a more comprehensive data representation. The textural mapping weaknesses of EfficientNet were effectively compensated by the spatial robustness of VGG16 and the deep feature extraction of ResNet50V2. By feeding this enriched, multi- dimensional feature vector into a newly initialized, fully connected Dense classification head, the network was able to construct a highly optimal decision hyperplane separating the Positive (Sick) and Negative (Healthy) classes. The resulting reduction in both False Positive and False Negative errors achieved by this ensemble strategy is explicitly detailed in Figure 6.

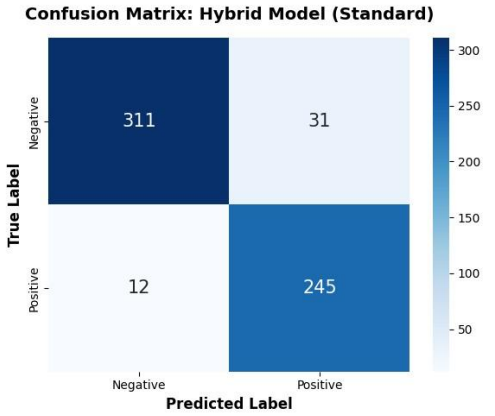

Figure 6. Confusion Matrix of Hybrid Model (Standard)

The primary methodological novelty of this research was evaluated by injecting a Convolutional Block Attention Module (CBAM) directly into the EfficientNetV2B3 pathway of the Hybrid model. This particular configuration yielded the highest performance observed in the study, achieving an accuracy of 94.00%. This enhancement, from the standard ensemble's 92.82% to 94.00%, is not a trivial increment; it signifies a substantial alteration in the network's approach to clinical data processing. In the absence of attention mechanisms, a conventional CNN uniformly weights all pixels within the 256x256 spatial grid, thereby permitting anatomical noise or extraneous background tissues to compromise the classification. The incorporation of CBAM effectively furnishes the model with "visual attention" to mitigate this constraint. Specifically, the Channel Attention sub-module selectively amplifies feature maps that are most indicative of cardiac pathology, whereas the Spatial Attention sub-module directs the network's focus toward crucial pixel coordinates, such as cardiac muscle anomalies or valve irregularities, while concurrently diminishing the influence of dark, irrelevant regions. Consequently, the feature vectors generated by the EfficientNet pathway become exceptionally refined and discriminative, thereby fundamentally enhancing the overall fused vector prior to the final classification stage.

4. Clinical Significance and Error Trade-off Analysis

In the domain of medical image analysis, relying solely on high global accuracy is an insufficient metric for clinical deployment because prediction errors carry asymmetrical risks. Specifically, while a False Positive (diagnosing a healthy patient as sick) causes psychological distress and necessitates further testing, a False Negative (diagnosing a sick patient as healthy) is potentially fatal due to missed early interventions. Therefore, minimizing False Negatives remains the absolute priority for any Computer-Aided Diagnosis (CAD) system. The diagnostic behavior of the proposed Hybrid + CBAM model deeply validated through its Confusion Matrix and Classification Report excels in this critical aspect by achieving an optimal diagnostic sensitivity.

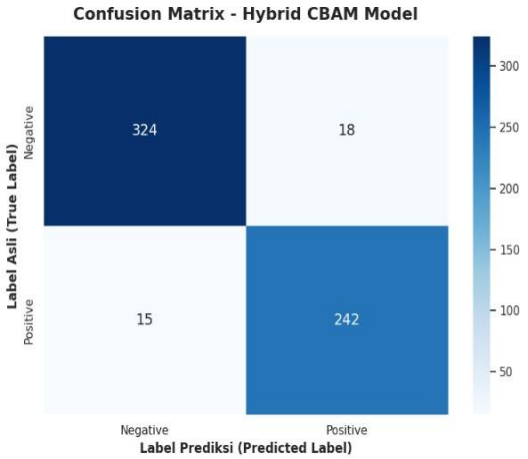

Figure 7. Confusion Matrix of Proposed Model (Hybrid CBAM)

The model demonstrated a positive recall of 0.94, indicating its capacity to accurately identify 94% of the heart disease cases within the testing group; this resulted in a false negative rate of only 6%, thereby satisfying the rigorous sensitivity criteria essential for a reliable early-stage screening instrument. Simultaneously, the model exhibited considerable diagnostic confidence, as evidenced by a negative precision of 0.96, which ensured that the clinical reliability of the "Healthy" designation was exceptionally high. The harmonious interplay

between this precision and recall yielded an impressive F1-score of 0.94, thereby demonstrating the model's lack of bias toward any predominant class. Consequently, this comprehensive metric profile substantiates the methodological soundness of the study, confirming that the initial data restructuring specifically, the pooling and80:10:10 stratified splitting effectively eliminated the extreme class imbalance and completely averted the accuracy paradox that had previously affected the raw dataset.

E. Conclusion

This research provides compelling evidence that a Feature-Level Ensemble architecture, enhanced by a Convolutional Block Attention Module (CBAM), substantially addresses the limitations of individual Convolutional Neural Networks in the context of cardiovascular image classification. Through a methodical restructuring of a 5,977-image medical dataset, employing a rigorous 80:10:10 stratified split, the accuracy paradox was completely eliminated, thereby establishing a strong and impartial evaluation framework. Although individual baseline models, including EfficientNetV2B3 and ResNet50V2, achieved peak accuracies of 68.78% and 88.98% respectively, the concatenated hybrid feature space improved overall diagnostic accuracy to 92.82%. Furthermore, the strategic integration of CBAM within the EfficientNet pathway enhanced the model's diagnostic precision by actively diminishing irrelevant anatomical noise, resulting in a peak accuracy of 94.00%. Most significantly, the proposed CAD system reached a crucial clinical benchmark, recording a 0.94 Recall for positive pathological cases and a 0.96 Precision for negative cases.

References

[1] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, “Convolutional neural networks for medical image analysis: State- of-the-art, comparisons, improvement and perspectives,” *Neurocomputing*, vol. 444, pp. 92–110, Jul. 2021, doi: 10.1016/j.neucom.2020.04.157.

[2] Y.-R. Wang *et al.*, “Screening and diagnosis of cardiovascular disease using artificial intelligence-enabled cardiac magnetic resonance imaging,” *Nat Med*, vol. 30, no. 5, pp. 1471–1480, May 2024, doi: 10.1038/s41591-024-02971-2.

[3] L. V. P. G, M. A, B. R. Y V, H. Marturi, and V. V. Paul, “Deep Learning in Medical Imaging: Image Processing - From Augmenting Accuracy to Enhancing Efficiency,” in *Proceedings of the 2024 7th International Conference on Digital Medicine and Image Processing*, in DMIP ’24. New York, NY, USA: Association for Computing Machinery, 2025, pp. 101–106. doi: 10.1145/3705927.3705945.

[4] D. Ouyang *et al.*, “Efficient Multi-Scale Attention Module with Cross-Spatial Learning,” in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece: IEEE, Jun. 2023, pp. 1–5. doi: 10.1109/ICASSP49357.2023.10096516.

[5] I. Pacal, O. Celik, B. Bayram, and A. Cunha, “Enhancing EfficientNetv2 with global and efficient channel attention mechanisms for accurate MRI-Based brain tumor classification,” *Cluster Comput*, vol. 27, no. 8, pp. 11187–11212, Nov. 2024, doi: 10.1007/s10586-024-04532-1.

[6] M. Jafari *et al.*, “Automated diagnosis of cardiovascular diseases from cardiac magnetic resonance imaging using deep learning models: A review,” *Computers in Biology and Medicine*, vol. 160, p.106998, Jun. 2023, doi: 10.1016/j.compbiomed.2023.106998.

[7] R. Gu *et al.*, “CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation,” *IEEE Trans. Med. Imaging*, vol. 40, no. 2, pp. 699–711, Feb. 2021, doi: 10.1109/TMI.2020.3035253.

[8] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical Image Analysis using Convolutional Neural Networks: A Review," *J Med Syst*, vol. 42, no. 11, p. 226, Nov. 2018, doi: 10.1007/s10916-018-1088-1.

[9] G. Trimarchi *et al.*, "Charting the Unseen: How Non-Invasive Imaging Could Redefine Cardiovascular Prevention," *JCDD*, vol. 11, no. 8, p. 245, Aug. 2024, doi: 10.3390/jcdd11080245.

[10] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[11] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," 2021, doi: 10.48550/ARXIV.2104.00298.

[12] A. P. Setianto, C. Damarjati, and A. Asroni, "Automatic Measurement Application of Heart Area from Chest X-Ray Images Using the U-Net Deep Learning Method," *EIST*, vol. 2, no. 1, pp. 16–23, Jan. 2023, doi: 10.18196/eist.v2i1.16864.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015, *arXiv*. doi: 10.48550/ARXIV.1512.03385.

[14] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep Convolutional Neural Networks for breast cancer screening," *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 19–30, Apr. 2018, doi: 10.1016/j.cmpb.2018.01.011.

[15] F. Fitrianto, E. Rouza, and Basorudin, "Implementation Of Transfer Learning In Cat Breed Detection Using Web-Based Convolutional Neural Network (CNN)," *EIST*, vol. 6, no. 1, pp. 43–53, May 2025, doi: 10.18196/eist.v6i1.27104.

[16] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," 2018, *arXiv*. doi: 10.48550/ARXIV.1807.06521.