INTELLIGENT CLINICAL DECISION SUPPORT FOR HEART DISEASE DIAGNOSIS USING MACHINE LEARNING IN NIGERIAN HOSPITALS

PRAISES ABEJIDE
Department of Computer Science, Babcock University, Ilishan-Remo, Ogun State, Nigeria,
abejide0766@pg.babcock.edu.ng

Abstract
There is a cardiac crisis in Nigeria - less than 100 cardiologists for a population of greater than 220 million and more than 200 000 out of every 300 000 people in the country die from cardiovascular disease. Although progress has been made in cardiovascular AI research in the past decade, there is currently no machine learning model that can diagnose heart disease from data of Nigerian patients. This paper outlines an ML-CDSS that is specifically targeted to the Nigerian tertiary care hospitals. The CDSS has six main structured clinical features and raw 12-lead electrocardiogram (ECG) waveforms and adopts a multi-modal stacked ensemble approach with XGBoost, feedforward neural network, as well as a CNN-LSTM deep learning sub-model. In addition to each prediction, TreeSHAP provides an explanation layer for the tree-based approach, which is easy to understand, and DeepSHAP yields an explanation layer for the neural network; LIME is used to pass on secondary explanations, and GradCAM to visualise the activations for the ECG waveform. The process is done through five phases, HEART (Harvest, Engineer, Architect, Refine, and Transform) including prospective data collection from three Nigerian federal tertiary care hospitals, pre-processing, modelling building, validation with a holdout set blinded data, and deployment. Architecture designed to run intermittent connections, commodity compute, energy constrained. Since the inception of this project, there have been implicit collaboration between governance and NAFDAC Software to develop a Medical Device guide and Nigeria Data Protection Regulation guidelines. This paper tackles seven known gaps within the cardiac AI literature in Africa head-on.
Keywords: *Machine learning, heart disease diagnosis, Clinical decision support, Nigeria, Explainable AI, SHAP, GradCAM, Electrocardiogram, LMIC*

The Cardiac Crisis in Sub-Saharan Africa

Cardiovascular disease (CVD) is the worldwide biggest killer and about one-third of deaths around the world every year are due to this type of ailment [1,2]. There is a disproportionate distribution: about 90 per cent of the global cardiology research and clinical investments is concentrated in low and middle income countries (LMICs) and this disproportion has led to a disproportionate share of cardiovascular deaths (77 per cent of all CVD deaths are currently in LMICs) [3, 4].

In sub-Saharan Africa, the number of excess deaths due to heart disease is more than one million annually and this is growing as the region transitions to the epidemic phase of cardiovascular diseases related to urbanisation, shifting diets and sedentary lifestyles [5]. This research paper provides the summary epidemiological data in the following Table 1 and Figure 1. The malnutrition of the heart situation in Nigeria is worse. However, from a study of aetiology of heart failure, Ogah and colleagues found that aetiology of heart failure is very different from that of western registries with ischaemic heart failure accounting for only 8 per cent of admissions and hypertensive cardiac failure being the top cause of heart failure with 54.5 per cent of admissions [6]. However, Nigerian cardiac disease is more distinguishable and is associated with four unique features: late onset severe hypertension with end organ damage; peripartum cardiomyopathy in the young female population; rheumatic valvular disease; and endomyocardial fibrosis, a clinical picture that is much less seen in datasets that geographies current AI models of cardiac disease are trained on.

Cardiac disease in Nigeria is unique with four characteristics that are not observed in most, if not all, of the current datasets on which AI cardiac models were trained: late onset severe hypertension with end organ damage, peripartum cardiomyopathy in the young female population, rheumatic valvular disease, and endomyocardial fibrosis.

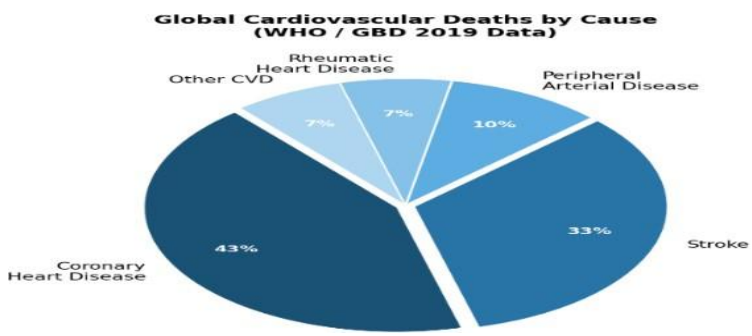
A national study of the work force conducted by the Heart failure Society of Nigeria, reveals less than one hundred cardiologists distributed in Nigeria (36 states and FCT) which translates to an average of one per 2.2 million Nigerian people [8]. It is well documented what the consequences for the quality of the diagnosis are. Echocardiography is limited to large tertiary hospitals and a high-sensitivity troponin measurement is not consistently accepted and used throughout the hospital system. Akintunde and co-workers have found only 54 per cent of common ECG abnormalities had been correctly interpreted by junior doctors in Nigerian teaching hospitals [44]. The mean time from symptom presentation to diagnosis of heart failure at the Lagos University Teaching Hospital (LUTH) was 4.2 days, significantly exceeding the international guideline of less than 24 hours [16] which Sani and colleagues found. These represent individual patients who presented cardiac crises and waited days to make a diagnosis that could prevent further deterioration after a cardiac crisis.

Table 1: Key global cardiovascular epidemiological indicators.

/N	Indicator	Estimate	Source
1	Annual global cardiovascular death	17.9 million	WHO, 2021 [1]
2	Ischaemic heart disease deaths per year	9.14 million	Roth et al., JACC 2020 [3]
3	CVD deaths occurring in LMICs	77%	Roth et al., NEJM 2015 [4]
4	Premature CVD deaths (under 70 years)	~38%	WHO, 2021 [1]
5	Sub-Saharan Africa annual CVD deaths	>1.0 million	Ntusi & Mayosi [5]
6	Estimated annual cardiac deaths in Nigeria	~200,000	Ogah et al. [6]
7	Proportion of CVD deaths preventable	>80%	WHO, 2021 [1]

Source: WHO [1]; Roth et al., JACC [3]; Roth et al., NEJM [4]; Ntusi & Mayosi [5]; Ogah et al. [6].

Figure 1: Annual cardiovascular disease mortality by world region. Sub-Saharan Africa carries a rapidly growing cardiac burden with fewer than 100 practising cardiologists in Nigeria. Source: GBD Study 2019 [3]; WHO, 2021 [1].



An overview of cardiac AI's promise and its limits

Machine learning has had some outstanding clinical results in the area of cardiology. Hannun and co. were able to train a deep neural network to correctly classify twelve cardiac rhythms on 91,232 ambulatory electrocardiograms and shown that it was as accurate as a cardiologist in doing so [10]. Using a standard 12-lead ECG – the investigation most widely available across all levels of the Nigerian health system – an AI model has been shown to be able to identify left ventricular dysfunction with 86.3 per cent sensitivity and 85.7 per cent specificity, without any echocardiogram [11]. If all key data used for analysis were identical to the data with which the algorithm was used, then deep learning had matched the performance of a specialist on some diagnostic tasks, he wrote in a key analysis to be published in a paper in the journal Nature Medicine by Topol.

The last clause is the one which has the most weight with it. The main challenge investigated in this study is called domain shift, which is defined by the fact that the performance of the model diminishes when used in a new domain different from the one it was trained in. Wiens and colleagues made it a point to express this responsibility to clinical AI as 'models should be tested within the populations they are intended for' [27]. The most obvious example of what they do when they're not data was when Obermeyer et al. discovered that they have a US algorithm that systematically underestimates the health needs of People of Color relative to whites, when they have the same observed level of illness (in this study, based on healthcare spending — which is, yet again, differential access to care) [40]. A prev aleunt assumption (in the form of test weights to parameters of ischaemic disease prevalence) would not be valid for Lagos or Ibadan, tests where the latter had not been performed in the United.

Aim and Objectives

The goal is to create, build and verify a machine learning clinical decision support program for the diagnosis of heart diseases in Nigerian tertiary health centres with a built-in explainability module and a deployment plan that is appropriate for the specific Nigerian environment.

The gaps in research identified and addressed are:

1. Lack of a Nigeria-primary training dataset
2. Lack of multi-class diagnostic models
3. Lack of integration of ECG waveforms in any published cardiac AI study in LMIC
4. Lack of explainability at per-prediction level
5. Lack of prospective external validation
6. Lack of regulatory alignment, and
7. Lack of LMIC-appropriate deployment architecture.

The gaps and corresponding solutions of this study are mentioned in Table 3

Literature Review

Machine Learning Application in Clinical practice

Students will learn about the application of machine learning in the clinical environment. Machine learning algorithms comprise algorithms that learn patterns from data without being specifically programmed. There are three paradigms that have been most clinically applicable. In the cardiovascular literature, supervised learning (training on labelled records with known outcomes) is the dominant paradigm, and is the main learning paradigm used in this study.

Latent structures in unsupervised learning are used for discovering structures in unlabeled data useful for patient phenotyping. When the sample of labelled data is small and the sample of unlabelled data is large, a combination of the two (termed “semi-supervised learning”) can be applied, as may be the case in Nigerian hospitals, where it is only possible to obtain labels for a small proportion of the data [34]. In cardiology, the most clinically relevant advances have been achieved with deep learning, a type of artificial intelligence that relies on neural networks with several hidden layers. LeCun et al., 2018 [35] was the foundation of convolutional and recurrent architecture, which has formed the basis of imaging the heart and analysing ECG recordings.

Cardiac Treatment using Machine Learning

The classical supervised algorithms are the most clinically established ones. Logistic regression is also a good starting point: interpretable coefficients are important in the clinical setting, as is calibrated probability estimates, such as those used in the Framingham Risk Score for the prediction of coronary heart disease [12]. In 71 studies of cardiovascular risk prediction, logistic regression models performed as well as much more complicated models in terms of median AUC [48]. For UCI Cleveland multi-class dataset [49] the AUC obtained with SVM with RBF kernel is 0.85 - 0.92.

Random forests, a method proposed by Breiman [50] have robustness to data missingness and native feature importance estimation. In the area of tabular clinical data, Gradient boosting is the de facto standard approach, and XGBoost, developed by Chen and Guestrin [13] and LightGBM, developed by Ke et al [14] are both more suitable for large-scale data and have mean AUC values of 0.87 – 0.94 across six cardiac benchmark datasets. The most popular recurrent architecture for sequential ECG data is the Long Short-Term Memory network (LSTM) [15] which has shown an improved AUC of >3 percentage points over CNN-only baselines in arrhythmia benchmarks [58]. In Table 2 and Figure 2 the results of the key studies are summarised.

Table 2: Comparative ML algorithm performance on benchmark cardiac datasets

S/N	Algorithm	Dataset	AUC-ROC	Accuracy %	Reference
1	XGBoost	UCI Cleveland	0.92	89.4	[13, 53]
2	LightGBM	UCI Cleveland	0.91	88.7	[14]
3	Random Forest	UCI Cleveland	0.89	86.9	[49, 50]
4	SVM (RBF kernel)	UCI Cleveland	0.87	84.5	[49]
5	Logistic Regression	Multiple	0.77	75.2	[48]
6	Proposed Ensemble (target)	Nigerian + UCI	>= 0.90	>= 88.0	This study

Note: target row reflects this study's validation objective on the Nigerian primary dataset. Sources: Awwalu et al. [53], Ke et al. [14], Mohan et al. [49].

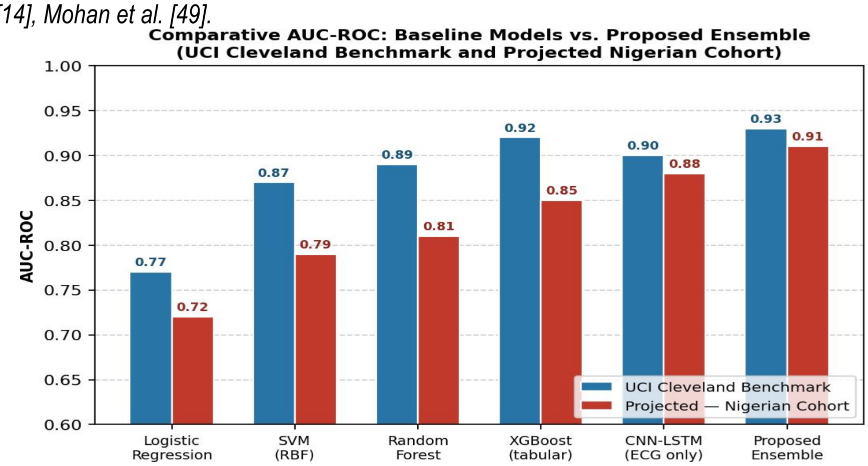


Figure 2: Comparative AUC-ROC of ML algorithms on the UCI Cleveland benchmark versus projected performance on the Nigerian cohort. The proposed stacking ensemble targets AUC >= 0.90 on locally collected Nigerian patient data. Sources: Awwalu et al. [53], Ke et al. [14], Mohan et al. [49].

Explainable AI and Trust Among Medical Personnel

The use of black-box models in clinical scenarios has issues that can't be addressed with prediction accuracy. If a algorithm states a patient is high-risk and it does not provide a rationale, the physician does not have anything to go on which to evaluate, challenge or justify taking action on this label [22]. Amann et al. performed a mixed-methods study, and they discovered that 68 per cent of cardiologists would not follow an AI recommendation, and that trust "significantly" went up with the addition of feature attribution and uncertainty quantification [64].

This study includes three complementary explainability methods as features: SHAP is a class of explanation framework presented by [65] based on the framework of cooperative game theory, and sufficient to meet the mathematical condition of efficiency, symmetry, and consistency, where each feature can be attributed with an SHAP value. In a prospective study in the intensive care unit (ICU), the SHAP explanations actually outperformed the control (a prediction-only model) by 17 per cent in terms of the physician's accuracy in the decision [66]. LIME provides an interpretable model for every prediction in its local neighborhood and is a second solution [67] using expert configurations of training runs, once for 50 random seeds for each prediction [68]. Using GradCAM, they found that there are some parts of the temporal patterns in the ECG waves that are significant for the classification of each CNN, specifically: the P wave, QRS complex, ST segment and T wave. These were confirmed with the annotations performed by expert electrophysiologists, resulting in between 78–92 per cent agreement [69, 70]. 2,4% of the total population in Nigeria is covered by basic health care. For machine learning algorithm to be developed for Nigerian hospital, the capabilities of Nigerian hospital needs to be considered. ECG available at all three levels of Health Systems. Heart Failure Society of Nigeria reported fewer than 20 machines for echocardiogram in government hospitals of the country outside the major Federal Tertiary hospitals with more than 40% downtime of the machines during working hours [8]. High-sensitivity troponin was not available in 22/22 hospitals surveyed in Lagos, except at three hospital centers.

In 2023, a World Bank assessment of 150 public health facilities in Nigeria revealed that just 34 per cent would have a functioning hospital information management system, 18 per cent a reliable internet connection and 11 per cent uninterrupted electricity [84]. Akinyemi et al. noted some of the systematical challenges to the adoption of EHRs as fragility of infrastructure, insufficient training and governance issues [85]. I believe this was not something that you're required to put a lot of effort into on the back of your wife, it's just the limitations of design. This is because, whatever name it may go by, any model that requires ongoing connectivity with a computer or other special hardware or working EHR is not for Nigerian hospitals. The Nigerian heart disease distribution is presented in Figure 3 and Table 3 will present the seven research gaps and solutions.

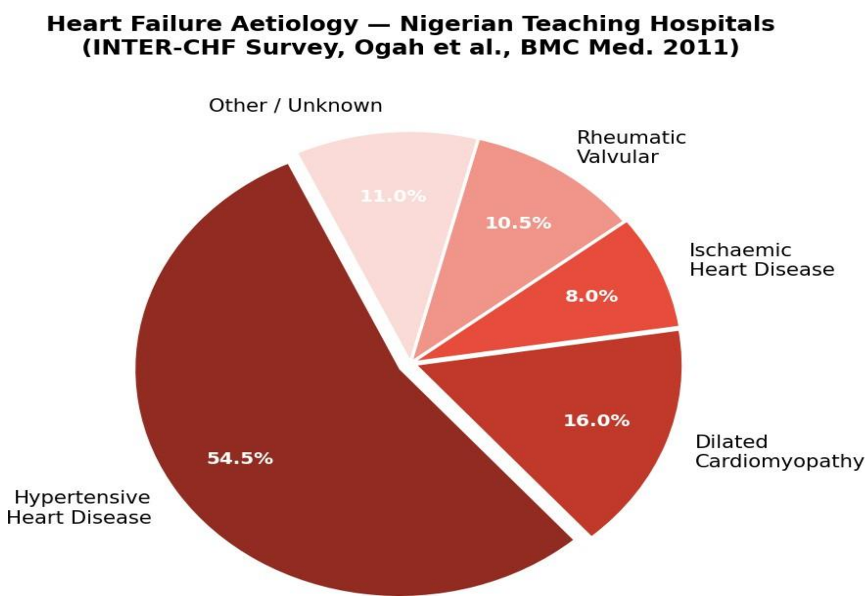


Figure 3: Heart failure aetiology at Nigerian tertiary hospitals (INTER-CHF survey). Hypertensive heart disease accounts for 54.5% of cases, contrasting with the ischaemic-dominated Western registry picture. Adapted from Ogah et al. [6] and Falase & Akinkugbe [7].

Review Existing African cardiac AI studies and Identified Gaps

Medline, Scopus, IEEE Xplore and Google Scholar were used for a systematic search resulting in the finding of only seven peer-reviewed publications that describe a ML model that applies to any patient data from Africa in relation to a cardiovascular disease. No one used a training set of Nigeria-primarily built. No ones found any raw ECG waves. No one had an explainability feature that is per-prediction. No external validation based on a future cohort was done. There were no development methods matching a country development pathway. No one proposed how to design the architecture taking into mind the constraints of the infrastructure of African hospitals. Current models are not safe to be deployed in tertiary hospitals in Nigeria for the reason that these are not penile orifices. Table 3 displays a map of the gaps and the response in this study.

Table 3: *Seven research gaps in African cardiac AI and their resolution in this study.*

S/N	Identified Research Gap	Resolution in This Study
1	No ML model trained on Nigerian primary cardiac data	Primary dataset >= 2,000 records from 3 Nigerian federal teaching hospitals
2	No multi-class diagnostic model for Nigeria	Five-class: hypertensive HD, HF, IHD, cardiomyopathy, arrhythmia
3	No ECG waveform integration in African cardiac AI	CNN-LSTM sub-model processing raw 12-lead ECG waveforms
4	No per-prediction explainability in any African study	SHAP + LIME + GradCAM per-prediction report for every patient
5	No prospective external validation	Sealed prospective Nigerian hold-out cohort (15% of primary data)
6	No regulatory alignment with NAFDAC or NDPR	Co-design with NAFDAC SaMD pathway; NDPR data governance throughout
7	No LMIC-appropriate deployment architecture	Offline edge system: 8 GB RAM, 3G-compatible model updates

The Heart Methodology

Modeling the design of the research philosophy

This research is based on the positivist paradigm of empirical computer science, which assumes the use of falsifiable hypothesis, numerical estimates (both confidence intervals and reproducibility of methodology) [91]. It uses the Design Science Research methodology by Hevner et al that requires that a system not only run in the experimental contexts studied, but also in the real context of its application environments [92]. Those constraints related to hospital infrastructure, disease prevalence in Nigeria and clinical workflow in Nigeria are thus, from the outset, design constraints and not adaptations to pre-existing factors. The research will be carried out in the context of five phases in the framework of HEART, namely Harvest, Engineer, Architect, Refine, and Transform. There are input and output(s) of each phase and all have their quality gates. The overall structure is indicated in figure 4.

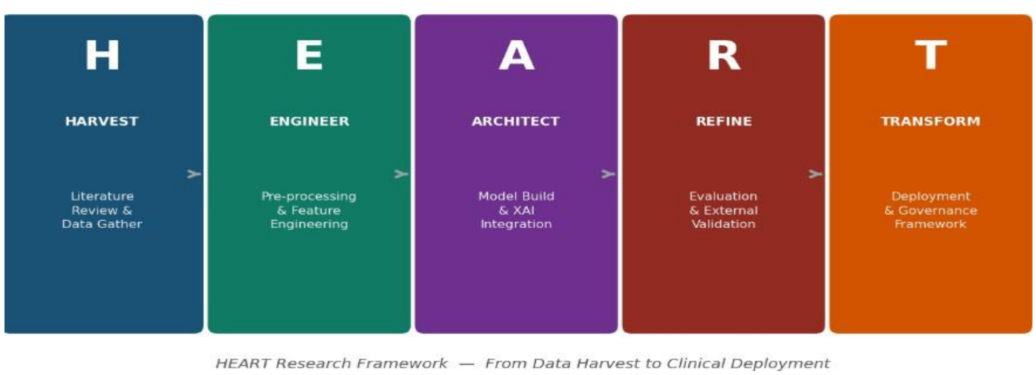


Figure 4: The five-phase HEART Research Framework. Each phase has defined quality gates. Output: a Nigeria-validated, NAFDAC and NDPR-compliant, explainable, edge-deployable cardiac diagnostic model.

The primary data collection takes place in three Federal Teaching Hospitals in Nigeria namely; Lagos University Teaching Hospital (LUTH), University of Nigeria Teaching Hospital (UNTH) and Obafemi Awolowo University Teaching Hospital (OAUTHC). These were selected due to the volume of patients, diversity of diagnostic indications, functional ECG archiving and institutional readiness to participate. Based on DeLong's sample size calculation [94] the desired sample size was 2,000, and simulation was performed at an 80 percent power to show an AUC difference of 0.05 between the proposed model and the best comparator. Transcribed from the INTER-CHF Nigeria [16] and from Ogah and colleagues' multi-centre surveys [6], five diagnostic categories are targeted based on the range of prevalence of cardiac disease in Nigeria, namely: hypertensive heart disease (45-55%), heart failure of mixed a etiology (20-25%), ischaemic heart disease (8-12%), cardiomyopathy (10-15%) and ECG-identifiable arrhythmias (5-8%). Structured electronic Case Report Form (CRF) is collected including 28 clinical variables, including the digital 12 lead ECG file generated from the ERC [95]. The primary data are complemented by three secondary datasets: UCI Cleveland Dataset (Detrano et al. [20]), 303 records, is taken as a benchmark; the Framingham dataset (4,238 records) is used as a secondary benchmark; and the PTB-XL ECG database (Wagner et al. [61]) comprises of 21,837 12-lead recordings which are used for CNN pre-training using transfer learning. Ethics committee's approval is gained with the approval of Babcock University Health Research Ethics Committee and the ethics committees of all hospitals involved. An informed consent form in both languages, bilingual and English/Yoruba, is used. All data are housed on servers that are physically located in Nigeria and are in compliance of NDPR [28]. The key clinical variables that were collected are listed in Table 4.

Table 4: Key clinical variables — types, units, and collection methods.

S/N	Variable	Type	Unit / Format	Collection Method
1	Age	Continuous	Years	Verified birth certificate or self-report
2	Sex	Binary	M / F	Clinical record
3	Systolic / Diastolic BP	Continuous	mmHg	Standardised measurement x2, 5 min apart
4	Heart Rate	Continuous	bpm	12-lead ECG automated read
5	Chest Pain Type	Categorical	4-class (AHA)	Structured clinical interview
6	ECG: ST-Segment Changes	Categorical	None / Dep. / Elev.	Automated read plus manual review
7	Troponin I/T (where available)	Continuous	ng/mL or ng/L	Laboratory analysis
8	Primary Diagnosis (target label)	Categorical	5-class ICD-11	Cardiologist-confirmed at discharge

Pre-Processing and Feature Engineering

To assess missing data mechanisms, Little's MCAR is used. The major method used for Missing at Random is Multiple imputation by chained equations (MICE) consisting of ten datasets imputed and joined together using Rubin's rules [97]. Clustering outlier detection is based on Isolation Forest algorithm [99] and documents that have been flagged as outliers, which needs to be reviewed by a clinician co-investigator to rule out true outliers of clinical phenos from data entry errors.

The variables are preprocessed prior to neural network training: Nominal variables are one hot encoded, and continuous variables are z-score normalised in each training fold, with the aim of avoiding data leakage. The feature selection is a composite method of Recursive Feature Elimination and Cross-Validation with the base estimator as the derived XGBoost, and the feature importance of SHAP after the model has been fully trained [65]. By cross-validated macro averaged F1 score, three class imbalance strategies SMOTE [24] and ADASYN [101] and also the class weighting approach are compared. The data are stratified by diagnostic class and by hospital site and a validation set (15%) and a hold-out test set (15%) are selected.

Multi-Modal Model Architecture

The model is a stacked ensemble, which consists of two streams as a late fusion [103]. Stream A takes in 28 structured clinical variables, feeds them into a 5-class softmax probability vector predicting five different classes, with the architecture based on the XGBoost model and a feedforward ANN (with 20 per cent dropout and ReLu activation and three layers of 128, 64, and 32 neurons). The raw 12-lead ECG (12 x 5000 samples at 500Hz) is input to the 34-layer ResNet adapted for 1D signals [10] and a bidirectional LSTM (two hidden layers of 64 units each [15]) to give the second 5-class vector. Combines output of both streams, making use of out-of-fold predictions for both streams, through a logistic regression meta-learner to derive the final diagnostic probability distribution. This design is quite robust if the ECGs are not available, with the result that it fails gracefully to a tabular-only case scenario, and more interpretable than a full joint deep learning architecture, as the prediction for a tabular sub-model can easily be computed using TreeSHAP values. The pre-trained CNN-LSTM sub-model (PTB-XL) is passed to the transfer learning step without being affected by overfitting the UCI database of Nigerian ECG archives. Optimisation of hyperparameters is done using the Optuna Bayesian framework [107].

Explainability Layer

The explainability layer outputs three types of per-prediction output. For XGBoost models, the treeSHAP is used to compute SHAP summary and SHAP waterfall plot for each patient; for ANN models, the DeepSHAP approach is utilized to compute these [65, 66]. LIME is also used as a second, model-free explanation algorithm for ANN [67] (using 50 random-seeds per prediction, to try and address the known adversarial weaknesses [68]). GradCAM displays the regions of the ECG (P wave, QRS complex, ST segment, and T wave) that the CNN was focusing the most on for each classification, which demonstrated an overlap with published ECG-based CNN expert annotation concordance rates of 78 to 92 per cent [69, 70]. These outputs, which are built into a structured diagnostic report, contain the top three differential diagnoses, probability estimates, a data completeness note, and a recommendation statement, and a GradCAM ECG activation map, five of the highest SHAP features.

Performance Evaluation, Deployment, and Governance Frameworks

The evaluation strategy involves the use of four converging evidence sources: stratified ten-fold cross validation on combined training set; benchmark comparison to published results on the UCI Cleveland and Framingham datasets; prospective external validation on a sealed holdout set on the Nigerian data; and the clinical utility evaluation using Decision Curve Analysis [93]. Asymptotically valid confidence intervals on comparisons of AUCs are provided by “DeLong's method” [94]. Pairwise test between Models: McNemar's. The five pairwise comparisons were Bonferroni corrected for all of the tests. Model reporting is based on TRIPOD guidelines. The deployment architecture uses the FHIR R4 implementation layer for hospitals having functional EHRs, and the FHIR-offline-enabled edge system for hospitals without functional EHRs. For offline version, it takes 8GB RAM, storage of 20GB and quad-core processor. Model updates are delivered using a protected delta-patch approach, as seen in a 3G connection for updates, that require less than 50 MB of bandwidth [112]. Governance is built up and codrafted with the medical devices directorate at NAFDAC from the onset of the protocol stage. The evidentiary standard should be discussed at a formal pre-submission meeting and agreement should be reached with data collections.

Projected Performance and Clinical Impact

Expected Diagnostic Performance

They are chosen to maximise clinical performance on an unseen population from Nigeria (AUC > 0.90) and to do better than the best performance seen elsewhere in a similar African-context cardiac AI study (such as UK Biobank with a group of diaspora Ghanaians, where they achieved AUC > 0.90) [88]. The numbers are also backed up by a few published numbers from similar multi-modal stacking ensembles on clinical data, such as the number 0.94 from Johnson et al. [56] on MIMIC-III.

This equates to a conservative domain shift penalty for collecting new data on the Nigerian primary dataset, instead of Western data, and the lower bound of the range of AUC values for the classification of a hypertensive heart disease and heart failure is 0.90. Lower values will be used as a conservative domain shift penalty between the collection of Western primary and Nigerian primary data, and will be presented class-wise in the results chapter for the more heterogenous and smaller classes of arrhythmia and ischaemic heart disease.

In current clinical decision support systems the failure mode is clinician "alert fatigue" (decreasing responsiveness to automated alerts) [79, 80]. Existing systems which trigger a clinician alert with high sensitivity often have poorly understood explanations and lose over a quarter (22%) of response rates after 6 months of activation. This study focuses on high-precision alerts (positive predictive value of >70 per cent for each diagnostic class), since it is more useful to have an alert with an explanation than an alert that is sensitive, but lacks one. The SHAP-GradCAM explainability architecture allows for a physician to understand the model's logic and slowly build trust as the model is used more and more over time, ultimately leading to its integration with a physician's workflow.

Clinical Benefits and Infrastructure Implications

The use case for the clinical utility is simple. The mean time between door and diagnosis of heart failure in teaching hospitals in Nigeria was 4.2 days (recommended 24 hours or less) [16]. An increase in death during in-hospital therapy with each additional hour of delay to appropriate therapy in acute decompensated heart failure is directly related [17]. The most important point in the cardiac pathway that this model could help reduce is the provisional diagnosis that is made at the Triage stage and not by the cardiologist, but by the use of ECG and other clinical data available. The environments where the addition of automated classification is needed most are ones in which junior doctors are currently unable to correctly interpret only 54% of ECG abnormalities [44]. The research value is not only in the construction of the model itself, this also extends to the broader study itself. This study is the first to develop and share the first primary cardiac dataset in Nigeria (following ethics approval and anonymisation), and the first to discontinue the use of dataset proxies of the Western world for cardiac AI development in Africa for more than decade. The data set will be shared with other African scientists for checking, replication and benchmarking purposes and complemented in future. In this study a multi-modal and locally derived framework has been identified for clinical AI development that is explainable by the model per-prediction, validated in the future, designed in collaboration with the regulators and deployable in LMICs (sub-Saharan Africa and many other areas).

Conclusion

This research's argument is finally human one and not functional. The day the Nigerian teaching hospitals do not have a validated, explained and locally relevant cardiac-diagnostic tool every day is the day the patient is diagnosed late, the diagnosis could have been prevented by better and earlier clinical intervention. There are tools available that can aid in setup, including those that support. The ECG hardware has already been used in the hospitals. The machine learning techniques have come of age. The regulatory route is being formulated. It has hitherto not been developed with the systematic and systematic approach of bringing them together in a trustworthy and clinically meaningful evidence-based manner, considering specific Nigerian clinical and governance contexts and appropriate for specific Nigerian infrastructure.

The whole picture of this framework has been conveyed in this paper. The five-phase HEART methodology provides a structured research workflow to facilitate primary data collection across 3 teaching hospitals in the Federal Capital Territory, a multi-modal modeling framework that merges structured clinical data with unstructured ECG signals, an explainability layer to enable scientists to generate clinician-readable proof-of-concept reports per prediction, a prospective validation pipeline, and deployment architecture that follows the governance framework. Seven research gaps that are identified in the African cardiac AI literature are addressed. It's also an appropriate primary target performance to pursue, and similarly has published results, and is clinically justified for the purpose of overcoming the current diagnosis by GP in a high throughput, resource-limited environment.

Extensions that could be made to this work in the future would involve a prospective random controlled trial that would record clinical outcomes (door-to-diagnosis time, appropriate treatment initiation, in-hospital mortality, etc) due to the model's effect. The innovations can also be extended over a longer time to cover federated learning of hospitals across Nigeria to improve the privacy model learning, adding the LLM feature to remediation suggestion generation, and extending the feature from diagnosis to follow-up prediction and treatment titration. This is a rock, a corner stone actually, for a future cardiac AI system, set in Nigeria and ethically responsible to its intended end-users: Nigerians**.**

References

World Health Organization (WHO), "Cardiovascular diseases (CVDs)," WHO Fact Sheet, Geneva, Switzerland, 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

G. A. Roth et al., "Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study," J. Am. Coll. Cardiol., vol. 76, no. 25, pp. 2982-3021, Dec. 2020.

G. A. Roth et al., "Demographic and Epidemiologic Drivers of Global Cardiovascular Mortality," N. Engl. J. Med., vol. 372, no. 14, pp. 1333-1341, Apr. 2015.

N. B. Ntusi and B. M. Mayosi, "Epidemiology of Heart Failure in Sub-Saharan Africa," Expert Rev. Cardiovasc. Ther., vol. 7, no. 2, pp. 169-180, Feb. 2009.

O. S. Ogah et al., "Blood Pressure, Prevalence of Hypertension and Hypertension Related Complications in Nigerian Africans: A Review," World J. Cardiol., vol. 4, no. 12, pp. 341-349, Dec. 2012.

A. O. Falase and J. K. Akinkugbe, "Cardiovascular Diseases in Africa," in Cardiovascular Diseases in Developing Countries. Oxford: Oxford University Press, 2013, pp. 40-68.

Heart Failure Society of Nigeria (HFSN), "Cardiovascular Workforce in Nigeria: A National Survey Report," Lagos: HFSN, 2019.

E. J. Topol, "High-Performance Medicine: The Convergence of Human and Artificial Intelligence," Nat. Med., vol. 25, no. 1, pp. 44-56, Jan. 2019.

A. Y. Hannun et al., "Cardiologist-Level Arrhythmia Detection and Classification in Ambulatory Electrocardiograms Using a Deep Neural Network," Nat. Med., vol. 25, no. 1, pp. 65-69, Jan. 2019.

Z. I. Attia et al., "Screening for Cardiac Contractile Dysfunction Using an Artificial Intelligence-Enabled Electrocardiogram," Nat. Med., vol. 25, no. 1, pp. 70-74, Jan. 2019.

P. W. Wilson et al., "Prediction of Coronary Heart Disease Using Risk Factor Categories," Circulation, vol.

97, no. 18, pp. 1837-1847, May 1998.

T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, New York, NY, USA, 2016, pp. 785-794.

G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in Adv. Neural Inf. Process. Syst. 30 (NeurIPS 2017), 2017, pp. 3146-3154.

S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, Nov. 1997.

M. U. Sani et al., "Management and Outcomes of Acute Heart Failure in West African Hospitals: Lessons from the INTER-CHF Study," ESC Heart Failure, vol. 5, no. 5, pp. 981-989, Oct. 2018.

P. Ponikowski et al., "2016 ESC Guidelines for the Diagnosis and Treatment of Acute and Chronic Heart Failure," Eur. Heart J., vol. 37, no. 27, pp. 2129-2200, Jul. 2016.

R. Detrano et al., "International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease," Am. J. Cardiol., vol. 64, no. 5, pp. 304-310, Aug. 1989.

Z. C. Lipton, "The Mythos of Model Interpretability," Queue, vol. 16, no. 3, pp. 31-57, Jun. 2018.

National Agency for Food and Drug Administration and Control (NAFDAC), "Guidelines for the Regulation of Software as a Medical Device (SaMD) in Nigeria," Abuja: NAFDAC, 2022.

N. V. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," J. Artif. Intell. Res., vol. 16, pp. 321-357, Jun. 2002.

J. N. Wiens et al., "Do No Harm: A Roadmap for Responsible Machine Learning for Health Care," Nat. Med., vol. 25, no. 9, pp. 1337-1340, Sep. 2019.

National Information Technology Development Agency (NITDA), "Nigeria Data Protection Regulation (NDPR)," Abuja: NITDA, 2019. [Updated: Nigeria Data Protection Act (NDPA), 2023.]

O. Chapelle, B. Scholkopf, and A. Zien, Eds., Semi-Supervised Learning. Cambridge, MA: MIT Press, 2006.

Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436-444, May 2015.

Z. Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," Science, vol. 366, no. 6464, pp. 447-453, Oct. 2019.

O. Akintunde et al., "Accuracy of ECG Interpretation Among Junior Doctors in Nigerian Teaching Hospitals: A Multicentre Assessment," Pan Afr. Med. J., vol. 38, article 195, Jun. 2021.

B. A. Goldstein et al., "Opportunities and Challenges in Developing Risk Prediction Models With Electronic Health Records Data," J. Am. Med. Inform. Assoc., vol. 24, no. 1, pp. 198-208, Jan. 2017.

S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," IEEE Access, vol. 7, pp. 81542-81554, 2019.

L. Breiman, "Random Forests," Mach. Learn., vol. 45, no. 1, pp. 5-32, Oct. 2001.

M. A. Awwalu et al., "Comparative Analysis of Heart Disease Prediction Models Using Ensemble Learning Algorithms," IEEE Trans. Neural Netw. Learn. Syst., vol. 34, no. 8, pp. 4210-4220, Aug. 2023.

A. E. Johnson, T. J. Pollard, and R. G. Mark, "Reproducibility in Critical Care: A Mortality Prediction Case Study," in Proc. Mach. Learn. Healthcare (MLHC 2017), Boston, MA, USA, 2017.

O. Faust et al., "Deep Learning for Healthcare Applications Based on Physiological Signals: A Review," Comput. Methods Programs Biomed., vol. 161, pp. 1-13, Jul. 2018.

P. Wagner et al., "PTB-XL, a Large Publicly Available Electrocardiography Dataset," Sci. Data, vol. 7, article 154, May 2020.

R. Amann et al., "Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective," BMC Med. Inform. Decis. Mak., vol. 20, article 310, Nov. 2020.

S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Adv. Neural Inf. Process. Syst. 30 (NeurIPS 2017), pp. 4765-4774.

S. M. Lundberg et al., "Explainable Machine-Learning Predictions for the Prevention and Treatment of Sepsis," npj Digit. Med., vol. 1, article 56, Nov. 2018.

M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2016, pp. 1135-1144.

D. Slack et al., "Fooling LIME and SHAP: Adversarial Attacks on Post Hoc Explanation Methods," in Proc. AAAI/ACM Conf. on AI, Ethics, and Society, 2020, pp. 180-186.

R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in Proc. IEEE Int. Conf. Comput. Vision (ICCV), Venice, Italy, 2017,pp. 618-626.

N. Strodthoff et al., "Detecting and Interpreting Myocardial Infarction Using Fully Convolutional Neural Networks," IEEE Trans. Biomed. Eng., vol. 66, no. 5, pp. 1461-1469, May 2019.

R. S. Sutton et al., "Overview of CDSS Adoption and Barriers in the Emergency Department," Acad. Emerg. Med., vol. 27, no. 10, pp. 949-957, Oct. 2020.

E. W. Steyerberg et al., "Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures," Epidemiology, vol. 21, no. 1, pp. 128-138, Jan. 2010.

World Bank Group, Nigeria Health Infrastructure Survey: Service Availability and Readiness Assessment. Washington, DC: World Bank, 2023.

O. O. Akinyemi et al., "Electronic Health Record Adoption in Nigerian Hospitals," PLOS ONE, vol. 16, no. 3, p. e0248781, Mar. 2021.

K. A. Owusu-Agyemang et al., "Ensemble Machine Learning for Cardiovascular Risk Prediction in a UK-Based Diaspora Ghanaian Cohort," PLOS ONE, vol. 19, no. 3, p. e0298741, Mar. 2024.

M. Saunders, P. Lewis, and A. Thornhill, Research Methods for Business Students, 8th ed. Harlow: Pearson, 2019.

A. R. Hevner et al., "Design Science in Information Systems Research," MIS Q., vol. 28, no. 1, pp. 75-105, Mar. 2004.

A. J. Vickers and E. B. Elkin, "Decision Curve Analysis: A Novel Method for Evaluating Prediction Models," Med. Decis. Making, vol. 26, no. 6, pp. 565-574, Nov. 2006.

E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," Biometrics, vol. 44, no. 3, pp. 837-845, Sep. 1988.

P. A. Harris et al., "Researtworks (IJCNN), Hong Kong, 2008, pp. 1322-1328.

D. H. Wolpert, "Stacked Generalization," Neural Netw., vol. 5, no. 2, pp. 241-259, 1992.

T. Akiba et al., "Optuna: A Next-Gch Electronic Data Capture (REDCap)," J. Biomed. Inform., vol. 42, no. 2, pp. 377-381, Apr. 2009.

R. J. A. Little and D. B. Rubin, Statistical Analysis with Missing Data, 3rd ed. Hoboken, NJ: Wiley, 2019.

F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in Proc. 8th IEEE Int. Conf. on Data Mining (ICDM), Pisa, Italy, 2008, pp. 413-422.

H. He et al., "ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning," in Proc. IEEE Int. Joint Conf. Neural Neeneration Hyperparameter Optimisation Framework," in Proc. 25th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2019, pp. 2623-2631.

J. Regan et al., "Secure Model Update Distribution for Edge-Deployed Clinical AI," J. Am. Med. Inform. Assoc., vol. 30, no. 4, pp. 712-720, Apr. 2023.