Closing the Gap Between Lab and Clinic: Validating Lightweight AI for ECG Arrhythmia Classification

Research Report submitted in partial fulfillment of the requirements
for the **Postgraduate Diploma in Data Analytics**
at The Independent Institute of Education.

Makabongwe Lwethu Sibisi
ST10145439

Supervisor: Prof. Mohammed Vawda

13 / 06 / 2025

DECLARATION

I hereby declare that the **research proposal** submitted for the **Postgraduate Diploma in Data Analytics** to The Independent Institute of Education (The IIE) is my own work and has not previously been submitted to another University or Higher Education Institution for a postgraduate qualification.

Signature L. Sibisi____________________ Date____13. / 06 / 2025____

Table of Contents

INTRODUCTION

TITLE:

Closing the Gap Between Lab and Clinic: Validating Lightweight AI for ECG Arrhythmia Classification

INTRODUCTION:

Arrhythmia classification is a labour intensive and time-consuming process that requires specialised clinicians and ECG readings for an accurate diagnosis(Sajad & Fatemeh, 2019). Many studies have explored automating heartbeat classification based on ECG data, aiming to identify intricate patterns and irregularities from the norm. Among these approaches, deep learning algorithms have demonstrated significant potential for automating arrhythmia detection. However, despite achieving high accuracy in controlled laboratory settings, many of these models fail to maintain performance in real-world clinical environments. This is due to issues such as data variability, the presence of noise, and a lack of external validation (Yaqoob, et al., 2023). The aim of this proposed study is to address the gap between laboratory accuracy and real-world clinical performance This will be achieved by implementing a Recursive Depthwise Separable Convolutions (RDSCNN) and evaluating it using robust external validation strategies. Ultimately, this model is intended as a decision-support tool for clinicians, automating initial arrhythmia categorisation so that specialists can focus on confirming diagnosis and planning treatment.

BACKGROUND:

The World Health Organization (WHO) reports that cardiovascular diseases (CVDs) were the leading cause of death globally in 2019, accounting for an estimated 17.9 million fatalities, or 32% of all global deaths. Notably, over 75% of these deaths occurred in middle- and low-income countries, underscoring the critical need for accessible diagnostic tools across diverse healthcare settings (WHO, 2021). Cardiovascular diseases (CVDs) encompass a range of conditions affecting the circulatory system, including the heart and blood vessels (Robinson, 2021). These can manifest as structural issues like coronary artery disease, congenital heart defects, and heart failure, or as functional problems such as arrhythmia (Peek & Buczinski, 2018). Arrhythmia is a condition where the heart beats in an irregular rhythm (too slow, too fast, or erratically). As a cardiovascular condition, arrhythmia impacts the heart's ability to efficiently pump blood, which can lead to complications and reduce overall cardiovascular well-being (Wang, et al., 2021).

Diagnosing cardiac arrhythmias typically involves a multi-faceted approach, incorporating blood pressure measurements, detailed analysis of ECG readings, and the identification of irregular heartbeats. Clinicians also consider patient-reported symptoms like weakness, fatigue, dizziness, and a decline in normal activity levels. The electrocardiogram (ECG) is a non-invasive diagnostic tool that records the electrical activity of the heart over time. It plays a crucial role in detecting abnormalities in heart rhythm and functions (Luz, et al., 2016). Although fully automated systems have been proposed, clinical guidelines and liability concerns favour tools that augment clinician expertise rather than replace it (Price II WN, 2024).

RATIONALE:

Given the global burden of cardiovascular diseases and the disproportionately high impact in middle- and low-income countries, there is a pressing need for scalable, cost-effective diagnostic tools that can assist clinicians in early and accurate arrhythmia detection. While deep learning models have shown promise in automating arrhythmia classification using ECG data, many existing models are trained and evaluated in controlled laboratory environments, limiting their generalisability in diverse clinical settings (Yaqoob, et al., 2023).

This study will aim to address that limitation by implementing a deep learning-based model that will be both robust and externally validated. By adhering to the Inter-Patient Paradigm and AAMI Standards, this study will not only assess generalisability but also will position the system as an assistive technology. Although fully automated arrhythmia detectors have been explored, prevailing clinical guidelines and liability considerations favour tools that augment, rather than replace, human expertise (Crigger, et al., 2021).

The proposed model, based on Recursive Depthwise Separable Convolutions (RDSCNN), will be designed to support clinicians in accurately classifying arrhythmias from ECG data. Due to its lightweight architecture and efficiency, RDSCNN will offer a promising balance between accuracy and computational cost, making it especially suitable for deployment in diverse healthcare settings (Chollet, 2017). This study will contribute meaningfully to the fields of Data Science and Artificial Intelligence by advancing current machine learning approaches in medical diagnostics. Moreover, it will offer a viable framework for integrating assistive diagnostic tools into real-world clinical workflows.

PROBLEM STATEMENT:

Despite the success of deep learning models in classifying arrhythmias from ECG data under controlled conditions, their real-world clinical reliability remains limited due to poor generalisability and insufficient external validation. There is a lack of models developed specifically to function as assistive diagnostic tools that are both computationally efficient and clinically aligned with prevailing guidelines (Sajad & Fatemeh, 2019) (Yaqoob, et al., 2023).

RESEARCH QUESTION(S):

Primary Research Question:

How does a Recursive Depthwise Separable Convolutional Neural Network (RDSCNN) perform in classifying arrhythmias under the Inter-Patient Paradigm and AAMI standards, compared to existing deep learning models?

Subsidiary Research Questions:

1. Can the RDSCNN maintain classification accuracy across multi-institutional ECG datasets with varying noise levels and patient demographics?
2. To what extent does external validation using the Inter-Patient Paradigm and adherence to AAMI standards improve the generalizability and clinical applicability of the RDSCNN model?
3. What accuracy, computational efficiency, and interpretability metrics indicate the RDSCNN's suitability as an assistive tool in resource-constrained settings?

RESEARCH OBJECTIVES:

1. To design and implement a deep learning model using Recursive Depthwise Separable Convolutions for arrhythmia classification.
2. To evaluate the model's performance under the Inter-Patient Paradigm to assess generalisability.
3. To benchmark the model against existing ECG classification methods in terms of accuracy, computational efficiency, and clinical applicability.
4. To demonstrate the relevance of adhering to AAMI standards in creating assistive, clinically acceptable diagnostic tools.

LITERATURE REVIEW

THEORETICAL APPROACH:

This proposed study is fundamentally grounded in Hierarchical Feature Learning Theory, which provides the core theoretical framework for understanding how deep neural networks, particularly Convolutional Neural Networks (CNNs), efficiently learn and extract meaningful representations from complex data such as electrocardiogram (ECG) signals (Zhao, et al., 2017). This theory posits that deep learning models develop a hierarchical understanding of data by learning progressively abstract and complex features from raw input through multiple layers of processing (Ihsanto, et al., 2020). Each successive layer builds upon the features learned by the previous layer, transforming raw data into high-level representations suitable for classification.

In the context of ECG arrhythmia classification, Hierarchical Feature Learning Theory is crucial because it enables the automated extraction of discriminative features from raw waveform data, eliminating the need for manual and often subjective feature engineering (Aziz, et al., 2021). Convolutional layers within CNNs are inherently designed to detect local patterns (e.g., specific waveform morphologies, QRS complexes, P-waves, T-waves) across the input signal, regardless of their precise location. As information propagates through deeper layers, these local features are combined to form more abstract and comprehensive representations of cardiac rhythms and abnormalities.

Thus, Hierarchical Feature Learning Theory explains how the RDSCNN model can efficiently learn complex patterns from raw ECG data to achieve accurate arrhythmia classification. This efficient and generalized feature learning is the foundational technical capability that underpins the study’s aim to develop a robust, assistive diagnostic tool capable of bridging the gap between laboratory performance and real-world clinical reliability (Ramli, et al., 2020).

LITERATURE REVIEW:

The precise diagnosis and categorization of cardiac arrhythmias from electrocardiogram (ECG) data are foundational to cardiovascular medicine. However, traditional manual analysis remains a labour-intensive, time-consuming, and notably error-prone process (Sajad & Fatemeh, 2019). This reliance on specialized clinicians and extensive visual inspection introduces bottlenecks in patient care, particularly in settings with limited access to cardiology experts. Recognizing these inherent limitations, researchers have consistently sought automated methods for heartbeat classification.

Early studies in automated ECG analysis primarily leveraged conventional Machine Learning (ML) techniques. These approaches typically involved extensive hand-crafted feature engineering, extracting specific morphological characteristics (such as QRS complex duration, ST-segment elevation) and heartbeat interval features from ECG signals. Classifiers such as Support Vector Machines (SVMs) or K-Nearest Neighbours (KNN) then utilized these features (Chazal, et al., 2004). While these methods represented significant initial advancements in diagnostic support, their performance and generalizability were often constrained. Their effectiveness was directly tied to the quality and exhaustiveness of the manually engineered features. This reliance on domain expertise for feature extraction limited their scalability and adaptability across diverse patient populations and varying ECG recording conditions.

The development of Deep Learning (DL) has fundamentally transformed the landscape of automated ECG analysis. DL models enable end-to-end learning directly from raw ECG waveforms. This approach, rooted in Hierarchical Feature Learning Theory, allows DL models, particularly Convolutional Neural Networks (CNNs), to automatically extract complex, hierarchical features. This capability bypasses the need for laborious and potentially subjective manual feature engineering. Numerous studies consistently demonstrate the significant potential of CNNs and their architectural variants in automated arrhythmia detection (Luo, et al., 2017)(Yildirim, 2018)(Midani, et al., 2023).

Recent advancements in deep learning for ECG analysis highlight continued innovation. Architectures integrating Bidirectional Long Short-Term Memory (BiLSTM) with CNNs have shown robust performance. These models excel at classifying multiple heartbeats by capturing complex temporal dependencies across ECG sequences (Yang, et al., 2022). Furthermore, the emergence of Transformer architectures, exemplified by models like ECGformer, represents a frontier in capturing long-range dependencies within ECG signals. This leads to enhanced arrhythmia classification accuracy (Akan, et al., 2023). These deep learning methodologies have consistently reported impressive accuracy metrics. They often exceed 95% on benchmark datasets like the MIT-BIH Arrhythmia Database. This success has fuelled considerable optimism regarding their transformative potential in cardiac diagnostics, offering the promise of rapid, consistent, and potentially more accessible arrhythmia screening.

Despite these compelling advancements and the high-performance metrics consistently reported in controlled laboratory environments, a critical and persistent gap remains. This gap exists between reported laboratory accuracy and the consistent reliability and practical utility of these models in real-world clinical settings. This enduring disparity is primarily attributed to pervasive methodological shortcomings in how many DL models are developed and evaluated. Recent systematic reviews explicitly underscore this concern (Silva, et al., 2025) (Yaqoob, et al., 2023).

Firstly, a significant and widely acknowledged issue is the non-adherence to the Association for the Advancement of Medical Instrumentation (AAMI) EC57 standards (ANSI/AAMI EC57:1998/®2003, 1998–2008). These standards provide precise, internationally recognized definitions for five core arrhythmia classes: Normal (N), Supraventricular Ectopic Beat (S), Ventricular Ectopic Beat (V), Fusion Beat (F), and Unknown Beat (Q). These guidelines are crucial. They ensure consistency in beat classification and direct comparability across different research studies. However, a substantial number of published works either do not explicitly state their adherence to AAMI standards or instead utilize custom classification schemes (Silva, et al., 2025). This prevalent lack of standardization introduces significant variability in reported performance metrics. It makes meaningful comparison and synthesis of findings across the literature exceedingly challenging. Ultimately, it undermines the progression towards clinically robust and interoperable automated diagnostic solutions. Without AAMI compliance, models developed in one setting may be incompatible with data from another, hindering their widespread adoption and regulatory approval. This raises concerns about safety and efficacy in diverse clinical applications.

Secondly, the failure to implement rigorous inter-patient evaluation paradigms remains a critical issue. This directly impacts real-world applicability and model generalizability. Many studies inadvertently adopt training and testing splits where data from the same patient can appear in both sets. This leads to a phenomenon known as data leakage or patient-specific overfitting (Sajad & Fatemeh, 2019) (Xiao, et al., 2023). Data leakage results in artificially inflated accuracy metrics. The model effectively "memorizes" characteristics of specific patients rather than learning generalizable arrhythmia patterns. These learned patterns should be transferable to novel individuals. When such models are subsequently deployed in actual clinical environments to analyse ECGs from previously unseen patients, their performance often degrades significantly. They fail to maintain the high accuracy observed during controlled validation (Xiao, et al., 2023). This deficiency highlights a crucial point: reported high accuracies frequently do not reflect a model's true generalizability. This refers to its capacity to perform reliably on new data from diverse, unseen patient populations. Generalizability is paramount for fostering clinical trust and enabling widespread deployment of AI systems. The literature increasingly emphasizes that strict inter-patient partitioning, where no patient's data is present in both the training and testing sets, is a minimum requirement for credible model evaluation and a prerequisite for real-world applicability (Silva, et al., 2025) (Yaqoob, et al., 2023). This ensures that the model truly learns underlying physiological patterns, not just individual patient peculiarities.

Furthermore, while maximizing diagnostic accuracy is a primary research focus in academic settings, the practical implementation of automated diagnostic tools in clinical environments necessitates broader considerations. This is especially true in resource-constrained healthcare settings. Here, factors beyond pure performance become critical. Computational efficiency that encompassing model size, inference speed, and overall resource consumption is often underreported or not adequately prioritized in research. Models with a substantial number of parameters, while potentially delivering high accuracy, may not be feasible for practical real-time analysis on edge devices such as wearable sensors or portable ECGs, and in clinical settings with constrained computing resources. This creates a tangible barrier to translating laboratory successes into widespread clinical benefits. Such models may require expensive hardware or cloud computing resources, making them inaccessible to low-resource settings. This directly contradicts the global need for scalable and cost-effective diagnostic tools.

Moreover, prevailing clinical guidelines and liability considerations increasingly favour assistive diagnostic tools. These tools augment human expertise rather than replace it entirely (Crigger, et al., 2021). This paradigm requires automated systems to be not only accurate and computationally efficient but also capable of providing reliable and interpretable outputs. Clinicians must be able to seamlessly integrate these outputs into their complex diagnostic workflows. Trust in AI systems in healthcare is built on transparency, reliability, and the ability of human experts to retain ultimate control and responsibility. A model that merely provides a "black box" prediction, regardless of its accuracy, is less likely to be adopted or trusted in critical diagnostic scenarios.

Crucially, recent systematic reviews, such as that by (Silva, et al., 2025), explicitly identify these persistent methodological limitations. More importantly, they propose a clear pathway forward for robust evaluation of ECG classification systems. This specific systematic review introduces the EC3 framework. EC3 comprises Standards Adherence (E1), Fair Evaluation (E2), and Embedded Feasibility (E3). These are presented as essential pillars for assessing the translational readiness of ECG classification systems. This framework underscores that developing truly clinically relevant automated arrhythmia classification systems depends on adopting standardized evaluation approaches. These approaches extend beyond basic accuracy metrics. Such frameworks consistently emphasize robust AAMI compliance for consistent beat classification (aligned with E1: Standards Adherence), stringent inter-patient validation to ensure generalizability (aligned with E2: Fair Evaluation), comprehensive assessment of computational efficiency metrics (aligned with E3: Embedded Feasibility), and explicit consideration of clinical relevance criteria. These comprehensive evaluation standards are vital for accelerating the transition of AI research from laboratories to effective clinical practice (ANSI/AAMI EC57:1998/(R)2003, 1998–2008.).

This proposed study directly addresses the identified niche. It focuses on the development and rigorous evaluation of a lightweight Recursive Depthwise Separable Convolutional Neural Network (RDSCNN) model. The selection of RDSCNN is a direct response to the critical need for computational efficiency. This capability is deeply rooted in its Hierarchical Feature Learning via Depthwise Separable Convolutions (Chollet, 2017). Simultaneously, the study aims for high classification accuracy, representing a critical balance for resource-constrained settings. Furthermore, this research will meticulously adhere to AAMI EC57 standards and implement a strict inter-patient validation paradigm across multiple publicly available and diverse ECG datasets. By doing so, this study aims to contribute novel insights into the generalizability and real-world applicability of DL models for arrhythmia classification. This fills a significant gap in the extant literature. It also provides a viable framework for the development of clinically trustworthy assistive diagnostic tools, paving the way for more accessible and efficient cardiac care globally.

CONCEPTUALISATION:

To ensure clarity and precision throughout this research, key concepts and terminology integral to the study are defined, and their operationalisation within the methodology is outlined. These definitions are grounded in authoritative sources to provide a robust conceptual framework for the proposed work.

Arrhythmia Classification:

Refers to the automated process of categorizing individual heartbeats within an ECG signal into specific arrhythmia classes (e.g., Normal, Supraventricular Ectopic Beat, Ventricular Ectopic Beat, Fusion Beat, Unknown Beat) as defined by AAMI EC57 standards. This process is central to providing diagnostic support (Gai, 2022).

The primary output of the RDSCNN model will be a class prediction for each detected heartbeat, evaluated against expert-annotated ground truth labels.

ECG Data:

Refers to the digitized, time-series electrical signals recorded from the heart's electrical activity. These signals provide a non-invasive window into cardiac function (Moody & Mark, 2001).

In relation to the study publicly available, expert-annotated ECG datasets (e.g., MIT-BIH Arrhythmia Database, PTB Diagnostic ECG Database) will be utilized as input for model training and evaluation.

Deep Learning (DL):

A subfield of machine learning that utilizes multi-layered artificial neural networks (deep neural networks) to learn complex patterns and representations directly from raw data, bypassing the need for explicit feature engineering (Goodfellow, et al., 2016). DL forms the technological core of the proposed solution.

In relation to the study the RDSCNN model, a type of deep neural network, will be implemented using Python and a deep learning framework (e.g., TensorFlow/Keras).

Recursive Depthwise Separable Convolutional Neural Network (RDSCNN):

A Lightweight CNN architecture combining depthwise separable convolutions and recursion for efficient ECG feature learning (Chollet, 2017).

In relation to the study the architecture will be precisely defined and implemented in code, with specific parameters (e.g., number of layers, kernel sizes, recursion depth) experimentally determined during model development.

Inter-Patient Paradigm:

An evaluation strategy where the training and testing datasets are strictly partitioned such that no data from a patient present in the training set is also present in the testing set (Sajad & Fatemeh, 2019). This paradigm is critical for ensuring the model's ability to generalize to unseen individuals and for reflecting real-world performance.

In relation to the study dataset splitting will be performed at the patient level (not at the beat level), ensuring complete separation of patient data between training, validation, and test sets.

AAMI EC57 Standards:

The guidelines established by the Association for the Advancement of Medical Instrumentation (ANSI/AAMI EC57:1998/(R)2003, 1998–2008.) for performance testing of ambulatory ECG analysers. These standards include precise definitions for five primary arrhythmia beat classes. Adherence to these standards is essential for clinical comparability and validation (Silva, et al., 2025).

In relation to the study all ECG beat annotations in the datasets will be mapped to the AAMI EC57 classes, and model performance metrics will be calculated specifically for these standardized classes.

Generalisability:

The ability of a trained model to maintain its performance accuracy and reliability when applied to new, unseen data from diverse patient populations or clinical environments, beyond the data it was trained on (Yu, et al., 2022). This concept is central to the clinical utility of any AI diagnostic tool.

In relation to the study generalisability will be assessed quantitatively by the model's performance metrics (e.g., accuracy, precision, recall, F1-score) on independent, externally validated test datasets, specifically those partitioned using the inter-patient paradigm.

External Validation:

The process of evaluating a trained model on a dataset that was not used during any part of the model development (training, validation, hyperparameter tuning) and ideally originates from a different source or population (Yu, et al., 2022). This process is vital for confirming a model's real-world robustness.

In relation to the study a specific, publicly available ECG datasets distinct from the training set will be reserved exclusively for final external validation.

Assistive Diagnostic Tool / Decision-Support Tool:

A system designed to provide clinicians with automated analysis and categorizations to aid in their diagnostic process, rather than providing a definitive, unverified diagnosis (Crigger, et al., 2021). The final diagnostic responsibility remains with the human expert, emphasizing human-AI collaboration.

In relation to the study the RDSCNN model will output arrhythmia classifications along with confidence scores, presenting information in a manner intended to complement, not replace, clinical judgment. Its potential utility will be discussed considering its performance and efficiency metrics.

Computational Efficiency:

Refers to the model's resource requirements, including its memory footprint (number of parameters), inference speed (time taken to process a single ECG beat/segment), and overall computational complexity (Cheng, et al., 2020). This is a crucial factor for deploy ability, especially in resource-constrained settings.

Computational efficiency will be measured quantitatively by the number of model parameters, floating-point operations (FLOPs) per inference, and inference time on a standardized hardware configuration.

EC3 Framework (Standards Adherence, Fair Evaluation, Embedded Feasibility):

A structured methodological framework (Silva, et al., 2025) proposed for the comprehensive evaluation and translation of ECG classification systems into clinical practice. This framework guides rigorous and clinically relevant assessment.

This framework will guide the overall methodological design of the study, ensuring rigorous adherence to standards (E1), fair and unbiased model evaluation through techniques like inter-patient validation (E2), and assessment of practical considerations such as computational efficiency and deploy ability (E3). Specific metrics and procedures detailed in the methodology section will align with the principles of the EC3 framework.

RESEARCH METHODOLOGY

RESEARCH DESIGN/APPROACH:

This study will adopt an empirical, quantitative research design aimed at developing and rigorously evaluating a novel deep learning model for ECG arrhythmia classification. The research will utilize a retrospective computational experiment approach, leveraging publicly available, expertly annotated ECG datasets such as the MIT-BIH Arrhythmia Database and PTB Diagnostic ECG Database, which are widely recognized standards in arrhythmia research (Jianwei Zheng, 2022). These datasets will provide diverse, multi-patient ECG recordings, enabling robust training and evaluation of the model across heterogeneous patient populations.

The study's design will be cross-sectional in nature, analysing existing ECG data collected at specific points in time without new data collection. This approach will be appropriate for benchmarking and validating machine learning models in biomedical signal analysis (Silva, et al., 2025). The focus will be on predictive modelling, where the Recursive Depthwise Separable Convolutional Neural Network (RDSCNN) will be trained to classify arrhythmia types directly from raw ECG signals, consistent with the theoretical framework of hierarchical feature learning that underpins deep learning architectures (Yaqoob, et al., 2023).

To ensure clinical relevance and generalizability, the study will implement the Inter-Patient Paradigm for dataset splitting, ensuring no overlap of patient data between training, validation, and test sets (Sajad & Fatemeh, 2019). This paradigm will simulate real-world deployment scenarios by assessing the model's ability to generalize to unseen patients, addressing a key limitation in many existing studies that use intra-patient splits and report overly optimistic performance (Silva, et al., 2025).

Furthermore, the model's performance will be evaluated according to the AAMI EC57 standards, which define standardized arrhythmia beat classes and performance metrics critical for clinical comparability (Silva, et al., 2025). This adherence will ensure that the model's outputs align with clinical diagnostic categories and will facilitate benchmarking against prior work and regulatory expectations.

The research process will include systematic model development, hyperparameter optimization, and rigorous quantitative evaluation. Performance metrics such as accuracy, sensitivity, specificity, F1-score, and computational efficiency (e.g., inference time, parameter count) will be measured on the standardized test sets. These results will then be compared against performance benchmarks reported by state-of-the-art models in the existing literature, particularly those that also utilized rigorous inter-patient validation and adhered to AAMI standards. This comparison will primarily involve a synthesis and critical analysis of reported results from relevant prior studies, acknowledging any methodological differences in their specific implementations or evaluation setups. Additionally, noise reduction and class imbalance handling techniques will be applied during preprocessing and model training to enhance robustness and fairness, reflecting best practices in ECG signal classification research.

This design will support the study's objectives by providing a structured, reproducible framework to develop a computationally efficient, clinically applicable arrhythmia classification model. The retrospective computational experiment approach, combined with standardized validation protocols, will offer a rigorous pathway to bridge the gap between laboratory model performance and real-world clinical utility (Silva, et al., 2025).

POPULATION:

The target population for this study comprises all human cardiac electrical activity, specifically focusing on individuals experiencing various arrhythmias as detectable by standard 12-lead or single-lead ECG recordings. This broad population includes patients across diverse demographics (age, sex, ethnicity), health conditions (e.g., cardiovascular disease, structural heart disease), and geographical locations, all of whom may present with varying ECG signal qualities due to factors such as noise, baseline drift, and amplitude variations. The goal of the developed model is to generalize effectively across this heterogeneous clinical population to accurately classify heartbeats into relevant arrhythmia categories. The "parameters" of this population are therefore the complex and variable morphological and temporal features of ECG waveforms, which the RDSCNN model aims to interpret automatically.

SAMPLING:

Given that this study will utilize existing secondary datasets, the sampling method will involve a form of purposive sampling of publicly available, expertly annotated ECG databases. This approach is chosen to ensure the selection of datasets that are widely recognized, of high quality, and conducive to rigorous machine learning model training and evaluation, particularly concerning arrhythmia classification. The primary datasets to be sampled will include:

- **MIT-BIH Arrhythmia Database:** This database is a widely accepted standard for arrhythmia research, providing two-channel ambulatory ECG recordings from 47 subjects, totalling over 100,000 annotated heartbeats. It offers a diverse range of arrhythmias and noise levels.
- **PTB Diagnostic ECG Database:** This database contains 15-lead ECG recordings from 290 subjects, offering a broader spectrum of cardiac conditions beyond arrhythmias, which can contribute to the model's generalizability in differentiating normal from pathological rhythms.

The sample size for model training and evaluation will be determined by the entire collective data available from these selected databases. This comprehensive inclusion of data from chosen datasets ensures maximal diversity in patient demographics, ECG characteristics, and arrhythmia patterns, enhancing the representativeness of the sample to the broader clinical population.
The justification for this sampling method is threefold:

1. **Clinical Relevance:** The selected databases are widely used in medical research and provide real-world ECG data with expert annotations, crucial for developing clinically relevant models.
2. **Generalizability:** Utilizing multiple diverse datasets will enhance the model's ability to generalize across different recording conditions and patient variability, directly addressing a key limitation highlighted in the literature.
3. **Reproducibility and Comparability:** Using publicly accessible and standardized databases ensures the reproducibility of the study's findings and facilitates direct comparison with existing research that has utilized the same benchmarks.

Crucially, within these sampled datasets, strict inter-patient partitioning will be implemented during data splitting for training, validation, and testing. This ensures that no data from a single patient will be present in more than one split (e.g., a patient's data used for training will not appear in the test set). This methodological "sampling" approach rigorously simulates real-world clinical deployment by testing the model's ability to classify arrhythmias from completely unseen patients, providing a more reliable assessment of its generalizability than intra-patient splitting methods.

DATA COLLECTION METHOD:

As this study employs a retrospective computational experiment design, the data collection method will involve accessing and downloading publicly available, expertly annotated secondary ECG datasets. There won't be any primary data collection involving human participants. The primary data sources will be the MIT-BIH Arrhythmia Database and the PTB Diagnostic ECG Database, primarily accessed via the Kaggle platform. These datasets are widely recognized and provide well-documented recordings essential for arrhythmia research. The data will be acquired in their native formats and subsequently processed using Python-based scripts. The research "instruments" in this context are the carefully curated and annotated digital ECG databases themselves, along with the computational environment (hardware and software) that will process and analyse this data. Specifically, Python libraries such as wfdb for signal loading and NumPy, SciPy, and Pandas for data manipulation and preprocessing will be utilized. Detailed descriptions of how these datasets are accessed and processed will be provided to ensure transparency and reproducibility.

DATA ANALYSIS METHOD:

The data analysis in this study will primarily involve computational analysis using deep learning techniques, meticulously applied to the prepared ECG datasets. The analysis will be conducted in a structured, multi-stage process to address the research questions and objectives:

1. **Data Preprocessing:**
    - Raw ECG signals from the selected datasets will undergo standard signal processing techniques. This will include noise reduction (e.g., baseline wander removal, powerline interference filtering) and normalisation to ensure consistent signal characteristics.
    - Individual heartbeats will be automatically detected and segmented from the continuous ECG recordings based on R-peak detection.
    - Beat annotations from the datasets will be mapped to the five standardized AAMI EC57 classes (N, S, V, F, Q) to ensure consistent classification targets and clinical comparability.
    - Techniques to address class imbalance (e.g., oversampling minority classes, under sampling majority classes, or using weighted loss functions during training) will be applied to prevent bias towards common arrhythmia types and enhance the model's fairness and robustness for rarer but clinically critical events.

2. **Model Training and Optimization:**
    - The pre-processed and segmented heartbeat data will be split into training, validation, and test sets following the strict Inter-Patient Paradigm. This ensures no patient data overlaps between these sets, allowing for a robust assessment of generalizability.
    - The Recursive Depthwise Separable Convolutional Neural Network (RDSCNN) will be implemented using a deep learning framework (e.g., TensorFlow/Keras in Python).
    - The model will be trained on the training set, with hyperparameter optimization (e.g., learning rate, batch size, number of recursive layers, kernel sizes) performed using the validation set to achieve optimal performance and computational efficiency.
3. **Quantitative Performance Evaluation:**
    - The trained RDSCNN model's performance will be rigorously evaluated on the independent, unseen test sets derived from the selected datasets, and potentially on an entirely separate external validation dataset.
    - Performance metrics will be calculated for each of the AAMI EC57 classes, as well as macro-averaged or weighted averages, to provide a comprehensive view of the model's classification accuracy. Key metrics will include:
        - **Accuracy:** Overall proportion of correctly classified beats.
        - **Precision (Positive Predictive Value):** Proportion of true positive predictions among all positive predictions.
        - **Recall (Sensitivity):** Proportion of true positive predictions among all actual positive instances.
        - **Specificity:** Proportion of true negative predictions among all actual negative instances.
        - **F1-score:** The harmonic mean of precision and recall, providing a balanced measure of performance.
    - **Computational Efficiency Metrics:** To assess suitability for resource-constrained settings, the following will be measured:
        - **Number of trainable parameters:** Indicating model size and memory footprint.
        - **Inference time:** The time required to classify a single ECG heartbeat or sequence, assessing real-time capability.
        - **Floating Point Operations (FLOPs):** Estimating computational complexity.

4. **Comparative Analysis and Benchmarking:**
    - The RDSCNN's performance and efficiency metrics will be compared against performance benchmarks reported by state-of-the-art models in the existing literature. This comparison will prioritize studies that have also utilized rigorous inter-patient validation and adhered to AAMI standards.
    - This benchmarking will primarily involve a synthesis and critical analysis of reported results from relevant prior studies. Any methodological differences in their specific implementations or evaluation setups will be acknowledged and discussed to contextualize the comparisons. This will directly address the primary research question regarding the RDSCNN's performance compared to existing models.

VALIDITY, RELIABILITY, OR TRUSTWORTHINESS:

As this is a quantitative study focused on the development and evaluation of a machine learning model, the emphasis will be on enhancing the validity and reliability of the findings. These measures will ensure the robustness and trustworthiness of the research outcomes.

To enhance the validity of the study:

- **Internal Validity:** Measures will be taken to minimize confounding factors and ensure that observed performance accurately reflects the model's capabilities. This will be achieved through the use of standardized, expertly annotated public datasets (MIT-BIH, PTB Diagnostic ECG Database), which reduce variability in ground truth. Crucially, the Inter-Patient Paradigm will be strictly implemented for dataset splitting, preventing data leakage and ensuring the model learns generalizable patterns rather than memorizing patient-specific features. Adherence to AAMI EC57 standards for arrhythmia classification will further standardize the output categories, aligning them with widely accepted clinical definitions and enhancing the interpretability of results.
- **External Validity (Generalizability):** The ability to generalize findings to unseen populations will be maximized by utilizing multiple and diverse ECG datasets. These datasets encompass varying patient demographics, recording conditions, and noise levels. The rigorous Inter-Patient Paradigm directly contributes to assessing generalizability to novel patients. Furthermore, the focus on computational efficiency in the RDSCNN design will enhance its practical deployability across various clinical settings, thus supporting its real-world generalizability.

- **Construct Validity:** The constructs of "arrhythmia classes" will be measured consistently with clinical definitions through strict adherence to AAMI EC57 standards. The chosen performance metrics (accuracy, F1-score, sensitivity, specificity) are also standard, well-defined measures widely accepted in machine learning and medical AI evaluation.

To enhance the reliability of the study:

- **Reproducibility:** The entire research methodology will be meticulously documented. This will include detailed descriptions of data preprocessing steps, the exact RDSCNN model architecture, hyperparameter optimization strategies, training procedures, and evaluation protocols. All code will be developed using open-source programming languages (Python) and widely used deep learning frameworks (e.g., TensorFlow/Keras). The use of publicly available datasets will allow other researchers to replicate the experiments. Additionally, random seeds will be fixed during data splitting and model initialization to ensure that experimental results are consistent across multiple runs.
- **Consistency:** The model's ability to produce consistent performance metrics under identical input conditions will be implicitly evaluated through the repeated application of the rigorous testing protocol on the designated test sets. The stability of results across different folds of the inter-patient validation will also contribute to demonstrating consistency.

CONTRIBUTION, ETHICAL CONSIDERATIONS AND LIMITATIONS

PROPOSED CONTRIBUTION:

This study is poised to make three significant contributions to the fields of Data Science, Artificial Intelligence (AI), and ultimately, cardiovascular diagnostics. The anticipated contributions are directly aligned with the study's core purpose: to bridge the gap between laboratory accuracy and real-world clinical performance in automated ECG arrhythmia classification by developing a robust, generalizable, and computationally efficient assistive diagnostic tool.

The most important contribution of this study lies in providing a rigorous and reproducible framework for evaluating deep learning models for ECG arrhythmia classification that explicitly addresses critical limitations in existing literature. By meticulously adhering to AAMI EC57 standards and implementing a strict Inter-Patient Paradigm for dataset splitting, this research will move beyond inflated laboratory-centric performance metrics. It will instead offer a more reliable and clinically meaningful assessment of model generalizability. This approach will provide a blueprint for future research aiming for true clinical translation, thereby advancing methodological best practices in biomedical AI evaluation.

Furthermore, this study will contribute by developing and evaluating a lightweight Recursive Depthwise Separable Convolutional Neural Network (RDSCNN) model specifically designed for computational efficiency without sacrificing accuracy. This addresses a crucial gap in the literature regarding deployability in resource-constrained healthcare environments. By demonstrating that a model with optimized architecture can achieve robust performance under stringent validation, this research will offer practical insights into designing AI solutions that are not only effective but also accessible and scalable across diverse clinical settings globally. This contributes to the practical implementation of AI in healthcare, moving from theoretical possibility to tangible utility.

Finally, by positioning the RDSCNN model as an assistive diagnostic tool rather than a fully autonomous system, the study contributes to the evolving understanding of Human-AI Collaboration in clinical decision-making. It will provide evidence of how a computationally efficient and rigorously validated AI can augment, rather than replace, clinician expertise. This aligns with prevailing clinical guidelines and liability considerations, fostering greater trust and facilitating the integration of AI into complex diagnostic workflows. This contribution supports the development of responsible AI in medicine, ensuring that technology serves to empower clinicians and improve patient outcomes without diminishing human oversight.

This research will contribute to the current body of knowledge by validating a novel model under real-world-simulating conditions, providing a practical framework for generalizable and efficient arrhythmia classification, and illustrating a viable pathway for the responsible integration of AI as a decision-support tool in healthcare.

ETHICAL CONSIDERATIONS:

This study involves the use of publicly available, de-identified ECG datasets and does not involve direct interaction with human participants. Nevertheless, six important ethical considerations have been identified and will be addressed to uphold ethical research principles.

Firstly, ethical clearance will be obtained from the university's campus ethics committee prior to commencing data analysis, ensuring that the study complies with institutional and national ethical standards. Although this study does not involve direct human participation, formal approval is necessary because it uses human-derived data.

Secondly, the study will strictly adhere to data privacy and confidentiality principles as outlined in the Protection of Personal Information Act (POPIA) of South Africa. The datasets used are anonymized and publicly accessible, minimizing risks related to participant identification or privacy breaches. All data handling and storage will comply with relevant data protection regulations to safeguard sensitive information.

Thirdly, the study acknowledges the emerging ethical challenges surrounding liability and accountability in AI-driven medical devices. While the RDSCNN model is intended as a decision-support tool rather than an autonomous diagnostic system, clear communication about its assistive role will be emphasized to avoid overreliance or misinterpretation by clinicians. This aligns with current ethical guidelines advocating human oversight in AI-assisted healthcare.

Fourthly, ethical approval for dataset use will be confirmed by verifying that the original data collection complied with ethical standards and participant consent. This ensures respect for the rights of individuals whose data contribute to the research.

Fifthly, the study will maintain transparency and reproducibility by documenting all methodological steps and sharing code where possible, fostering trust and accountability in AI research.

Lastly, the study will consider the ethical implications of potential biases in AI models, including demographic or clinical biases that could affect fairness or equity in diagnostic support. Where possible, performance across diverse patient subgroups will be evaluated, and limitations will be openly discussed.

Although the study does not involve new data collection or direct participant interaction, these ethical considerations are vital to ensure responsible research conduct and respect for human dignity.

LIMITATIONS:

This study acknowledges five methodological limitations inherent to its design. Firstly, the use of publicly available, expertly annotated ECG datasets (e.g., MIT-BIH, PTB) ensures reproducibility but represents a fixed sample of the broader patient population. While these datasets are diverse, they may not fully capture rare arrhythmia presentations, global demographic variability, or real-world signal noise encountered in clinical practice. To mitigate this, the study will apply the Inter-Patient Paradigm and prioritize datasets with documented diversity.

Secondly, the retrospective computational design and time constraints limit real-time clinical deployment and direct user feedback from clinicians. This restricts the assessment of practical workflow integration or long-term clinical impact, though theoretical frameworks (e.g., Human-AI Collaboration Theory) will contextualize potential adoption.

Thirdly, computational resource limitations may constrain exhaustive hyperparameter optimization or architectural exploration for the RDSCNN. To address this, optimization will focus on established lightweight techniques, and results will be contextualized within these constraints.

Fourthly, the study's scope is confined to ECG-based arrhythmia classification, excluding other diagnostic modalities (e.g., echocardiography) or broader cardiovascular conditions. Findings should be interpreted strictly within ECG contexts.

Finally, the retrospective reliance on pre-existing datasets may not fully replicate real-world clinical environments, such as heterogeneous noise levels or equipment differences. While noise reduction techniques will be applied, this limitation underscores the need for prospective validation in future work.

Bibliography

Sajad, M. & Fatemeh, A., 2019. *Inter- and Intra- Patient ECG Heartbeat Classification for Arrhythmia Detection: A Sequence to Sequence Deep Learning Approach.* [Online]

Available at: https://ieeexplore.ieee.org/abstract/document/8683140/references#references

[Accessed 8 June 2025].

Taymaz, A., Sait, A. & Mohammad, A. N. B., 2023. *ECGformer: Leveraging Transformer for ECG Heartbeat Arrhythmia Classificatio.* [Online]

Available at: https://ieeexplore.ieee.org/abstract/document/10590116

[Accessed 8 June 2025].

Yaqoob, A., Omar, M., Khalid, Q. & Erchin, S., 2023. *Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023.* [Online]

Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10542398/

[Accessed 8 June 2025].

WHO, 2021. *Cardiovascular diseases (CVDs).* [Online]

Available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[Accessed 9 June 2025].

Robinson, S., 2021. *Priorities for health promotion and public health.* [Online]

Available at:

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=S.+Robinson%2C+"Cardiovascular+disease%2C"+Priorities+for+Health+Promotion+and+Public+Health%2C&btnG=

[Accessed 9 June 2025].

Peek, S. F. & Buczinski, S., 2018. *Cardiovascular Diseases.* [Online]

Available at: https://doi.org/10.1016/B978-0-323-39055-2.00003-6

[Accessed 10 June 2025].

Wang, J. et al., 2021. *Automated ECG classification using a non-local convolutional block attention module.* [Online]

Available at: https://doi.org/10.1016/j.cmpb.2021.106006

[Accessed 9 June 2025].

Luz, E. J. d. S., Schwartz, W. R., Cámara-Chávez, G. & Menotti, D., 2016. *ECG-based heartbeat classification for arrhythmia detection: A survey.* [Online]

Available at: https://doi.org/10.1016/j.cmpb.2015.12.008

[Accessed 9 June 2025].

Luo, K., Li, J., Wang, Z. & Cuschieri, A., 2017. *Patient-Specific Deep Architectural Model for ECG Classification.* [Online]

Available at: https://doi.org/10.1155/2017/4108720
[Accessed 10 June 2025].
Yildirim, Ö., 2018. *A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification.* [Online]
Available at: https://doi.org/10.1016/j.compbiomed.2018.03.016
[Accessed 10 June 2025].
Midani, W., Ouarda, W. & Ayed, M. B., 2023. *DeepArr: An investigative tool for arrhythmia detection using a contextual deep neural network from electrocardiograms (ECG) signals.* [Online]
Available at: https://doi.org/10.1016/j.bspc.2023.104954
[Accessed 10 June 2025].
ANSI/AAMI EC57:1998/(R)2003, 1998–2008.. *Testing And Reporting Performance Results Of Cardiac Rhythm And ST Segment Measurement Algorithms.* s.l.: AAMI.
Price II WN, G. S. C. I., 2024. Liability for use of artificial intelligence in medicine. In: B. S. a. I. G. Cohen., ed. *Research Handbook on Health, AI and the Law.* Cheltenham, UK: Edward Elgar Publishing Ltd, p. Chapter 9.
Crigger, E. et al., 2021. *Trustworthy Augmented Intelligence in Health Care.* [Online]
Available at: https://doi.org/10.1007/s10916-021-01790-z
[Accessed 11 June 2025].
Chollet, F., 2017. *Xception: Deep Learning with Depthwise Separable Convolutions.* [Online]
Available at: https://doi.org/10.48550/arXiv.1610.02357
[Accessed 11 June 2025].
Zhao, S., Song, J. & Ermon, S., 2017. *Learning Hierarchical Features from Generative Models.* [Online]
Available at: https://doi.org/10.48550/arXiv.1702.08396
[Accessed 11 June 2025].
Ihsanto, E., Ramli, K., Sudiana, D. & Gunawan, T. S., 2020. *Fast and Accurate Algorithm for ECG Authentication Using Residual Depthwise Separable Convolutional Neural Networks.* [Online]
Available at: https://doi.org/10.3390/app10093304
[Accessed 12 June 2025].
Aziz, S., Ahmed, S. & Alouini, M.-S., 2021. *ECG-based machine-learning algorithms for heartbeat classification.* [Online]
Available at: https://doi.org/10.1038/s41598-021-97118-5
[Accessed 12 June 2025].
Ramli, I. E., Sudiana, K. & Teddy, D. G., 2020. *Fast and Accurate Algorithm for ECG Authentication Using Residual Depthwise Separable Convolutional Neural Networks.* [Online]
Available at:
https://www.researchgate.net/publication/341300354_Fast_and_Accurate_Algorithm_for_ECG_Aut

hentication_Using_Residual_Depthwise_Separable_Convolutional_Neural_Networks
[Accessed 11 June 2025].
Chazal, P. d., O'Dwyer, M. & Reilly, R., 2004. *Automatic classification of heartbeats using ECG morphology and heartbeat interval features.* [Online]
Available at:
https://www.researchgate.net/publication/8459885_Automatic_Classification_of_Heartbeats_Using_ECG_Morphology_and_Heartbeat_Interval_Features
[Accessed 12 June 2025].
Yang, M., Liu, W. & Zhang, H., 2022. *A robust multiple heartbeats classification with weight-based loss based on convolutional neural network and bidirectional long short-term memory.* [Online]
Available at: https://doi.org/10.3389/fphys.2022.982537
[Accessed 12 June 2025].
Akan, T., Alp, S. & Bhuiya, M. A. N., 2023. *ECGformer: Leveraging Transformer for ECG Heartbeat Arrhythmia Classification.* [Online]
Available at: https://doi.org/10.48550/arXiv.2401.05434
[Accessed 12 June 2025].
Silva, G. et al., 2025. *A SYSTEMATIC REVIEW OF ECG ARRHYTHMIA CLASSIFICATION: ADHERENCE TO STANDARDS, FAIR EVALUATION, AND EMBEDDED FEASIBILITY.* [Online]
Available at: https://arxiv.org/abs/2503.07276
[Accessed 10 June 2025].
Xiao, Q. et al., 2023. *Deep Learning-Based ECG Arrhythmia Classification:.* [Online]
Available at: https://doi.org/10.3390/app13084964
[Accessed 12 June 2025].
Gai, N. D., 2022. *ECG beat classification using machine learning and pre-trained convolutional neural networks.* [Online]
Available at: https://doi.org/10.48550/arXiv.2207.06408
[Accessed 12 June 2025].
Moody, G. & Mark, R., 2001. *The impact of the MIT-BIH Arrhythmia Database.* [Online]
Available at: https://doi.org/10.1109/51.932724
[Accessed 12 June 2025].
Goodfellow, I., Bengio, Y., Courvill, A. & Bengio, Y., 2016. *Deep learning. Vol. 1, no. 2..* Cambridge: MIT press.
Yu, A. C., Mohajer, B. & Eng, J., 2022. *External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review.* [Online]
Available at: https://doi.org/10.1148/ryai.210064
[Accessed 12 June 2025].

Cheng, Y., Wang, D., Zhou, P. & Zhang, T., 2020. *A Survey of Model Compression and Acceleration for Deep Neural Networks.* [Online]

Available at: https://doi.org/10.48550/arXiv.1710.09282

[Accessed 12 June 2025].

Jianwei Zheng, H. G. ,. H. C., 2022. *A large scale 12-lead electrocardiogram database for arrhythmia study.* [Online]

Available at: https://doi.org/10.13026/wgex-er52

[Accessed 12 June 2025].