International Journal of Scientific Research and Management (IJSRM)
||Volume||13||Issue||09||Pages||2324-2338||2025||
Website: https://ijsrm.net ISSN (e): 2321-3418
DOI: 10.18535/ijsrm/v13i09.mp03

Machine Learning Prediction of BNP, HS-Troponin, and CRP from Routine Blood Tests in Heart Failure Patients: Evidence from Abakaliki, Ebonyi State, Nigeria

[1]Ikechukwu Ogbo PhD, [2]Nancy Chiatogu Ibeh, PhD, [3]Grace Ifechukwu Amilo, PhD, [4]Chizoba Okechukwu Okeke, PhD, [5]Joelson Friday Onwe, M.Sc

1. Department of Haematology and Blood Transfusion Science, Faculty of Health Science and Technology, Ebonyi State University (EBSU), Nigeria.
2. Department of Haematology and Blood Transfusion Science, Faculty of Medical Laboratory Science, Nnamdi Azikiwe University (NAU), Nigeria
3. Department of Haematology, Faculty of Basic Clinical Medicine, Nnamdi Azikiwe University (NAU), Nigeria
4. Department of Haematology and Blood Transfusion Science, Faculty of Medical Laboratory Science, Nnamdi Azikiwe University (NAU), Nigeria
5. Haematology and Blood Group Serology Unit, Department of Medical Laboratory Services, Alex Ekwueme Federal University Teaching Hospital, Abakaliki, Nigeria

Abstract
Heart failure (HF) is a significant cause of morbidity and mortality worldwide, with a hefty burden in low-resource settings where diagnostic assays remain costly and limited in availability. This study examined whether routine haematology, coagulation, and biochemistry parameters can predict B-type Natriuretic Peptide, High-sensitivity Troponin and C-reactive Protein in HF patients using machine learning (ML).

Methods:
We prospectively enrolled 579 adults with HF at Alex Ekwueme Federal University Teaching Hospital, Abakaliki. We collected 10 mL of venous blood for laboratory investigations and obtained demographic and clinical data from their medical records. We processed the data in Python 3.12 and applied feature selection techniques including correlation thresholds, recursive feature elimination, its standard libraries. We trained and evaluated nine models, choosed the best model for each biomarker, and conducted sex-stratified analyses to compare performance between male and female participants.

Results:
Important predictors included Urea, creatinine, eGFR, D-dimer, fibrinogen, and neutrophil-to-lymphocyte ratio (NLR). Models that combined all the parameters outperformed single-domain models. CatBoost produced the best results for BNP (R^2 0.30–0.33), ElasticNetCV for hs-Troponin (R^2 0.09–0.12), and Ridge/ElasticNetCV for CRP (R^2 0.53–0.54). SHAP analysis indicated that Urea and D-dimer strongly influenced BNP, while NLR, eGFR, and fibrinogen contributed most to the predictions of hs-Troponin and CRP. Sex-stratified models showed consistent behaviour across algorithms, with only minor differences in predictive strength.

Conclusion:
Routine laboratory data can estimate BNP, hs-Troponin, and CRP using ML in patients with HF from low-resource settings.

Keywords: Heart failure, Machine learning, BNP, High-sensitivity troponin, C-reactive protein, Routine laboratory tests, Cardiac biomarker prediction, Low-resource settings

Introduction

Machine Learning Prediction of B-type Natriuretic Peptide (BNP), High-sensitivity Troponin (TnT), and C-reactive Protein (CRP) from Routine Blood Tests in Heart Failure Patients: Evidence from Abakaliki,

Ebonyi State, Nigeria examines a critical challenge in cardiovascular diagnostics within low-resource environments. Machine learning (ML), a branch of artificial intelligence, enables algorithms to identify complex data patterns and generate predictions without requiring task-specific programming. In cardiovascular medicine, ML has shown potential as a practical alternative to direct biomarker assays, such as BNP, high-sensitivity troponin, and CRP (1,2). These routinely available tests hold particular value in sub-Saharan Africa, where standardised biomarker assays remain inaccessible or unaffordable.

Coagulation indices, inflammatory markers, and renal function measures show strong associations with BNP, troponin, and CRP, reflecting myocardial stress, systemic inflammation, and impaired clearance (3,4). However, most existing ML models have been developed in high-income countries, with minimal application in African populations, where HF typically presents earlier, progresses more severely, and occurs under significant diagnostic and socioeconomic limitations (5,6).

This study aimed to evaluate whether routine haematological, coagulation, and biochemical parameters can predict BNP, hs-Troponin, and CRP in Nigerian HF patients using ML. Our objective was to develop and validate prediction models tailored to this population and to determine whether surrogate biomarker estimation can be applied in clinical care under resource-constrained conditions.

2.0 Methods

2.1 Study Design and Setting

We conducted a prospective, single-centre observational study at Alex Ekwueme Federal University Teaching Hospital, Abakaliki (AEFUTHA), a tertiary referral hospital in southeastern Nigeria. The facility provides care to a predominantly low-resource population and offers essential diagnostic laboratory services. The Research Ethics Committee approved the protocol (Reference Number: AEFUTHA/REC/VOL 3/2020/119). We obtained written informed consent from all participants.

2.2 Study Population

We enrolled 579 adult patients (≥18 years) with a clinical diagnosis of HF who visited the cardiology unit at AEFUTHA between January 2021 and October 2024. We recruited participants consecutively during clinic visits to limit selection bias. We excluded patients with active infections, hematologic malignancies, a history of blood transfusion within the previous 14 days, or an inability to provide informed consent.

2.3 Sample Size Estimation

We determined the sample size based on both statistical power and the requirements of the ML model. Detecting a modest correlation ($r = 0.20$) between routine laboratory parameters and BNP, hs-Troponin, and CRP at 80% power and a 5% significance level required at least 200 participants. For ML training, we applied the rule of at least 10 observations per predictor variable, which indicated a larger cohort. Allowing for 20% attrition or incomplete data, we targeted recruitment at 579 participants.

2.4 Data Collection

We collected 10 mL of venous blood from each participant into appropriate anticoagulated tubes. Laboratory investigations—including complete blood count, coagulation profile, renal and liver function tests, lipid profile, and cardiac biomarkers (BNP, hs-Troponin, and CRP)—were performed using validated automated analysers under strict internal and external quality control protocols. We extracted demographic data, comorbidities, vital signs, and body mass index (BMI) from patient records and entered all data into a password-protected Excel file.

2.5 Laboratory Analysis

We performed laboratory tests in the Clinical Chemistry, Haematology, and Coagulation laboratories of AEFUTHA. We measured haematological parameters on the Sysmex XN-500 automated analyser, conducted coagulation assays on the Sysmex CA-1500 system, and analysed renal, hepatic, electrolyte, and lipid parameters on the Selectra Pro S chemistry analyser. We strictly followed the manufacturer's instructions and applied internal quality control measures to ensure accuracy.

2.6 Data Analysis

We conducted all data analyses in Python (version 3.12) using core libraries, including scikit-learn, pandas, NumPy, XGBoost, LightGBM, CatBoost, SHAP, and StatsModels. We preprocessed the data by examining continuous predictors for distributional properties with the Shapiro test, log-transforming skewed variables to reduce skewness, and standardising them with the RobustScaler. We removed missing values from both predictors and covariates before analysis. We defined BNP, hs-troponin, and CRP as the primary study outcomes, while routine haematological, coagulation, and biochemical parameters—excluding the outcome biomarkers—served as predictor variables with vital covariates.

2.7 Exploratory Analysis and Features Selection

We began our exploratory analysis by applying descriptive statistics and generating visualisations such as Spearman correlation heatmaps to explore pairwise relationships among predictors. To manage multicollinearity and reduce dimensionality without losing interpretability, we implemented a multi-step feature selection strategy. This strategy combined correlation-based screening, recursive feature elimination with cross-validation, tree-based variable importance ranking, L1-regularised regression, and model-agnostic methods such as Shapley Additive exPlanations (SHAP). We prioritised features consistently identified as important across at least five methods and retained the top ten predictors for each biomarker model. This approach balanced predictive accuracy with clinical interpretability and minimised the risk of overfitting in subsequent machine learning models.

2.8 Machine Learning Model Development

We built separate regression models for each biomarker—BNP, hs-troponin, and CRP—using three predictor sets: haematology alone, biochemistry alone, and a combined dataset. To capture both simple and complex relationships, we trained a diverse set of machine learning algorithms, ranging from linear models to non-linear methods such as ensemble trees, regularised regression, and kernel-based approaches. Model training followed a nested 5×5 cross-validation framework, with inner loops dedicated to hyperparameter tuning and outer loops used for unbiased performance evaluation. To enhance generalisability and minimise spurious associations, we adjusted all models for prespecified clinical confounders.

2.9 Model Evaluation

We assessed model performance using a set of complementary metrics, including explained variance, R^2 (with adjusted R^2), RMSE, MAE, and median absolute error. To examine calibration, we estimated slopes and intercepts to test how well predicted values matched observed outcomes. We also analysed residuals to check assumptions of homoscedasticity and normality and to identify any systematic biases. By combining these approaches, we ensured that the models were not only accurate but also reliable and well-calibrated for clinical use.

2.10 Model Interpretability

We used SHAP values to assess the contribution of each predictor to the prediction of each biomarker. By comparing results across different algorithms, we identified which models performed best for each biomarker and predictor combination. Predictors that consistently appeared as important were noted as potential surrogate or supplementary biomarkers, especially in settings where direct measurement of BNP, hs-troponin, or CRP may not be readily available. By integrating interpretability with predictive modelling, we maximised translational value and strengthened the link between methodological rigour and clinical applicability.

3.0 Results

Table 1. Baseline Continuous Variables

Category	Variable	Median [IQR]
Clinical & Demographic	Age (years)	66.00 [20.00]

Category	Variable	Median [IQR]
	Body temperature (°C)	36.30 [0.30]
	Pulse (beats/min)	85.00 [26.00]
	Respiration (breaths/min)	19.00 [2.00]
	Systolic blood pressure (mmHg)	130.00 [34.00]
	Diastolic blood pressure (mmHg)	75.00 [21.00]
	Mean arterial pressure (mmHg)	93.33 [22.00]
	Weight (kg)	50.00 [15.00]
	Height (m)	1.57 [0.14]
	BMI (kg/m²)	20.54 [5.07]
	Charlson Comorbidity Index (score)	2.00 [2.00]
Hematology & Coagulation	White blood cell count (×10³/μL)	6.69 [4.16]
	Neutrophil count (×10³/μL)	4.97 [3.69]
	Lymphocyte count (×10³/μL)	0.90 [0.67]
	Monocyte count (×10³/μL)	0.45 [0.30]
	Basophil count (×10³/μL)	0.03 [0.02]
	Eosinophil count (×10³/μL)	0.05 [0.11]
	RBC (×10⁶/μL)	3.90 [0.91]
	Hemoglobin (g/L)	118.00 [28.00]
	Hematocrit (%)	36.00 [8.00]
	RDW-CV (%)	14.40 [1.90]
	RDW-SD (fL)	48.10 [6.25]
	MCV (fL)	93.40 [7.45]
	MCH (pg)	30.50 [2.95]
	MCHC (g/L)	326.00 [16.00]
	MPV (fL)	12.00 [2.20]
	Platelet count (×10³/μL)	142.00 [79.00]
	PDW (%)	16.30 [0.70]
	NLR	5.53 [5.96]
	PLR	159.00 [131.71]
	MLR	0.50 [0.48]
	LMR	2.00 [1.70]
	PCT (%)	0.17 [0.08]
	D-dimer (mg/L)	1.31 [1.71]
	INR	1.22 [0.22]
	APTT (s)	34.20 [6.85]
	Thrombin time (s)	17.20 [1.50]
	Prothrombin activity (%)	67.37 [19.93]
	Prothrombin time ratio	1.22 [0.22]
	Fibrinogen (g/L)	3.01 [1.25]
Biochemistry	Creatinine (μmol/L)	85.90 [52.10]
	Urea (mmol/L)	7.71 [5.26]
	Uric acid (μmol/L)	440.00 [203.50]
	eGFR (mL/min/1.73 m²)	66.72 [45.62]
	Cystatin C (mg/L)	1.59 [0.98]

Category	Variable	Median [IQR]
	Calcium (mmol/L)	2.28 [0.21]
	Potassium (mmol/L)	3.84 [0.75]
	Chloride (mmol/L)	102.60 [7.70]
	Sodium (mmol/L)	139.00 [5.75]
	Albumin (g/L)	36.60 [5.95]
	Globulin (g/L)	28.20 [6.60]
	Indirect bilirubin (µmol/L)	11.60 [9.50]
	Direct bilirubin (µmol/L)	6.60 [6.20]
	Total bilirubin (µmol/L)	18.20 [15.45]
	Alkaline phosphatase (U/L)	80.00 [38.00]
	Total protein (g/L)	64.60 [9.80]
	Total cholesterol (mmol/L)	3.62 [1.34]
	LDL-C (mmol/L)	1.74 [0.98]
	Triglyceride (mmol/L)	0.94 [0.56]
	HDL-C (mmol/L)	1.08 [0.47]
	Total cholesterol/HDL-C ratio	3.26 [1.57]
Cardiac Biomarkers	High-sensitivity troponin (ng/mL)	0.06 [0.09]
	Brain natriuretic peptide (pg/mL)	770.76 [1596.92]
	CRP (mg/L)	9.40 [23.25]

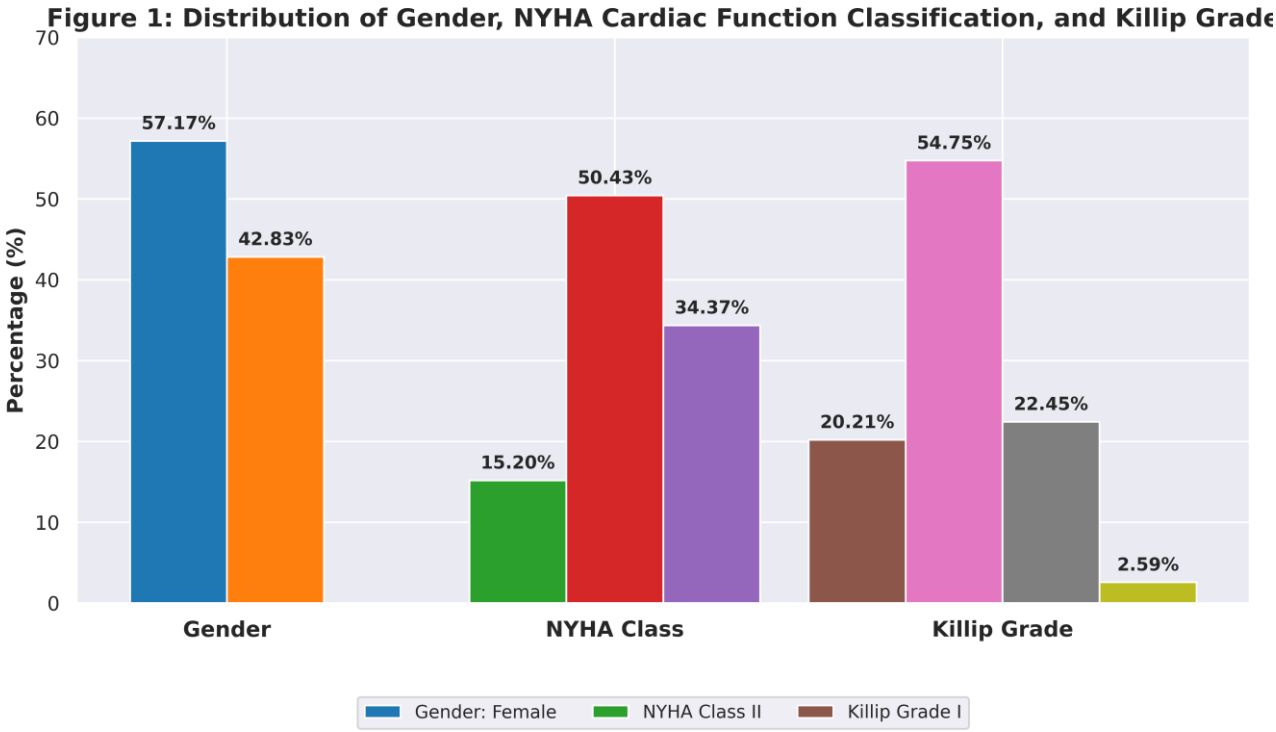
Values are expressed as median [interquartile range]. Abbreviations: BMI, body mass index; RBC, red blood cell; Hb, hemoglobin; RDW-CV, red cell distribution width–coefficient of variation; RDW-SD, red cell distribution width–standard deviation; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MPV, mean platelet volume; PDW, platelet distribution width; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; LMR, lymphocyte-to-monocyte ratio; PCT, plateletcrit; INR, international normalized ratio; APTT, activated partial thromboplastin time; eGFR, estimated glomerular filtration rate; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; CRP, C-reactive protein.

Table 1 summarises the baseline characteristics of the study population. We reported the continuous variables as median values with corresponding interquartile ratios (IQR). The cohort had a median age of 74 ± 20 years and a BMI of 20.5 ± 5.1 kg/m². Vital signs included a resting pulse of 85 ± 26 beats/min, a respiration rate of 19 ± 2 breaths/min, and blood pressures of 130 ± 34 mmHg (systolic) and 75 ± 21 mmHg (diastolic), yielding a mean arterial pressure of 93 ± 22 mmHg. Median body weight was 50 ± 15 kg and height 1.57 ± 0.14 m.

Renal indices showed creatinine 85.9 ± 52.1 µmol/L, eGFR 66.7 ± 45.6 mL/min/1.73 m², and cystatin C 1.59 ± 0.98 mg/L. Hematologic markers included RDW-CV 14.4 ± 1.9 %, NLR 5.53 ± 5.96, and MPV 12.0 ± 2.2 fL. Coagulation parameters were D-dimer 1.31 ± 1.71 mg/L, INR 1.22 ± 0.22, and fibrinogen 3.01 ± 1.25 g/L. Median hs-Troponin was 0.06 ± 0.09 ng/mL and BNP 770.8 ± 1596.9 pg/mL.

Electrolytes and proteins included sodium 139 ± 5.8 mmol/L, albumin 36.6 ± 6.0 g/L, and globulin 28.2 ± 6.6 g/L. Lipid parameters were as follows: cholesterol, 3.62 ± 1.34 mmol/L; LDL-C, 1.74 ± 0.98 mmol/L; and HDL-C, 1.08 ± 0.47 mmol/L.

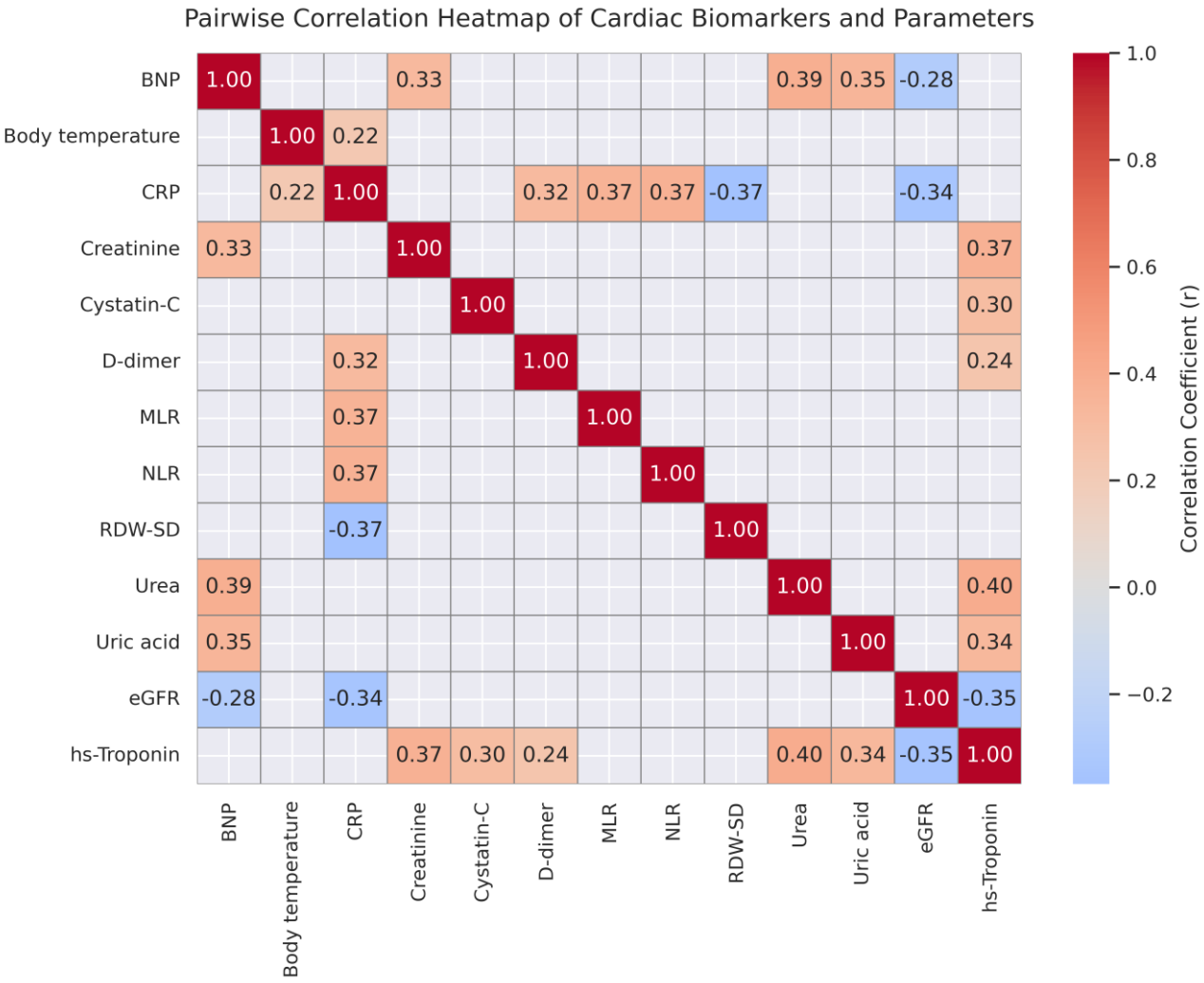
Figure 1: Distribution of Gender, NYHA Cardiac Function Classification, and Killip Grade



Percentages represent the distribution of study participants by gender, New York Heart Association (NYHA) functional class, and Killip grade. NYHA classification: II = mild limitation, III = marked limitation, IV = symptoms at rest. Killip grading: I = no signs of heart failure, II = S3 gallop or mild pulmonary congestion, III = acute pulmonary oedema, IV = cardiogenic shock.

Figure 1 displays the percentage distribution of participants by gender, NYHA functional class, and Killip grade. Females comprised 57.17% of the cohort, and males 42.83%. The majority were in NYHA Class III (50.43%), followed by Class IV (34.37%) and Class II (15.20%). The Killip grade was predominantly Grade II (54.75%), with fewer cases in Grades III (22.45%), I (20.21%), and IV (2.59%).

Figure 2: Correlation Heatmap of Cardiac Biomarkers and Significant Laboratory Parameters



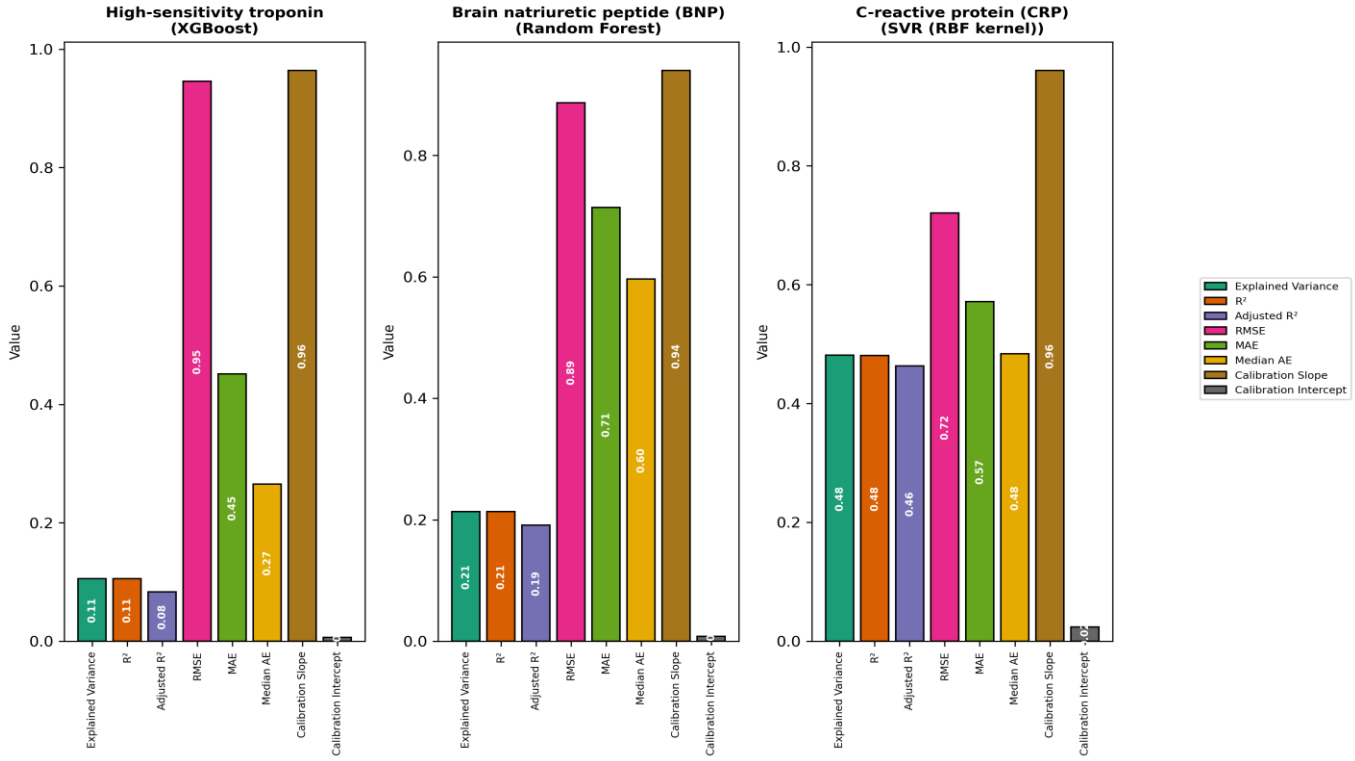
Values represent Spearman rank correlation coefficients (r). Only parameters with |r| > 0.2 and p < 0.05 are reported. BNP, B-type natriuretic peptide; CRP, C-reactive protein; RDW-SD, red cell distribution width–standard deviation; MLR, monocyte-lymphocyte ratio; NLR, neutrophil-lymphocyte ratio; eGFR, estimated glomerular filtration rate. We showed only statistically significant correlations (p < 0.05) with a correlation coefficient greater than 0.2 or less than -0.2.

Figure 2 summarises the significant Spearman correlations ($p < 0.05$, $|r| > 0.2$) between three cardiac biomarkers—High-Sensitivity Troponin, Brain Natriuretic Peptide (BNP), and C-Reactive Protein (CRP)—and various clinical and laboratory parameters.

High-sensitivity troponin strongly correlates with BNP ($r = 0.60$) and exhibits positive associations with renal markers (serum Urea), inflammation (WBC, ALT), and negative correlations with glomerular filtration rate and prothrombin activity. BNP correlates positively with respiratory rate, potassium, serum Urea, WBC, and ALT, and negatively with glomerular filtration rate. CRP is positively associated with body temperature, sodium, haemoglobin, globulin, and serum Urea, while negatively correlating with heart rate, albumin, diastolic blood pressure, D-dimer, and random Glucose. These findings highlight the interconnected roles of cardiac injury, renal function, and inflammation in the clinical profile.

Figure 3: Model Performance Metrics for Prediction of Cardiac Biomarkers with Haematological and Coagulation Parameters

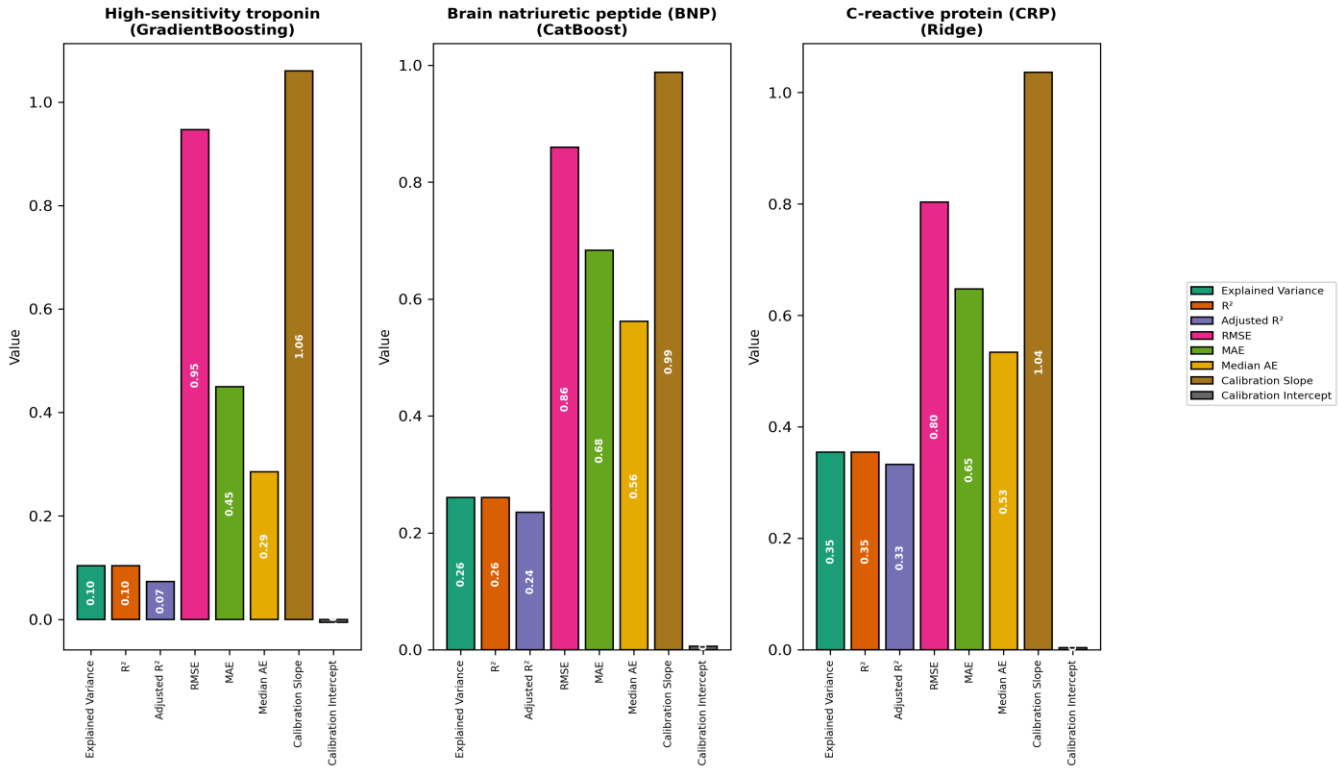
Figure 3. Model Performance Metrics for Prediction of Cardiac Biomarkers with Haematological and Coagulation Parameters



RMSE = root mean squared error; MAE = mean absolute error; Median AE = median absolute error; MAPE = mean absolute percentage error; SVR = support vector regression; RBF = radial basis function; ElasticNetCV = elastic net regression with cross-validation.

Figure 4: Model Performance Metrics for Prediction of Cardiac Biomarkers with Biochemistry Parameters

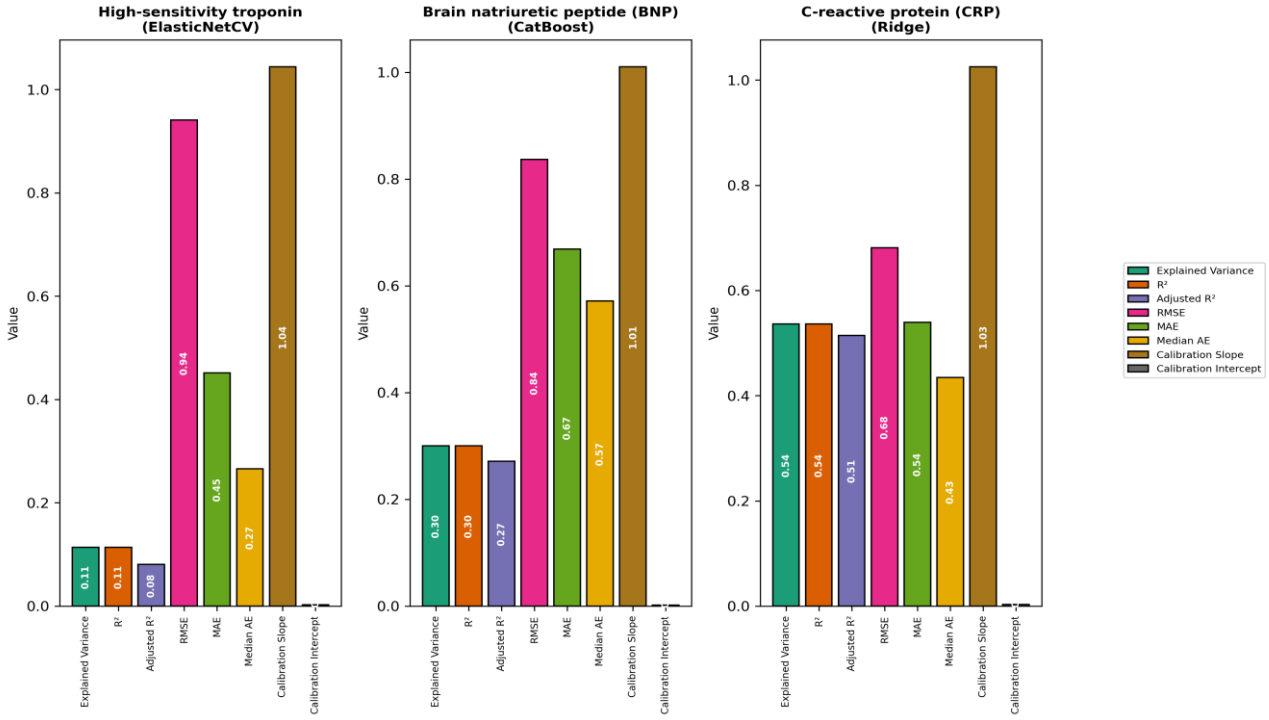
Figure 4. Model Performance Metrics for Prediction of Cardiac Biomarkers with Biochemistry Parameters



RMSE = root mean squared error; MAE = mean absolute error; AE = absolute error; MAPE = mean absolute percentage error; Calib. = calibration. Models trained on routine biochemistry parameters to predict BNP, hs-troponin, and CRP in heart failure patients.

Figure 5. Model Performance Metrics for Prediction of Cardiac Biomarkers using Haematology, Coagulation, and Biochemistry Parameters

Figure 5. Model Performance Metrics for Prediction of Cardiac Biomarkers using Hematology, Coagulation, and Biochemistry Parameters



Explained variance, R^2, and adjusted R^2 reflect model fit. RMSE = root mean squared error, MAE = mean absolute error, Median AE = median absolute error, MAPE = mean absolute percentage error. Calibration slope and intercept indicate model calibration. Unfavourable explained variance indicates poor model generalisation.

Across the various modelling strategies, we observed notable variations in the predictive performance of the three cardiac biomarkers (Figure 3). For hs-troponin, models leveraging combined feature sets achieved their best results with XGBoost and ElasticNetCV, while Random Forest and LightGBM optimised BNP prediction. CRP, in contrast, was better captured by non-linear approaches, with SVR, CatBoost, and Random Forest consistently surpassing linear models. When restricted to biochemistry parameters (Figure 4), Gradient Boosting and XGBoost provided the most accurate estimates for hs-troponin, CatBoost emerged as the strongest model for BNP, and Ridge, together with ElasticNetCV, produced stable outputs for CRP. Expanding to a multimodal framework that incorporated haematology, coagulation, and biochemistry features (Figure 5) further enhanced accuracy across all biomarkers. Specifically, ElasticNetCV was most effective for hs-troponin, CatBoost remained strongest for BNP, and Ridge, alongside ElasticNetCV, yielded the most reliable CRP predictions.

Table 3: Best-Performing Models for cardiac Biomarker using both haematology, Coagulation and biochemistry parameters based on Gender

Cardiac biomarkers	Gender	Best Model	R^2	EVS	RMSE	MAE
hs-Troponin	Male	ElasticNetCV	0.0897	0.0901	0.8481	0.4116

Cardiac biomarkers	Gender	Best Model	R^2	EVS	RMSE	MAE
hs-Troponin	Female	ElasticNetCV	0.1242	0.1244	1.0056	0.4812
BNP	Male	CatBoost	0.3290	0.3295	0.8479	0.6809
BNP	Female	CatBoost	0.2700	0.2705	0.8280	0.6596
CRP	Male	ElasticNetCV	0.5363	0.5374	0.6825	0.5467
CRP	Female	LinearRegression	0.5391	0.5397	0.6778	0.5346

***Abbreviations**: BNP = brain natriuretic peptide; CRP = C-reactive protein; EVS = explained variance score; MAE = mean absolute error; RMSE = root mean squared error; R^2 = coefficient of determination.*

Sex-stratified analyses (Table 3) confirmed that ElasticNetCV is consistently superior for hs-troponin prediction in both males and females, CatBoost maintaining its dominance for BNP, and CRP prediction optimised by ElasticNetCV in males and Linear Regression in females.

Figure 6A. SHAP Feature Importance Ranking for Brain Natriuretic Peptide (BNP).

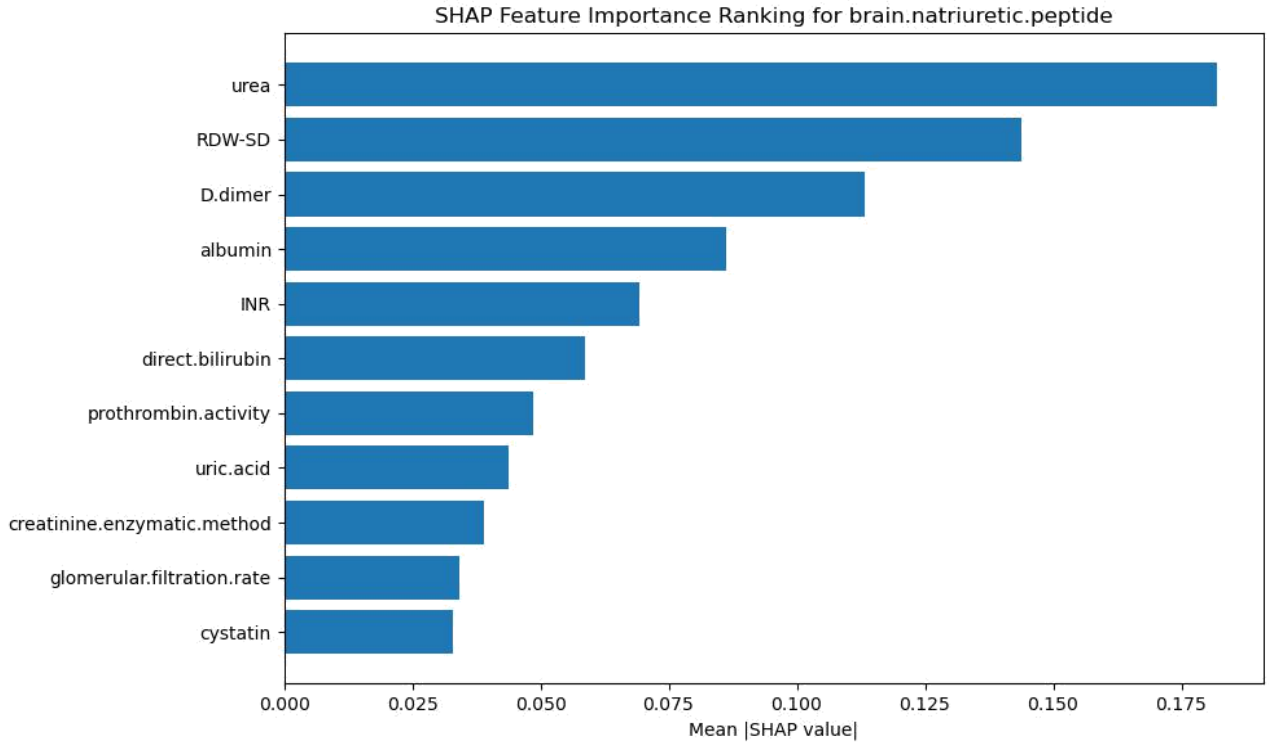


Figure 6A. SHAP feature importance ranking for brain natriuretic peptide (BNP). The bar chart displays the relative contribution of individual predictors, with higher SHAP values indicating stronger influence on BNP prediction. Urea and D-dimer were the most influential variables, followed by RDW-SD, uric acid, and INR.

Figure 6B: Sharp Feature importance Ranking for CRP.

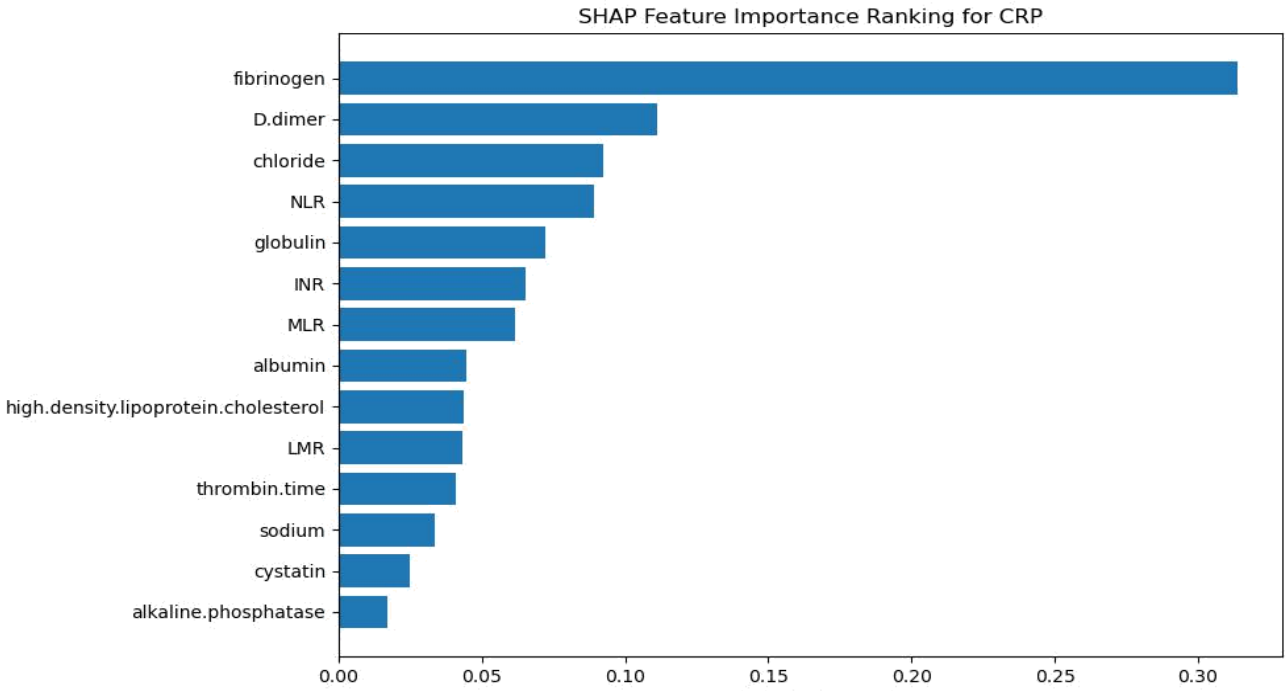


Figure 6B. SHAP feature importance ranking for C-reactive protein (CRP). The bar chart illustrates the relative contribution of predictors, with higher SHAP values denoting a more substantial influence on CRP prediction. Fibrinogen was the most influential feature, followed by D-dimer, chloride, and neutrophil-to-lymphocyte ratio (NLR).

Figure 6C: Sharp Feature Importance Ranking for Highly sensitive troponin

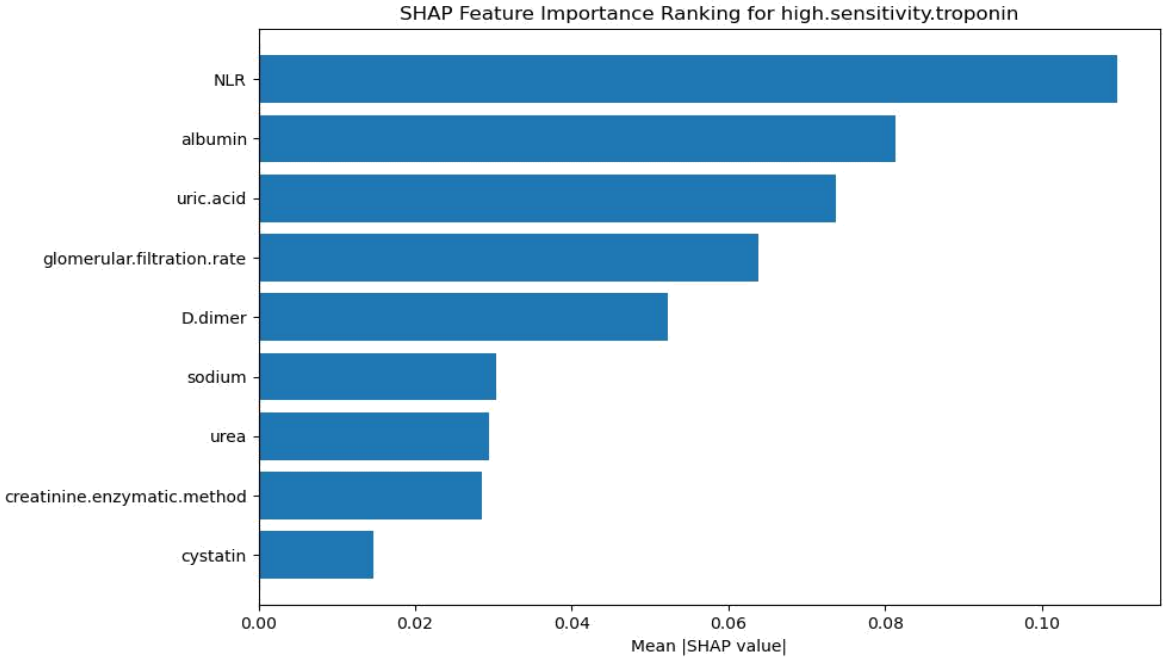


Figure 6C. SHAP feature importance ranking for high-sensitivity troponin (hs-Troponin). The bar chart shows the relative contribution of predictors, where higher SHAP values indicate a more substantial influence on hs-Troponin prediction. Neutrophil-to-lymphocyte ratio (NLR) was the most influential feature, followed by albumin, uric acid, estimated glomerular filtration rate (eGFR), and creatinine.

3. Discussion

Our study demonstrates that the demographic and clinical characteristics of heart failure (HF) patients in Abakaliki are mainly consistent with those reported in prior studies. The median age of 66 years and the distribution of vital signs, including pulse, respiratory rate, and blood pressure, align with typical HF cohorts

(7–11). Similarly, comorbidity profiles, as reflected by a median Charlson comorbidity index of 2 and a high prevalence of hypertension, diabetes, and coronary artery disease, mirror literature reports from diverse populations (12–15). Functional classifications using the NYHA and Killip scales also showed predominant class II–III and II–III distributions, consistent with prior studies of acute and chronic HF (8, 16, 17).

Laboratory findings in our cohort corroborate established HF pathophysiology. Renal function markers reflect moderate renal involvement, consistent with their prognostic importance in HF (18,19). Inflammatory and hematologic indices, including NLR, RDW-CV, and MPV, demonstrate relationships with disease severity in line with prior observations (20,21). Cardiac biomarkers, such as BNP and hs-Troponin, correlated with renal function and inflammatory markers, confirming the interplay between hemodynamic stress, myocardial injury, and systemic inflammation (22–24). Moreover, the positive associations between BNP, troponin, and functional severity (NYHA/Killip classes) are in agreement with prior literature highlighting natriuretic peptides as reliable indicators of HF progression (24,25).

Machine learning (ML) analyses in our study further support the predictive potential of routine laboratory data. Models integrating haematology, biochemistry, and coagulation markers effectively predicted BNP, hs-Troponin, and CRP, demonstrating that multimodal laboratory integration enhances biomarker prediction compared to single-domain approaches(26–29). SHAP-based interpretability identified NLR, fibrinogen, renal indices, D-dimer, and INR as key contributors, consistent with the literature, which emphasises the value of renal, inflammatory, and hematologic parameters in cardiovascular risk modelling (30–32). Routine laboratory parameters offer a cost-effective and scalable approach, supporting early risk stratification, particularly in resource-limited settings, and align with prior studies demonstrating comparable predictive accuracy to specialised biomarkers such as NT-proBNP (2,33–35).

Despite general concordance, some findings diverge from previously reported HF cohorts. Our patients exhibited lower median BMI and lower absolute weight (50 kg), contrasting with the literature indicating overweight or obese HF populations (7,11). The median age of 66 years is slightly younger than the cohorts reporting means up to 74–82 years (14,36). The gender distribution in our study showed a female predominance, differing from most studies that have shown male predominance or near-equal distribution (7,37).

Biomarker concentrations also exhibited some deviation. BNP levels and hs-Troponin were lower than the thresholds commonly associated with adverse outcomes in hospitalised HF patients (38,39). Similarly, CRP levels were lower than those reported in studies of acute HF with pronounced systemic inflammation (40,41). Lipid profiles were also lower than those of typical HF cohorts, reflecting potential population-specific differences in metabolic status (21). Furthermore, we did not specifically analyse HFpEF prevalence, which prevented a direct comparison to studies reporting a 60–69% predominance of HFpEF (42,43). The low proportion of Killip IV patients may reflect differences in hospital admission criteria relative to prior cohorts reporting (17,31).

The machine learning performance in our study, while demonstrating the benefit of multimodal integration, achieved moderate predictive power, which is lower than some reports in larger or more homogeneous datasets (30, 33, 44). Key predictors in our study—NLR, fibrinogen, renal indices, and D-dimer—differ somewhat from prior SHAP analyses, which emphasise proBNP, troponin, CK-MB, albumin, and electrolytes as dominant features (8, 32, 45). Additionally, our sex-stratified modelling revealed modest differences in predictive accuracy between males and females, whereas other studies report more pronounced sex-specific performance disparities (44,46).

Finally, unlike most previous studies that primarily predict HF outcomes, our work focuses on the direct prediction of cardiac biomarkers from routine laboratory data. This approach, although less commonly reported, extends the application of ML in HF and demonstrates the potential for routine laboratorys to serve as surrogates for expensive biomarker assays (30, 47, 48).

Conclusion

This study highlights the feasibility of leveraging routine laboratory data and machine learning to complement conventional biomarker assessment, providing a scalable, cost-effective strategy for HF management while extending the application of predictive modeling beyond traditional outcome prediction.

References

1. Parasar D, Ali A, Pillai NM, Shahi A, Alfurhood BS, Pant K. Detailed review on Integrated Healthcare Prediction System Using Artificial Intelligence and Machine Learning. In: 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) [Internet]. 2023 [cited 2025 Aug 27]. p. 682–5. Available from: https://ieeexplore.ieee.org/document/10182709
2. James N, Gerrish L, Rokotyan N, Gladding PA. Machine Learning Applied to Routine Blood Tests and Clinical Metadata to Identify and Classify Heart failure [Internet]. medRxiv; 2021 [cited 2025 Aug 27]. p. 2021.07.26.21261115. Available from: https://www.medrxiv.org/content/10.1101/2021.07.26.21261115v1
3. Li J, Zhu P, Tang X, Jiang L, Li Y, Yan K, et al. Combined effect of D-dimer, hs-CRP, and Lp(a) on 5-year clinical outcomes after percutaneous coronary intervention: A large real-world study in China. iScience [Internet]. 2023 July 21 [cited 2025 Aug 27];26(7). Available from: https://www.cell.com/iscience/abstract/S2589-0042(23)01107-0
4. Duan S, Li Y, Yang P. Predictive value of blood urea nitrogen in heart failure: a systematic review and meta-analysis. Front Cardiovasc Med [Internet]. 2023 July 31 [cited 2025 Aug 27];10. Available from: https://www.frontiersin.org/journals/cardiovascular-medicine/articles/10.3389/fcvm.2023.1189884/full
5. Ogunmodede JA, Kolo PM, Bojuwoye MO, Dele-Ojo BF, Ogunmodede AJ, Omotoso AB. Characteristics of patients with acute heart failure in North Central Nigeria. Research Journal of Health Sciences. 2021 July 7;9(3):221–35.
6. Karaye KM, Dokainish H, ElSayed A, Mondo C, Damasceno A, Sliwa K, et al. Clinical Profiles and Outcomes of Heart Failure in Five African Countries: Results from INTER-CHF Study. Global Heart [Internet]. 2021 July 30 [cited 2025 Aug 27];16(1). Available from: https://globalheartjournal.com/articles/10.5334/gh.940
7. Feola M, Lombardo E, Testa M, Avogadri E, Piccolo S, Vado A. Prognostic factors of mid-term clinical outcome in congestive heart failure patients discharged after acute decompensation. Arch Med Sci. 2012 July 4;8(3):462–70.
8. Zhang Z, Cao L, Chen R, Zhao Y, Lv L, Xu Z, et al. Electronic healthcare records and external outcome data for hospitalized patients with heart failure. Sci Data. 2021 Feb 5;8(1):46.
9. McGranaghan P, Saxena A, Düngen HD, Rubens M, Appunni S, Salami J, et al. Performance of a cardiac lipid panel compared to four prognostic scores in chronic heart failure. Sci Rep. 2021 Apr 14;11(1):8164.
10. Hidayaturrohman QA, Hanada E. Impact of Data Pre-Processing Techniques on XGBoost Model Performance for Predicting All-Cause Readmission and Mortality Among Patients with Heart Failure. BioMedInformatics. 2024 Dec;4(4):2201–12.
11. Méndez-Bailón M, Iguarán-Bermúdez R, López-García L, Sánchez-Sauce B, Pérez-Mateos P, Barrado-Cuchillo J, et al. Prognostic Value of the PROFUND Index for 30-Day Mortality in Acute Heart Failure. Medicina (Kaunas). 2021 Oct 23;57(11):1150.
12. Chou RH, Huang PH, Hsu CY, Chang CC, Leu HB, Huang CC, et al. Circulating Fibroblast Growth Factor 21 is Associated with Diastolic Dysfunction in Heart Failure Patients with Preserved Ejection Fraction. Sci Rep. 2016 Sept 21;6(1):33953.
13. Kadowaki S, Watanabe T, Otaki Y, Narumi T, Honda Y, Takahashi H, et al. Combined assessment of myocardial damage and electrical disturbance in chronic heart failure. World Journal of Cardiology. 2017 May 26;9(5):457–65.
14. Spoto S, Argemi J, Di Costanzo R, Gavira Gomez JJ, Salterain Gonzales N, Basili S, et al. Mid-Regional Pro-Adrenomedullin and N-Terminal Pro-B-Type Natriuretic Peptide Measurement: A Multimarker Approach to Diagnosis and Prognosis in Acute Heart Failure. Journal of Personalized Medicine. 2023 July;13(7):1155.
15. Olasińska-Wiśniewska A, Urbanowicz T, Perek B, Misterski M, Grodecki K, Grygier M, et al. Predictive Value of Monocyte-To-Lymphocyte Ratio in Differentiating Heart Failure with Reduced Ejection Fraction in Patients with Severe Aortic Stenosis—A Retrospective Analysis. Journal of Clinical Medicine. 2024 Jan;13(20):6249.
16. Tuersun R, Abudouwayiti A, Li YX, Pan Y, Aimaier S, Wen ZY, et al. Serum CA125: a prognostic biomarker for mortality in chronic heart failure. BMC Cardiovascular Disorders. 2025 Mar 27;25(1):227.

17. Liyis BG de, Ciaves AF, Intizam MH, Jusuf PJ, Artha IMJR. Hematological Biomarkers of Troponin, Neutrophil-to-Lymphocyte Ratio, and Monocyte-toLymphocyte Ratio Serve as Effective Predictive Indicators of High-Risk Mortality in Acute. BioMedicine. 2023 Nov 27;13(4):32–43.
18. Ochi Y, Kubo T, Baba Y, Nakashima Y, Ueda M, Takahashi A, et al. Prediction of Medium-Term Mortality in Japanese Patients With Wild-Type Transthyretin Amyloidosis. Circulation Reports. 2020;2(6):314–21.
19. Xu Y, Song C, Wei W, Miao R. Development and validation of a cardiac surgery-associated acute kidney injury prediction model using the MIMIC-IV database. PLOS ONE. 2025 June 12;20(6):e0325151.
20. Nyawo TA, Dludla PV, Mazibuko-Mbeje SE, Mthembu SXH, Nyambuya TM, Nkambule BB, et al. A systematic review exploring the significance of measuring epicardial fat thickness in correlation to B-type natriuretic peptide levels as prognostic and diagnostic markers in patients with or at risk of heart failure. Heart Fail Rev. 2022 Mar 1;27(2):665–75.
21. Pérez-Herrero S, Lorenzo-Villalba N, Urbano E, Sánchez-Sauce B, Aguilar-Rodríguez F, Bernabeu-Wittel M, et al. Prognostic Significance of Lung and Cava Vein Ultrasound in Elderly Patients Admitted for Acute Heart Failure: PROFUND-IC Registry Analysis. Journal of Clinical Medicine. 2022 Jan;11(15):4591.
22. Rehman SU, James L. Januzzi J. Natriuretic Peptide Testing in Primary Care. Current Cardiology Reviews. 4(4):300–8.
23. Noveanu M, Breidthardt T, Potocki M, Reichlin T, Twerenbold R, Uthoff H, et al. Direct comparison of serial B-type natriuretic peptide and NT-proBNP levels for prediction of short- and long-term outcome in acute decompensated heart failure. Critical Care. 2011 Jan 5;15(1):R1.
24. Peeters JMPWU, Sanders-van Wijk S, Bektas S, Knackstedt C, Rickenbacher P, Nietlispach F, et al. Biomarkers in outpatient heart failure management; Are they correlated to and do they influence clinical judgment? Neth Heart J. 2014 Mar 1;22(3):115–21.
25. de Boer RA. Clinical evaluation of heart failure patients: what good are biomarkers bringing us? Neth Heart J. 2014 Mar 1;22(3):113–4.
26. Choi JS, Trinh TX, Ha J, Yang MS, Lee Y, Kim YE, et al. Implementation of Complementary Model using Optimal Combination of Hematological Parameters for Sepsis Screening in Patients with Fever. Sci Rep. 2020 Jan 14;10(1):273.
27. Truslow JG, Goto S, Homilius M, Mow C, Higgins JM, MacRae CA, et al. Cardiovascular Risk Assessment Using Artificial Intelligence-Enabled Event Adjudication and Hematologic Predictors. Circulation: Cardiovascular Quality and Outcomes. 2022 June;15(6):e008007.
28. Goriki Y, Tanaka A, Nishihira K, Kawaguchi A, Natsuaki M, Watanabe N, et al. A Novel Predictive Model for In-Hospital Mortality Based on a Combination of Multiple Blood Variables in Patients with ST-Segment-Elevation Myocardial Infarction. Journal of Clinical Medicine. 2020 Mar;9(3):852.
29. Yan Q, Chen S. Hemoconcentration is a valuable predictor of prognosis in patients with acute heart failure. Experimental and Therapeutic Medicine. 2020 Apr 1;19(4):2792–8.
30. Kim KH, Park JH, Ro YS, Hong KJ, Song KJ, Shin SD. Emergency department routine data and the diagnosis of acute ischemic heart disease in patients with atypical chest pain. PLOS ONE. 2020 Nov 5;15(11):e0241920.
31. Zhang Z, Shao B, Liu H, Huang B, Gao X, Qiu J, et al. Construction and Validation of a Predictive Model for Coronary Artery Disease Using Extreme Gradient Boosting. JIR. 2024 July 1;17:4163–74.
32. Vinci R, d'Aiello A, Costantino S, Capocchiano ND, Matamalas JT, Pedicino D. Editorial: Artificial intelligence applied to coronary artery diseases: from pathophysiology to precision medicine. Front Cardiovasc Med [Internet]. 2023 Aug 9 [cited 2025 Aug 27];10. Available from: https://www.frontiersin.org/journals/cardiovascular-medicine/articles/10.3389/fcvm.2023.1247620/full
33. Wang Z, Gu Y, Huang L, Liu S, Chen Q, Yang Y, et al. Construction of machine learning diagnostic models for cardiovascular pan-disease based on blood routine and biochemical detection data. Cardiovascular Diabetology. 2024 Sept 28;23(1):351.
34. Cadamuro J. Rise of the Machines: The Inevitable Evolution of Medicine and Medical Laboratories Intertwining with Artificial Intelligence—A Narrative Review. Diagnostics. 2021 Aug;11(8):1399.
35. Wickramaratne SD, Shaad Mahmud M. Bi-Directional Gated Recurrent Unit Based Ensemble Model for the Early Detection of Sepsis. In: 2020 42nd Annual International Conference of the IEEE Engineering

in Medicine & Biology Society (EMBC) [Internet]. 2020 [cited 2025 Aug 27]. p. 70–3. Available from: https://ieeexplore.ieee.org/document/9175223
36. Olasińska-Wiśniewska A, Urbanowicz T, Perek B, Misterski M, Grodecki K, Grygier M, et al. Predictive Value of Monocyte-To-Lymphocyte Ratio in Differentiating Heart Failure with Reduced Ejection Fraction in Patients with Severe Aortic Stenosis—A Retrospective Analysis. Journal of Clinical Medicine. 2024 Jan;13(20):6249.
37. Pereira-Barretto AC, Oliveira Junior MT de, Strunz CC, Del Carlo CH, Scipioni AR, Ramires JAF. O nível sérico de NT-proBNP é um preditor prognóstico em pacientes com insuficiência cardíaca avançada. Arq Bras Cardiol. 2006;87:174–7.
38. Ikonomidis I, Michalakeas CA, Lekakis J, Paraskevaidis I, Th. Kremastinos D. Multimarker Approach in Cardiovascular Risk Prediction. Disease Markers. 2009;26(5–6):135423.
39. Oommen SG, Man RK, Talluri K, Nizam M, Kohir T, Aviles MA, et al. Heart Failure With Improved Ejection Fraction: Prevalence, Predictors, and Guideline-Directed Medical Therapy. Cureus. 16(6):e61790.
40. Akhtar M, Elliott PM. Risk Stratification for Sudden Cardiac Death in Non-Ischaemic Dilated Cardiomyopathy. Curr Cardiol Rep. 2019 Nov 25;21(12):155.
41. Sadahiro T, Kohsaka S, Okuda S, Inohara T, Shiraishi Y, Kohno T, et al. MRI and serum high-sensitivity C reactive protein predict long-term mortality in non-ischaemic cardiomyopathy. Open Heart [Internet]. 2015 Oct 19 [cited 2025 Aug 27];2(1). Available from: https://openheart.bmj.com/content/2/1/e000298
42. Méndez-Bailón M, Iguarán-Bermúdez R, López-García L, Sánchez-Sauce B, Pérez-Mateos P, Barrado-Cuchillo J, et al. Prognostic Value of the PROFUND Index for 30-Day Mortality in Acute Heart Failure. Medicina. 2021 Nov;57(11):1150.
43. Dong M, Wang L, Tse G, Dai T, Wang L, Xiao Z, et al. Effectiveness and safety of transcatheter aortic valve replacement in elderly people with severe aortic stenosis with different types of heart failure. BMC Cardiovascular Disorders. 2023 Jan 18;23(1):34.
44. Stock EO, Asztalos BF, Miller JM, He L, Creasy KT, Schwemberger R, et al. High-Density Lipoprotein Particles, Inflammation, and Coronary Heart Disease Risk. Nutrients. 2025 Jan;17(7):1182.
45. Soni N, Kaur P, Gurjar V, Bhargava A, Tiwari R, Chouksey A, et al. Unveiling the Interconnected Dynamics of Mitochondrial Dysfunction Associated With Age-Related Cardiovascular Risk: A Cross-Sectional Pilot Study. Cureus. 17(4):e82961.
46. Meyer L, Mulder D, Wallace J. A supervised machine learning approach with feature selection for sex-specific biomarker prediction. npj Syst Biol Appl. 2025 July 1;11(1):69.
47. Shou L, Huang WW, Barszczyk A, Wu SJ, Han H, Waese-Perlman A, et al. Blood Biomarkers Predict Cardiac Workload Using Machine Learning. BioMed Research International. 2021;2021(1):6172815.
48. Pan Y, Wei M, Jin M, Liang Y, Yi T, Tu J, et al. An interpretable machine learning model based on optimal feature selection for identifying CT abnormalities in patients with mild traumatic brain injury. eClinicalMedicine [Internet]. 2025 Apr 1 [cited 2025 Aug 27];82. Available from: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(25)00124-5/fulltext