JIRRS
Journal International Review of Research Studies

Received: 01/06/2026 - **Accepted:** 06/07/2026
Vol: 01.09
DOI: 10.66104/bwtkpv61
Pages: 1-11

MACHINE LEARNING-BASED PREDICTION OF HEART FAILURE MORTALITY USING CLINICAL BIOMARKERS AND IMBALANCED DATA STRATEGIES

Abiodun OJO [1]
1 Sheffield Hallam University; jeoph8501@gmail.com; https://orcid.org/0009-0002-5679-7373

Corresponding author: jeoph8501@gmail.com

Abstract

Heart failure continues to account for high global mortality rates despite improvements in cardiovascular medicine. Early mortality risk detection is essential for healthcare resource planning and personalised intervention. This study develops a machine learning framework for mortality prediction in patients with chronic heart failure, using demographic, clinical and biochemical data. A dataset containing 299 patient records and 13 biomarkers was pre-processed, normalised, and evaluated to identify feature-outcome relationships. Several modelling strategies were assessed, including Random Forest and Logistic Regression. Class imbalance was handled using Tomek Links, SMOTE and Random Oversampling to reduce decision boundary distortion. Random Forest trained on Tomek Links rebalanced data produced the strongest performance, achieving test accuracy of 0.91 and a ROC-AUC of 0.94. Logistic Regression achieved lower accuracy (0.82) and weaker minority recall. The results highlight the effect of biomarker depth on model separability, where serum creatinine, ejection fraction and age were the strongest predictors. The outcomes demonstrate that machine learning models can effectively identify high-risk patients and that minority data resampling significantly improves mortality classification reliability. The approach offers potential value for clinical decision support systems and prioritised care pathways.

Keywords: Artificial Intelligence; Business; Cyberattack; Cyber security

1. Introduction

Heart failure (HF) is a complex and life-threatening syndrome that profoundly affects quality of life and is a significant contributor to morbidity and mortality worldwide (Serianu Limited, 2025; Africa Data Protection Association, 2024; Savarese, Becher & Lund, 2023; African Union, 2020). In contrast to heart failure with reduced ejection fraction (HFrEF), heart failure with preserved ejection fraction (HFpEF) is characterized by compromised diastolic function, resulting in elevated filling

JIRRS
Journal International Review of Research Studies
Received: 01/06/2026 - **Accepted:** 06/07/2026
Vol: 01.09
DOI: 10.66104/bwtkpv61
Pages: 1-11

pressure and pulmonary congestion (Check Point Software Technologies, 2025; African Union 2024; Redfield & Borlaug, 2023).

Divorced or separated individuals face a significantly higher risk of adverse cardiosvascular outcomes and early mortality compared to continuously married individuals. For heart failure patients, being unmarried is independently associated with an approximately 50% to 80% higher risk of cardiovascular death and re-hospitalization (Akinloye, 2024).

For heart failure among older adults, active regular communication and society connection with family including grandchildren are powerful protective factors (Akinloye, 2026). They improve emotional well-being and are tied to better medication (Dare, 2024). It is expedient to note that heart failure mortality prediction is improved by sustainable logistics, such that standardizing the quality of clinical data and into predictive models and supporting real-time, and remote patient care (Olaniyi, Bamise, Gbadegesin & Oloyede 2026; Adeniran, Olorunfemi & Akanmu, 2025).

The prevalence of HFpEF is increasing; the condition accounts for nearly one-half of all HF cases (Deloitte, 2026; CIO Africa, 2025; Shah, Xu, & Matsouaka, 2023). Individuals with HFpEF are at increased risks of recurrent HF-related hospitalizations and poor clinical outcomes and have a 5-year survival rate of <50%, similar to that of patients with HFrEF; HFpEF is thus a substantial public health burden (Deloitte, 2026; CIO Africa, 2025; Shah, Xu & Matsouaka, 2023). Due to the distinct pathophysiological mechanisms of HFpEF and HFrEF, the risk factors influencing clinical outcomes must be assessed separately for each condition.

Numerous risk calculation models are available for predicting the clinical outcomes of patients with HF; however, few specifically focus on patients with HFpEF (Digital Policy Alert 2026; Oladunmade, 2026; Adeniran, Adeniran, Familusi & Adedayo, 2024; Peterson, Rumsfeld, & Liang, 2010; Omigbodun & Akinloye, 2026; ESET 2026; IT News Africa 2026; Interpol 2025; African Union 2022; Pocock, Ariti & McMurray, 2013); Esentry, 2025; Kreso, 2025; Ken Research, 2025; Interpol, 2024; Makoshi, 2025; Lagu, Pekow & Shieh, 2016); (Mordor Intelligence, 2026; Omigbodun, Familusi, Akinsola, Adeniran & Adeniran 2026; Suzuki, Yoshihisa & Sato, 2018).

The I-PRESERVE (Irbesartan in Heart Failure with Preserved Ejection Fraction Study) is one of the few predictive investigations focusing on patients with HFpEF (Ecofin Agency 2026; Olaniyi, Ayomide, Ogieva, & Ogwuche 2026; Smart Security Solutions. 2026; TechAfrica News 2026; Dare, 2025; Komajda, Carson & Hetzel, 2011). The I-PRESERVE trial utilized multivariable Cox regression models to identify factors associated with the outcomes of HFpEF.

Nevertheless, the study was limited because it excluded patients with moderate or severe renal dysfunction and because it did not perform validation in an independent cohort. In another study, data from the TOPCAT (Treatment of Preserved Cardiac Function Heart Failure with an Aldosterone Antagonist) trial were utilized to formulate models for predicting mortality and HF hospitalization of

JIRRS
Journal International Review of Research Studies
Received: 01/06/2026 - **Accepted:** 06/07/2026
Vol: 01.09
DOI: 10.66104/bwtkpv61
Pages: 1-11

outpatients with HFpEF (Olaniyi, Ayomide, Ogieva, & Ogwuche 2026; SA Profile. 2025; Angraal, Mortazavi & Gupta, 2020). However, that study was limited by the lack of follow-up data and did not include time-to-event analysis. Additionally, neither the I-PRESERVE nor the TOPCAT trial studies included echocardiographic parameters in their prediction models. Consequently, a substantial gap regarding the prediction of clinical outcomes for patients with HFpEF remains in the published data.

Novel machine learning methods, such as random survival forest (RSF), can accommodate high-dimensional and nonlinear effects of variables, enabling identification of the effects of individual variables from large variable sets (World Economic Forum. 2026; SOCRadar. 2025; KnowBe4, 2025; Usercentrics, 2025; World Bank, 2021; Dietrich, Floegel & Troll, 2016). These methods enable the calibration and integrated analysis of factors encompassing demographic, clinical, biological, and echocardiographic variables to identify predictors associated with adverse clinical outcomes. The present study employed RSF to formulate and validate models predicting HF hospitalization and cardiovascular (CV) death in patients with HFpEF.

Heart failure mortality prediction is a long-standing problem in cardiovascular medicine. Traditional scoring approaches such as Cox regression struggle to generalize across heterogeneous patient cohorts because they assume linear feature outcome interactions. Machine learning techniques, by contrast, operate without strict statistical assumptions and are more robust to non-linear biomarker variation and multicollinearity.

The dataset used in this study comprises 299 patients diagnosed with heart failure, with death outcomes recorded at follow-up. The dataset includes key biomarkers: ejection fraction, serum creatinine, serum sodium, platelet count, high blood pressure, diabetes status and creatinine phosphokinase levels. These features are known to influence cardiovascular function, renal performance and metabolic stress.

A key limitation in classification tasks within medical domains is the imbalance between event and non-event outcomes: approximately 32% of cases in this study correspond to patient mortality, introducing model bias favouring survival classification. Class imbalance tends to weaken minority recall in standard classifiers, causing unacceptable false-negative risk in clinical applications.

This paper proposes a machine learning framework that applies targeted resampling and data scaling strategies to improve mortality classification accuracy. The investigation focuses exclusively on the death outcome, allowing the modelling task to be tuned for binary discrimination precision.

2. Literature Review

Existing work such as Savarese, Becher and Lund (2023); Lagu, Pekow and Shieh (2016); Yoshihisa and Sato (2018), demonstrates the viability of machine learning in heart failure prognosis; however, many studies such as Komajda, Carson and

JIRRS
Journal International Review of Research Studies
Received: 01/06/2026 - **Accepted:** 06/07/2026
Vol: 01.09
DOI: 10.66104/bwtkpv61
Pages: 1-11

Hetzel (2011), Redfield and Borlaug (2023), rely on logistic regression and do not address imbalance with advanced techniques. Ensemble methods such as Random Forest have shown improved generalisation, but the literature lacks consistent comparison between undersampling and oversampling on mortality outcomes. In addition, several studies report model accuracy alone without deeper inspection of ROC-AUC, recall or false-negative behaviour. This research builds on prior methodologies and introduces a multi-strategy imbalance comparison to evaluate boundary refinement effects.

3. Methods

A. Dataset:

The patient dataset contains 13 continuous and categorical features and one binary output variable (death_event). No missing values were detected.

B. Biomarker Correlation Matrix

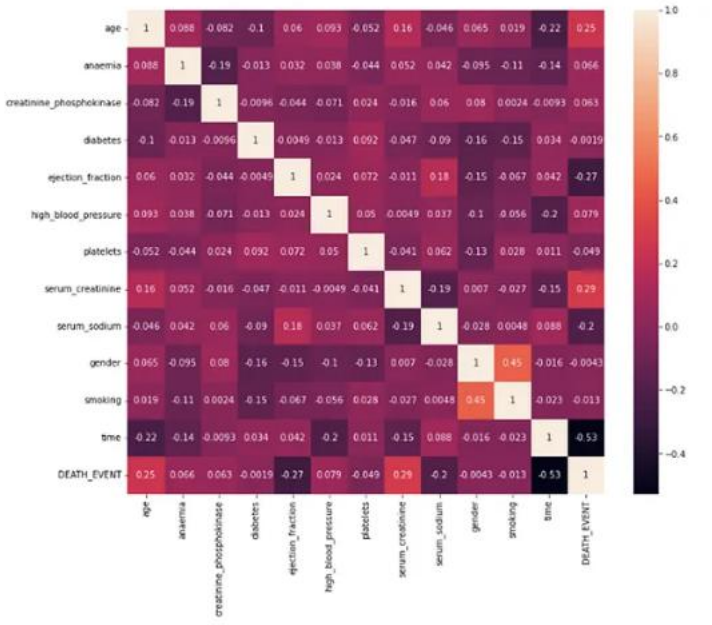


fig-1 correlation matrix

The correlation heatmap analysis indicates that mortality is positively associated with age (r = 0.25) and serum creatinine (r = 0.29), suggesting renal impairment and ageing strongly influence death outcomes. In contrast, ejection fraction shows a negative correlation with mortality (r = - 0.27), confirming reduced cardiac output as a major contributor to patient death. Serum sodium also demonstrates a mild negative trend (r = -0.20), supporting its clinical relevance in heart failure progression.

JIRRS
Journal International Review of Research Studies
Received: 01/06/2026 - **Accepted:** 06/07/2026
Vol: 01.09
DOI: 10.66104/bwtkpv61
Pages: 1-11

C. Data Splitting
X and y were separated and partitioned into 70% training and 30% testing sets.

D. Data Normalisation
Variables exhibited scale disparities between clinical biomarkers, so StandardScaler was applied to ensure stable model convergence.

E. Imbalance Processing
Multiple imbalance techniques were tested:

Tomek Links undersampling removed ambiguous majority-class survivors near minority-class decision boundaries.
Random Oversampling increased minority event density.
SMOTE synthetically interpolated minority cases.

F. Trained Models
Two primary classifiers were evaluated:
1. Random Forest Classifier (with parameter optimisation)
2. Logistic Regression

All models were implemented using scikit-learn libraries.

G. Performance Metrics
Evaluation included:

- Accuracy
- Precision
- Recall
- F1-score
- ROC-AUC

IV. Results

A. Random Forest (Tomek Links)
The Random Forest model trained on Tomek Links processed data delivered:

- Accuracy: 0.91
- ROC-AUC: 0.94
- Balanced precision and recall

JIRRS
Journal International Review of Research Studies
Received: 01/06/2026 - **Accepted:** 06/07/2026
Vol: 01.09
DOI: 10.66104/bwtkpv61
Pages: 1-11

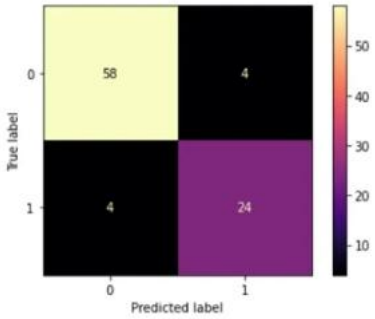


fig-2. Confusion matrix

B. Random Forest (Oversampling)

Oversampling yielded:

- Accuracy: 0.88
- ROC-AUC: 0.93

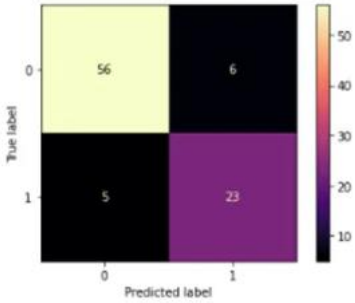


fig-3. Confusion matrix

C. Logistic Regression (baseline)

Logistic Regression tested on original data achieved:

- Accuracy: 0.82
- ROC-AUC: 0.86
- Lower recall for mortality class

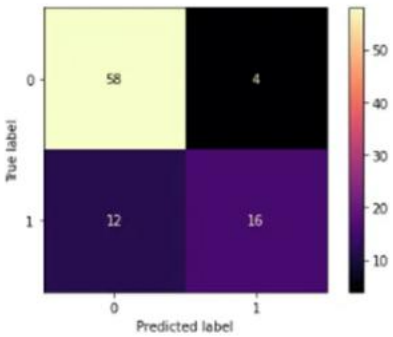


fig-4. Confusion matrix

JIRRS
Journal International Review of Research Studies
Received: 01/06/2026 - **Accepted:** 06/07/2026
Vol: 01.09
DOI: 10.66104/bwtkpv61
Pages: 1-11

4. Discussion

The findings confirm mortality prediction feasibility using clinical biomarkers. The Random Forest algorithm demonstrated superior classification strength due to its ability to capture non-linear biomarkers. Imbalance processing significantly changed model sensitivity to the minority class, with Tomek Links refinement providing the best risk detection.

From a clinical perspective, strong mortality detection at high accuracy could enable triage support, planning of treatment priority and patient counselling.

5. Conclusion

This research shows that mortality prediction in heart failure can be effectively achieved using machine learning. The strongest configuration integrated Random Forest modelling with Tomek Links boundary cleaning. Future work should include larger patient samples, XAI visualisation and cross-cohort generalisation.

Dataset

The patient dataset contains 13 continuous and categorical features and one binary output variable (death_event). No missing values were detected.

Funding

This research received no external funding.

Conflicts of Interest

The author declare no conflict of interest.

References

Adeniran, A. A., Adeniran, A. O., Familusi, O. B. & Adedayo, O. (2024). The outlook of cyber security in African businesses: issues and way-out. Management Analytics and Social Insights, 1 (2), 260-271. https://doi.org/10.22105/SA.2021.281500.1061

Adeniran, A. O., Olorunfemi, S.O., & Akanmu, A. A. (2025). Exploring Sustainable E-Commerce Logistics Services for Customer Satisfaction. In: Emmanuel, M. Ogochukwu, U. and Chinebuli, W. (Eds.) Transport Management Services in Africa, Vol. 1, Taylor & Francis, Pp. 30. https://doi.org/10.4324/9781003560074-6

JIRRS
Journal International Review of Research Studies
Received: 01/06/2026 - **Accepted:** 06/07/2026
Vol: 01.09
DOI: 10.66104/bwtkpv61
Pages: 1-11

Africa Data Protection Association (2024). Africa Data Protection Report: Nigeria Fines Meta a Record $220 Million. Douala: Africa Data Protection Association. https://www.africadataprotection.org/report-dec-2024

African Union (2020). The Digital Transformation Strategy for Africa (2020-2030). Addis Ababa: African Union Commission. https://au.int/en/documents/20200518/digital-transformation-strategy-africa-2020-2030

African Union. (2022). AU Data Policy Framework. Addis Ababa: African Union Commission. https://au.int/sites/default/files/documents/42078-doc-DATA-POLICY-FRAMEWORKS-2024-ENG-V2.pdf

African Union. (2024). African Digital Compact: A Unified Resolve for Africa's Digital Future (V4). Addis Ababa: African Union Commission. https://au.int/sites/default/files/documents/44005-doc-AU_Digital_Compact_V4.pdf

Akinloye, D. N. (2024). Grandparent-grandchild communication: synthesis of integrative model. Psychology Nexus, 1(1), 1-15

Akinloye, D. (2026). Exploring the magnitude of Artificial Intelligence on family communication. Journal International Review of Research Studies, 1(06), 1-21.

Angraal, S., Mortazavi, B. J., & Gupta, A. (2020). Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. JACC Heart Fail., 8:12–21. https://doi.org/10.1016/j.jchf.2019.06.013

Check Point Software Technologies. (2025). African Perspectives on Cyber Security Report 2025. San Carlos: Check Point Software Technologies Ltd. https://www.checkpoint.com/security-report

CIO Africa. (2025). Africa's Digital Growth Outpacing Cybersecurity. Nairobi: Chief Information Officer Africa Ltd. https://cioafrica.co/africas-digital-growth-outpacing-cybersecurity/

Dare, A. (2024). Communication and online platforms: pitfalls, possibilities, and pressures.ttu-ir.tdl.org.

Dare, A. (2025). Communication pattern and the influence of social media contents on young adults' wellbeing in Texas, United States of America. Journal of Communication, 6(12), 1-22. https://doi.org/10.46303/ressat.2023.13

Deloitte (2026). Nigeria Cybersecurity Outlook 2026. Lagos: Deloitte Nigeria. https://www2.deloitte.com/ng/en/pages/risk/articles/cybersecurity-outlook.html

Dietrich, S., Floegel, A., & Troll, M. (2016). Random Survival Forest in practice: a method for modeling complex metabolomics data in time to event analysis. Int J Epidemiol., 45:1406–1420. https://doi.org/10.1093/ije/dyw145

JIRRS
Journal International Review of Research Studies

Received: 01/06/2026 - **Accepted:** 06/07/2026
Vol: 01.09
DOI: 10.66104/bwtkpv61
Pages: 1-11

Digital Policy Alert (2026). The Year of the Teeth: Data Protection in Africa Roundup 2025 and Projections for 2026. St. Gallen: SGEPT. https://digitalpolicyalert.org/blog/data-protection-in-africa-roundup

Ecofin Agency (2026). Cyber risks top concerns for African businesses in 2026 report. Geneva: Ecofin Agency. https://www.ecofinagency.com/news-digital/1201-51832-cyber-risks-top-concerns-for-african-businesses-in-2026-report

Esentry (2025). Eagle's Eyes Q3 2025 Report: Monitoring Data Breaches and Identity Intrusions. Lagos: esentry.io. https://www.esentry.io/reports/q3-2025

ESET (2026). Cybersecurity for renewable energy: Best practices to protect critical infrastructure. Bratislava: ESET Research. https://www.eset.com/blog/business/renewable-energy-cybersecurity

Interpol (2024). Interpol African Cyberthreat Assessment Report 2024 Outlook. Lyon: INTERPOL Cybercrime Directorate. https://www.interpol.int/en/Crimes/Cybercrime/Africa-Cyber-Threat-Assessment-Report-2024

Interpol (2025). Africa Cyberthreat Assessment Report 2025 (4th Edition). Lyon: INTERPOL Cybercrime Directorate. https://www.interpol.int/en/How-we-work/Criminal-Intelligence/Africa-cyberthreat-report-2025

IT News Africa (2026). 8 Key Trends that will Define Africa's Cyber Security Landscape in 2026. Johannesburg: IT News Africa. https://www.itnewsafrica.com/2026/01/8-key-trends-that-will-define-africas-cyber-security-landscape-in-2026/

Ken Research. (2025). Morocco Cybersecurity (MDR & SOC) Market Analysis. Gurgaon: Ken Research Pvt. Ltd. https://www.kenresearch.com/morocco-cybersecurity-mdr-soc-market

KnowBe4. (2025). African Cybersecurity & Awareness Report 2025. Cape Town: KnowBe4 Africa. https://www.knowbe4.com/hubfs/Africa-Annual-Survey_Whitepaper_US_EN-F.pdf.

Komajda, M., Carson, P. E., & Hetzel, S. (2011). Factors associated with outcome in heart failure with preserved ejection fraction: findings from the Irbesartan in Heart Failure with Preserved Ejection Fraction Study (I-PRESERVE). Circ Heart Fail., 4:27-35. https://doi.org/10.1161/circheartfailure.109.932996

Kreso, I. (2025). Cyber threats and vulnerability mapping in the energy sector: Laying the groundwork for smart grid resilience. Applied Cybersecurity & Internet Governance, 4 (1). https://doi.org/10.60097/ACIG/211124

Lagu, T., Pekow, P. S., & Shieh, M. S. (2016). Validation and comparison of seven mortality prediction models for hospitalized patients with acute decompensated heart failure. Circ Heart Fail., 9(8):e002912. https://doi.org/10.1161/circheartfailure.115.002912

JIRRS
Journal International Review of Research Studies

Received: 01/06/2026 - **Accepted:** 06/07/2026
Vol: 01.09
DOI: 10.66104/bwtkpv61
Pages: 1-11

Makoshi, S. M. (2025). Cybersecurity Challenges in Nigeria: Threat Landscape, Economic Impact, and Strategic Solutions. Zenodo. https://doi.org/10.5281/zenodo.15144077

Mordor Intelligence. (2026). Nigeria Cybersecurity Market Size & Share Analysis - Growth Trends and Forecast (2026 - 2031). Hyderabad: Mordor Intelligence. https://www.mordorintelligence.com/industry-reports/nigeria-cybersecurity-market

Oladunmade, M. (2026). Exclusive: CinetPay customers owed over $1 million months after alleged cyberattack. TechCabal. https://techcabal.com/2026/02/01/cinetpay-cyberattack-settlement-backlog/

Olaniyi, A. A., Ayomide, A. A., Ogieva, M. O., & Ogwuche, G. (2026). Adoption of Intelligent Transport Systems (ITS) in Urban Transportation Planning. Discover Global Society, 4(13), 1-23. https://doi.org/10.1007/s44282-025-00312-3

Olaniyi, A. A., Bamise, E. O., Gbadegesin, A. A., & Oloyede, F. S. (2026). Green Practices, Guest Satisfaction and Loyalty in the Hotel Industry. Journal International Review of Research Studies, 01(04), 1-https://doi.org/10.66104/cnpnrz3

Omigbodun, O. T., Familusi, O. B., Akinsola, S. A., Adeniran, A. A., & Adeniran, A. O. (2026). AI-Powered Workflow Automation in Small Businesses. Open Journal of Economics and Commerce, 7(1): 13-45. https://doi.org/10.22259/2638-549X.0701003

Omigbodun, O. T., & Akinloye, D. (2026). Context and Application of Agnew's General Strain Theory to Explain Crime. Journal International Review of Research Studies,1(06), 1-11.

Peterson, P. N., Rumsfeld, J. S., & Liang, L. (2010). A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association Get With The Guidelines program. Circ Cardiovasc Qual Outcomes, 3:25-32. https://doi.org/10.1161/circoutcomes.109.854877

Pocock, S. J., Ariti, C. A., & McMurray, J. J. (2013). Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. Eur Heart J. 2013;34:1404-1413. https://doi.org/10.1093/eurheartj/ehs337

Redfield, M. M., & Borlaug, B. A. (2023). Heart failure with preserved ejection fraction: a review. JAMA., 329:827–838. https://doi.org/10.1001/jama.2023.2020

SA Profile. (2025). South African Businesses Are Underestimating Cyber Risks. Cape Town: Media Xpose. https://www.saprofilemagazine.co.za/business-news/south-africa-cybersecurity-warning-wwise-2025/

Savarese, G., Becher, P. M., & Lund L. H. (2023). Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res., 118:3272-3287. https://doi.org/10.1093/cvr/cvac013.

JIRRS
Journal International Review of Research Studies

Received: 01/06/2026 - **Accepted:** 06/07/2026
Vol: 01.09
DOI: 10.66104/bwtkpv61
Pages: 1-11

Serianu Limited. (2025). Africa Cybersecurity Report 2024/2025: From Risk to Resilience. Nairobi: Serianu. https://www.serianu.com/downloads/Africa%20Cybersecurity%20Report%20-%20Kenya.pdf

Shah, K. S., Xu, H., & Matsouaka, R. A. (2023). Heart failure with preserved, borderline, and reduced ejection fraction: 5-year outcomes. J Am Coll Cardiol., 70:2476–2486. https://doi.org/10.1016/j.jacc.2017.08.074

Smart Security Solutions. (2026). Factories, grids, and finance: Critical infrastructure cyber lessons of 2025. Johannesburg: Technews Publishing. https://www.securitysa.com/26388r

SOCRadar. (2025). Africa Threat Landscape Report 2025. Middletown: SOCRadar Extended Threat Intelligence. https://socradar.io/resources/report/africa-threat-landscape-report-2025/

Suzuki, S., Yoshihisa, A., & Sato, Y. (2018). Clinical significance of Get With the Guidelines-Heart Failure Risk Score in patients with chronic heart failure after hospitalization. J Am Heart Assoc., 7:e008316. https://doi.org/10.1161/jaha.117.008316

TechAfrica News. (2026). Cybersecurity Tops Risk List for African Businesses in 2026. Accra: TechAfrica News. https://techafricanews.com/2026/01/12/cybersecurity-tops-risk-list-for-african-businesses-in-2026-says-new-report/

Usercentrics. (2025). Global Data Privacy Laws: Your 2025 Guide (GDPR, CCPA, More). Munich: Usercentrics GmbH. https://usercentrics.com/guides/data-privacy-laws/

World Bank. (2021). Data for Better Lives. Washington, D.C.: World Bank Publications. https://doi.org/10.1596/978-1-4648-1600-0

World Economic Forum. (2026). Global Cybersecurity Outlook 2026. Geneva: World Economic Forum in collaboration with Accenture. https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2026.pdf

Author Responsibility Statement

AO: Conceptualization, Writing - original draft, Introduction, Methods, Analysis, Discussion, Conclusion. The author read and approved the final manuscript.