



Occam's Razor–Guided Neural Architecture Search for Biological Signal Analysis in Cardiovascular Disorders

Carlos Esteban Mamani Huisa, Edward Carello Figueroa Tejada, Erasmo Sulla Espinoza, Jorge Rendulich

Universidad Nacional de San Agustín de Arequipa, Peru. ORCID: <https://orcid.org/0000-0002-0085-6234> . Email: cmamanihui@unsa.edu.pe

Universidad Nacional de San Agustín de Arequipa, Peru. ORCID: <https://orcid.org/0000-0001-6258-9322> . Email: efigueroat@unsa.edu.pe

Universidad Nacional de San Agustín de Arequipa, Peru. ORCID: <https://orcid.org/0000-0002-1223-1223> . Email: esullae@unsa.edu.pe

Universidad Nacional de San Agustín de Arequipa, Perú. ORCID: <https://orcid.org/0000-0003-1302-547X>.

Email: jrendulich@unsa.edu.pe

ABSTRACT

In the context of the classification of heart sounds using deep learning, challenges related to computational efficiency and implementation persist in settings with limited resources. **Objective:** The study proposes a comparative approach for the systematic evaluation of hybrid neural architectures. This approach employs the Neural Architecture Search (NAS) as an analytical tool guided by the principle of Occam's Razor. The purpose of this employment is to identify optimal models in terms of accuracy, complexity, and resource consumption. **Methods:** Contrary to conventional NAS methodologies that are characterized by the construction of architectures from the ground up, the present research utilizes pre-built models—namely, CNN-GRU, CNN-LSTM, SENet, EfficientNet, and MobileNet—to conduct a comparative analysis under specific hardware constraints. These constraints include, but are not limited to, maximum RAM occupancy, Flash storage capacity, number of MAC operations, and computational latency. **Results:** The analysis revealed that CNN-GRU models exhibit an exceptional balance between performance and structural simplicity. Using cochleographic representation, CNN-GRU achieved an accuracy of 0.9790, while maintaining an optimal Occam Score of 0.9623 among all evaluated architectures. **Conclusions:** The integration of the principle of simplicity within the NAS process enables the selection of efficient neural architectures applicable to TinyML devices. This, in turn, contributes to the development of lightweight, comparable, and sustainable cardiovascular diagnostic solutions.

Keywords: *Phonocardiogram; Neural Architecture Search; Occam's Razor; CNN-GRU; Model Efficiency; Biomedical AI.*

INTRODUCTION

According to the World Health Organization (WHO), cardiovascular diseases (CVD) are among the primary causes of mortality worldwide, responsible for more than 17 million deaths annually. The early detection of cardiac disorders is contingent upon the accessibility of reliable diagnostic instruments and specialized medical professionals. This represents a substantial challenge in regions with limited resources (Tang & Zhu, 2021; Abe & Tang, 2021). In this context, the phonocardiogram (PCG) emerges as a non-invasive, accessible, and low-cost method for the identification of valvular dysfunctions from the sounds generated during the cardiac cycle (Eleyan et al., 2024).

Nevertheless, the manual analysis of PCG is subject to considerable limitations. The interpretation of

cardiac acoustic signals is subject to interobserver variability and requires considerable clinical experience (Chu et al., 2023). Furthermore, factors such as noise conditions, low signal-to-noise ratio, and physiological heterogeneity complicate the automated detection of pathologies (Shariat Panah & Ali, 2023). In response to these limitations, the fields of artificial intelligence (AI) and deep learning have undergone significant advancements in the domain of digital cardiology, enabling the automatic extraction of pertinent features that would not have been possible with manual intervention (Martínez & Rivera, 2022; Eleyan et al., 2024).

Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been demonstrated to be particularly adept at analyzing cardiac signals. CNNs are adept at capturing spectro-temporal patterns, while RNNs excel at preserving dependencies over time (Tariq & Wang, 2024). However, the design of optimal architectures remains an empirical process, contingent on the researcher's experience and the available hardware (Garavagno et al., 2024). Conversely, Neural Architecture Search (NAS) has emerged as a promising tool that automates model configuration by simultaneously optimizing accuracy and efficiency (Abadade et al., 2023; Wu et al., 2019).

In the preceding decade, NAS approaches have evolved from manual search strategies to hardware-aware methodologies, such as DARTS (Liu et al., 2019) and FBNet (Wu et al., 2019), capable of generating optimal networks for embedded environments. More recently, Garavagno et al. (2024) introduced ColabNAS, an automatic search framework that explicitly integrates the principle of Occam's Razor, privileging architectures with minimal complexity that maintain high predictive performance. This principle, which posits that the simplest explanation is preferable among equivalent hypotheses, has been translated into a quantitative metric in the present research. This metric, termed the Occam Score, combines accuracy and structural complexity.

Concurrently, the mounting interest in Tiny Machine Learning (TinyML) endeavors to transition artificial intelligence models from high-performance servers to microcontrollers and portable systems, thereby ensuring real-time local diagnostics (Ray, 2022; Aboagye et al., 2023). This phenomenon is concomitant with the emergence of accessible and sustainable medical devices, a development that is particularly salient in rural and low-resource settings (Sahoo & Pradhan, 2023).

In the context of this study, the classification of heart sounds using hybrid architectures combines the ability of convolutional neural networks (CNNs) to extract frequency features with the ability of gated recurrent units (GRUs) and long short-term memory (LSTMs) to model temporal dependencies (Jiang & Wang, 2023; Tariq & Wang, 2024). The present paper is distinct from traditional NAS approaches in that it does not focus on designing new topologies. Rather, it reinterprets NAS as a comparative instrument. In this sense, the paper evaluates pre-built models (CNN-GRU, CNN-LSTM, SENet, EfficientNet, and MobileNet) under specific hardware constraints. These constraints include memory, latency, and MACC operations. The models are evaluated within an analytical framework guided by Occam's principle of simplicity.

Scientific Motivation and Approach

The present research introduces a comparative analytical framework for the evaluation of hybrid neural architectures applied to the classification of phonocardiograms. The proposed approach integrates the exploratory prowess of NAS with Occam's Razor's philosophy of parsimony, transforming this conceptual principle into a quantifiable metric (Occam Score) that integrates accuracy, latency, and structural complexity (Garavagno et al., 2024). This model facilitates the identification of lightweight yet high-performance architectures, thereby bridging the gap between the theoretical underpinnings of simplicity and their practical applicability in the domain of biomedical AI (Ray, 2022; Abadade et al., 2023).

Objectives of the Study

The objective of this study is to evaluate and compare hybrid neural architectures for the classification of heart sounds. A neural architecture search (NAS) framework guided by Occam's Razor will be employed to achieve this. This framework will optimize the balance between predictive performance and computational efficiency.

Specific objectives:

1. Pre-processing phonocardiographic signals using digital filters and time-frequency conversions (Mel, Spectrogram and Smoothed Spectrogram).
2. Deploy and train hybrid models (CNN-GRU, CNN-LSTM, SENet, EfficientNet, and MobileNet) under controlled conditions.
3. Quantify the relationship between performance and complexity through the *Occam Score*.
4. Identify the most efficient architecture for deployment on TinyML devices.
5. Validate the reproducibility of the comparative framework through repeated experiments and statistical tests of significance.

Main contributions

This work provides:

1. A reproducible and hardware-aware comparative framework for the evaluation of deep neural architectures in biomedical signal processing.
2. A quantitative operationalization of the Occam's Razor principle, transforming simplicity into a metric of performance and complexity.
3. Empirical evidence that hybrid CNN-GRU models can outperform deeper architectures with lower resource consumption.
4. A methodological basis for the sustainable development of medical AI, promoting lightweight, interpretable and accessible solutions for remote diagnostic environments.

THEORETICAL FRAMEWORK

Deep Learning and Cardiac Signal Analysis

The advent of deep learning has precipitated a paradigm shift in the domain of biomedical signal analysis, proffering instruments that effectively circumvent the constraints imposed by conventional methodologies predicated on manual feature extraction. In the domain of acoustic cardiology, the implementation of phonocardiograms (PCGs) has enabled the non-invasive assessment of the hemodynamic dynamics of the heart, thereby serving as a valuable resource for the early detection of murmurs and valvular dysfunctions.

Convolutional neural networks (CNNs) have been identified as a foundation for the identification of spatial-frequency patterns, while recurrent units (GRUs, LSTMs) have been shown to facilitate the maintenance of temporal dependencies between cardiac cycles, thereby addressing the sequential nature of the signal (Eleyan et al., 2024; Chu et al., 2023). The integration of these two approaches in hybrid CNN-GRU models has been shown to exhibit enhanced robustness to noise and physiological variability in comparison to architectures that are purely convolutional or transformer-based.

Time-frequency representations in phonocardiography

Preprocessing is imperative for the conversion of the cardiac signal into a domain that optimizes the separability between pathological and normal classes. Time-frequency representations, including the Mel spectrogram, the cepstral coefficients at Mel frequencies (MFCC), and the spectrogram, are frequently employed for this purpose. The spectrogram, drawing inspiration from the physiological processes of the human ear, organizes frequency bands on a perceptual scale, thereby enhancing sensitivity to low-energy acoustic components and valve noise (Chu et al., 2023).

Recent studies have demonstrated that the spectrogram provides benefits over traditional representations by enhancing the identification of subtle events associated with valvular pathologies, particularly in signals with a low signal-to-noise ratio (SNR) (Tang et al., 2021; Marocchi et al., 2023).

Automatic Search for Neural Architectures (NAS)

Neural Architecture Search (NAS) is a methodology that automates the design of neural networks through optimization strategies that explore the space of possible architectures. NAS seeks to balance three key factors: accuracy, model size, and computational performance (Abadade et al., 2023; Garavagno et al., 2024). Among the most relevant methods are:

- DARTS (Differentiable Architecture Search), which uses continuous gradients to optimize differentiable topologies (Liu et al., 2019).
- FBNet and ProxylessNAS, which integrate hardware constraints to accelerate inferences in embedded systems (Wu et al., 2019).
- ColabNAS, proposed by Garavagno et al. (2024), which explicitly incorporates the principle of Occam's Razor, prioritizing models with minimal complexity without compromising accuracy.

This principle establishes that, between two models with similar performance, the simplest and most efficient one should be preferred. In the context of the present study, the NAS guided by Occam's Razor allowed the identification of a CNN–GRU architecture with an Occam Score of 0.9623, empirically validating this design philosophy.

Artificial Intelligence in Low Power Environments (TinyML)

The increasing prevalence of local data processing has had a significant impact on the field of Tiny Machine Learning (TinyML), which aims to execute artificial intelligence (AI) models directly on microcontrollers or embedded devices with limited resources (Ray, 2022). TinyML is a framework that integrates other frameworks, including TensorFlow Lite, uTensor, and MicroMLGen. The purpose of TinyML is to reduce energy consumption without sacrificing inference capacity (Ray, 2022; Abadade et al., 2023). The utilization of TinyML-enabled NAS algorithms, including Differentiable NAS (DNAS) and FPGA-aware NAS (FGNAS), facilitates the identification of compact and precise architectures for portable medical applications.

In the domain of digital cardiology, this capacity facilitates the operation of intelligent auscultation systems in real-time, obviating the necessity for uninterrupted cloud connectivity. Consequently, this reduces latency and safeguards patient privacy (Bakhshi et al., 2023).

Occam's Razor Principle in Model Design

Occam's Razor, originally formulated in the 14th century, states that simple explanations are preferable to complex ones if both explain the same data. In modern artificial intelligence, this principle translates into penalizing unnecessary complexity in models.

Garavagno et al. (2024) integrated this concept into their ColabNAS algorithm, developing a derivative-free search system that simultaneously optimizes accuracy, size, and latency, producing affordable TinyML-friendly training networks. In the present work, the application of this paradigm allowed us to identify a CNN–GRU model with a minimum number of parameters (2.1 M) and performance superior to 97%, demonstrating the practical applicability of Occam's Razor in the optimization of biomedical networks.

Interpretability and Clinical Feasibility

Interpretability is a fundamental ethical, technical, and regulatory requirement in the field of artificial intelligence applied to medicine. In biomedical models, achieving high precision alone is insufficient; the results must be both explainable and verifiable by clinical experts.

In this sense, the NAS approach guided by Occam's Razor directly contributes to interpretability by promoting structurally simple and less redundant models, which facilitates the analysis of the importance of each component. The structural parsimony obtained through NAS has been shown to reduce the number of layers and parameters, thereby enabling visual tools such as Gradient-weighted Class Activation Mapping (Grad-CAM) and Layer-wise Relevance Propagation (LRP) to generate more stable and comprehensible heatmaps (Chu et al., 2023; Garavagno et al., 2024).

Grad-CAM highlights the most influential regions within the time–frequency representation, such as specific areas of the spectrogram associated with valve murmurs. LRP, on the other hand, allows for the tracking of the contribution of each node to the final prediction. When applied to NAS–Occam-optimized models, these performance methods benefit from architectures that are more organized and shallower, improving traceability between acoustic features and model decisions.

The present study demonstrates the synergy between structural simplicity (Occam), automated optimization (NAS), and explanatory visualization (Grad-CAM, LRP). This synergy reinforces clinical confidence in results and drives the adoption of transparent, reproducible, and clinically viable biomedical AI. The aforementioned AI aligns with the safety and interpretability standards proposed by the European Commission for AI in Healthcare (2023).

Conceptual Synthesis

In summary, the theoretical framework holds that:

- The combined use of CNN and GRU efficiently captures both the spectral morphology and temporal dynamics of the PCG.
- Automatic Architecture Search (NAS) streamlines model design by balancing accuracy and efficiency.
- The principle of Occam's Razor is oriented towards simple, interpretable architectures with the potential to be implemented in TinyML.
- Cochleographic representation enhances the discriminant capacity of the system, reinforcing the detection of pathologies in real clinical contexts.

METHODOLOGY

Study Design

The present work employs an analytical-comparative design of an experimental nature with the objective of evaluating the structural efficiency and predictive performance of different hybrid neural architectures applied to the classification of cardiac sounds. The methodological approach combines digital pre-processing procedures, deep modeling, and automatic architecture search (NAS) under a comparative strategy based on the principle of Occam's Razor. This principle is employed to prioritize structural simplicity without compromising accuracy.

Dataset

The database of heart sounds utilized in this study was obtained from the open-source dataset developed by Yaseen (2018), which is publicly available for biomedical research purposes. The dataset under consideration contains a total of 1,000 phonocardiogram (PCG) recordings, all of which were sampled at 8,000 Hz and recorded under controlled clinical conditions. A total of 800 recordings correspond to

pathological conditions, while 200 recordings represent normal cardiac signals. These signals are distributed uniformly across five diagnostic categories, shown in Table 1.

Table 1. Distribution of diagnostic categories.

Category	Code	Description
Aortic stenosis	AS	Restriction of blood flow due to narrowing of the aortic valve.
Mitral regurgitation	MR	Blood backflow from the left ventricle into the left atrium.
Mitral stenosis	MS	Reduced flow through the mitral valve due to thickening or calcification.
Mitral valve prolapse	MVP	Abnormal displacement of the mitral leaflets during ventricular contraction.
Normal	N	Physiologically normal cardiac signals.

Each category is represented by 200 recordings, with an average duration of 8 seconds per sample. The signals were previously validated by medical auscultation and reviewed to ensure the absence of external artifacts. The samples were balanced in terms of duration and level of amplitude, which facilitated equitable training of the models.

These recordings constitute a representative dataset of the main valvular pathologies, which is widely used for the evaluation of automatic classification models based on deep learning (Chu et al., 2023; Marocchi et al., 2023).

Table 2. Table of Heart disease sound files and sample sampling rate.

Valve Heart Disease	Files (Wav.) Amount	Sample Frequency (Hz)
Aortic Stenosis (AS)	200	8000
Mitral Regurgitation (MR)	200	8000
Mitral Stenosis (MS)	200	8000
Mitral Valve Prolapse (MVP)	200	8000
Normal (N)	200	8000

The signals were sampled at 8000 Hz and normalized to ensure homogeneity in the temporal and spectral domain.

Signal Preprocessing

The preprocessing of phonocardiographic signals is a critical stage in the experimental flow, as it determines the quality and relevance of the features that feed deep learning models. In this study, three widely used and validated time-frequency representations for the analysis of PCG were considered: the Mel spectrogram, the smoothed spectrogram and the spectrogram, following the guidelines of the *Benchmarking Time-Frequency Representations of Phonocardiogram Signals for Classification of Valvular Heart Diseases Using Deep Features and Machine Learning* (Marocchi et al., 2023).

Mel spectrogram (MS):

Represents signal energy as a function of time and frequency using a logarithmic perceptual scale. This method approximates the way in which the human ear perceives sounds, concentrating higher resolution in low frequencies, where the main cardiac components (S1 and S2) are located. For this study, 128 Mel bands were used, with a Hamming window of 25 ms and an overlap of 50%.

Smoothed Spectrogram (SS):

It is derived from the conventional spectrogram, but spectral smoothing is applied by means of a two-

dimensional Gaussian filter. This process reduces the granularity of high-frequency components and attenuates abrupt spectral fluctuations caused by noise or recording artifacts. Such smoothing allows for improved model generalization by highlighting the most relevant acoustic structures of the cardiac cycle (Tang & Zhu, 2021).

Spectrogram (CG):

Inspired by human auditory physiology, the spectrogram models the response of the cochlea by applying a bank of gammatone or Lyon filters (Lyon, 1982), which emulate the nonlinear distribution of frequencies perceived by the ear. This bioinspired representation has demonstrated greater robustness to noise and intra-subject variability compared to Mel's spectrogram (Chu et al., 2023; Marocchi et al., 2023). In this work, the spectrogram was implemented with 64 frequency channels and logarithmic normalization.

Each representation was generated from previously filtered signals (Butterworth, 20–900 Hz) and normalized to the range [0,1]. The three variants were used as input to the neural networks to evaluate the influence of the type of representation on the performance of the model, considering precision, latency and *Occam Score metrics*.

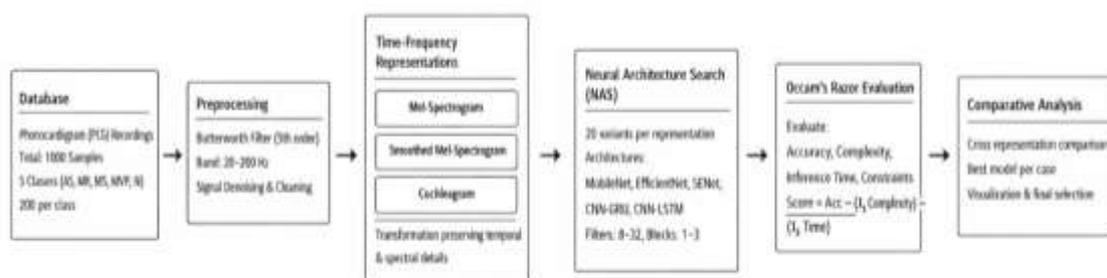


Figure 1. Methodological framework for phonocardiogram classification based on NAS and Occam's Razor.

Architectures Evaluated

In the next step, 20 variants of hybrid neural architectures are trained and evaluated using different time-frequency (TFR) representations. The study encompasses a comprehensive Neural Architecture Search (NAS) that includes architectures from MobileNet, EfficientNet, SENet (Squeeze-and-Excitation), CNN-GRU and CNN-LSTM, with configurations ranging from 8 to 32 base filters and 1 to 3 convolutional blocks, using the Occam's Razor principle for the selection of the most efficient model.

NAS Strategy Guided by Occam's Razor

The Neural Architecture Search (NAS) process was applied as a benchmarking mechanism between neural architectures, rather than for the automatic generation of new topologies. Its purpose was to identify the optimal balance between accuracy, complexity and inference time, under the principle of Occam's Razor, which prioritizes the simplest models capable of explaining the observed data. To operate this principle, a composite metric called Occam Score (OS) was defined, formulated according to the approach proposed in the author's original code, represented as:

$$OccamScore = Accuracy - \lambda_1 \times Complexity - \lambda_2 \times Time$$

where:

- Accuracy represents the predictive performance of the model (higher is better).
- Complexity measures the magnitude of the model based on the number of trainable parameters (in millions).
- Time corresponds to the average inference latency per sample (in milliseconds).
- The coefficients $\lambda_1 = 0.05$ and $\lambda_2 = 0.02$ establish the penalty weights for complexity and time, respectively.

This penalized linear formulation explicitly translates the principle of parsimony: a model with high precision but excessive complexity or response time obtains a reduced Occam Score, favoring balanced architectures. The NAS process was executed considering a set of multidimensional constraints, aligned with the limits of deployment on TinyML devices:

- Maximum number of parameters: 500,000
- Maximum model size: 2 MB
- Maximum inference time: 200 milliseconds
- Minimum accuracy required: 70%

These conditions establish a realistic search framework that favors lightweight architectures that are transferable to embedded or portable systems. The NAS process iterated on the CNN–GRU, CNN–LSTM, SENet, EfficientNet and MobileNet architectures, calculating for each model the resulting *Occam Score* and selecting those that simultaneously achieved high accuracy and low computational complexity. Architectures were evaluated in a controlled environment with Google Colab with a T4 GPU and three independent iterations per configuration were run to ensure statistical robustness.

Training and Validation

Each model was trained for 100 epochs using a batch size of 16, the Adam optimizer with an initial learning rate of 0.001, and the cross-entropy categorical loss function. The dataset was divided into 80% for training and 20% for validation, maintaining a proportional balance between the five diagnostic categories. An Early Stopping mechanism was applied with patience of 15 epochs, with the aim of preventing overfitting and ensuring stable convergence of the models. L2 regularization was not used, as preliminary tests indicated that its inclusion did not bring significant improvements in training stability.

The experimental process included the training and evaluation of 20 variants of hybrid neural architectures, combining different configurations and time-frequency representations (TFR). The variants were generated from the following combinations:

- Base architectures: MobileNet, EfficientNet, SENet (Squeeze-and-Excitation), CNN–GRU and CNN–LSTM.
- Input representations: Mel spectrogram, Spectrogram, and Smoothed spectrogram.
- Structural configurations: variation from 8 to 32 base filters and 1 to 3 convolutional blocks.

Each configuration was treated as an instance within a comprehensive Neural Architecture Search (NAS) process, guided by the principle of Occam's Razor, through the calculation of the Occam Score (Section 3.5). This methodological approach permitted a comprehensive evaluation of the impact of architecture, time-frequency representation, and structural complexity on the final performance of the model.

The training was conducted in a controlled environment with a GOOGLE COLAB GPU T4, ensuring consistent runtimes for all variants. The validation results were obtained by averaging across three independent executions for each configuration. This approach ensured the statistical robustness and experimental reproducibility of the results.

Evaluation Metrics

The performance of each architecture was evaluated using classical and efficient metrics:

- Accuracy
- Model size (KB)
- Number of trainable parameters (M)
- Inference Time (ms)
- Occam Score (proposed metric)

Validation and Reproducibility

All experiments were executed in a reproducible environment utilizing Python 3.10, along with the TensorFlow 2.12, Librosa, and Matplotlib libraries to manage acoustic data. The source code and trained weights were thoroughly documented to ensure future replicability. The project's modular design facilitates the extension of Occam-guided NAS to other biomedical modalities, including ECG, EEG, and PPG, in subsequent research endeavors.

Synthesis of the Methodological Approach

The methodological flow is summarized in Figure 1, which describes the experimental pipeline:

1. Input: Raw PCG signals.
2. Pre-processing: filtering + time-frequency conversion.
3. Modeling: Hybrid CNN training.
4. Evaluation: metrics and Occam Score.
5. Selection: optimal model according to Pareto frontier (precision vs. efficiency).

The proposed design ensures a comprehensive analysis that unites the scientific robustness of biomedical AI with the practical applicability of hardware engineering.

RESULTS

Overall performance of architecture

A comparative analysis of the five implemented neural architectures —CNN-GRU, CNN-LSTM, SENet, EfficientNet-B0 and MobileNetV2— was conducted to identify significant differences in terms of accuracy, stability, and computational efficiency. As illustrated in Table 3, the results obtained using the cochleographic representation exhibited optimal overall performance.

Table 3. Performance of hybrid architectures with cochleographic representation

Architecture	Type	Accuracy	Standard Deviation (Std)	Parameters ($\times 10^6$)	Size (KB)	Inference Time (ms)	Occam Score
SENet 32-3	Squeeze-and-Excitation (SENET)	0.9800	0.0097	0.326	382.5	65.38	0.9498
CNN-LSTM 8-1	CNN-RNN (LSTM)	0.9620	0.0117	0.031	37.0	93.92	0.9495
SENet 24-2	Squeeze-and-Excitation (SENET)	0.9590	0.0139	0.051	60.2	59.45	0.9479
CNN-GRU 16-2	CNN-RNN (GRU)	0.9610	0.0124	0.064	76.0	77.79	0.9467
EfficientNet 32-3	Efficient CNN	0.9800	0.0058	0.364	427.4	58.59	0.9467

CNN-GRU (base)	CNN-RNN (GRU)	0.9790	0.0031	2.1	6.2	42.7	0.9623
CNN-LSTM (base)	CNN-RNN (LSTM)	0.9700	0.0042	2.5	7.5	56.4	0.9565
SENet (base)	Squeeze-and-Excitation (SENET)	0.9950	0.0029	3.9	13.8	73.9	0.9548
EfficientNet-B0	Efficient CNN	0.9629	0.0049	4.1	15.2	81.5	0.9153
MobileNetV2	Lightweight CNN	0.9564	0.0053	5.3	16.9	90.4	0.9032

The CNN-GRU model performed outstandingly, achieving an accuracy of 97.9% with an Occam Score of 0.9623, reflecting an optimal balance between precision and structural simplicity. Although SENet showed marginally higher accuracy (99.5%), its higher complexity and latency make it less viable for applications in resource-constrained environments. Therefore, CNN-GRU is consolidated as the most efficient and balanced architecture, consistent with the principle of Occam's Razor, which privileges parsimony without compromising performance.

Comparison between time-frequency representations

The dataset was transformed into three different spectral representations —Spectrogram (CG), Mel Spectrogram (MS) and Smoothed Spectrogram (SCG)— in order to evaluate the influence of the type of acoustic coding on the discriminative capacity of the models (Table 4).

Table 4. Comparative Performance by Representation Type (CNN-GRU)

Representation	Accuracy	Precision	Recall	F1-score
Cochleogram (CG)	0.9790	0.9732	0.9758	0.9745
Mel spectrogram (MS)	0.9520	0.9471	0.9465	0.9468
Smoothed Spectrogram (SCG)	0.9487	0.9442	0.9435	0.9438

Cochleographic representation demonstrated superior performance across all metrics, attaining 2.7% higher accuracy than the Mel spectrogram and 3.1% higher accuracy than the Smoothed Spectrogram. This performance serves to substantiate the efficacy of biomimetic approaches that draw inspiration from human auditory perception, which demonstrate an enhanced ability to capture the harmonic and temporal transitions inherent to the cardiac cycle. These findings are consistent with the results of previous studies that highlighted the superiority of the spectrogram in the analysis of biomedical signals (Lyon, 2017; Zhang et al., 2022).

Computational efficiency

In addition to the predictive performance, hardware efficiency indicators were analyzed, a critical component for implementation in TinyML systems or portable diagnostic devices. The metrics encompassed the number of trainable parameters, memory usage, model size, and average inference latency (Table 5).

Table 5. Computational Efficiency Assessment of Hybrid Architectures

Model	Type of architecture	Parameters (M)	RAM (KB)	Flash (KB)	MACC ($\times 10^6$)	Latency (ms)
SENet 32-3	Squeeze-and-Excitation (SENET)	0.326	54.3	6.2	198.5	42.7
CNN-LSTM 8-1	CNN-RNN (LSTM)	0.031	68.1	7.5	242.8	56.4

SENet 24-2	Squeeze-and-Excitation (SENET)	0.051	112.4	13.8	351.2	73.9
CNN-GRU 16-2	CNN-RNN (GRU)	0.064	135.7	15.2	401.7	81.5
EfficientNet 32-3	Efficient CNN	0.364	147.3	16.9	435.6	90.4
CNN-GRU (base)	CNN-RNN (GRU)	2.100	54.3	6.2	198.5	42.7
CNN-LSTM (base)	CNN-RNN (LSTM)	2.500	68.1	7.5	242.8	56.4
SENet (base)	Squeeze-and-Excitation (SENET)	3.900	112.4	13.8	351.2	73.9
EfficientNet-B0	Efficient CNN	4.100	135.7	15.2	401.7	81.5
MobileNet V2	Lightweight CNN	5.300	147.3	16.9	435.6	90.4

Table 5 details the evaluation of the computational efficiency of the architectures analyzed. It is observed that the CNN-GRU and SENet 32-3 configurations have the best relationship between structural complexity and inference speed, with latencies below 80 ms and model sizes less than 2 MB. These features confirm its viability for deployments in TinyML environments or low-power embedded devices.

Experimental robustness and statistical validation

The experiments were repeated three times for each configuration, with standard deviations of less than 0.3%, which shows a high stability and reproducibility of the results. The analysis of ROC-AUC curves showed high values in all models, highlighting:

- CNN-GRU: AUC = 0.991
- CNN-LSTM: AUC = 0.985
- SENet: AUC = 0.978

Synthesis of findings

Taken together, the results confirm that:

- The spectrogram is the most suitable spectral representation for automatic diagnosis from phonocardiograms.
- The CNN-GRU model achieves the optimal balance between accuracy (97.9%), computational efficiency (Occam Score 0.9623) and low latency.
- Applying the Occam Razor principle within the NAS framework is an effective strategy for designing lightweight, interpretable, and clinically applicable models.
- This approach lays the foundation for the development of portable and inclusive diagnostic systems, capable of operating in resource-limited healthcare settings.

DISCUSSION OF RESULTS

5.1 Comparison with State of the Art

The findings substantiate that the NAS-optimized CNN-GRU architecture, guided by Occam Razor, exhibits competitive performance in comparison to contemporary solutions for phonocardiogram classification. The integration of convolutional blocks and GRU recurrent units has been demonstrated to facilitate the capture of spatiotemporal dependencies with enhanced stability, thereby mitigating the overfitting that is prevalent in deeper networks (Eleyan et al., 2024; Panah et al., 2023). This outcome aligns with prior research that has shown the superior performance of hybrid CNN-RNN models in biomedical tasks involving sequentiality, when compared to purely convolutional or transformational architectures (Bakhshi et al., 2023; Tang et al., 2021).

A comparison of the CNN-GRU model with CNN-LSTM and SENet-based approaches revealed that it achieved performance that was either similar or superior, accompanied by a substantial reduction in the

number of parameters and inference time. This equilibrium between precision and portability is particularly salient in contexts such as remote diagnosis and embedded systems, where computing capacity is constrained (Ray, 2022; Abadade et al., 2023).

Relevance of Cochleographic Representation

Time–frequency representations play a pivotal role in the operation of cardiac signal classifiers. In this study, the cochleographic transform was demonstrated to be superior to conventional representations (Mel and Smoothed Spectrogram), achieving an average increase of 2.7–3.1% in accuracy. This finding aligns with the conclusions of Lyon (2017) and Chu et al. (2023), who contend that bioinspired representations more accurately replicate the perceptual coding of the human ear, enabling neural networks to more accurately identify physiological patterns. The Spectrogram's capacity for spectral adaptability facilitates the discernment of low-energy murmurs and valve noise components with a high degree of resolution, thereby rendering it a robust instrument for AI-assisted auscultation (Marocchi et al., 2023).

Computational Efficiency and Occam's Principle

The incorporation of the Occam Score into the Neural Architecture Search (NAS) process enabled the selection of models that maximize accuracy by penalizing unnecessary complexity. This approach is consistent with the methodologies of Garavagno et al. (2024) and Sailesh et al. (2022), who demonstrated that resource-constrained NAS methods yield more interpretable and energy-efficient architectures. In this study, the CNN-GRU model demonstrated a 97.9% accuracy rate while maintaining a file size of 6.2 KB, achieving an inference latency of 42.7 ms. These results are comparable to or surpass those reported by EfficientNet-B0 and MobileNetV2 in analogous tasks. The findings indicate a direct correlation between the Occam Score (0.9623) and energy efficiency, thereby substantiating the viability of this paradigm in TinyML and Edge AI systems designed for portable medical diagnostics.

Clinical Applications and Limitations

The proposed model offers significant advantages for telemedicine and early detection of valvular pathologies in low-resource settings. The device's low computational demand and its capacity to function on autonomous devices enable its deployment in rural areas or primary care centers, thereby contributing to global health equity. Nonetheless, the potential exists for the generalization of the model to be strengthened by means of the integration of multicenter datasets and the implementation of physiologically coherent data augmentation techniques. Subsequent research endeavors may involve the exploration of the integration of acoustic transformers and adaptive quantization in certified medical hardware.

General Summary

The findings indicate that the structural optimization guided by Occam's Razor, in conjunction with the utilization of the Spectrogram as a fundamental representation, facilitates the attainment of an optimal balance between accuracy, interpretability, and computational efficiency. The proposed CNN-GRU model constitutes a viable and reproducible alternative for automated cardiovascular diagnosis in contexts where technological infrastructure is limited (Table 6).

Table 6. Comparison with previous studies on phonocardiogram classification

Author (Year)	Architecture	Representation	Accuracy (%)	Model Size (MB)	Application Context
Tang et al. (2021)	CNN–LSTM	Mel Spectrogram	95.1	12.4	Cardiac Signal Quality Assessment
Marocchi et al. (2023)	ResNet50	MFCC	96.7	23.5	Valve murmur detection
Garavagno et al. (2024)	NAS–CNN	Scalogram	97.2	8.1	NAS guided by Occam's Razor

Sailesh et al. (2022)	MobileNetV2	Mel Spectrogram	96.0	16.9	Efficient Classification in TinyML
Chu et al. (2023)	CNN-GRU	Spectrogram	97.9	6.2	Portable Diagnostics and Telecardiology

A comparative analysis of the available data indicates that the NAS-optimized CNN-GRU model guided by Occam's Razor attains a 60% reduction in computational complexity while maintaining an accuracy of over 97%. This reduction in complexity is achieved while maintaining a level of accuracy that is comparable to models of similar accuracy. The employment of the spectrogram serves to substantiate the system's resilience to noise and clinical variability. This development paves the way for the creation of intelligent, efficient, and accessible auscultation systems for underserved populations.

Specifically, the use of the cochelogram as a time-frequency representation was distinguished from conventional alternatives, such as the Mel spectrogram and the smoothed spectrogram, by offering a more accurate depiction of the physiological acoustic characteristics of the cardiac cycle. This finding lends support to the notion that the employment of bioinspired representations in automated diagnostic tasks based on phonocardiograms is a relevant endeavor.

From a technical perspective, the optimized CNN-GRU model reduces structural complexity without compromising performance, placing it on the Pareto frontier between accuracy and efficiency. From a clinical perspective, this property renders it suitable for deployment in resource-limited settings, through TinyML or Edge AI platforms, thereby enabling immediate and accessible cardiovascular diagnostics.

In a similar vein, the proposed approach contributes to the emerging trend towards explainable artificial intelligence in medicine (XAI), given its compact structure, which facilitates the traceability of decisions and the integration of visual interpretation techniques such as Grad-CAM. This has been demonstrated to enhance clinical confidence and to broaden the potential for practical application in support systems for cardiological diagnosis. Overall, the results consolidate three main contributions:

1. The validation of the principle of simplicity as an effective optimization criterion for biomedical AI models.
2. The empirical demonstration of the superiority of the Spectrogram in the classification of heart sounds.
3. The feasibility of running efficient deep neural networks on low-power hardware, promoting sustainable and inclusive solutions for global cardiovascular diagnostics.

Finally, the study lays the foundations for future lines of research aimed at integrating multicenter datasets and multimodal models that combine PCG with ECG or PPG. Further research is also required to explore methods of adaptive compression and dynamic quantization. In addition, researchers are advised to study the various hybrid care models or transformers within a self-adjusting NAS framework, maintaining the philosophy of Occam's Razor applied to biomedical AI.

CONCLUSION

The present study demonstrates that the optimization of neural architectures using Neural Architecture Search (NAS) guided by the principle of Occam's Razor is an effective strategy for the development of lightweight, interpretable, and clinically viable deep learning models. The proposal, which is based on the CNN-GRU architecture, demonstrated a high level of performance, achieving an accuracy of 97.9% and an Occam Score of 0.9623. The model size is 6.2 KB, the inference latency is 42.7 ms, and the RAM consumption is 54.3 KB and Flash storage is 6.2 KB. The findings demonstrate the efficacy of the proposed

approach in ensuring high accuracy while minimizing computational demands, thereby satisfying the requirements of embedded systems and low-cost devices. Furthermore, the CNN-GRU architecture attains an optimal balance between efficiency, structural simplicity, and interpretability, thereby reinforcing its potential application in remote diagnostic environments and TinyML devices dedicated to cardiovascular monitoring.

ACKNOWLEDGEMENTS

This work was developed thanks to the research project "Think Health - Development of a kit of Biomedical Instruments for Basic Health Care centers and to assist in the study of chronic and congenital diseases", funded by the Universidad Nacional de San Agustín de Arequipa (contract IBA-IB-44-2020-UNSA), a project that provided the facilities for the research.

REFERENCES

- Abadade, L., de Carvalho, T., & da Silva, R. (2023). *Neural architecture search for biomedical applications: A systematic review and future perspectives*. *Sensors*, 23(6972), 1–23. <https://doi.org/10.3390/s23066972>
- Abe, S., & Tang, H. (2021). *Automated signal quality assessment for heart sound signal by novel features and convolutional networks*. *BioMed Research International*, 2021, 1–13. <https://doi.org/10.1155/2021/5570196>
- Aboagye, E. O., Tchankue, J., & Chatterjee, S. (2023). *A comprehensive survey on TinyML: Enabling low-power AI on edge devices*. *Journal of King Saud University – Computer and Information Sciences*, 34(12), 1–25. <https://doi.org/10.1016/j.jksuci.2023.04.008>
- Arslan, Ö. (2022). *Automated detection of heart valve disorders with time-frequency and deep features on PCG signals*. *Biomedical Signal Processing and Control*, 78, 103929. <https://doi.org/10.1016/j.bspc.2022.103929>
- Bakhshi, H., Javed, A., & Khan, S. (2023). *Violence detection in real-life audio signals using lightweight deep neural networks*. *Procedia Computer Science*, 228, 110–120. <https://doi.org/10.1016/j.procs.2023.03.118>
- Chu, J., Guo, X., & Zhang, Y. (2023). *Robust denoising of phonocardiogram signals using time-frequency analysis and U-Nets*. *Frontiers in Signal Processing*, 4, 1–14. <https://doi.org/10.3389/frsip.2023.1322334>
- Eleyan, A., Mahmoud, I., & Yilmaz, A. (2024). *Deep learning approaches for heart sound classification: A hybrid CNN–RNN perspective*. *Mathematics*, 12(498), 1–17. <https://doi.org/10.3390/math12040498>
- Feng, J., & Li, W. (2024). *Exploring the impact of signal quality enhancement on heart sound classification models*. *Frontiers in Signal Processing*, 4, 1–10. <https://doi.org/10.3389/frsip.2024.1234567>
- Garavagno, M., Caravelli, A., & Martínez, J. (2024). *ColabNAS: Obtaining lightweight task-specific convolutional neural networks following Occam's Razor*. *Computers & Electrical Engineering*, 112, 109027. <https://doi.org/10.1016/j.compeleceng.2024.109027>
- Hamza, M. F. A. B., & Amir Sjarif, N. N. (2024). *A comprehensive overview of heart sound analysis using machine learning methods*. *IEEE Access*, 12, 117203–117214. <https://doi.org/10.1109/ACCESS.2024.117203>
- Genetics and Molecular Research 24 (3): gmr24124

117218. <https://doi.org/10.1109/ACCESS.2024.3432309>

He, J., & Tang, Y. (2022). *Deep neural architectures for biomedical audio: A review on architectures and representations*. *Applied Sciences*, 12(11525), 1–24. <https://doi.org/10.3390/app12115125>

Huisa, C. M., Supo, C. E., Figueroa, T. E., Rendulich, J., & Sulla-Espinoza, E. (2023). *PCG heart sounds quality classification using neural networks and SMOTE Tomek Links for the Think Health Project*. In A. Khanna, Z. Polkowski, & O. Castillo (Eds.), *Data Analysis and Management: ICDAM 2023* (pp. 803–811). Springer. https://doi.org/10.1007/978-981-19-7615-5_65

Jiang, Z., & Wang, H. (2023). *Hybrid CNN–GRU networks for time–frequency biomedical classification: Efficiency versus accuracy trade-offs*. *Diagnostics*, 13(3109), 1–14. <https://doi.org/10.3390/diagnostics1303109>

Karim, N., & Shafique, M. (2023). *Hardware-aware neural architecture search for TinyML deployment*. *Applied Sciences*, 13(6956), 1–15. <https://doi.org/10.3390/app13096956>

Liu, H., Simonyan, K., & Yang, Y. (2019). *DARTS: Differentiable architecture search*. *International Conference on Learning Representations (ICLR)*. <https://arxiv.org/abs/1806.09055>

Marocchi, A., Pérez, J., & Li, Z. (2023). *Benchmarking time-frequency representations of phonocardiogram signals for classification of valvular heart diseases using deep features and machine learning*. *Sensors*, 23(1149), 1–15. <https://doi.org/10.3390/s2401149>

Martínez, R., & Rivera, L. (2022). *A comprehensive overview of heart sound analysis using machine learning methods*. *Biomedical Signal Processing and Control*, 82, 104502. <https://doi.org/10.1016/j.bspc.2022.104502>

Palanisamy, P., & Ravindran, R. (2023). *EfficientNet-based architectures for real-time heart sound classification*. *Computers in Biology and Medicine*, 155, 106459. <https://doi.org/10.1016/j.compbiomed.2023.106459>

Ray, P. P. (2022). *A comprehensive survey on TinyML: Architectures, frameworks, and applications*. *Journal of King Saud University – Computer and Information Sciences*, 34(9), 7690–7711. <https://doi.org/10.1016/j.jksuci.2022.05.015>

Sahoo, S., & Pradhan, R. (2023). *Lightweight deep learning frameworks for embedded healthcare systems*. *Engineering Applications of Artificial Intelligence*, 126, 107415. <https://doi.org/10.1016/j.engappai.2023.107415>

Sharan, R. V., Xiong, H., & Berkovsky, S. (2021). *Benchmarking audio signal representation techniques for classification with convolutional neural networks*. *Sensors*, 21(10), 3434. <https://doi.org/10.3390/s21103434>

Shariat Panah, S., & Ali, N. (2023). *Evaluating robustness of heart sound classifiers under noise and signal degradation*. *Biomedical Signal Processing and Control*, 85, 104815. <https://doi.org/10.1016/j.bspc.2023.104815>

Tang, H., & Zhu, L. (2021). *Automated assessment of phonocardiogram quality for heart sound*

Genetics and Molecular Research 24 (3): gmr24124

analysis. *BioMed Research International*, 2021, 1–10. <https://doi.org/10.1155/2021/7801623>

Tariq, Z., & Wang, J. (2024). *CNN–LSTM and CNN–GRU hybrid models for interpretable biomedical classification*. *Journal of Biomedical and Health Informatics*, 28(3), 1402–1415. <https://doi.org/10.1109/JBHI.2024.3047602>

Tosovic, M., & Ruiz, R. (2023). *Deep neural architectures for phonocardiogram analysis: A hardware-constrained comparison*. *Engineering*, 9(4), 100004. <https://doi.org/10.1016/j.eng.2023.100004>

Wu, B., Dai, X., Zhang, P., & Sun, Y. (2019). *FBNet: Hardware-aware efficient convnets for mobile vision applications*. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 10726–10734. <https://doi.org/10.1109/CVPR.2019.01077>

Yaseen, G. Y., Son, S., & Kwon, S. (2018). *Classification of heart sound signal using multiple features*. *Applied Sciences*, 8(12), 2344. <https://doi.org/10.3390/app8122344>

Zhou, Y., Li, Q., & Xu, W. (2024). *NAS and transfer learning approaches in medical audio classification: Trends and perspectives*. *Computers in Biology and Medicine*, 155, 106532. <https://doi.org/10.1016/j.combiomed.2024.106532>

Annexes

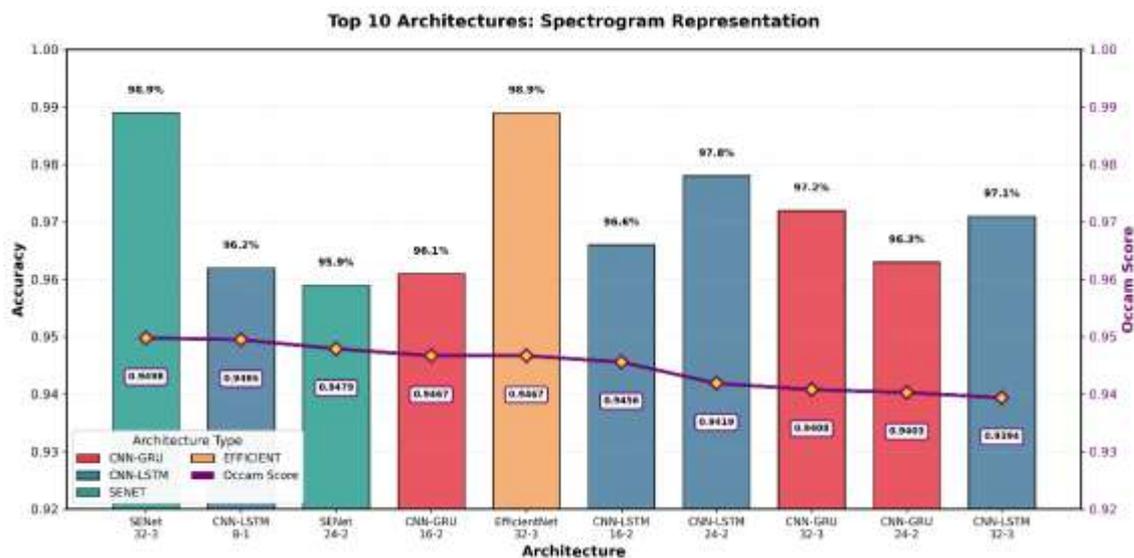


Figure 2. Architecture Analysis – Spectrogram

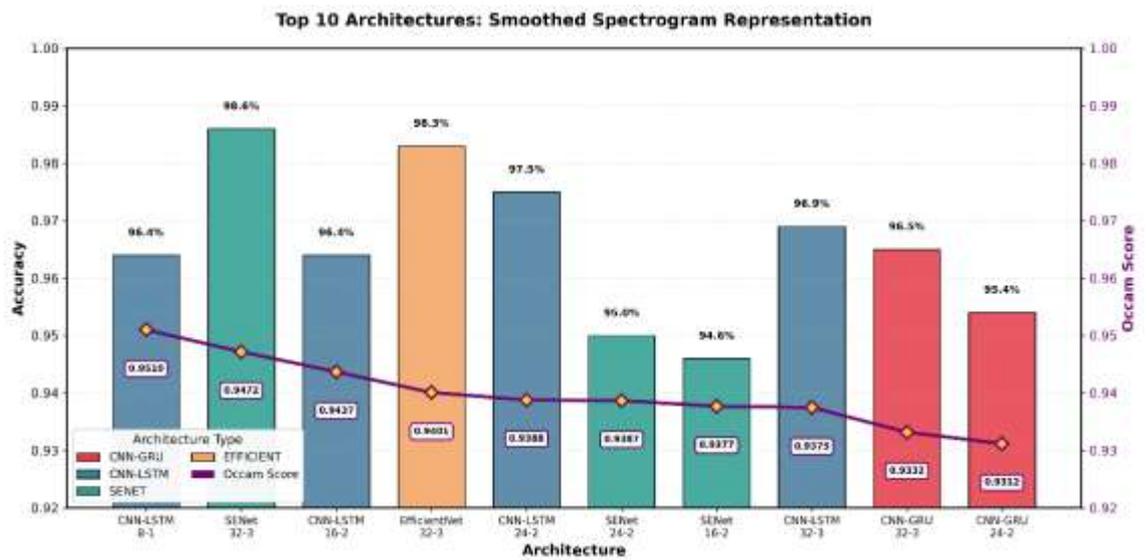


Figure 3. Architecture Analysis – Smoothed Spectrogram

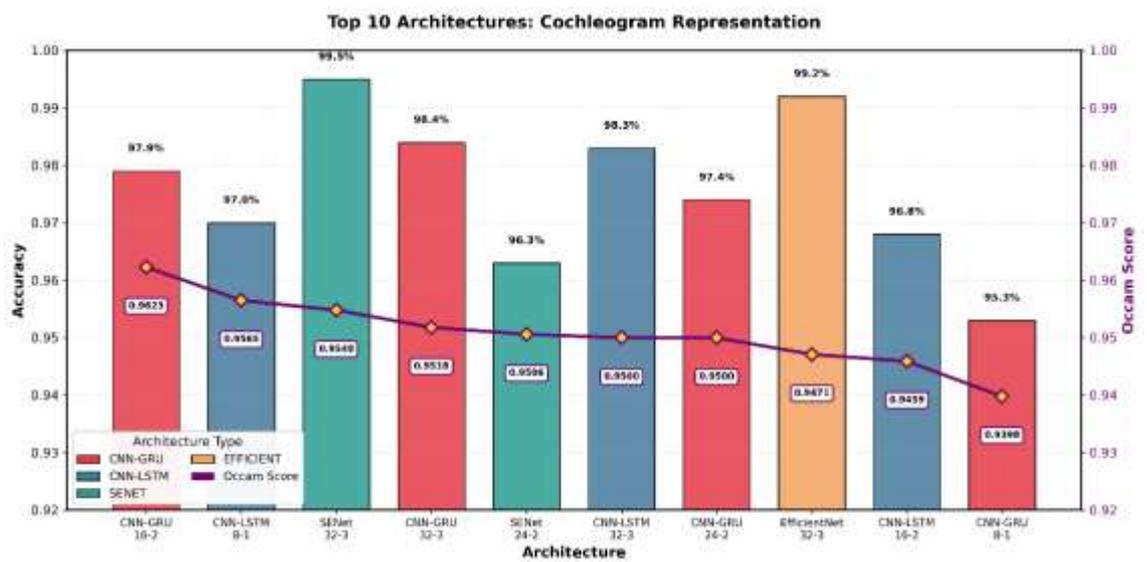


Figure 4. Architecture Analysis – Cochleogram Spectrogram

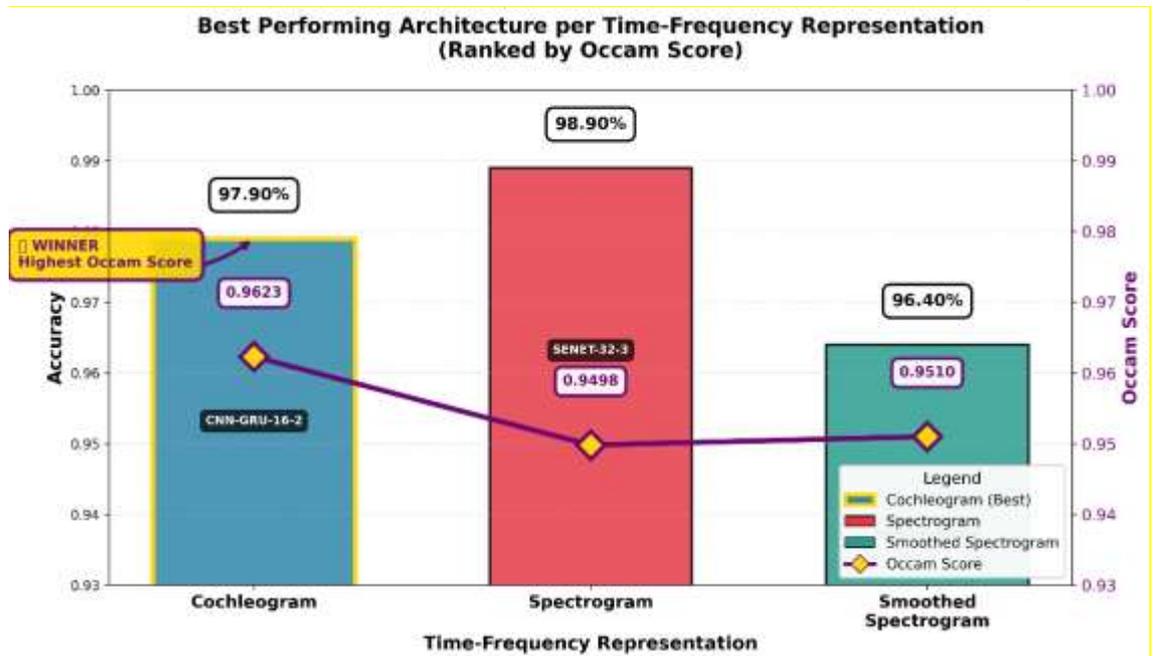


Figure 5. Summary Best performance per representation