



Research Article

# Predicting Hypertension Medication Uptake Using Explainable Artificial Intelligence: Evidence from a Kenyan Population-based Study

Eliud Koech<sup>1, 2, \*</sup> , Charles Kipkoech Mutai<sup>2</sup> , Gregory Kerich<sup>2</sup> 

<sup>1</sup>Department of Mathematics, Physics, and Computer Science, Alupe University, Busia, Kenya

<sup>2</sup>Department of Mathematics, Physics and Computing, Moi University, Eldoret, Kenya

## Abstract

Hypertension is a major contributor to cardiovascular morbidity and mortality worldwide, more so in Kenya, with limited progress towards achieving Africa's 2030 fast-track hypertension targets, especially in management. This study aimed to build a machine learning model to predict hypertension medication uptake in Kenya. Using data from 4,687 female and 5,269 male respondents from the 2022 Kenya Demographic and Health Survey, we applied Extreme Gradient Boosting, Support Vector Machine, Random Forest, and Elastic Net models. Data from 15 counties were split into training (80%) and testing (20%) sets, with class imbalance addressed using the Synthetic Minority Oversampling Technique and validation through leave-one-county-out cross-validation. The best-performing model, based on mean f1-score, was retrained using features selected through Sequential Forward Floating Selection. SHapley Additive exPlanations were used to interpret feature importance and directionality by sex. Treatment coverage remained suboptimal, with 26.6% of hypertensive males and 32.4% of females untreated. The XGBoost model achieved the best performance (78% males; 81% females). The most predictive features in both sexes were age, household size, sedentary time, income, exercise, wealth, residence duration, television viewership, and reproductive preferences among females. Interpretable machine learning revealed distinct sex-specific socio-behavioural predictors of hypertension treatment uptake in Kenya. Incorporating such data-driven insights can inform targeted, equitable interventions and strengthen hypertension control, especially in resource-limited settings where routine survey data can complement clinical assessments.

## Keywords

Hypertension, Medication Uptake, Socio-behavioural Factors, Machine Learning, Predictive Modelling

## 1. Introduction

Hypertension is a condition characterised by elevated blood pressure above normal levels (below 140/90 mmHg) and is a significant public health challenge due to its increasing prevalence and global impact. Due to the extensive utilisation of

antihypertensive drugs, the global mean blood pressure has either stayed stable or experienced a minor decline over the past forty years. Conversely, the incidence of hypertension has risen, particularly in low and middle-income countries [1].

\*Correspondence: Eliud Koech (elkoech4@gmail.com)

Received: 6 May 2026; Accepted: 15 May 2026; Published: 2 June 2026



Around 1.28 billion people aged 30 to 79 worldwide have hypertension, with nearly two-thirds in low- and middle-income countries. Furthermore, only 42% of hypertensive individuals have received a diagnosis and are currently taking medication, while over 46% remain unaware of their ailment.

In Kenya, elevated systolic blood pressure contributes to nearly 60% of cardiovascular disease (CVD) deaths in both men and women, highlighting its significance as a major modifiable risk factor [2]. In addition, hypertension affects more than 24% of the Kenyan population, indicating a considerable public health burden nationwide [3].

Consequently, there is an urgent need for effective large-scale treatments aimed at preventing or curing hypertension to reverse this trend. Between 2010 and 2030, the global target for non-communicable disease (NCD) prevention aims to achieve a 33% reduction in hypertension prevalence [2]. Therefore, as new and promising interventions emerge daily, the necessity for thorough evaluation of these interventions to guide evidence-based policies and clinical practice is increasingly imperative. It is believed that early detection, treatment, and control of hypertension can reduce health risks. Strategies include providing access to health care practitioners who detect and treat high blood pressure, lower drug costs (insurance coverage, plan design, cost sharing), and support hypertension control [3].

However, it may be beneficial to limit and manage the risk of hypertension by early identification of patients with interpretable risk factors. Consequently, the early identification of hypertension patients by the recognition of interpretable risk factors is crucial, as it facilitates prompt prevention and intervention for the patients. It is thus imperative to recognise and identify the interpretable risk factors of hypertension at an early stage.

Several risk factors linked to hypertension in low- and middle-income nations, including Kenya, have been identified by numerous studies and empirical research [2-4]. However, the previously conducted association studies suffered from a number of limitations. Primarily, prior investigations employed conventional linear models, including logistic regression and the Cox proportional hazards model, to find significantly linked risk variables for hypertension [5, 6]. Traditional linear models struggle with high-dimensional non-linear data, and their limited precision hinders patient-level usage.

Machine learning and its widespread application in public health research may help overcome constraints in complex actual data. In machine learning, algorithms use past experiences and data patterns to predict and perform tasks, such as classification or identification. Progress in artificial intelligence is propelled by machine learning. Academics and industry successfully apply it to create intelligent products that can generate accurate predictions using various data sources [7]. Various types of learning algorithms exist in machine learning, with supervised learning being the most prevalent and broadly applicable. The objective of supervised learning algorithms is to utilise datasets to construct models capable of predicting system outputs based on incoming inputs. Previous studies

have developed multivariable prediction models using various machine learning and explainable artificial intelligence techniques [5, 6, 8].

In Ethiopia, Islam et al. [9] utilised machine learning techniques to predict hypertension and reduce related mortality. The study analysed data from 612 participants across 27 lifestyle and health variables, achieving an accuracy of 88.81%, with XGBoost emerging as the best-performing model due to its enhanced interpretability using SHapley Additive exPlanations (SHAP). Key predictors included age, weight, obesity, income, body mass index (BMI), diabetes, salt intake, alcohol consumption, smoking, and prior hypertension history. The findings underscore the potential of machine learning-based models for effective hypertension risk prediction in resource-limited settings.

Similarly, Islam et al. [10] analysed data from over 818,000 individuals across Bangladesh, Nepal, and India to predict hypertension and its determinants using XGBoost, Gradient Boosting Machine (GBM), Logistic Regression, Random Forest, and Decision Tree models. Despite methodological limitations, XGBoost, GBM, Logistic Regression, and Linear Discriminant Analysis (LDA) achieved 100% recall and approximately 90% prediction accuracy. While age and body mass index (BMI) consistently emerged as significant predictors, the models lacked integration of key socio-behavioural and clinical factors, including family history, alcohol consumption, physical activity, dietary patterns, and biochemical indicators, which may have constrained their explanatory power.

To address these limitations, the present study aimed at applying machine learning approaches to identify predictors of hypertension medication status, incorporating socio-behavioural, demographic, and clinical characteristics to better capture contextual influences. The overarching goal is to develop a risk prediction framework tailored to Kenya's unique population, behavioural patterns, and health system context, enabling more targeted identification and management of high-risk individuals.

## 2. Methods

### 2.1. Data

This study utilised data from the Demographic and Health Surveys (DHS) Program, which implements nationally representative cross-sectional household surveys to assess key health indicators [11]. Specifically, data from the 2022 Kenya Demographic and Health Survey (KDHS) were used for this analysis, and include only individuals aged 15 to 49 in accordance with the DHS study design.

In this study, individual datasets were merged with household data separately for males and females, and then resampled using individual sample weights to account for non-coverage, non-response, and population-level adjustments, resulting in 5,930 variables for females and 564 for males. Variables with more than 30% missing values were

first excluded, resulting in the removal of 5,355 female and 303 male variables.

Subsequently, non-informative features were eliminated by removing constant variables (i.e., those with a single unique value) and exact duplicate columns, accounting for a further reduction of 150 variables in the female dataset and 11 in the male dataset. To minimise the influence of low-variability noise, a variance threshold of 0.01 was applied, leading to the exclusion of 98 female and 15 male variables. Also, variables were encoded for both the nominal and ordinal variables using the label-code and one-hot encoding methods appropriately based on the information from the survey [12]. To address multicollinearity, variables exhibiting high pairwise correlations (absolute correlation coefficient  $> 0.8$ ) were identified and removed, resulting in the exclusion of an additional 50 female and 22 male variables, as shown in Table 1.

Multiple imputations with chained equations (MICE) [13]

were utilised in imputing the missing values in each of these categories. Finally, data was harmonised and scaled by standardising to ensure a fair penalisation of the scheme used for all the regressors. This yielded a refined set of 277 variables for females and 213 variables for males, which were retained for subsequent modelling and analysis.

The awareness rate of hypertension was also determined among individuals who self-reported having hypertension by assessing whether participants were aware of their hypertensive status following diagnosis by a health professional. Treatment uptake was evaluated among individuals with self-reported hypertension based on whether participants reported currently taking prescribed antihypertensive medication or receiving treatment for blood pressure control. Descriptive statistics, including frequencies and percentages, were used to summarise awareness and treatment uptake across counties and sex categories.

**Table 1.** Variables excluded in pre-processing.

Pre-processing step	Females	Males
Total available variables	5930	564
More than 30% missing	5355	303
Constant or duplicate columns	150	11
Non-informative (low variance)	98	15
Above 0.8 correlated features	50	22
Total excluded	5653	351
Final variables	277	213

## 2.2. Model Development

A supervised machine learning framework was implemented as a binary classification task, where individuals with self-reported hypertension who were on medication constituted the positive class, and those not on medication formed the negative class. Data from 15 counties (Embu, Homa Bay, Kilifi, Laikipia, Meru, Migori, Mombasa, Nairobi, Nyamira, Nyandarua, Nyeri, Tharaka-Nithi, Uasin Gishu, and Vihiga) were randomly sampled without replacement, then randomly partitioned into training (80%) and test (20%) subsets, maintaining proportional representation of the outcome classes. A leave-one-county-out cross-validation approach (a separate left-out sample, not used during model training or tuning, was reserved for external validation) was used to assess the models' capacity to generalise across geographical areas. The procedure was iteratively applied across all counties, whereby data from one county were excluded during model training and reserved exclusively for testing. This process was conducted

separately for male and female datasets.

A grid of 50 hyperparameter control values was randomly sampled and combined with five-fold cross-validation for model training and validation. Four supervised learning algorithms were evaluated: Extreme Gradient Boosting (XGBoost) [14] a tree-based ensemble method optimized for speed and performance, Support Vector Machine (SVM) [15] using a radial basis function (RBF) kernel to capture non-linear relationships, Random Forest (RF) [16] an ensemble of decision trees that reduces variance through bootstrap aggregation, and Elastic Net (EN) [17] a regularized regression model that combines L1 and L2 penalties for feature selection and shrinkage.

The average f1-scores were computed for each hyperparameter set using a five-fold cross-validation scheme on the validation samples, and the optimal hyperparameter configuration was selected. The f1-score, defined as the harmonic mean of precision and recall, was used as the primary evaluation metric to balance the trade-off between false positives and false negatives, particularly given the class imbalance in hy-

pertensive medication uptake [18]. Comparative model performance across algorithms and data partitions was summarised in a table, while the distribution of f1-scores across the training, test, and left-out samples for males and females was visualised in a figure. Additionally, a Precision-Recall curve for the best-performing model in each sex category was plotted to illustrate precision across varying sensitivity levels [19].

### 2.3. Feature Importance and Direction of Association

To identify the most parsimonious and informative set of predictors for hypertensive medication uptake, the Sequential Forward Floating Selection (SFFS) algorithm was employed. The SFFS is an iterative feature selection technique that dynamically adds and removes variables to optimise model performance while preventing overfitting. Unlike traditional forward selection, the SFFS algorithm allows conditional exclusion of previously included variables, thereby maintaining flexibility in exploring the search space of predictor subsets [20].

The procedure began with an empty feature set. At each iteration, the variable whose inclusion maximised the f1-score on the training data (using five-fold cross-validation) was added. After each inclusion step, the algorithm performed conditional backward elimination to remove any variable whose exclusion improved the f1-score. The process continued until no further improvement was observed, resulting in an optimal subset of predictors for each sex-specific model.

The relationship between the number of selected variables and the model's f1-score was evaluated separately for males and females to assess how predictive performance changed with increasing model complexity. Results showing the progression of f1-score with the number of selected variables for males and females are plotted, illustrating the optimal point beyond which additional predictors yielded minimal improvement in predictive performance.

To interpret the contribution and directionality of each predictor in the final model, SHapley Additive exPlanations (SHAP) were employed [21]. SHAP provides a unified, model-agnostic framework for explaining complex machine learning models by assigning each feature an additive importance value that reflects its marginal contribution to the prediction outcome.

Following model training, SHAP values were computed separately for males and females to assess differences in the determinants of hypertensive medication uptake. The SHAP analysis enabled the quantification of both the magnitude and direction of each feature's influence on the model output. Features with higher absolute SHAP values were considered to have a greater overall impact on the prediction.

The SHAP summary plots were generated to visualise the ranking and directional effect of the features. Variables were ordered in descending order based on their mean absolute SHAP values, with the most influential features displayed at

the top. Each point in the SHAP plot represents an observation, and its horizontal position indicates the direction and strength of its effect on the predicted probability. Points on the left represent observations that shift the predicted probability in the negative direction (reducing the likelihood of medication uptake or infection risk). Points on the right represent observations that shift the prediction in the positive direction (increasing the likelihood of medication uptake or infection risk). Colour gradients were used to indicate feature values, where red denotes higher feature values associated with increased risk or probability, and blue denotes lower values associated with reduced risk.

This visualisation approach provided an intuitive summary of how explanatory variables influenced the model's predictions, allowing comparison of the dominant predictors and their directional effects between males and females.

## 3. Results

### 3.1. Awareness Rate of Individuals with Self-reported Hypertension

The distribution of hypertension awareness by sex was analyzed as it is shown in Table 2. It was observed that a total of 9,956 individuals were included in the analysis, comprising 5,269 females and 4,687 males. Among females, 534 (10.1%) were classified as hypertensive aware, while 229 (4.9%) of males were aware. Correspondingly, 4,735 (89.9%) females and 4,458 (95.1%) males were classified as not aware.

Similarly, a county-level distribution of hypertension awareness was assessed, and a substantial variability in hypertension awareness was observed across counties and by sex (Table 2). In females, the highest rate was recorded in Tharaka-Nithi (15.5%), followed by Laikipia (14.3%) and Nyeri (13.8%). In contrast, the lowest rate was observed in Kilifi (6.4%), Nyamira (6.4%), and Migori (7.7%). In males, hypertension awareness was comparatively lower across all counties, with the rate ranging between 3.1% in Uasin Gishu and 6.6% in Homa Bay. Other counties with low male hypertension awareness rates included Tharaka-Nithi (3.7%), Kilifi (3.5%), and Laikipia (5.4%).

Across all counties, females consistently exhibited higher hypertension awareness rates compared to males. The female-to-male gap was most pronounced in Tharaka-Nithi (15.5% versus 3.7%), Laikipia (14.3% versus 5.4%), and Nyeri (13.8% versus 5.5%). Counties such as Kilifi (6.4% versus 3.5%) and Nyamira (6.4% versus 5.3%) displayed narrower sex differences.

Overall, the results indicate that the hypertension awareness rate is notably higher among females compared to males across all counties. Geographic variability was evident, with counties in Central Kenya regions (for example, Tharaka-Nithi, Laikipia, Nyeri) showing elevated rates, while Coastal and Western counties (for example, Kilifi, Nyamira, Migori)

exhibited lower rates.

**Table 2.** Distribution of individuals aware of their hypertension status across counties and sex categories.

Characteristics	Levels	Overall		Hypertension awareness	
		Females	Males	Females	Males
n (Total number of individuals,%)		5,269	4,687	534 (10.1)	229 (4.9)
	Baringo	359	314	29 (8.1)	13 (4.1)
	Embu	296	305	30 (10.1)	19 (6.2)
	Homa Bay	374	271	37 (9.9)	18 (6.6)
	Kilifi	393	344	25 (6.4)	12 (3.5)
	Laikipia	300	259	43 (14.3)	14 (5.4)
	Meru	299	325	33 (11.0)	17 (5.2)
	Migori	403	313	31 (7.7)	15 (4.8)
County, n (%)	Mombasa	393	390	42 (10.7)	18 (4.6)
	Nairobi	484	374	51 (10.5)	17 (4.5)
	Nyamira	327	264	21 (6.4)	14 (5.3)
	Nyandarua	323	275	38 (11.8)	16 (5.8)
	Nyeri	275	289	38 (13.8)	16 (5.5)
	Tharaka-Nithi	264	297	41 (15.5)	11 (3.7)
	Uasin Gishu	391	355	44 (11.3)	11 (3.1)
	Vihiga	388	312	31 (8.0)	18 (5.8)

### 3.2. Treatment Uptake in Individuals with Self-reported Hypertension

Overall, among self-reported hypertensive males ( $n = 279$ ), 168 (73.4%) were on medication, while 61 (26.6%) were not on treatment (Table 3). Among hypertensive females ( $n = 1,324$ ), 173 (32.4%) were untreated, and 361 (67.6%) were on medication. Overall, treatment coverage was relatively high among the reported hypertensive individuals. However, a significant treatment gap existed, with females having a substantially higher proportion of untreated cases (32.4%) compared to males (26.6%).

At the county-level distribution, a substantial variability in treatment uptake was observed (Table 3). For instance, in males, the highest proportion of untreated cases occurred in Kilifi (50.0% treated versus 50.0% untreated) and Nairobi (41.2% treated versus 58.8% untreated). The lowest proportions of untreated males were in Embu (15.8% untreated), Homa Bay (16.7%), and Nyandarua (18.8%), suggesting

higher treatment coverage in these areas. For females, the greatest share of untreated females was found in Migori (58.1% untreated) and Kilifi (40.0% untreated). Conversely, higher treatment coverage was observed in Nyandarua (81.6% treated, 18.4% untreated), Vihiga (80.6% treated, 19.4% untreated), and Homa Bay (75.7% treated, 24.3% untreated).

Comparing sex differences, while untreated proportions were high in both sexes, females were generally more likely to remain untreated compared to males (68.7% versus 64.9%). In counties such as Migori, this disparity was particularly pronounced, with 58.1% of females untreated compared to 26.7% of males. In contrast, in Embu, a higher proportion of males remained untreated (84.2% treated versus 56.7% treated in females).

Overall, across all counties, the majority of hypertensive individuals were not on medication, highlighting significant gaps in hypertension management. Regional disparities were evident, with counties such as Nyandarua, Vihiga, and Homa Bay showing relatively higher treatment uptake, while Kilifi, Migori, and Nairobi reported the lowest coverage.

**Table 3.** County- and sex-specific distribution of hypertensive individuals according to treatment status.

Characteristics	Levels	Males on medication			Females on medication		
		Total No.	Yes	No	Total No.	Yes	No
n (Total number of individuals,%)	Overall	229	168 (73.4)	61 (26.6)	534	361 (67.6)	173 (32.4)
	Baringo	13	10 (76.9)	3 (23.1)	29	20 (69.0)	9 (31.0)
	Embu	19	16 (84.2)	3 (15.8)	30	17 (56.7)	13 (43.3)
	Homa Bay	18	15 (83.3)	3 (16.7)	37	28 (75.7)	9 (24.3)
	Kilifi	12	6 (50.0)	6 (50.0)	25	15 (60.0)	10 (40.0)
	Laikipia	14	10 (71.4)	4 (28.6)	43	27 (62.8)	16 (37.2)
	Meru	17	12 (70.6)	5 (29.4)	33	25 (75.8)	8 (24.2)
	Migori	15	11 (73.3)	4 (26.7)	31	13 (41.9)	18 (58.1)
	Mombasa	18	12 (66.7)	6 (33.3)	42	26 (61.9)	16 (38.1)
County, n (%)	Nairobi	17	10 (58.8)	7 (41.2)	51	30 (58.8)	21 (41.2)
	Nyamira	14	11 (78.6)	3 (21.4)	21	13 (61.9)	8 (38.1)
	Nyandarua	16	13 (81.2)	3 (18.8)	38	31 (81.6)	7 (18.4)
	Nyeri	16	12 (75.0)	4 (25.0)	38	28 (73.7)	10 (26.3)
	Tharaka-Nithi	11	9 (81.8)	2 (18.2)	41	30 (73.2)	11 (26.8)
	Uasin Gishu	11	8 (72.7)	3 (27.3)	44	33 (75.0)	11 (25.0)
	Vihiga	18	13 (72.2)	5 (27.8)	31	25 (80.6)	6 (19.4)

### 3.3. Model Performance

The predictive performance of four machine learning algorithms, XGBoost, Random Forest (RF), Support Vector Machine (SVM), and Elastic Net (EN), was evaluated across 15 Kenyan counties, stratified by sex (male and female). Performance was rigorously assessed using three sampling strategies: training, test, and leave-one-out cross-validation (LOOCV), with the f1-score as the primary metric, complemented by recall and precision (Figure 5-12, Tables 4-8). The analysis revealed pronounced differences in model performance, with tree-based models (XGBoost and RF) consistently outperforming SVM and EN across most settings.

#### 3.3.1. Model Performance in Males

Across all validation strategies, XGBoost demonstrated the highest overall performance in males, achieving a mean LOOCV f1-score of 82.6%, followed closely by RF (81.3%, SD = 7.8) and SVM (81.3%, SD = 7.9). In contrast, Elastic Net performed substantially worse, with a mean LOOCV f1-score of only 70.3% (SD = 9.7). This pattern was consistent in the held-out test set, where SVM exhibited a slight advantage (mean F1 = 76.3%, SD = 5.8), with XGBoost (76.0%, SD = 5.3) and RF (75.5%, SD = 2.6) performing comparably. EN

again lagged (mean F1 = 70.5%, SD = 4.7).

Notable county-level heterogeneity was observed. The highest male LOOCV f1-scores for XGBoost were recorded in Homa Bay (93.8%), Migori (91.7%), and Nyandarua (89.7%), indicating excellent generalizability in these populations. Conversely, the lowest performance was in Kilifi (66.7% for XGBoost, SVM, and RF), suggesting contextual factors limiting model transferability. A similar pattern was observed for RF, with exceptional LOOCV performance in Laikipia (90.9%) and Nyandarua (92.9%), but again poor performance in Kilifi (66.7%). Recall values for XGBoost and SVM in males were exceptionally high across many counties (often 100% in LOOCV), indicating near-perfect identification of positive cases, albeit sometimes at the cost of precision (e.g., Kilifi: Recall = 100%, Precision = 50%).

#### 3.3.2. Model Performance in Females

In the male cohort, XGBoost again achieved the highest mean LOOCV f1-score (79.9%, SD = 9.0), followed by RF (78.5%, SD = 10.1). SVM (69.6%, SD = 13.6) and Elastic Net (60.5%, SD = 15.2) demonstrated markedly inferior and more variable performance. On the test set, RF slightly outperformed XGBoost (mean F1 = 80.7%, SD = 1.1 vs. 80.4%, SD = 0.9), while EN remained the weakest (mean F1 = 68.3%, SD

= 2.8).

As with females, significant between-county variability was evident. Female LOOCV f1-scores for XGBoost were highest in Nyandarua (90.2%), Meru (90.9%), and Vihiga (88.4%), but notably low in Migori (62.1%) and Kilifi (63.6%). RF performance mirrored this trend, with excellent generalizability in Nyandarua (90.2%) and Meru (90.9%), but poor performance in Migori (60.0%) and Nyamira (60.9%). Strikingly, SVM and EN displayed severe performance degradation in specific counties. For SVM in females, LOOCV f1-scores fell below 50% in Migori (44.4%) and Nyamira (44.4%). Similarly, EN performance collapsed in several counties, including Meru (40.0%), Migori (43.5%), Mombasa (51.2%), and Tharaka-Nithi (36.4%), with precision fluctuating dramatically from 35.7% to 100%.

### 3.3.3. Comparative Summary and Key Observations

Across all datasets and evaluation strategies, a consistent hierarchy of model performance emerged: XGBoost  $\approx$  RF >

SVM >> EN. Tree-based ensemble methods demonstrated robust and balanced performance, characterised by high recall (often exceeding 95% in training and test sets) and moderate-to-good precision, suggesting their suitability for this predictive task. Notably, XGBoost exhibited superior generalizability in LOOCV for both sexes, a critical indicator of real-world utility.

In contrast, while SVM achieved strong LOOCV recall (frequently 100% in females), its performance was brittle, with pronounced f1-score variability across counties (e.g., female LOOCV F1 ranging from 44.4% to 89.5%). Elastic Net was consistently the poorest performer across all metrics and sample types, with mean test f1-scores below 70% for both sexes, and exhibited a striking failure in LOOCV for several female cohorts (e.g., Nyamira, Tharaka-Nithi, Migori), where f1-scores dropped below 45%. These findings confirm the unsuitability of linear models (EN) and the context-dependent limitations of kernel-based methods (SVM) for this complex, spatially heterogeneous prediction problem, while affirming the relative strength and stability of gradient-boosted and random forest approaches.

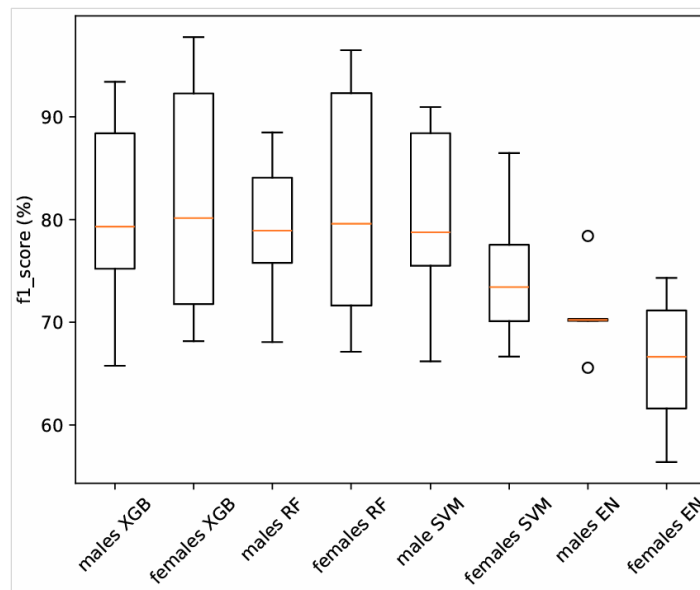


Figure 1. Box plots for model performance measured by f1-score in leave-one-out and test samples.

Table 4. Predictive performance of the four models based on f1-score.

Model	Train F1	Train Recall	Train Precision	Test F1	Test Recall	Test Precision	LOOC F1	LOOC Recall	LOOC Precision
males XGB	78.69	94.13	67.78	75.98	90.31	65.76	82.63	93.41	74.95
females XGB	81.43	97.04	70.21	80.42	96.23	69.07	79.87	97.76	68.16
males RF	79.91	87.11	74.15	75.53	85.00	68.07	81.31	88.48	76.52
females RF	83.55	96.47	73.77	80.67	96.49	69.33	78.51	96.19	67.13

Model	Train F1	Train Recall	Train Precision	Test F1	Test Recall	Test Precision	LOOC F1	LOOC Recall	LOOC Precision
male SVM	79.90	90.54	73.25	76.26	90.95	66.19	81.25	90.79	75.24
females SVM	82.56	80.62	84.78	78.31	86.47	71.61	69.60	75.23	66.65
males EN	68.60	65.13	73.01	70.16	70.13	70.29	70.33	65.57	78.40
females EN	69.47	65.24	74.61	68.31	64.95	72.09	60.49	56.39	74.32

### 3.4. Variable Selection and Predictive Performance

The Sequential Forward Floating Selection (SFFS) algorithm was employed to identify the most parsimonious set of predictors for hypertensive medication uptake. The relationship between the number of variables and the model's f1-score for males and females is presented in Figure 2.

For both sexes, the f1-score increased rapidly with the initial addition of variables before reaching a clear plateau. The point of diminishing returns, where adding more variables provided no substantial improvement in performance, was

identified at 8 variables for both males and females (Figures 2A and 2B). This set of variables was therefore selected for the final models.

The eight predictive variables selected for males were: number of household members (total listed), current age, number of minutes per week doing physical exercise, number of hours per day seated, how much was paid in the last month, use of the internet, highest educational level, and type of place of residence. Similarly, the eight predictive variables selected for females were: current age, years lived in place of residence, number of hours per day seated, number of household members (total listed), wealth index for urban/rural, frequency of watching television, ideal number of children, and number of minutes per week of exercise.

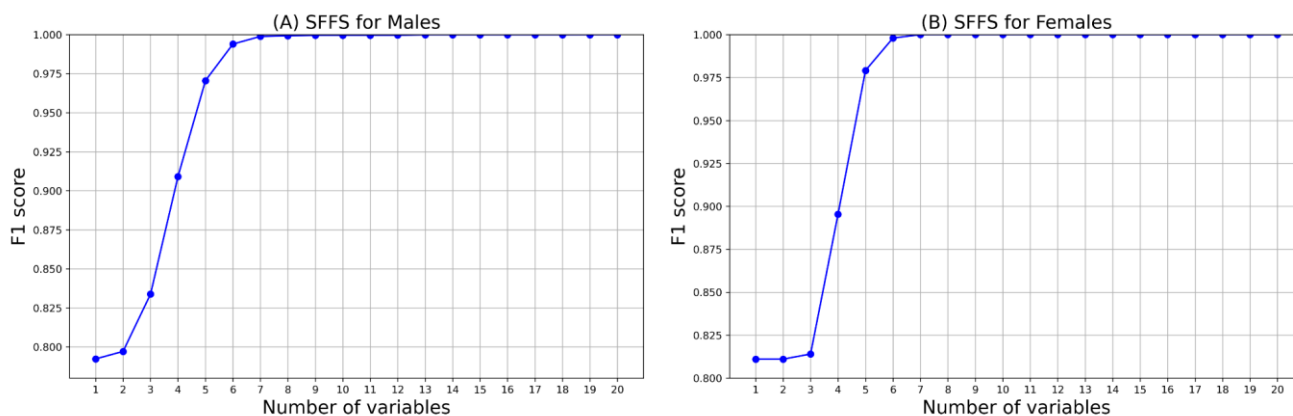


Figure 2. Variable selection from SFFS (A: Males, B: Females).

### 3.5. Final Model and Variable Associations with Treatment Uptake

Following variable selection, the final model was trained and evaluated. The performance in classifying hypertensive medication uptake is summarised by the Precision-Recall (PR) curves shown in Figure 3. The model demonstrated exceptional predictive accuracy for both sexes. For males (Figure

3A), the model achieved an f1-score of 0.94 and a high Area Under the PR Curve (AUC) of 0.96. Performance was even stronger for females (Figure 3B), with a near-perfect f1-score of 0.97 and an AUC of 0.99.

These results indicate that the models, built upon the selected sets of eight variables for each sex, have a very high ability to correctly identify individuals on hypertensive medication, with an outstanding balance between precision and recall.

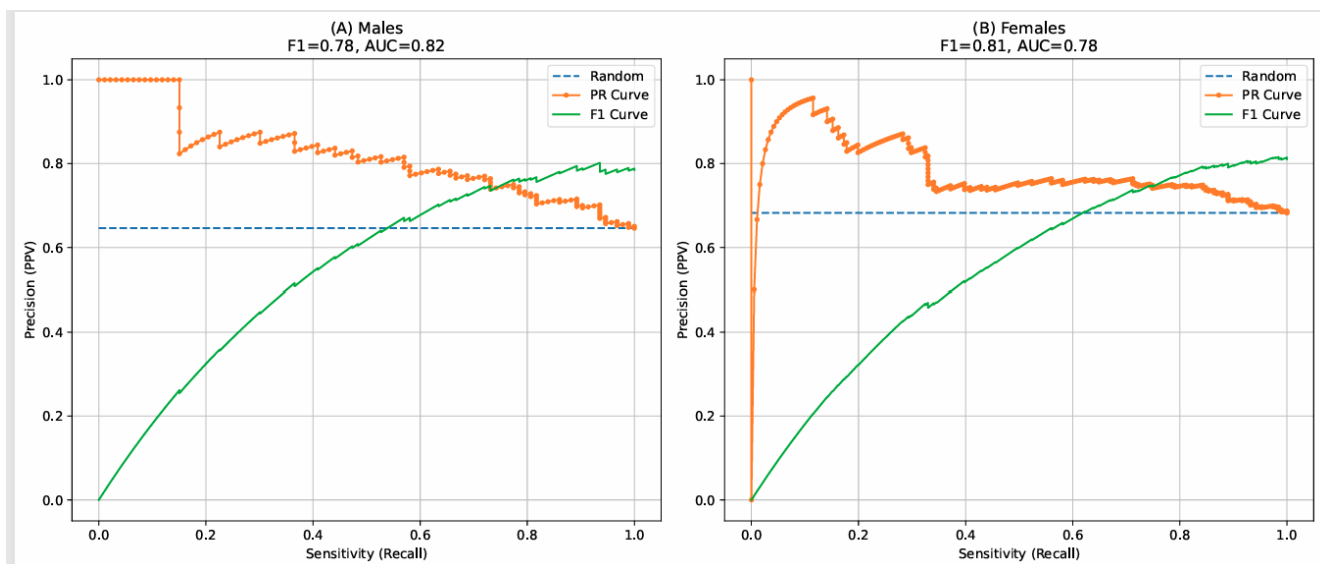


Figure 3. Precision–Recall (PR) curves illustrating the performance of the final model: (A) males and (B) females.

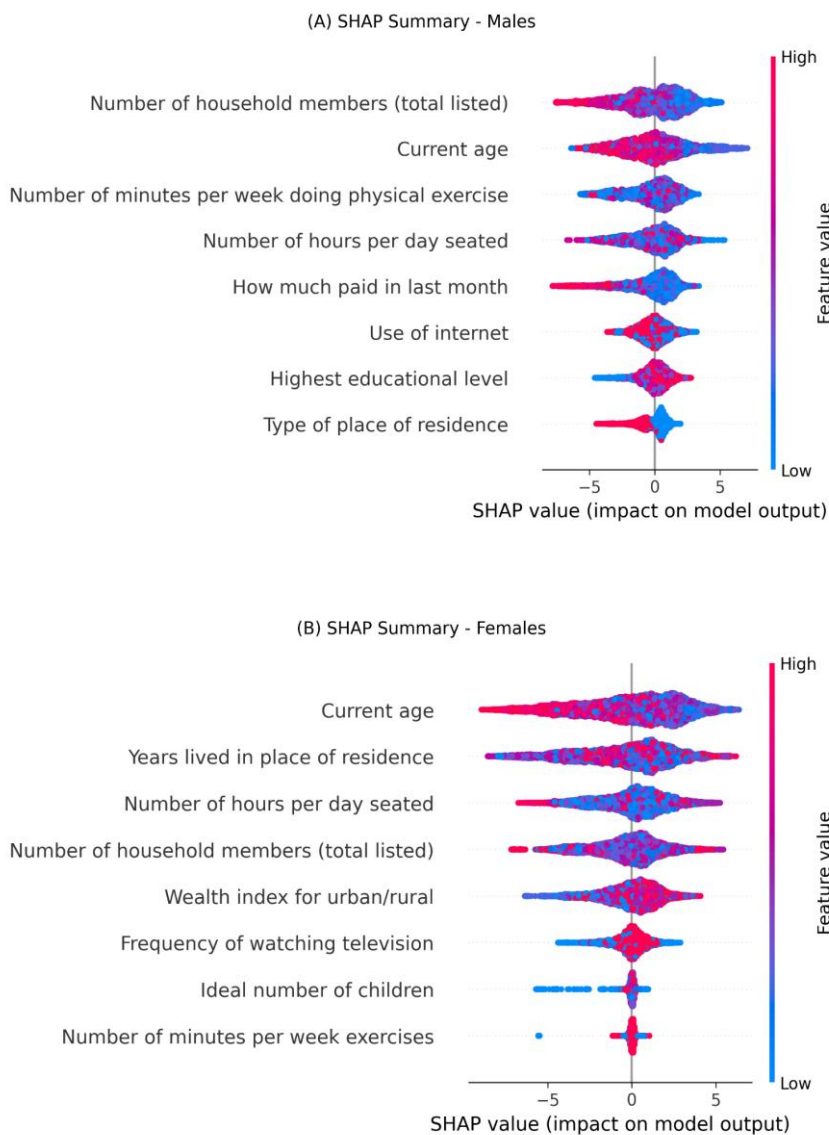


Figure 4. SHAP value plots for direction of associations (A: males, B: females; red indicates positive, blue indicates negative).

Finally, the feature importance and direction of the association were determined with SHAP (SHapley Additive exPlanations). The analysis of feature importance and directionality using SHAP revealed distinct patterns in the factors associated with hypertensive medication uptake between males and females, as illustrated in Figure 4. The graph summarises the impact of explanatory features on the model output and indicates the relative contribution of predictors to the prediction outcome rather than implying causal relationships.

For males (Figure 4A), the strongest positive drivers for medication uptake were older current age and a higher number of household members. This suggests that older men and those with larger households were more likely to be on medication. Conversely, a higher number of minutes per week doing physical exercise was the most prominent negative driver, indicating that men who engaged in more physical activity were less likely to be on medication. Other notable factors associated with a lower likelihood of medication use included a greater number of hours per day seated and higher earnings (How much was paid in the last month). Socioeconomic and infrastructural factors such as use of internet, highest educational level, and type of place of residence also featured among the top predictors, but with a comparatively lower mean impact on the model output.

For females (Figure 4B), the model identified a different set of key predictors. Current age was again the most influential feature, with older age strongly predicting medication use. A longer duration of residence (Years lived in place of residence) and a higher wealth index for urban/rural were also positive drivers for medication uptake. In contrast, a greater number of hours per day spent seated was associated with a reduced likelihood of being on medication. Interestingly, while the ideal number of children and a higher frequency of watching television were among the top features, their impact on the model output was negative. The number of minutes per week of exercise and the number of household members were also identified as relevant predictors for females, though with a lower mean compared to the top features.

In summary, while advancing age was a consistent and strong predictor for medication uptake across both sexes, the other major drivers exhibited significant sexual dimorphism. For males, household size and physical inactivity were paramount, whereas for females, stability factors (years in residence, wealth) and distinct sociodemographic measures (ideal number of children, TV viewing) were more influential.

## 4. Discussion

We analysed data from 9,956 respondents in a sample of 15 counties in Kenya to examine self-reported hypertension awareness and treatment uptake, using socio-behavioural and demographic factors. Among the participants, 10.1% of females and 4.9% of males were aware of their hypertension sta-

tus. In general, treatment coverage among self-reported hypertensive individuals was relatively high; however, a notable treatment gap remained, with 26.6% of hypertensive males and 32.4% of hypertensive females not receiving medication.

At the county level, substantial geographic variability was evident: females in Tharaka-Nithi (15.5%), Laikipia (14.3%), and Nyeri (13.8%) recorded the highest awareness rates, while Kilifi and Nyamira had the lowest (6.4%). Male awareness remained consistently lower across all counties, ranging from 3.1% in Uasin Gishu to 6.6% in Homa Bay, with the female-male gap most pronounced in Central counties such as Tharaka-Nithi and Laikipia.

Similarly, treatment uptake exhibited notable spatial and sex differences. Among hypertensive males, the lowest treatment coverage occurred in Kilifi (50.0% untreated) and Nairobi (58.8% untreated), whereas Embu, Homa Bay, and Nyandarua showed better treatment coverage. In females, the largest share of untreated cases was recorded in Migori (58.1%) and Kilifi (40.0%), while Nyandarua, Vihiga, and Homa Bay exhibited the highest treatment coverage. In general, females were more likely to remain untreated than males (68.7% versus 64.9%), with pronounced disparities observed in Migori, underscoring persistent geographic and sex-related inequities in hypertension management.

The XGBoost model was selected among the other three models (SVM, RF, and EN) as the optimal model for identifying socio-behavioural predictors of hypertension treatment uptake, given its superior accuracy (78%) among males and comparably strong performance (81%) among females, and was subsequently trained using the full dataset to enhance predictive robustness and generalizability.

Using the SFFS procedure, the study identified a parsimonious set of socio-behavioural predictors associated with hypertension medication uptake, highlighting key differences by sex. Among males, the most influential predictors included household size, current age, duration of physical exercise per week, hours spent seated per day, monthly earnings, internet use, education level, and type of residence. These factors collectively reflect both lifestyle behaviours and socioeconomic context, suggesting that sedentary patterns, economic capacity, and access to information may significantly influence treatment adherence among men. For females, the selected predictors are current age, duration of residence, sedentary time, household size, wealth index, television viewership frequency, ideal number of children, and exercise time, highlighting the intersection of socioeconomic status, lifestyle, and reproductive preferences in shaping treatment behaviour. The inclusion of both media exposure and wealth indicators among women suggests that awareness and empowerment may play a critical role in medication uptake.

The SHAP analysis provided valuable insights into the relative importance and directionality of socio-behavioural predictors influencing hypertension medication uptake, revealing distinct sex-specific patterns. Among males, older age and

larger household size emerged as the strongest positive drivers of treatment uptake, suggesting that ageing and social support within households may enhance medication adherence [21, 22]. In contrast, greater weekly physical activity was the most influential negative predictor, indicating that physically active men may perceive themselves as healthier and thus less in need of pharmacologic treatment [23, 24]. Additional negative associations were observed with prolonged sedentary behaviour and higher monthly income, implying that lifestyle choices and economic priorities may modulate treatment decisions [25-28]. Factors such as internet use, educational attainment, and place of residence also contributed to prediction, though with smaller effects, reflecting the multidimensional nature of health behaviour among men [29].

For females, the SHAP analysis highlighted a different constellation of influential factors. Similar to males, age remained the most critical predictor, with older women more likely to be on medication [4, 30]. Longer duration of residence and higher wealth index values were also positively associated with treatment uptake, suggesting that social stability and economic capacity may facilitate access to healthcare services. Conversely, increased sedentary time, higher television viewership, and a greater ideal number of children were negatively associated with medication use, possibly reflecting competing domestic priorities, limited health engagement, or lower perceived risk. The number of minutes per week exercising and household size showed weaker but notable contributions. Together, these results highlight the complex interplay between demographic, socioeconomic, and behavioural determinants of hypertension management and emphasise the value of interpretable machine learning approaches such as SHAP in uncovering sex-specific pathways that can inform more tailored and equitable intervention strategies.

These findings highlight the importance of integrating socio-behavioural, demographic, and economic dimensions into hypertension control strategies. The observed sex-specific predictors highlight the need for differentiated intervention approaches that address the unique drivers of medication uptake among men and women. For males, initiatives that enhance health literacy and encourage sustained treatment adherence even among physically active individuals could improve outcomes. For females, interventions that consider household dynamics, reproductive priorities, and access to media and health information may be more effective. At the policy level, leveraging community health programs, digital health platforms, and sex-responsive outreach can help close the awareness and treatment gaps identified in this study. Ultimately, the integration of interpretable machine learning models such as SHAP offers a powerful framework for evidence-based targeting of hypertension interventions, enabling more precise, equitable, and sustainable public health responses.

Our approach offers a valuable complementary tool for

identifying individuals who are most likely to benefit from enhanced hypertension mitigation strategies, particularly in settings where clinical testing is unavailable, resource-limited, or prohibitively costly. By leveraging socio-behavioural and demographic data, the model provides a data-driven means of prioritising high-risk populations for early intervention, screening, and targeted health promotion. This not only strengthens population-level disease surveillance but also supports equitable allocation of healthcare resources, thereby enhancing the efficiency and inclusiveness of hypertension prevention and control programs.

This study has several limitations that should be acknowledged. First, the analysis was restricted to individuals aged 15 to 49 years, consistent with the Demographic and Health Survey framework. This age restriction excludes older adults who bear a greater burden of hypertension, thereby limiting the generalizability of the findings to the broader population. Second, the predictive model's validity may have been affected by missing data and reliance on self-reported measures, both of which are subject to recall bias and potential misclassification. These factors may have introduced uncertainty in model training and predictive performance. Future research should aim to validate these findings using clinically confirmed data and include wider age groups to enhance robustness, accuracy, and external validity. Despite these limitations, the study demonstrates the promise of interpretable machine learning approaches in identifying key socio-behavioural predictors and guiding targeted hypertension control strategies in data-limited settings.

## Abbreviations

KDHS	Kenya Demographic and Health Surveys
EN	Elastic Net
RF	RandomForest
SVM	Support Vector Machine
GAM	Generalized Additive Model
CVDs	Cardiovascular Diseases
NCDs	Noncommunicable Diseases
SFFS	Sequential Forward Floating Selection
SHAP	SHapley Additive exPlanations
WHO	World Health Organization
PR	Precision-Recall

## Author Contributions

**Eliud Koeh:** Conceptualization, Data curation, Software, Formal Analysis, Visualization, Writing – original draft, Writing – review & editing

**Charles Kipkoeh Mutai:** Conceptualization, Supervision, Validation, Writing – review & editing

**Gregory Kerich:** Validation, Methodology, Writing – review & editing

## Conflicts of Interest

The authors declare that they have no competing interests. All authors approved the final manuscript.

## Appendix

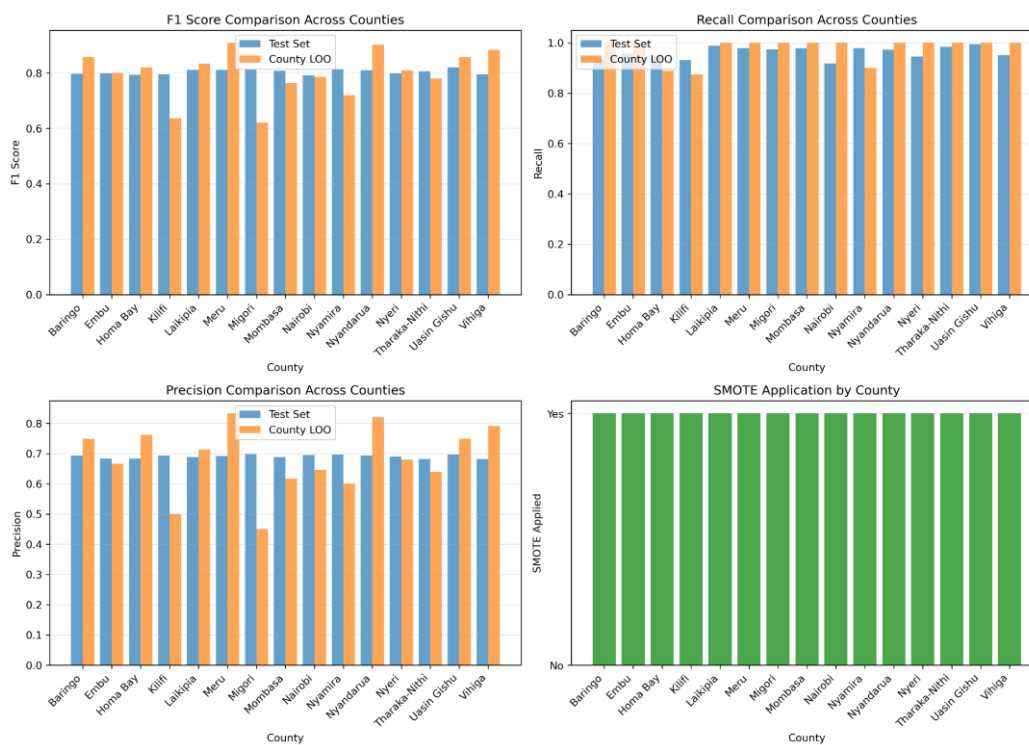


Figure 5. XGBoost in Females.

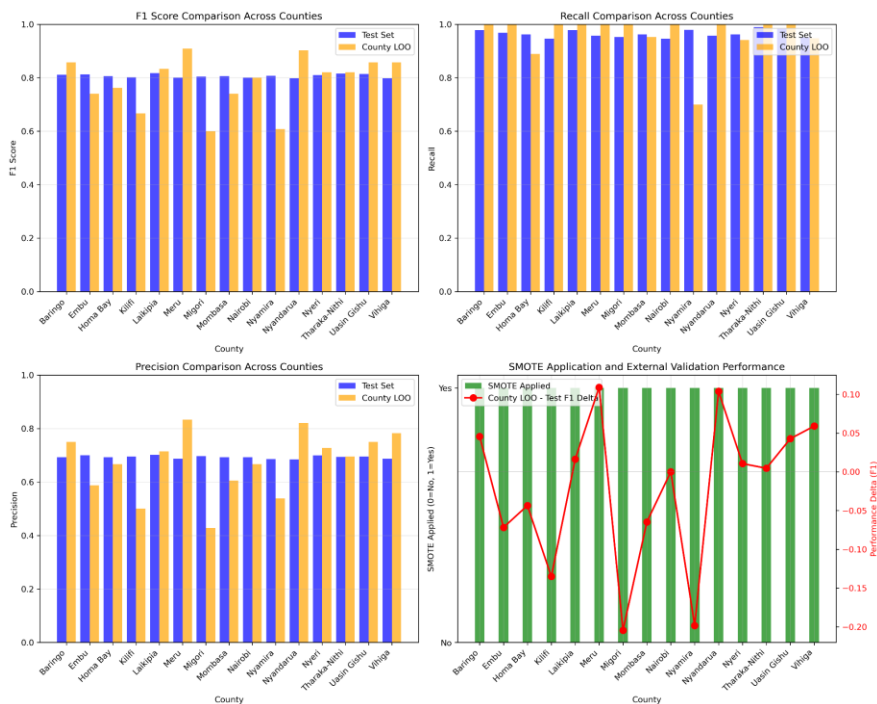


Figure 6. Random Forest in Females.

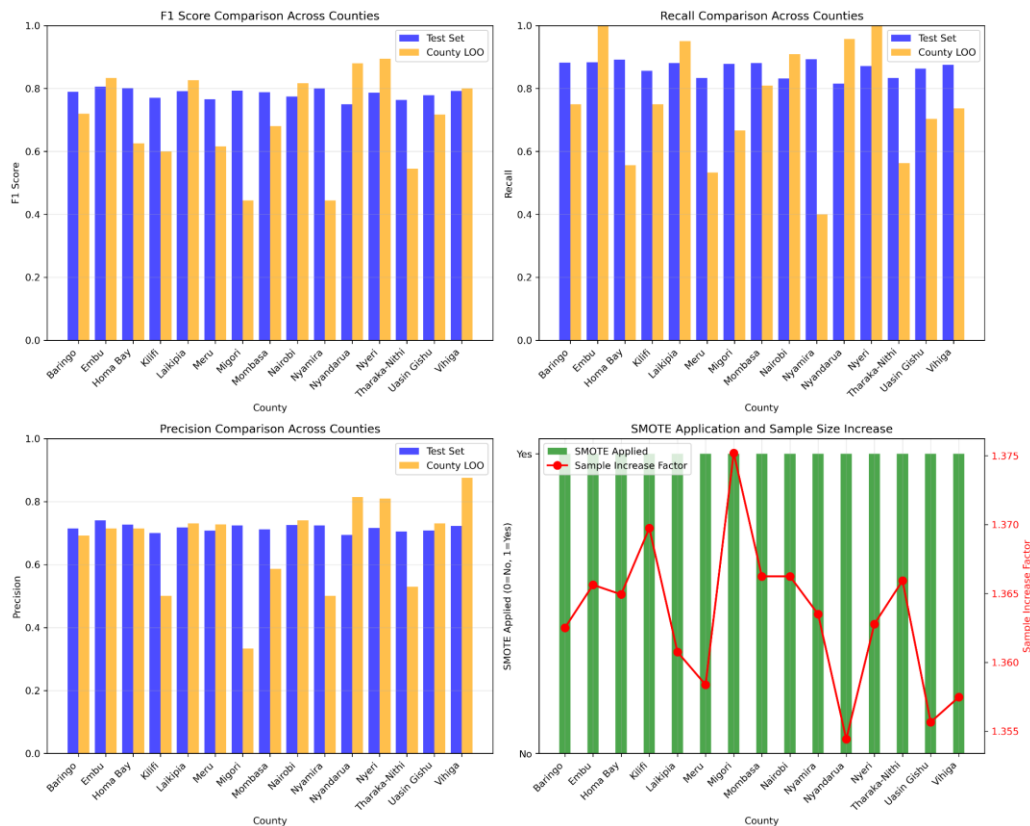


Figure 7. Support Vector Machine in Females.

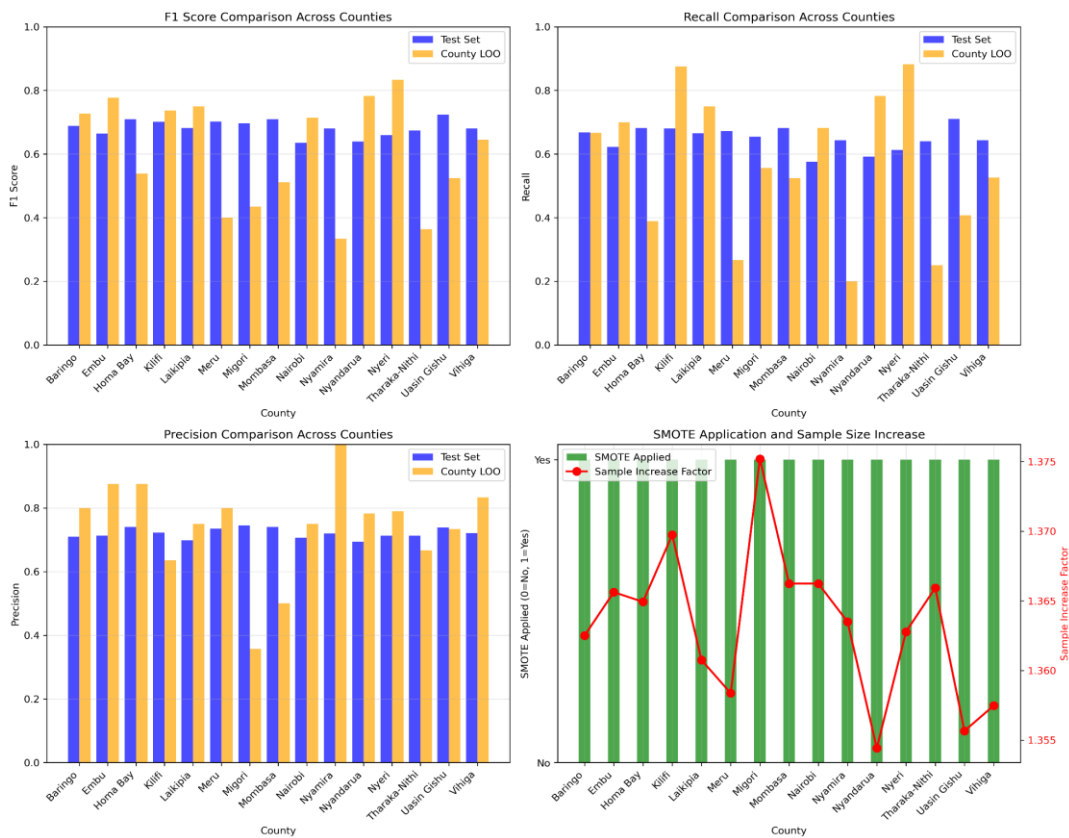


Figure 8. ElasticNet in Females.

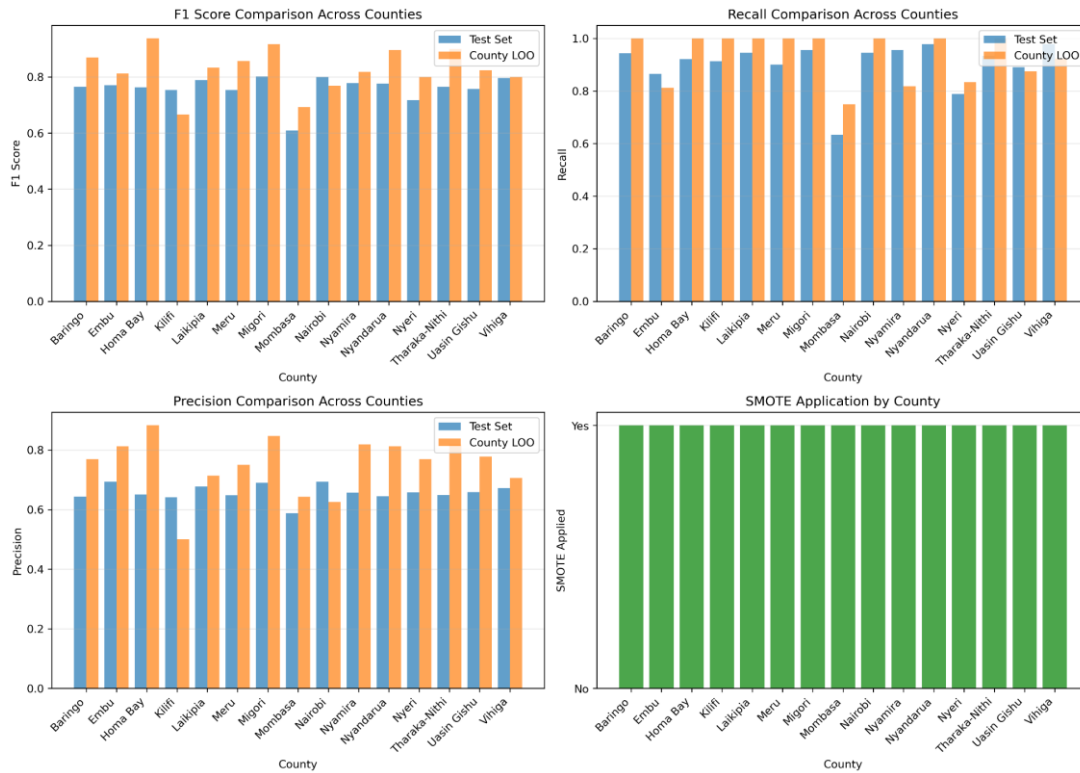


Figure 9. XGBoost in Males.

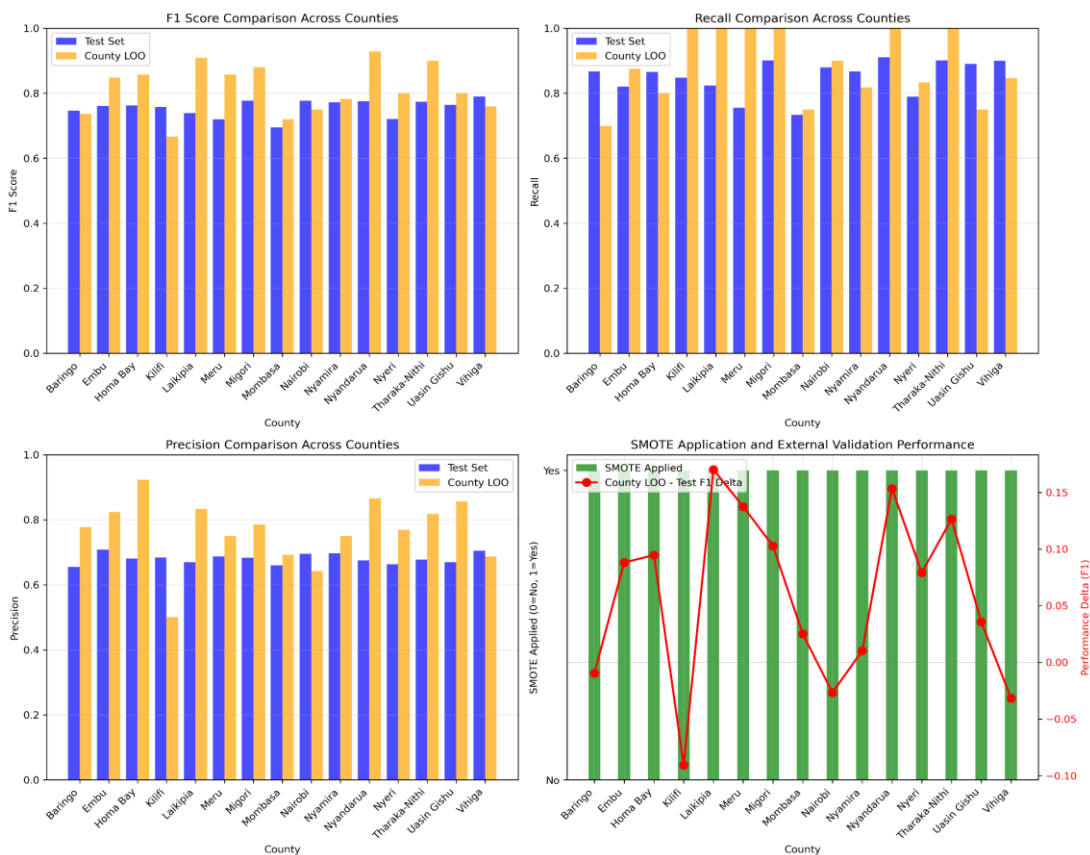


Figure 10. Random Forest in Males.

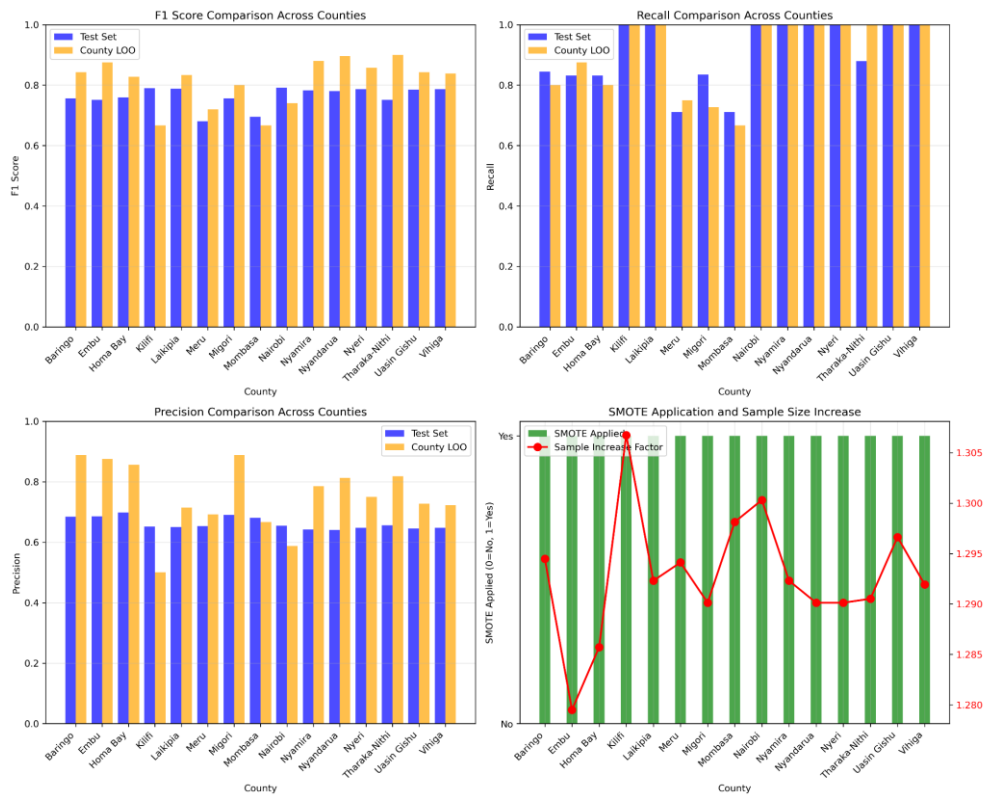


Figure 11. Support Vector Machine in Males.

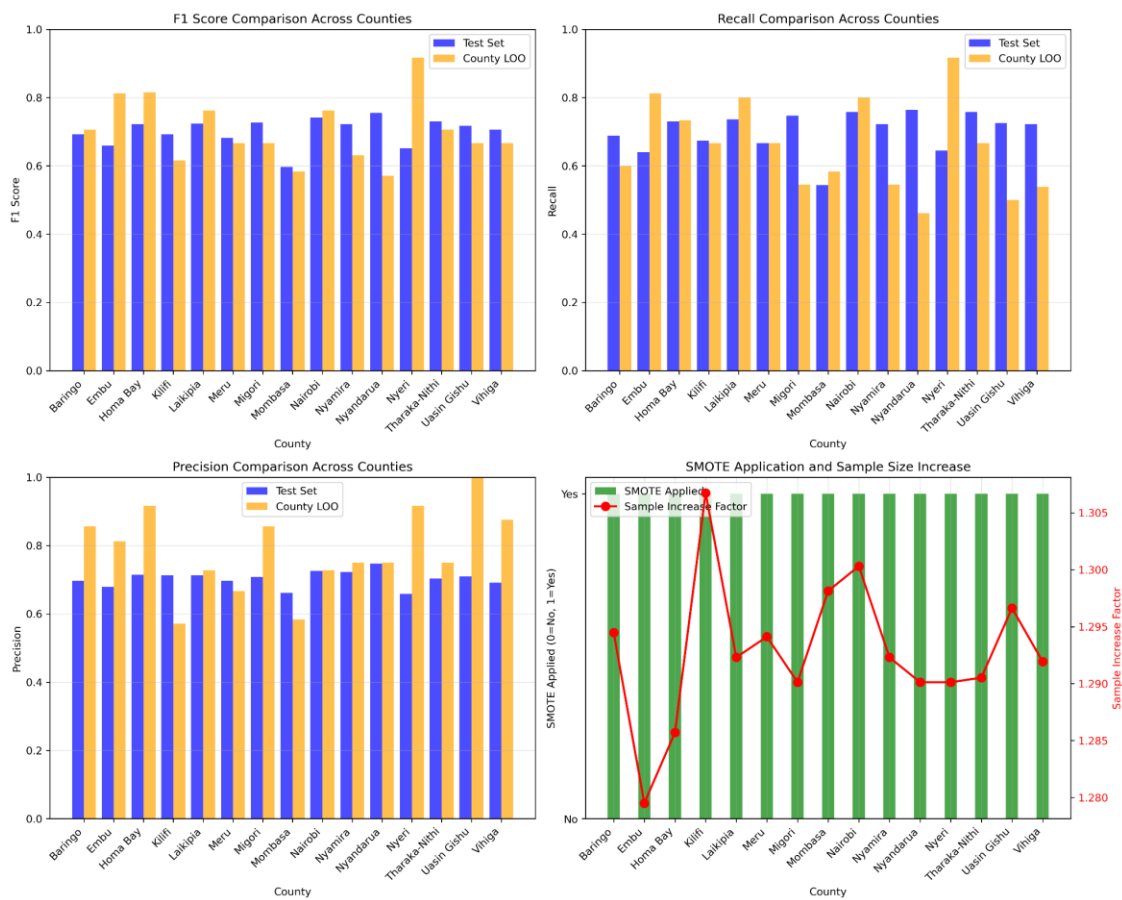


Figure 12. ElasticNet in Males.

**Table 5.** XGBoost Performance with SMOTE Across Counties and Sex.

County	Sex	Train F1	Train Rec	Train Prec	Test F1	Test Rec	Test Prec	LOOC F1	LOOC Rec	LOOC Prec
Baringo	Males	78.5	95.7	66.6	76.6	94.4	64.4	87.0	100	76.9
	Females	81.7	96.1	71.1	79.7	93.6	69.4	85.7	100	75.0
Embu	Males	77.2	89.3	68.1	77.0	86.5	69.4	81.2	81.2	81.2
	Females	82.4	97.5	71.5	79.8	95.7	68.4	80.0	100	66.7
Homa Bay	Males	78.8	93.3	68.3	76.3	92.1	65.1	93.8	100	88.2
	Females	81.3	97.0	70.0	79.4	94.6	68.4	82.1	88.9	76.2
Kilifi	Males	79.5	97.2	67.3	75.3	91.3	64.1	66.7	100	50.0
	Females	82.0	96.8	71.1	79.5	93.1	69.4	63.6	87.5	50.0
Laikipia	Males	78.1	95.7	66.1	78.9	94.5	67.7	83.3	100	71.4
	Females	81.4	98.8	69.2	81.2	98.9	68.8	83.3	100	71.4
Meru	Males	79.6	94.7	68.8	75.3	90.0	64.8	85.7	100	75.0
	Females	81.2	97.2	69.7	81.1	97.8	69.2	90.9	100	83.3
Migori	Males	77.5	91.4	67.3	80.2	95.6	69.0	91.7	100	84.6
	Females	82.2	98.4	70.6	81.3	97.3	69.8	62.1	100	45.0
Mombasa	Males	81.8	90.4	75.0	61.0	63.3	58.8	69.2	75.0	64.3
	Females	81.5	97.4	70.1	80.8	97.8	68.8	76.4	100	61.8
Nairobi	Males	77.8	95.2	65.9	80.0	94.5	69.4	76.9	100	62.5
	Females	80.8	94.9	70.4	79.2	91.8	69.5	78.6	100	64.7
Nyamira	Males	78.2	94.8	66.7	77.8	95.6	65.6	81.8	81.8	81.8
	Females	81.2	96.6	70.1	81.4	97.9	69.7	72.0	90.0	60.0
Nyandarua	Males	77.8	97.1	64.9	77.7	97.8	64.4	89.7	100	81.2
	Females	80.7	97.7	68.8	81.0	97.3	69.4	90.2	100	82.1
Nyeri	Males	78.6	93.3	68.0	71.7	78.9	65.7	80.0	83.3	76.9
	Females	80.7	94.9	70.3	79.8	94.6	69.0	81.0	100	68.0
Tharaka-Nithi	Males	79.3	95.7	67.7	76.6	93.4	64.9	90.0	100	81.8
	Females	81.7	97.7	70.3	80.6	98.4	68.3	78.0	100	64.0
Uasin Gishu	Males	79.1	94.3	68.2	75.7	89.0	65.9	82.4	87.5	77.8
	Females	81.2	97.2	69.8	82.0	99.5	69.7	85.7	100	75.0
Vihiga	Males	78.6	93.8	67.8	79.6	97.8	67.2	80.0	92.3	70.6
	Females	81.5	97.4	70.1	79.5	95.1	68.2	88.4	100	79.2

**Table 6.** Random Forest Performance with SMOTE Across Counties and Sex.

County	Sex	Train F1	Train Rec	Train Prec	Test F1	Test Rec	Test Prec	LOOC F1	LOOC Rec	LOOC Prec
Baringo	Males	79.8	86.7	74.1	74.6	86.7	65.5	73.7	70.0	77.8
	Females	83.1	96.6	73.0	81.2	97.9	69.3	85.7	100	75.0

County	Sex	Train F1	Train Rec	Train Prec	Test F1	Test Rec	Test Prec	LOOC F1	LOOC Rec	LOOC Prec
Embu	Males	78.0	84.4	72.6	76.0	82.0	70.9	84.8	87.5	82.4
	Females	83.7	97.2	73.5	81.2	96.8	70.0	74.1	100	58.8
Homa Bay	Males	80.5	89.8	72.9	76.2	86.5	68.1	85.7	80.0	92.3
	Females	83.7	95.8	74.4	80.5	96.2	69.3	76.2	88.9	66.7
Kilifi	Males	80.8	89.7	73.8	75.7	84.8	68.4	66.7	100	50.0
	Females	83.1	94.1	74.5	80.2	94.7	69.5	66.7	100	50.0
Laikipia	Males	80.3	86.7	75.5	73.9	82.4	67.0	90.9	100	83.3
	Females	83.5	95.8	74.1	81.7	97.8	70.2	83.3	100	71.4
Meru	Males	79.3	84.7	75.0	72.0	75.6	68.7	85.7	100	75.0
	Females	82.9	95.9	73.2	80.0	95.7	68.7	90.9	100	83.3
Migori	Males	79.0	86.6	73.1	77.7	90.1	68.3	88.0	100	78.6
	Females	84.0	96.8	74.3	80.4	95.2	69.6	60.0	100	42.9
Mombasa	Males	82.0	88.1	77.3	69.5	73.3	66.0	72.0	75.0	69.2
	Females	83.6	96.3	73.9	80.5	96.2	69.3	74.1	95.2	60.6
Nairobi	Males	79.6	88.6	72.5	77.7	87.9	69.6	75.0	90.0	64.3
	Females	83.6	96.7	73.6	80.0	94.6	69.3	80.0	100	66.7
Nyamira	Males	79.6	88.1	73.1	77.2	86.7	69.6	78.3	81.8	75.0
	Females	83.9	97.9	73.3	80.7	97.9	68.7	60.9	70.0	53.8
Nyandarua	Males	79.0	88.5	71.5	77.5	91.0	67.5	92.9	100	86.7
	Females	84.1	97.9	73.8	79.8	95.7	68.5	90.2	100	82.1
Nyeri	Males	80.6	87.5	75.0	72.1	78.9	66.4	80.0	83.3	76.9
	Females	83.7	94.9	75.0	81.0	96.2	69.9	82.1	94.1	72.7
Tharaka-Nithi	Males	80.0	85.3	75.6	77.4	90.1	67.8	90.0	100	81.8
	Females	83.2	96.8	73.1	81.6	98.9	69.4	82.1	100	69.6
Uasin Gishu	Males	80.3	85.8	75.5	76.4	89.0	66.9	80.0	75.0	85.7
	Females	83.2	97.6	72.6	81.4	98.4	69.5	85.7	100	75.0
Vihiga	Males	79.9	86.1	74.7	79.0	90.0	70.4	75.9	84.6	68.8
	Females	84.0	96.7	74.3	79.8	95.1	68.8	85.7	94.7	78.3

*Table 7. Support Vector Machine Performance with SMOTE Across Counties and Sex.*

County	Sex	Train F1	Train Rec	Train Prec	Test F1	Test Rec	Test Prec	LOOC F1	LOOC Rec	LOOC Prec
Baringo	Males	80.8	82.4	79.9	75.6	84.4	68.5	84.2	80.0	88.9
	Females	82.8	81.0	84.7	78.9	88.2	71.4	72.0	75.0	69.2
Embu	Males	77.3	76.7	78.5	75.1	83.1	68.5	87.5	87.5	87.5
	Females	83.0	81.7	84.5	80.6	88.3	74.1	83.3	100	71.4
Homa Bay	Males	80.1	80.2	80.4	75.9	83.1	69.8	82.8	80.0	85.7

County	Sex	Train F1	Train Rec	Train Prec	Test F1	Test Rec	Test Prec	LOOC F1	LOOC Rec	LOOC Prec
Kilifi	Females	81.9	80.3	83.8	80.1	89.2	72.7	62.5	55.6	71.4
	Males	80.7	100	67.6	79.0	100	65.2	66.7	100	50.0
Laikipia	Females	83.5	80.6	86.6	77.0	85.6	70.0	60.0	75.0	50.0
	Males	80.2	100	66.9	78.8	100	65.0	83.3	100	71.4
Meru	Females	81.5	80.7	82.4	79.1	88.1	71.8	82.6	95.0	73.1
	Males	79.4	78.5	80.8	68.1	71.1	65.3	72.0	75.0	69.2
Migori	Females	83.5	82.5	84.6	76.5	83.3	70.8	61.5	53.3	72.7
	Males	83.0	81.3	85.0	75.6	83.5	69.1	80.0	72.7	88.9
Mombasa	Females	85.0	82.6	87.5	79.3	87.8	72.4	44.4	66.7	33.3
	Males	81.5	79.9	84.1	69.6	71.1	68.1	66.7	66.7	66.7
Nairobi	Females	80.6	78.3	83.2	78.7	88.1	71.2	68.0	81.0	58.6
	Males	80.6	100	67.6	79.1	100	65.5	74.1	100	58.8
Nyamira	Females	82.0	79.9	84.3	77.5	83.2	72.5	81.6	90.9	74.1
	Males	78.7	100	64.9	78.3	100	64.3	88.0	100	78.6
Nyandarua	Females	82.8	81.0	84.9	80.0	89.4	72.4	44.4	40.0	50.0
	Males	77.6	100	63.4	78.1	100	64.0	89.7	100	81.2
Nyeri	Females	82.6	79.4	86.2	75.0	81.5	69.4	88.0	95.7	81.5
	Males	79.2	100	65.6	78.6	100	64.7	85.7	100	75.0
Tharaka-Nithi	Females	82.0	79.6	84.8	78.6	87.1	71.7	89.5	100	81.0
	Males	80.6	79.1	82.2	75.1	87.9	65.6	90.0	100	81.8
Uasin Gishu	Females	82.3	81.5	83.5	76.4	83.3	70.5	54.5	56.2	52.9
	Males	79.7	100	66.3	78.4	100	64.5	84.2	100	72.7
Vihiga	Females	83.3	80.9	86.0	77.8	86.3	70.9	71.7	70.4	73.1
	Males	79.1	100	65.5	78.6	100	64.7	83.9	100	72.2
	Females	81.6	79.3	84.7	79.2	87.6	72.3	80.0	73.7	87.5

**Table 8.** Elastic Net Performance with SMOTE Across Counties and Sex.

County	Sex	Train F1	Train Rec	Train Prec	Test F1	Test Rec	Test Prec	LOOC F1	LOOC Rec	LOOC Prec
Baringo	Males	68.1	64.4	73.1	69.3	68.9	69.7	70.6	60.0	85.7
	Females	69.3	65.8	73.3	68.9	66.8	71.0	72.7	66.7	80.0
Embu	Males	69.5	68.0	71.4	65.9	64.0	67.9	81.2	81.2	81.2
	Females	72.6	69.1	76.8	66.5	62.2	71.3	77.8	70.0	87.5
Homa Bay	Males	66.9	66.6	67.3	72.2	73.0	71.4	81.5	73.3	91.7
	Females	68.8	64.4	74.1	71.0	68.1	74.1	53.8	38.9	87.5
Kilifi	Males	68.5	63.9	74.4	69.3	67.4	71.3	61.5	66.7	57.1
	Females	70.1	64.5	76.8	70.1	68.1	72.3	73.7	87.5	63.6

County	Sex	Train F1	Train Rec	Train Prec	Test F1	Test Rec	Test Prec	LOOC F1	LOOC Rec	LOOC Prec
Laikipia	Males	69.6	65.7	74.5	72.4	73.6	71.3	76.2	80.0	72.7
	Females	66.6	61.2	73.4	68.1	66.5	69.9	75.0	75.0	75.0
Meru	Males	72.1	70.3	75.2	68.2	66.7	69.8	66.7	66.7	66.7
	Females	73.3	70.7	76.1	70.2	67.2	73.5	40.0	26.7	80.0
Migori	Males	69.8	64.1	77.1	72.7	74.7	70.8	66.7	54.5	85.7
	Females	70.9	64.4	79.5	69.7	65.4	74.5	43.5	55.6	35.7
Mombasa	Males	72.2	67.9	77.9	59.8	54.4	66.2	58.3	58.3	58.3
	Females	68.5	65.5	72.0	71.0	68.1	74.1	51.2	52.4	50.0
Nairobi	Males	62.0	58.6	66.4	74.2	75.8	72.6	76.2	80.0	72.7
	Females	66.4	62.9	70.4	63.5	57.6	70.7	71.4	68.2	75.0
Nyamira	Males	69.3	65.7	73.5	72.2	72.2	72.2	63.2	54.5	75.0
	Females	69.2	65.2	73.8	68.0	64.4	72.0	33.3	20.0	100
Nyandarua	Males	69.2	65.1	74.0	75.6	76.4	74.7	57.1	46.2	75.0
	Females	72.3	67.8	77.9	63.9	59.2	69.4	78.3	78.3	78.3
Nyeri	Males	69.0	64.6	74.2	65.2	64.4	65.9	91.7	91.7	91.7
	Females	69.7	66.0	74.8	65.9	61.3	71.2	83.3	88.2	78.9
Tharaka-Nithi	Males	68.8	65.8	72.2	73.0	75.8	70.4	70.6	66.7	75.0
	Females	68.7	64.0	74.4	67.4	64.0	71.3	36.4	25.0	66.7
Uasin Gishu	Males	64.2	60.4	69.4	71.7	72.5	71.0	66.7	50.0	100
	Females	68.6	64.9	72.7	72.4	71.0	73.9	52.4	40.7	73.3
Vihiga	Males	69.8	65.9	74.6	70.7	72.2	69.1	66.7	53.8	87.5
	Females	67.0	62.2	73.2	68.0	64.3	72.1	64.5	52.6	83.3

## References

- [1] K. T. Mills, A. Stefanescu, and J. He, 'The global epidemiology of hypertension', *Nat. Rev. Nephrol.*, vol. 16, no. 4, pp. 223–237, 2020, <https://doi.org/10.1038/s41581-019-0244-2>
- [2] WHO, 'Hypertension Kenya 2023 country profile', Technical document, 2023. Accessed: Nov. 28, 2024. Available: <https://www.who.int/publications/m/item/hypertension-ken-2023-country-profile>
- [3] Of M. (US) C. on P. H. P. to R. and C. Hypertension, 'Interventions Directed at Individuals with Hypertension', in *A Population-Based Policy and Systems Change Approach to Prevent and Control Hypertension*, National Academies Press (US), 2010. Accessed: Feb. 21, 2026. Available: <https://www.ncbi.nlm.nih.gov/books/NBK220091/>
- [4] S. F. Mohamed et al., 'Prevalence, awareness, treatment and control of hypertension and their determinants: results from a national survey in Kenya', *BMC Public Health*, vol. 18, no. 3, p. 1219, 2018, <https://doi.org/10.1186/s12889-018-6052-y>
- [5] R. Kurniawan et al., 'Hypertension prediction using machine learning algorithm among Indonesian adults', *IAES Int. J. Artif. Intell. IJ-AI*, vol. 12, no. 2, pp. 776–784, 2023, <https://doi.org/10.11591/ijai.v12.i2.pp776-784>
- [6] V. Visco et al., 'Artificial Intelligence in Hypertension Management: An Ace up Your Sleeve', *J. Cardiovasc. Dev. Dis.*, vol. 10, no. 2, p. 74, Feb. 2023, <https://doi.org/10.3390/jcdd10020074>
- [7] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, 'Machine learning in medicine: a practical introduction', *BMC Med. Res. Methodol.*, vol. 19, no. 1, p. 64, Dec. 2019, <https://doi.org/10.1186/s12874-019-0681-4>
- [8] M. M. Alsaleh et al., 'Prediction of disease comorbidity using explainable artificial intelligence and machine learning techniques: A systematic review', *Int. J. Med. Inf.*, vol. 175, p. 105088, 2023, <https://doi.org/10.1016/j.ijmedinf.2023.105088>

- [9] Md. M. Islam et al., 'Predicting the risk of hypertension using machine learning algorithms: A cross sectional study in Ethiopia', *PLOS ONE*, vol. 18, no. 8, p. e0289613, 2023, <https://doi.org/10.1371/journal.pone.0289613>
- [10] S. M. S. Islam et al., 'Machine Learning Approaches for Predicting Hypertension and Its Associated Factors Using Population-Level Data From Three South Asian Countries', *Front. Cardiovasc. Med.*, vol. 9, 2022, <https://doi.org/10.3389/fcvm.2022.839379>
- [11] DHS, 'The DHS Program - Kenya: Standard DHS'. Accessed: Apr. 05, 2026. Available: <https://dhsprogram.com/methodology/survey/survey-display-566.cfm>
- [12] M. Kuhn and K. Johnson, *Feature Engineering and Selection: A Practical Approach for Predictive Models*. New York: Chapman and Hall/CRC, 2019. <https://doi.org/10.1201/9781315108230>
- [13] S. van Buuren and K. Groothuis-Oudshoorn, 'mice: Multivariate Imputation by Chained Equations in R', *J. Stat. Softw.*, vol. 45, pp. 1–67, 2011, <https://doi.org/10.18637/jss.v045.i03>
- [14] T. Chen and C. Guestrin, 'XGBoost: A Scalable Tree Boosting System', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, in KDD '16*. New York, NY, USA: Association for Computing Machinery, 2016, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>
- [15] C. Cortes and V. Vapnik, 'Support-vector networks', *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, <https://doi.org/10.1007/BF00994018>
- [16] 'The random forest algorithm for statistical learning'. Accessed: May 05, 2026. Available: <https://journals.sagepub.com/doi/epub/10.1177/1536867X20909688>
- [17] '(PDF) Random Forest Algorithm Overview', ResearchGate, May 2026, <https://doi.org/10.58496/BJML/2024/007>
- [18] C. Goutte and E. Gaussier, 'A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation', in *Advances in Information Retrieval*, D. E. Losada and J. M. Fernández-Luna, Eds, Berlin, Heidelberg: Springer, 2005, pp. 345–359. [https://doi.org/10.1007/978-3-540-31865-1\\_25](https://doi.org/10.1007/978-3-540-31865-1_25)
- [19] T. Saito and M. Rehmsmeier, 'The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets', *PLOS ONE*, vol. 10, no. 3, p. e0118432, 2015, <https://doi.org/10.1371/journal.pone.0118432>
- [20] P. Pudil, J. Novovičová, and J. Kittler, 'Floating search methods in feature selection', *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994, [https://doi.org/10.1016/0167-8655\(94\)90127-9](https://doi.org/10.1016/0167-8655(94)90127-9)
- [21] W. Shahin, G. A. Kennedy, and I. Stupans, 'The association between social support and medication adherence in patients with hypertension: A systematic review', *Pharm. Pract.*, vol. 19, no. 2, p. 2300, 2021, <https://doi.org/10.18549/PharmPract.2021.2.2300>
- [22] E. Sarkodie, D. K. Afriyie, A. Hutton-Nyameaye, and S. K. Amponsah, 'Adherence to drug therapy among hypertensive patients attending two district hospitals in Ghana', *Afr. Health Sci.*, vol. 20, no. 3, pp. 1355–1367, Sep. 2020, <https://doi.org/10.4314/ahs.v20i3.42>
- [23] A. Ungar and G. Rivasi, 'Increasing awareness on frailty in the management of hypertensive older adults', *J. Hypertens.*, vol. 38, no. 11, p. 2148, 2020, <https://doi.org/10.1097/HJH.0000000000002538>
- [24] A. Fiuza-Luces et al., 'Exercise benefits in cardiovascular disease: beyond attenuation of traditional risk factors', *Nat. Rev. Cardiol.*, vol. 15, no. 12, pp. 731–743, 2018, <https://doi.org/10.1038/s41569-018-0065-1>
- [25] H. Ahrensberg, C. Bjørk Petersen, J. N. W. Jacobsen, M. Toftager, and A. Ernest Bauman, 'The Descriptive Epidemiology of Sedentary Behaviour', in *Sedentary Behaviour Epidemiology*, M. F. Leitzmann, C. Jochem, and D. Schmid, Eds, Cham: Springer International Publishing, 2023, pp. 45–80. [https://doi.org/10.1007/978-3-031-41881-5\\_2](https://doi.org/10.1007/978-3-031-41881-5_2)
- [26] A. Wondmieneh, G. Gedefaw, A. Getie, and A. Demis, 'Self-Care Practice and Associated Factors among Hypertensive Patients in Ethiopia: A Systematic Review and Meta-Analysis', *Int. J. Hypertens.*, vol. 2021, no. 1, p. 5582547, 2021, <https://doi.org/10.1155/2021/5582547>
- [27] S. Kimani, W. Mirie, M. Chege, O. T. Okube, and S. Muniu, 'Association of lifestyle modification and pharmacological adherence on blood pressure control among patients with hypertension at Kenyatta National Hospital, Kenya: a cross-sectional study', *BMJ Open*, vol. 9, no. 1, p. e023995, 2019, <https://doi.org/10.1136/bmjopen-2018-023995>
- [28] D. R. Hanna, J. A. Campbell, R. J. Walker, A. Z. Dawson, and L. E. Egede, 'Association between Health and Wealth among Kenyan Adults with Hypertension', *Glob. J. Health Sci.*, vol. 13, no. 4, pp. 86–94, 2021, <https://doi.org/10.5539/gjhs.v13n4p86>
- [29] R. Oyando, E. Barasa, and J. E. Ataguba, 'Socioeconomic Inequity in the Screening and Treatment of Hypertension in Kenya: Evidence From a National Survey', *Front. Health Serv.*, vol. 2, 2022, <https://doi.org/10.3389/frhs.2022.786098>
- [30] R. Oyando et al., 'Patient costs of hypertension care in public health care facilities in Kenya', *Int. J. Health Plann. Manage.*, vol. 34, no. 2, pp. e1166–e1178, 2019, <https://doi.org/10.1002/hpm.2752>