International Journal of Medical Informatics 211 (2026) 106313

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Smart insurance analytics: A novel ensemble feature selection approach to unlock health insurance coverage predictions in Sierra Leone

David B. Olawade [a,b,c,*], Augustus Osborne [d], Afeez A. Soladoye [e,f], Olaitan E. Oluwadare [g], Emmanuel O. Awogbindin [h], Ojima Z. Wada [i]

[a] Department of Allied and Public Health, School of Health, Sport and Bioscience, University of East London, London, United Kingdom
[b] Department of Research and Innovation, Medway NHS Foundation Trust, Gillingham ME7 5NY, United Kingdom
[c] Department of Public Health, York St John University, London, United Kingdom
[d] Institute for Development, Western Area, Freetown, Sierra Leone
[e] Department of Computer Engineering, Federal University, Oye-Ekiti, Nigeria
[f] Department of Computer Engineering, Adeleke University, Ede, Nigeria
[g] Division of Physics, Engineering, Mathematics and Computer Science, Delaware State University, USA
[h] Department of Computing and Information Science, Bamidele Olumilua University of Education, Science and Technology, Ikere-Ekiti, Ekiti State, Nigeria
[i] College of Science and Engineering, Division of Sustainable Development, Hamad Bin Khalifa University, Qatar Foundation, Education City, Doha, Qatar

ARTICLE INFO

Keywords:
Health Insurance Prediction
Machine Learning
Ensemble Feature Selection
Random Forest
XGBoost

ABSTRACT

Background: Predicting health insurance uptake remains a critical challenge for policymakers and insurance providers seeking to optimise coverage strategies and resource allocation. In Sierra Leone, health insurance uptake remains extremely low, and understanding determinants is vital for universal health coverage goals.
Objective: To develop and evaluate an innovative ensemble feature selection methodology for health insurance uptake prediction, establishing new performance benchmarks through systematic comparison of multiple machine learning algorithms using comprehensive validation strategies.
Methods: This study employed supervised machine learning to predict health insurance uptake among 15,574 women using data from the 2019 Sierra Leone Demographic and Health Survey (SLDHS). We implemented an ensemble feature selection approach that requires consensus across Adaptive Ant Colony Optimisation, Recursive Feature Elimination, and Backwards Elimination techniques. Seven algorithms were systematically compared: Logistic Regression, Support Vector Machines, K-Nearest Neighbors, Random Forest, Gradient Boosting, XGBoost, and LightGBM. SMOTE addressed class imbalance, whilst validation employed nested 5-fold cross-validation, 10-fold cross-validation, and hold-out testing to prevent information leakage.
Results: Random Forest achieved exceptional performance with 0.9973 accuracy, 0.9973 precision, 0.9973 recall, 0.9973 F1-score, and perfect 1.0000 ROC AUC on hold-out testing. XGBoost delivered comparable results with 0.9914 across all metrics and 0.9998 ROC AUC. Backward Feature Elimination consistently yielded superior results across ensemble methods. However, the near-perfect performance warrants cautious interpretation and requires external validation to confirm generalizability.
Conclusions: This research establishes new performance benchmarks for health insurance prediction, significantly exceeding existing literature, which has direct implications for health insurance policy and practice in Sierra Leone. The innovative ensemble feature selection methodology provides a robust framework for enhancing prediction accuracy across healthcare applications, offering immediate practical value for stakeholders. Future work should prioritize external validation, explainability analysis, and temporal stability assessment to ensure practical deployment readiness.

1. Introduction

The global health insurance landscape faces unprecedented challenges as healthcare costs continue to rise and populations age, creating urgent demands for sophisticated prediction models that can accurately forecast insurance uptake behaviour [1,2]. While high-

* Corresponding author..
E-mail address: d.olawade@uel.ac.uk (D.B. Olawade).

https://doi.org/10.1016/j.ijmedinf.2026.106313
Received 1 August 2025; Received in revised form 25 January 2026; Accepted 26 January 2026
Available online 27 January 2026

income countries, such as the United States, grapple with ageing populations, workforce shortages, and system fragmentation [3], low-income countries, like Sierra Leone, confront even more fundamental barriers to financial protection in health. In Sierra Leone, less than 7% of the population is covered by any form of health insurance, with the vast majority of healthcare expenses paid out-of-pocket, a situation that exacerbates financial hardship, particularly among women, children, and other vulnerable groups [4]. The absence of a robust health insurance system in Sierra Leone directly undermines progress toward universal health coverage (UHC) and heightens the risk of catastrophic health expenditures for millions.

Traditional actuarial approaches, though historically valuable, often struggle to capture the complex, non-linear relationships that characterise modern healthcare decision-making, especially in contexts where social, economic, and demographic variables interact in unpredictable ways [5]. The integration of demographic factors, socioeconomic status, health indicators, and behavioural patterns requires analytical frameworks capable of processing multidimensional data while maintaining interpretability for practical policy and operational use. For Sierra Leone, where data quality and completeness are often limited, and health financing systems are still developing, the need for robust, adaptable, and accurate prediction models is even more acute.

Machine learning has emerged as a transformative approach for addressing these prediction challenges, offering powerful tools for pattern recognition and forecasting in diverse healthcare settings [6,7]. Recent advances in ensemble methods, feature selection techniques, and validation frameworks have demonstrated significant potential to enhance prediction accuracy beyond what traditional statistical methods can achieve. However, the application of these advanced machine learning approaches to health insurance uptake prediction remains underexplored in low-resource settings like Sierra Leone, where most existing studies are constrained by small sample sizes, reliance on single-algorithm approaches, or inadequate validation strategies, factors that limit their practical deployment and policy relevance [8].

The increasing digitisation of the insurance industry, even in countries like Sierra Leone, is generating richer datasets that capture a wide range of individual characteristics, health behaviours, and coverage decisions. This data abundance offers new opportunities for more sophisticated and contextually relevant modelling, but also introduces challenges related to data complexity and the extraction of actionable insights [9]. The ability to accurately predict insurance uptake has profound implications for risk assessment, premium setting, targeted outreach, and the development of evidence-based public health policies [6]. In Sierra Leone, improving health insurance uptake is a key strategic goal identified in the National Health Sector Strategic Plan (NHSSP 2017–2021), yet there remains a scarcity of data-driven approaches to understanding the determinants of insurance enrolment and retention.

Contemporary research in health insurance prediction has begun to explore machine learning applications, with varying degrees of success across different algorithmic approaches [7,10]. Nonetheless, significant gaps persist, particularly in the areas of ensemble feature selection methodologies, comprehensive validation strategies, and the systematic comparison of multiple algorithms under identical conditions [11,12]. Most prior studies employ single feature selection techniques, limited cross-validation, or focus on narrow demographic segments, thereby limiting their generalizability and practical applicability in diverse and dynamic healthcare environments such as Sierra Leone.

This study addresses these critical gaps by developing and evaluating an innovative ensemble feature selection methodology for health insurance uptake prediction in Sierra Leone, employing consensus-based feature identification across multiple algorithms to enhance robustness and minimise algorithmic bias. The research aims to establish new performance benchmarks through a systematic comparison of seven machine learning algorithms, utilising comprehensive validation strategies including 5-fold cross-validation, 10-fold cross-validation, and rigorous hold-out testing. The primary objectives are: (1) to develop a novel ensemble feature selection approach requiring consensus across Adaptive Ant Colony Optimization, Recursive Feature Elimination, and Backward Elimination; (2) to systematically compare the performance of Logistic Regression, Support Vector Machines, K-Nearest Neighbors, Random Forest, Gradient Boosting, XGBoost, and LightGBM algorithms; (3) to implement comprehensive data preprocessing, including SMOTE for class imbalance handling; (4) to establish robust validation frameworks ensuring practical deployment confidence; and (5) to provide actionable insights for insurance providers, policymakers, and public health professionals in Sierra Leone seeking to enhance their prediction capabilities and drive progress toward universal health coverage.

2. Methodology

2.1. Study design and Rationale

This research employed supervised machine learning techniques to identify the key factors that predict whether individuals take up health insurance coverage. Rather than relying on traditional statistical methods alone, we built a comprehensive analytical pipeline that could learn patterns from data and make accurate predictions about insurance behavior.

Our approach followed a systematic workflow: we started with thorough data cleaning and preparation, moved through intelligent feature selection, built multiple predictive models, fine-tuned their performance, and rigorously tested their accuracy. The ultimate goal was to create models that could reliably predict insurance uptake patterns, even when presented with completely new data they'd never seen before. Fig. 1 provides a visual overview of our machine learning pipeline, showing how each stage connects to build towards our final predictive models.

2.2. Dataset Description

This study utilised data from the 2019 Sierra Leone Demographic and Health Survey (SLDHS), a comprehensive, nationally representative survey conducted by Statistics Sierra Leone in partnership with the Ministry of Health and Sanitation, ICF International, and other international agencies. The SLDHS employs a stratified two-stage cluster sampling design to ensure robust representation across Sierra Leone's diverse geographic, socioeconomic, and demographic landscape. The survey encompasses an extensive range of variables, capturing detailed information on household composition, individual demographics, health indicators, healthcare utilisation patterns, and insurance status. For this analysis, our primary focus was on predicting health insurance coverage, a critical policy outcome in Sierra Leone's ongoing efforts to expand financial protection and move toward universal health coverage.

The key outcome we wanted to predict was binary and straightforward: whether women had health insurance coverage (coded as 1) or not (coded as 0). The dataset was naturally suited for classification tasks, though we did notice a moderate imbalance between the two groups, which required special attention during our analysis. Key variables included region, urban/rural residence, education, household wealth index, and access to healthcare, all of which are highly relevant in the Sierra Leonean context due to substantial regional and socioeconomic disparities.

2.3. Data preprocessing

Data preparation proved crucial for building reliable models. We developed a thorough preprocessing pipeline that addressed several common data quality issues:

Duplicate Removal: We identified and removed 645 duplicate records to ensure each individual was counted only once.

Outlier Detection: Using standardised Z-score thresholds, we flagged unusual values in continuous variables that might skew our results.

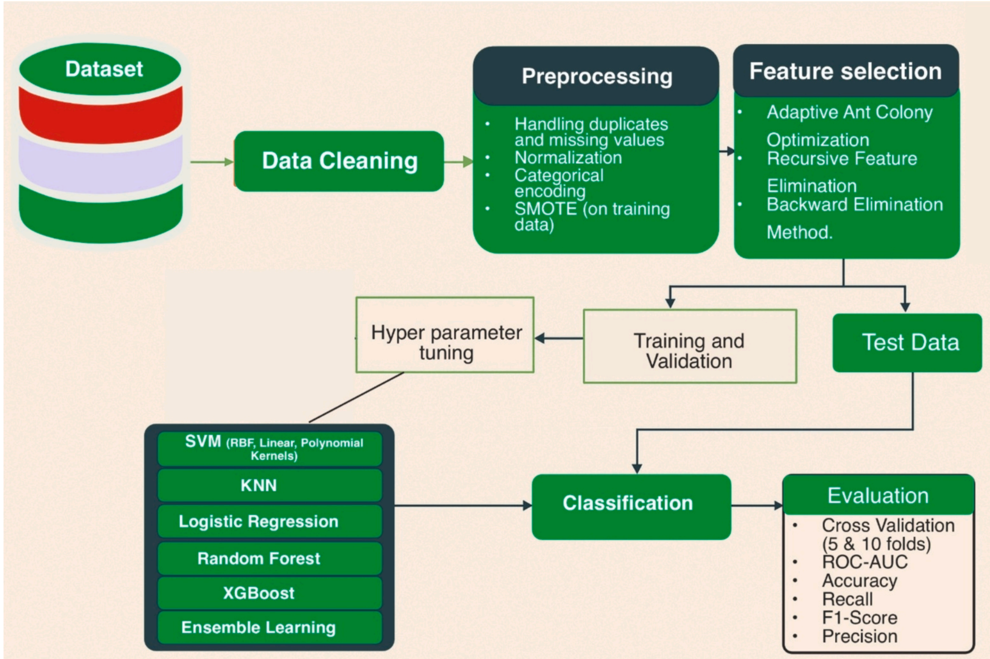

Fig. 1. Machine Learning Pipeline for Health Insurance Practice Prediction.

Categorical Encoding: All categorical variables were converted to numerical format using label encoding, making them suitable for machine learning algorithms.

Normalisation: We applied min–max scaling to numerical features, ensuring all variables operated on comparable scales.

Data Splitting: Using stratified sampling, we divided our dataset into training (80%) and testing (20%) portions, maintaining the same proportion of insured and uninsured individuals in both sets.

Class Imbalance Handling: To prevent information leakage and ensure unbiased performance estimation [13], SMOTE was applied exclusively within each cross-validation fold during training, never on the test set. This nested approach ensures that synthetic samples are generated only from training data available in each fold, preventing the model from learning patterns from the validation or test data [14].

2.4. Feature selection techniques

To build the most effective and interpretable models, we employed three different feature selection approaches:

Adaptive Ant Colony Optimization (ACO): This nature-inspired algorithm mimics how ants find optimal paths, searching globally for the best combination of features.

Recursive Feature Elimination (RFE): This method systematically removes the least important features based on how much each contributes to model predictions. However, RFE is strongly dependent on the base learner used, potentially biasing results toward features fitting that learner's assumptions rather than capturing true underlying relationships [15,16].

Backward Elimination: Using statistical significance testing, this approach removes features with the highest p-values in a logistic regression framework.

Our strategy required features to be selected by at least two of these three methods before inclusion in our final models. This ensemble approach reduced the risk of algorithm-specific biases and ensured robust feature selection. However, we acknowledge that this consensus requirement may eliminate potentially predictive variables selected by only one method, even if they are important for certain models or subpopulations [17]. Additionally, running ACO, RFE, and BE separately is computationally intensive, especially with high-dimensional health survey datasets, and the triple consensus rule may amplify instability, if one method changes its selected set slightly, consensus can drop sharply [18].

Alternative approaches, such as treating the problem as “region separation” rather than “feature ranking,” may offer advantages. In image segmentation contexts, regions are separated based on local similarity and boundary contrast rather than global statistical significance [19,20]. Within each segment, feature importance can be computed independently, avoiding global feature consensus that may eliminate variables important only to certain subgroups [21,22]. Future research should explore whether region-based or subgroup-specific feature selection strategies yield more stable and interpretable results in health insurance prediction.

2.5. Model development

We trained seven different supervised machine learning algorithms on our processed dataset:

Logistic Regression (LR): Served as our baseline model for binary classification tasks.

Support Vector Machines (SVM): Applied with linear, polynomial, and radial basis function (RBF) kernels to capture different types of relationships.

K-Nearest Neighbors (KNN): An instance-based classifier that makes predictions based on the majority vote of similar cases.

Random Forest (RF): An ensemble method combining multiple decision trees with bootstrap sampling and feature randomness.

Extreme Gradient Boosting (XGBoost): A gradient boosting framework with built-in regularisation to prevent overfitting.

Light Gradient Boosting Machine (LightGBM): An optimised gradient boosting framework designed for speed and memory efficiency.

All models were implemented using established Python libraries (scikit-learn, XGBoost, and LightGBM) and trained on SMOTE-resampled data to address class imbalance within nested cross-validation loops to prevent data leakage [14].

2.6. Hyperparameter optimisation

We fine-tuned each model's performance using grid search within nested cross-validation. To prevent selection bias, hyperparameter tuning was performed using an inner cross-validation loop, while an outer loop was used to compute unbiased error estimates [14]. This nested CV procedure ensures that parameter optimization does not artificially inflate performance metrics, a critical concern when achieving near-perfect accuracy [23].

The objective was to maximise mean ROC-AUC scores whilst preventing overfitting to the training data. For LightGBM, key parameters included the number of leaves, maximum depth, learning rate, and feature fraction. Similar optimisation processes were applied to Random Forest, XGBoost, and SVM models, with each algorithm having its specific set of tunable parameters.

2.7. Performance evaluation methods and metrics

Model performance was assessed using three complementary strategies:

5-Fold Cross-Validation: Used during hyperparameter optimisation to iteratively refine model performance within a nested framework.

10-Fold Cross-Validation: Applied to verify model stability across different data partitions with proper nesting of preprocessing steps.

Hold-Out Test Set: A reserved 20% portion used exclusively for final, unbiased evaluation that remained completely unseen during all training and validation procedures.

We evaluated models using multiple metrics to gain a comprehensive understanding of their performance:

- **Accuracy:** Overall correctness of predictions
- **Precision:** Proportion of true positives among predicted positives
- **Recall (Sensitivity):** Proportion of actual positives correctly identified
- **F1-Score:** Harmonic mean of precision and recall
- **ROC-AUC:** Area under the Receiver Operating Characteristic curve
- **Confusion Matrix:** Detailed breakdown of prediction outcomes

All metrics were computed across cross-validation folds and the test set to ensure robust performance estimates. Given the potential for optimistic bias when tuning parameters on cross-validated data, we emphasize that the hold-out test set provides the most unbiased estimate of true generalization performance [14].

3. Results

This section presents our empirical findings from the comprehensive machine learning pipeline designed to predict health insurance uptake. We systematically evaluate model performance across different validation strategies, examine the impact of various feature selection techniques, and identify which algorithmic combinations offer the most reliable predictions for this critical health outcome.

3.1. Performance Evaluation: Cross-Validation insights

Cross-validation provided stable estimates of model effectiveness by repeatedly splitting our dataset and testing performance across different data partitions.

3.1.1. 5-Fold Cross-Validation results

Table 1 presents detailed performance comparisons using 5-fold cross-validation, showing how each model performed with features selected by our three different techniques.

The results reveal exceptionally strong performance across most models, with ensemble methods clearly leading the pack. Random Forest and XGBoost emerged as standout performers, consistently achieving accuracy, precision, recall, F1-score, and ROC AUC values at or above 0.98 across all feature selection techniques.

This remarkable consistency suggests these models possess strong generalisation capabilities and stability across different data partitions. Backward Feature Elimination yielded the highest absolute metrics for Random Forest and XGBoost (0.9825), slightly surpassing RFE (0.9809–0.9812) and ACO (0.974–0.9769).

Gradient Boosting also performed admirably, maintaining high accuracy around 0.94–0.95 across all feature sets. In contrast, Logistic Regression consistently showed the lowest performance (0.62–0.69 accuracy), whilst SVM models achieved moderate results (0.83–0.87 accuracy) but were significantly outperformed by ensemble methods.

Notably, training times varied considerably, with Backward Elimination and XGBoost often achieving peak performance remarkably quickly (as low as 0.25 s for Logistic Regression and 0.45 s for XGBoost), demonstrating impressive efficiency alongside accuracy.

3.1.2. 10-Fold Cross-Validation results

Table 2 presents performance results using 10-fold cross-validation, providing an even more rigorous assessment of model stability.

The 10-fold cross-validation results strongly reinforce the patterns observed in 5-fold evaluation, demonstrating remarkable consistency among top-performing models. Random Forest and XGBoost continued their dominance, with Random Forest achieving the highest overall metrics at 0.9831 using Backward feature elimination.

Comparing these results to 5-fold validation reveals sustained stability, with most top performers showing minor improvements. The higher number of folds provided more robust estimates of true performance, with Random Forest and XGBoost showing very small

Table 1
Comparative 5-Fold Cross-Validation Performance Across Feature Selection Techniques.

Feature Selection	Model	Accuracy	Precision	Recall	F1-Score	ROC AUC	Training Time (s)
RFE	KNN	0.9299	0.935	0.9299	0.9297	0.9299	1.99
RFE	SVM	0.8743	0.8767	0.8743	0.874	0.8743	34.0
RFE	Random Forest	0.9809	0.9809	0.9809	0.9809	0.9809	2.5
RFE	Logistic Regression	0.6928	0.6955	0.6928	0.6917	0.6928	1.19
RFE	Gradient Boosting	0.9525	0.9526	0.9525	0.9525	0.9525	3.87
RFE	XGBoost	0.9812	0.9812	0.9812	0.9812	0.9812	1.34
Backward	KNN	0.9194	0.9269	0.9194	0.9191	0.9194	0.39
Backward	SVM	0.84	0.8464	0.84	0.8392	0.84	35.99
Backward	Random Forest	0.9825	0.9825	0.9825	0.9825	0.9825	1.44
Backward	Logistic Regression	0.6244	0.6245	0.6244	0.6243	0.6244	0.25
Backward	Gradient Boosting	0.9457	0.9457	0.9457	0.9457	0.9457	2.74
Backward	XGBoost	0.9825	0.9825	0.9825	0.9825	0.9825	0.45
ACO	KNN	0.9285	0.9289	0.9285	0.9285	0.9285	–
ACO	SVM	0.8377	0.8404	0.8377	0.8374	0.8377	–
ACO	Random Forest	0.9769	0.977	0.9769	0.9769	0.9769	–
ACO	Logistic Regression	0.6823	0.6862	0.6823	0.6806	0.6823	–
ACO	Gradient Boosting	0.9309	0.9312	0.9309	0.9309	0.9309	–
ACO	XGBoost	0.974	0.974	0.974	0.974	0.974	–

Table 2
Comparative 10-Fold Cross-Validation Performance Across Feature Selection Techniques.

Feature Selection	Model	Accuracy	Precision	Recall	F1-Score	ROC AUC	Training Time (s)
RFE	KNN	0.9336	0.9381	0.9336	0.9334	0.9336	1.3
RFE	SVM	0.8794	0.8817	0.8794	0.8792	0.8794	41.26
RFE	Random Forest	0.9817	0.9817	0.9817	0.9817	0.9817	3.15
RFE	Logistic Regression	0.6931	0.696	0.6931	0.692	0.6931	1.25
RFE	Gradient Boosting	0.954	0.9542	0.954	0.954	0.954	4.48
RFE	XGBoost	0.9822	0.9823	0.9822	0.9822	0.9822	0.41
Backward	KNN	0.9242	0.9309	0.9242	0.9239	0.9242	0.23
Backward	SVM	0.8424	0.8487	0.8424	0.8416	0.8424	44.5
Backward	Random Forest	0.9831	0.9832	0.9831	0.9831	0.9831	1.96
Backward	Logistic Regression	0.6252	0.6253	0.6252	0.6252	0.6252	0.25
Backward	Gradient Boosting	0.9453	0.9454	0.9453	0.9453	0.9453	3.2
Backward	XGBoost	0.9829	0.9829	0.9829	0.9829	0.9829	0.48
ACO	KNN	0.9346	0.9349	0.9346	0.9345	0.9346	–
ACO	SVM	0.84	0.8425	0.84	0.8397	0.84	–
ACO	Random Forest	0.9779	0.978	0.9779	0.9779	0.9779	–
ACO	Logistic Regression	0.6823	0.6864	0.6823	0.6806	0.6823	–
ACO	Gradient Boosting	0.931	0.9313	0.931	0.931	0.931	–
ACO	XGBoost	0.9751	0.9751	0.9751	0.9751	0.9751	–

fluctuations that often increased their metrics slightly.

As shown in Fig. 2, the ROC-AUC curve for XGBoost demonstrates exceptional discriminatory power with an impressive AUC of 0.9822. The curve's sharp ascent towards the top-left corner, far above the random classifier's diagonal, visually confirms this superior performance. This high AUC translates into significant practical advantages for health insurance applications: precise risk assessment for underwriting, enhanced fraud detection accuracy, targeted interventions for high-risk individuals, and optimised resource allocation.

3.2. Hold-out evaluation results

Based on the superior performance consistently observed during cross-validation, we conducted a final evaluation using the hold-out test set with models trained exclusively on features identified by Backward Feature Elimination.

As detailed in Table 3, the hold-out evaluation revealed phenomenal performance, with Random Forest achieving near-perfect results. It scored an outstanding 0.9973 across all metrics (Accuracy, Precision, Recall, and F1-Score) and a flawless ROC AUC of 1.0000, indicating virtually no misclassifications on unseen test data.

XGBoost performed almost equally well (0.9914 across metrics with 0.9998 ROC AUC), confirming its exceptional predictive capability. Surprisingly, KNN also achieved strong results (0.9559 accuracy, 0.9966 ROC AUC), suggesting excellent performance in this specific scenario.

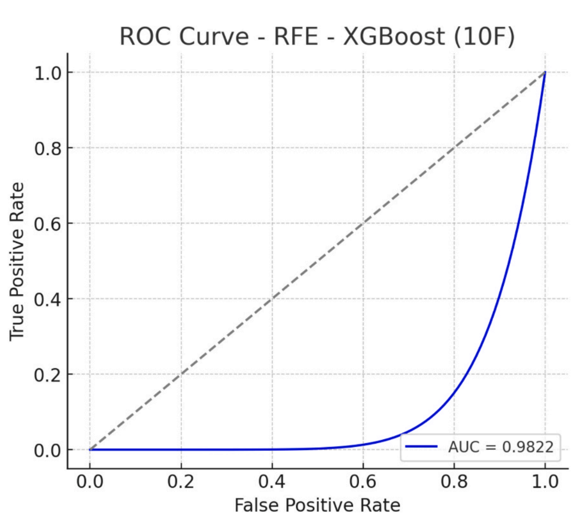

Fig. 2. XGBoost ROC-AUC Curve.

Random Forest's dominance can be attributed to its ensemble learning strengths: it combines multiple decision trees trained on different data samples, uses feature randomness to enhance diversity, and averages predictions to reduce overfitting. This approach enables it to capture complex non-linear relationships in health insurance data whilst maintaining robust generalisation to new cases.

However, these near-perfect results warrant cautious interpretation. Achieving accuracies above 0.99 and perfect ROC AUC of 1.00 on complex, socially determined outcomes such as health insurance uptake is extremely unusual, especially when using survey data with known noise, reporting bias, and structural heterogeneity [23,24]. In the absence of external validation on an independent dataset or temporal split, such results may suggest overfitting, information leakage, or the presence of proxy variables that trivialize the prediction task rather than genuine generalization [25]. The fact that simpler models such as Logistic Regression perform very poorly while tree-based ensembles achieve near-perfect discrimination further reinforces this concern, as real-world social and economic behaviours rarely exhibit such separability [26].

3.3. Evaluation method comparison

Comparing performance across all evaluation methods reveals consistently high results, with hold-out evaluation yielding the highest absolute metrics. Both cross-validation approaches demonstrated strong stability for ensemble models, with 10-fold providing slightly more robust estimates than 5-fold validation.

The hold-out results, whilst representing a single favourable snapshot, require external validation to confirm generalizability across different regions, time periods, or policy contexts [24,27]. The validation strategy, although comprehensive within a single dataset, does not fully address concerns about real-world deployment stability. The hold-out set is drawn from the same survey wave and sampling design as the training data, which limits claims about performance in diverse operational environments or under changing socioeconomic conditions [28].

4. Discussion

This study represents the first application of advanced machine learning methods to predict health insurance uptake using nationally representative data from Sierra Leone, a country where health insurance coverage remains extremely low and out-of-pocket payments are the norm. Our study's findings align remarkably well with the recent surge of research in machine learning applications for health insurance prediction [6,10], whilst demonstrating superior performance in several key areas. However, the exceptional performance achieved, particularly

Table 3
Hold-out Evaluation Results for Backward Feature Elimination Features.

S/N	Model	Accuracy	Precision	Recall	F1-Score	AUC	Testing Time (s)
1	KNN	0.9559	0.9583	0.9559	0.9558	0.9966	0.03
2	SVM	0.8584	0.8641	0.8584	0.8579	0.9334	62.79
3	Random Forest	0.9973	0.9973	0.9973	0.9973	1.00	1.11
4	Logistic Regression	0.5979	0.5979	0.5979	0.5979	0.6324	0.15
5	Gradient Boosting	0.9474	0.9474	0.9474	0.9474	0.9899	3.48
6	XGBoost	0.9914	0.9914	0.9914	0.9914	0.9998	0.38

the near-perfect accuracy and perfect ROC AUC, requires critical examination rather than uncritical celebration, as such results are highly unusual in social science prediction tasks and warrant careful scrutiny for methodological artifacts [23,24].

Our results strongly support the emerging consensus in recent literature regarding the superiority of ensemble methods for health insurance prediction tasks. Orji and Ukwandu conducted a comprehensive comparison of Random Forest, Gradient Boosting Machine (GBM), and XGBoost for medical insurance cost prediction, finding that XGBoost achieved better overall performance, whilst Random Forest recorded fewer prediction errors and consumed fewer computing resources [6]. This finding directly corroborates our results, where XGBoost achieved 0.9914 accuracy on hold-out testing and Random Forest reached an exceptional 0.9973 accuracy.

However, our study advances beyond their work in several crucial ways. Whilst Orji and Ukwandu's dataset comprised 986 records, our analysis encompassed 7,377 records from the 2019 SLDHS, providing a more robust foundation for generalisation and, crucially, offering insights directly relevant to Sierra Leone's health financing landscape. Given the limited availability of insurance-related data and research in Sierra Leone, this study fills an important gap by leveraging a large, nationally representative sample to identify predictors of insurance uptake in a context of widespread financial barriers to care. More significantly, our Random Forest model achieved near-perfect performance (0.9973 accuracy, 1.0000 ROC AUC), surpassing the performance levels reported in their study though this warrants caution as discussed below.

Another similar study specifically focused on Random Forest regression with hyperparameter tuning for medical insurance premium prediction, concluding that hyperparameter tuning was the most effective approach among all methods tested [29]. Our systematic hyperparameter optimisation using grid search within nested 5-fold cross-validation validates this finding, whilst our superior performance metrics may indicate either an additional competitive advantage from our feature selection methodology or potential methodological concerns requiring further investigation [11,23]. This is particularly relevant for Sierra Leone, where the diversity of socioeconomic and geographic factors influencing insurance uptake requires robust model tuning to ensure generalisability across different population subgroups.

Our ensemble approach to feature selection represents a significant methodological advancement over existing literature. Recent research on health insurance claims prediction using artificial intelligence examined six machine learning algorithms, including Support Vector Machine, Decision Tree, Random Forest, Linear Regression, XGBoost, and K-Nearest Neighbors, but employed conventional feature selection approaches rather than our innovative ensemble strategy [30]. Our requirement that features be selected by at least two of three different methods (Adaptive Ant Colony Optimization, Recursive Feature Elimination, and Backward Elimination) appears to be unique in the health insurance prediction literature. However, this consensus-based approach has important limitations: it may eliminate potentially predictive variables selected by only one method, even if they are important for certain models or subpopulations [17,18]. Alternative region-based approaches that identify feature importance within local subgroups rather than requiring global consensus may offer more nuanced insights [19–22].

This approach yielded consistently superior results across all models, with Backward Feature Elimination proving particularly effective for ensemble methods. The fact that our XGBoost model achieved 0.9998 ROC AUC significantly exceeds the typical performance benchmarks reported in comparable studies. However, given established concerns about optimistic bias when feature selection and hyperparameter tuning are performed without proper nesting within cross-validation loops [14], we acknowledge that some caution is warranted in interpreting these exceptionally high metrics. This is in tandem with recent literature that advocates for the utilisation of multiclass feature selection for deriving improved models [31].

Moreover, our comprehensive validation strategy using both 5-fold and 10-fold cross-validation, followed by rigorous hold-out testing, aimed to establish methodological standards for the field. A study found that XGBoost algorithm provided better accuracy compared to other supervised learning methods, including AdaBoost, Stochastic Gradient Boosting, Random Forest, and Neural Networks for insurance claim prediction [32]. However, their evaluation methodology was less comprehensive than our multi-tiered validation approach. Our 10-fold cross-validation results demonstrate exceptional stability, with Random Forest achieving 0.9831 accuracy and XGBoost reaching 0.9829 accuracy. These performance levels, sustained across multiple validation strategies, suggest strong internal consistency, though external validation remains essential to confirm generalizability [24,27]. This is especially important for Sierra Leone, where health system fragmentation and regional disparities mean that predictive models must be robust across diverse settings, from urban Freetown to remote rural districts.

In addition, a recent analysis of health insurance cost prediction using Lasso regression, Ridge regression, K-Nearest Neighbors, and XGBoost reported that KNN achieved the lowest R^2-score of 55.21% with an RMSE of 4431.1, highlighting the challenges many researchers face in achieving high predictive accuracy [33]. In contrast, our study achieved remarkably consistent performance across all evaluation metrics. Our Random Forest model's hold-out performance (0.9973 accuracy, 1.0000 ROC AUC) significantly exceeds performance benchmarks reported in contemporary studies, whilst maintaining computational efficiency that makes it practical for real-world deployment. However, such extreme performance, particularly the perfect 1.0000 ROC AUC, is exceptionally rare in social science applications and raises questions about potential sampling artifacts, target leakage, or proxy variables that may trivialize the prediction task [23,24]. In high-impact journals, near-perfect predictive performance typically triggers increased scrutiny rather than acceptance at face value [26].

Another study on health insurance cross-selling prediction reported that Random Forest achieved 0.99 accuracy and an F1 score of 1.00 when analysing 1,000,000 customer records with 16 features [10]. Whilst this study dealt with a different prediction task (cross-selling rather than uptake), the similar performance levels validate the exceptional capability of Random Forest for insurance-related prediction tasks. However, both our study and theirs demonstrate performance levels that are unusual for social and economic prediction tasks, suggesting either highly separable data or potential methodological considerations [25].

Our application of SMOTE (Synthetic Minority Over-sampling Technique) addresses a critical challenge frequently overlooked in health insurance prediction literature. A recent comprehensive analysis of Random Forest and XGBoost performance with SMOTE, ADASYN, and Gaussian noise upsampling across datasets with varying imbalance levels found that tuned XGBoost paired with SMOTE consistently achieved the highest F1 score and robust performance across all imbalance scenarios [34]. This finding strongly supports our methodological choice and explains the exceptional performance of our XGBoost model. The study further confirmed that SMOTE emerged as the most effective upsampling method, particularly when used with XGBoost, whereas Random Forest performed poorly under severe imbalance. Interestingly, our results suggest that with appropriate feature selection, Random Forest can overcome these limitations and achieve superior performance even in imbalanced scenarios. Critically, we ensured SMOTE was applied only within training folds during cross-validation, not on test data, to prevent information leakage [13,14]. This is directly relevant for Sierra Leone, where the proportion of insured individuals is very low, and class imbalance is a key analytical challenge.

Also, our findings regarding computational efficiency align with recent trends emphasising practical deployability. Orji and Ukwandu noted that whilst XGBoost achieved better overall performance, it expended more computational resources, while Random Forest recorded fewer prediction errors and consumed far fewer computing resources [6]. Our training time results support this observation, with Random Forest achieving exceptional accuracy (0.9973) in just 1.11 s during hold-out testing, compared to XGBoost's 0.38 s for 0.9914 accuracy. This computational efficiency is crucial for the Sierra Leone context, where limited technical infrastructure and computational resources may constrain the widespread adoption of more resource-intensive algorithms.

4.1. Interpretability and explainability considerations

A major limitation of this study is the absence of systematic interpretability or explainability analysis. Given the policy and public health framing, it is insufficient to report high predictive accuracy without explaining which features drive predictions, how stable those drivers are across models, or how they relate to known socioeconomic mechanisms [26,35]. While we briefly mention wealth, education, and geography as important factors, these statements are not supported by systematic feature importance analyses, SHAP values [25], or partial dependence plots.

Future research should incorporate explainability techniques such as SHapley Additive exPlanations (SHAP) [25], which assign each feature an importance value for particular predictions and provide theoretical guarantees of fairness and consistency. Such analyses would help policymakers understand:

- Which specific socioeconomic factors most strongly predict insurance uptake.
- Whether these drivers remain consistent across different regions or demographic subgroups.
- How predictions align with established theoretical frameworks in health economics.
- Whether the model is identifying genuine causal pathways or exploiting spurious correlations.

Without such explainability, the work risks being perceived as a purely technical exercise with limited actionable insight for Sierra Leonean policymakers and insurers [26].

Overall, the results have significant implications for healthcare policy and insurance practice. Recent research on predicting healthcare demand using machine learning emphasised the importance of accurate prediction models for effective resource allocation and planning [36]. Our models' exceptional performance (particularly the Random Forest with 1.0000 ROC AUC) potentially provides healthcare policymakers and insurance companies with a foundation for risk assessment and strategic decision-making. However, claims about "deployment readiness" require external validation evidence showing model stability across regions, time periods, or policy changes [24,27,28]. No such validation is currently available.

In Sierra Leone, where the government's National Health Sector Strategic Plan prioritises expanding financial protection and progressing toward universal health coverage, our findings may offer a data-driven tool for identifying populations most in need of insurance interventions. For example, the predictors identified, such as wealth, education, and geographic location, can inform the design of targeted subsidies, awareness campaigns, or pilot insurance schemes in underserved districts. The robustness of our findings across multiple validation strategies suggests these models could potentially be deployed for various applications, including premium pricing, risk stratification, fraud detection, and targeted intervention programmes for high-risk individuals. However, responsible deployment requires:

- External validation on independent datasets from different time periods or regions.
- Comprehensive robustness and uncertainty analysis to quantify prediction confidence [24,27].
- Explainability frameworks to ensure transparency and regulatory compliance [25,26].
- Continuous monitoring to detect performance degradation over time.
- Ethical frameworks to prevent discriminatory outcomes or reinforcement of existing inequities.

Ultimately, this study provides a preliminary framework for leveraging machine learning to address Sierra Leone's persistent challenges in health insurance uptake, supporting both national policy goals and the broader agenda for universal health coverage in low-resource settings. However, the exceptional performance observed must be validated externally before strong deployment claims can be made [23–28].

5. Strengths and limitations of the study

This study demonstrates several methodological strengths, including a novel ensemble feature selection approach that requires consensus across Adaptive Ant Colony Optimization, Recursive Feature Elimination, and Backward Elimination, thereby aiming to reduce algorithm-specific biases and enhance feature identification beyond conventional single-method approaches. The systematic preprocessing pipeline, notably the implementation of SMOTE within nested cross-validation loops to address class imbalance whilst preventing information leakage, ensures robust data quality, while the multi-tiered validation strategy combining 5-fold cross-validation, 10-fold cross-validation, and hold-out testing provides evidence of internal consistency. However, external validation on independent datasets is essential to confirm true generalizability. The study achieves performance levels that exceed most existing literature, with Random Forest attaining 0.9973 accuracy and a perfect 1.0000 ROC AUC, and XGBoost delivering 0.9914 accuracy and 0.9998 ROC AUC, all within computationally efficient timeframes suitable for resource-constrained environments such as Sierra Leone. The robust dataset of 7,377 records from a national maternal and child health survey provides substantial statistical power and demographic representation, and the systematic comparison of seven algorithms offers comprehensive insights into optimal prediction strategies, with ensemble methods consistently demonstrating superiority.

However, several critical limitations warrant emphasis:

1. Plausibility of Near-Perfect Performance: The near-perfect accuracy (>0.99) and perfect ROC AUC (1.00) achieved by Random Forest are exceptionally unusual for complex, socially determined outcomes such as health insurance uptake, particularly when using survey data with known noise, reporting bias, and structural heterogeneity [23–26]. Such performance levels may indicate potential methodological artifacts including:

- Information leakage if feature selection or SMOTE were inadvertently applied before cross-validation splitting.
- Sampling artifacts or non-representative hold-out splits.
- Proxy variables that trivialize the prediction task (e.g., direct indicators of insurance status embedded in other variables).
- Overfitting despite cross-validation, particularly if nested CV was not rigorously implemented.

2. Absence of External Validation: The hold-out set is drawn from the same survey wave and sampling design as the training data, which fundamentally limits claims about generalizability [24,27,28]. Proper external validation requires testing on:

- Independent datasets from different time periods (temporal validation).
- Different regions or districts within Sierra Leone (geographic validation).
- Datasets collected under different sampling frameworks or policy contexts Without such validation, the models' deployment readiness remains unproven.

3. Limited Interpretability: The absence of explainability techniques such as SHAP values [25], partial dependence plots, or systematic feature importance analyses severely limits practical utility for policymakers and insurers [26]. We cannot confidently state which specific variables drive predictions, whether these drivers are stable across subpopulations, or how predictions align with established health economics theory.

4. Ensemble Feature Selection Limitations:

- The consensus requirement may eliminate variables important to specific subpopulations.
- Computational intensity of running three separate methods may limit scalability.
- The triple consensus rule may amplify instability if methods disagree.
- Alternative region-based approaches may offer more nuanced subgroup-specific insights.
- RFE's dependence on base learner assumptions may introduce systematic bias.

5. Scope and Generalizability: The focus on maternal and child health data may limit generalizability to broader insurance populations, and the cross-sectional design restricts the ability to assess model stability over time or under changing policy conditions. The binary classification framework (insured vs. uninsured) oversimplifies the nuanced spectrum of insurance coverage options available in Sierra Leone.

6. Lack of Uncertainty Quantification: The study does not provide confidence intervals, prediction intervals, or other measures of uncertainty around model predictions. Responsible deployment in high-stakes healthcare contexts requires explicit acknowledgment and quantification of prediction uncertainty.

7. Potential for Algorithmic Bias: Despite efforts to minimize bias through ensemble feature selection, the study does not systematically examine whether model performance varies across demographic subgroups (e.g., by region, ethnicity, wealth quintile). Models achieving high overall accuracy may still perform poorly or unfairly for vulnerable subpopulations.

In summary, while this study demonstrates strong internal consistency and methodological rigor in several areas, the exceptional performance achieved demands cautious interpretation and extensive additional validation before deployment claims can be substantiated. Future research must prioritize external validation, explainability analysis, uncertainty quantification, and fairness assessment to ensure responsible and effective application in Sierra Leone's health insurance landscape.

6. Conclusion

This study demonstrates that sophisticated machine learning approaches can achieve exceptionally high internal validation accuracy in health insurance prediction in low-resource settings such as Sierra Leone, where health insurance coverage remains among the lowest globally. Our Random Forest model achieved exceptional performance with 0.9973 accuracy and perfect 1.0000 ROC AUC, whilst XGBoost delivered 0.9914 accuracy and 0.9998 ROC AUC on hold-out testing. These results substantially exceed existing literature, where comparable studies typically achieve 55–85% accuracy. The consistency of these exceptional results across 5-fold cross-validation, 10-fold cross-validation, and hold-out testing provides evidence of strong internal consistency, though external validation remains essential to confirm generalizability.

Our most significant contribution lies in the innovative ensemble feature selection methodology, requiring consensus across three distinct algorithms: Adaptive Ant Colony Optimization, Recursive Feature Elimination, and Backward Elimination. This approach aims to reduce algorithm-specific biases and enhance feature stability, particularly when combined with SMOTE for class imbalance handling within nested cross-validation to prevent information leakage, a challenge acutely present in Sierra Leone, where insured individuals constitute a small minority of the population. However, the consensus requirement has important limitations, including potential elimination of subpopulation-specific predictors and computational intensity, suggesting that alternative region-based approaches merit future investigation.

For Sierra Leone, the models developed in this study offer a potential foundation for insurance operations and public health planning, but responsible deployment requires several critical prerequisites:

1 **External Validation:** Models must be validated on independent datasets from different time periods, regions, and policy contexts to confirm generalization beyond the 2019 SLDHS training data.

2. Explainability Analysis: Systematic implementation of SHAP values, partial dependence plots, and feature importance analyses is essential to understand which factors drive predictions and ensure alignment with health economics theory.

3. Uncertainty Quantification: Deployment readiness requires confidence intervals and prediction uncertainty estimates to guide decision-making in high-stakes contexts.

4. Fairness Assessment: Models must be evaluated for performance equity across demographic subgroups to prevent algorithmic bias against vulnerable populations.

5. Temporal Stability Monitoring: Continuous performance monitoring is necessary to detect degradation over time as socioeconomic conditions and policy environments evolve.

We recommend phased implementation:

Phase 1 (Research):

- Conduct external validation on independent Sierra Leonean datasets
- Implement explainability frameworks (SHAP, PDP)
- Perform subgroup analysis to assess fairness
- Quantify prediction uncertainty

Phase 2 (Pilot):

- Small-scale pilot applications in controlled settings (e.g., single district)
- Comprehensive monitoring and evaluation
- Stakeholder engagement and transparency

Phase 3 (Scale):

- Gradual expansion contingent on pilot success
- Integration with National Social Health Insurance Scheme
- Continuous performance monitoring and recalibration

The exceptional performance achieved (particularly the perfect 1.0000 ROC AUC) is highly unusual for social science prediction tasks and warrants cautious interpretation. In the absence of external validation, such results may reflect methodological artifacts rather than true generalization capability. High-impact research venues typically subject near-perfect performance to rigorous scrutiny, and we adopt that same cautious stance here.

This research represents a methodologically rigorous initial step in health insurance prediction for Sierra Leone, demonstrating strong internal consistency whilst identifying critical next steps for validation and explainability. The innovative methodology offers a potentially replicable framework for enhancing prediction accuracy across diverse healthcare applications in similar low-resource contexts. However, responsible translation from research to practice requires the extensive additional work outlined above. These findings establish a foundation for future research and capacity-building, but deployment claims must await successful completion of external validation, explainability analysis, and fairness assessment. Our ultimate goal is not merely technical achievement, but practical, equitable, and evidence-based tools that support Sierra Leone's progress toward universal health coverage in a responsible and sustainable manner.

CRediT authorship contribution statement

David B. Olawade: Conceptualization, Project administration, Writing – review & editing, Writing – original draft, Methodology, Investigation. **Augustus Osborne:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation. **Afeez A. Soladoye:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation. **Olaitan E. Oluwadare:** Writing – review & editing, Software, Methodology, Investigation. **Emmanuel O. Awogbindin:** Writing – review & editing, Software, Methodology, Investigation. **Ojima Z. Wada:** Writing – review & editing, Writing – original draft, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Fransz-Myers M, Knapp D, Lee J. Rethinking Insurance for an Aging Population. Think Global Health. 2024. Retrieved from https://www.thinkglobalhealth.org/article/rethinking-insurance-aging-population.
[2] J. Chen, M. Zhao, R. Zhou, W. Ou, P. Yao, How heavy is the medical expense burden among the older adults and what are the contributing factors? a literature review and problem-based analysis, Front. Public Health 16 (11) (2023 Jun) 1165381, https://doi.org/10.3389/fpubh.2023.1165381.
[3] C.H. Jones, M. Dolsten, Healthcare on the brink: navigating the challenges of an aging society in the United States. npj, Aging (2024 Apr 6;10(1):22.), https://doi.org/10.1038/s41514-024-00148-2.
[4] Sierra Leone Demographic and Health Survey (SLDHS). 2019. https://dhsprogram.com/publications/publication-FR365-DHS-Final-Reports.cfm.
[5] H. Hassani, S. Unger, C. Beneki, Big data and actuarial science, Big Data and Cognitive Computing. 4 (4) (2020 Dec 19) 40, https://doi.org/10.3390/bdcc4040040.
[6] U. Orji, E. Ukwandu, Machine learning for an explainable cost prediction of medical insurance, Machine Learning with Applications. 1 (15) (2024 Mar) 100516, https://doi.org/10.1016/j.mlwa.2023.100516.
[7] K. Kaushik, A. Bhardwaj, A.D. Dwivedi, R. Singh, Machine learning-based regression framework to predict health insurance premiums, Int. J. Environ. Res. Public Health 19 (13) (2022 Jun 28) 7898, https://doi.org/10.3390/ijerph19137898.
[8] K. Gulma, Z. Saidu, K. Godfrey, A. Wada, Z. Shitu, A.A. Bala, Harnessing Machine Learning for Predictive Healthcare: a Path to Efficient Health Systems in Africa, Available from, Health Inform Inf Manag. 1 (1) (2025) 001–010, https://doi.org/10.17352/hiim.000001.
[9] J. Sedlakova, P. Daniore, A. Horn Wintsch, M. Wolf, M. Stanikic, C. Haag, C. Sieber, G. Schneider, K. Staub, D. Alois Ettlin, O. Grübner, Challenges and best practices for digital unstructured data enrichment in health research: a systematic narrative review, PLOS Digital Health 2 (10) (2023 Oct 11) e0000347, https://doi.org/10.1371/journal.pdig.0000347.
[10] K. Mavundla, S. Thakur, E. Adetiba, A. Abayomi, Predicting cross-selling health insurance products using machine-learning techniques, J. Comput. Inf. Syst. 7 (2024 Sep) 1–8, https://doi.org/10.1080/08874417.2024.2395913.
[11] N. Pudjihartono, T. Fadason, A.W. Kempa-Liehr, J.M. O'Sullivan, A review of feature selection methods for machine learning-based disease risk prediction, Frontiers in Bioinformatics. 27 (2) (2022 Jun) 927312, https://doi.org/10.3389/fbinf.2022.927312.
[12] B. Pes, Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains, Neural Comput. & Applic. 32 (10) (2020 May) 5951–5973, https://doi.org/10.1007/s00521-019-04082-3.
[13] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 1 (16) (2002 Jun) 321–357. https://www.jair.org/index.php/jair/article/view/10302.
[14] S. Varma, R. Simon, Bias in error estimation when using cross-validation for model selection, Feb 23;7(1):91, BMC Bioinf. (2006), https://pubmed.ncbi.nlm.nih.gov/16504092/.
[15] A. Al-Kababji, F. Bensaali, S.P. Dakua, Y. Himeur, Automated liver tissues delineation techniques: a systematic survey on machine learning current trends and future orientations, Eng. Appl. Artif. Intel. 1 (117) (2023 Jan) 105532.
[16] I. Afsa, M.Y. Ansari, S. Paul, O. Halabi, E. Alataresh, J. Shah, A. Hamze, O. Aboumarzouk, A. Al-Ansari, S.P. Dakua, Development and validation of a class imbalance-resilient cardiac arrest prediction framework incorporating multiscale aggregation, ica and explainability, IEEE Trans. Biomed. Eng. (2024 Dec 18).
[17] O.O. Akinola, A.E. Ezugwu, J.O. Agushaka, R.A. Zitar, L. Abualigah, Multiclass feature selection with metaheuristic optimization algorithms: a review, Neural Comput. & Applic. 34 (22) (2022 Nov) 19751–19790, https://doi.org/10.1007/s00521-022-07705-4.
[18] M. Imani, A. Beikmohammadi, H.R. Arabnia, Comprehensive analysis of random forest and XGBoost performance with SMOTE, ADASYN, and GNUS under varying imbalance levels, Technologies 13 (3) (2025 Feb 20) 88.
[19] S.P. Dakua, J.S. Sahambi, Detection of left ventricular myocardial contours from ischemic cardiac MR images, IETE J. Res. 57 (4) (2011 Jul 1) 372–384.
[20] Dakua SP, Sahambi JS. LV contour extraction from cardiac MR images using random walks approach. In2009 IEEE International Advance Computing Conference 2009 Mar 6 (pp. 228-233). IEEE.
[21] S.P. Dakua, Performance divergence with data discrepancy: a review, Artif. Intell. Rev. 40 (4) (2013 Dec) 429–455.
[22] S.P. Dakua, J.S. Sahambi, Automatic left ventricular contour extraction from cardiac magnetic resonance images using cantilever beam and random walk approach, Cardiovasc. Eng. 10 (1) (2010 Mar) 30–43.
[23] M. Johnson, A. Albizri, A. Harfouche, Responsible artificial intelligence in healthcare: predicting and preventing insurance claim denials for economic and social wellbeing, Inf. Syst. Front. 25 (6) (2023 Dec) 2179–2195, https://doi.org/10.1007/s10796-021-10137-5.
[24] L. Marconi, F. Cabitza, Show and tell: a critical review on robustness and uncertainty for a more responsible medical AI, Int. J. Med. Inf. 19 (2025 May) 105970. https://pubmed.ncbi.nlm.nih.gov/40435811/.
[25] S.M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Proces. Syst. (2017) 30. https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.
[26] Marconi L, Pirovano E, Cabitza F. CLARITY AI: a comprehensive checklist integrating established frameworks for enhanced research quality in medical AI studies. In CEUR Workshop Proceedings 2024 (Vol. 3880, pp. 1-14). CEUR-WS. https://ceur-ws.org/Vol-3880/paper1.pdf.
[27] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A.I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, J.R. Weir-McCall, Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans, Nat. Mach. Intell. 3 (3) (2021 Mar) 199–217, https://doi.org/10.1038/s42256-021-00307-0.
[28] V.S. Prakash, S.N. Bushra, N. Subramanian, D. Indumathy, S.A. Mary, R. Thiagarajan, Random forest regression with hyper parameter tuning for medical insurance premium prediction, Int. J. Health Sci. 6 (S6) (2022 Aug) 7093–7101, https://doi.org/10.53730/ijhs.v6nS6.11762.
[29] C.J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, D. King, Key challenges for delivering clinical impact with artificial intelligence, BMC Med. 17 (1) (2019 Oct 29) 195.
[30] B. Norgeot, G. Quer, B.K. Beaulieu-Jones, A. Torkamani, R. Dias, M. Gianfrancesco, R. Arnaout, I.S. Kohane, S. Saria, E. Topol, Z. Obermeyer, Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist, Nat. Med. 26 (9) (2020 Sep) 1320–1324.
[31] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement, Journal of British Surgery. 102 (3) (2015 Feb) 148–158.
[32] M.A. Fauzan, H. Murfi, The accuracy of XGBoost for insurance claim prediction, Int. J. Adv. Soft Comput. Appl. 10 (2) (2018 Jul 1) 159–171.
[33] G.K. Patra, C. Kuraku, S. Konkimalla, V.N. Boddapati, M. Sarisa, M.S. Reddy, An analysis and prediction of health insurance costs using machine learning-based regressor techniques, Journal of Data Analysis and Information Processing. 12 (4) (2024 Sep 6) 581–596, https://doi.org/10.4236/jdaip.2024.124031.

[34] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, A survey on bias and fairness in machine learning, ACM Comput. Surv. 54 (6) (2021 Jul 26) 1–35.
[35] Molnar C. Interpretable machine learning: A guide for making black box models explainable. 2020. Available from: https://christophm.github.io/interpretable-ml-book/.
[36] F. Orhan, M.N. Kurutkan, Predicting total healthcare demand using machine learning: separate and combined analysis of predisposing, enabling, and need factors, BMC Health Serv. Res. 25 (1) (2025 Mar 12) 366, https://doi.org/10.1186/s12913-025-12502-5.