

1 **Artificial Intelligence Methods to Detect Heart Failure with Preserved Ejection Fraction (AIM-**  
2 **HFpEF) within Electronic Health Records: An equitable disease prediction model**

3

4 Jack Wu<sup>1</sup>, Dhruva Biswas<sup>1,7</sup>, Samuel Brown<sup>1,2</sup>, Matthew Ryan<sup>1,4</sup>, Brett Bernstein<sup>1,2</sup>, Brian Tam To<sup>2</sup>,  
5 Tom Searle<sup>3</sup>, Maleeha Rizvi<sup>1,4</sup>, Natalie Fairhurst<sup>2</sup>, George Kaye<sup>2</sup>, Ranu Baral<sup>2</sup>, Dhanushan  
6 Vijayakumar<sup>6</sup>, Daksh Mehta<sup>6</sup>, Narbeh Melikian<sup>1,2</sup>, Daniel Sado<sup>1,2</sup>, Gerald Carr-White<sup>4</sup>, Phil  
7 Chowienczyk<sup>4</sup>, James Teo<sup>2,3</sup>, Richard JB Dobson<sup>3</sup>, Daniel I Bromage<sup>1,2</sup>, Thomas F Lüscher<sup>5</sup>, Ali  
8 Vazir<sup>4</sup>, Theresa A McDonagh<sup>1,2</sup>, Jessica Webb<sup>4</sup>, Ajay M Shah<sup>1,2</sup>, Kevin O’Gallagher<sup>1,2</sup>

9

10 Author Affiliations

11 <sup>1</sup>School of Cardiovascular and Metabolic Medicine & Sciences, British Heart Foundation Centre of  
12 Research Excellence, King’s College London, London, UK.

13 <sup>2</sup>King’s College Hospital NHS Foundation Trust, London, UK.

14 <sup>3</sup>Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and  
15 Neuroscience, King’s College London, London, UK.

16 <sup>4</sup>Guy’s and St Thomas’ Hospital, Guy’s and St Thomas’ NHS Foundation Trust, London, UK

17 <sup>5</sup>Royal Brompton and Harefield Hospitals, London, UK

18 <sup>6</sup>King’s College London GKT School of Medical Education, London, UK

19 <sup>7</sup>Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine,  
20 New Haven, CT, USA

21

22 Corresponding author:

23 Dr Kevin O’Gallagher, PhD

- 1 The James Black Centre, School of Cardiovascular and Metabolic Medicine & Sciences, British Heart
- 2 Foundation Centre of Research Excellence, King's College London, 125 Coldharbour Lane, London,
- 3 UK, SE5 9NU
- 4 Email: [kevin.o'gallagher@kcl.ac.uk](mailto:kevin.o'gallagher@kcl.ac.uk)
- 5

1 **Abstract**

2 Background and aims: Heart Failure with Preserved Ejection Fraction (HFpEF) accounts for  
3 approximately half of all heart failure cases, with high levels of morbidity and mortality. However,  
4 most cases of HFpEF are undiagnosed as conventional risk scores underestimate risk in non-White  
5 populations. Our aim was to develop and validate a diagnostic prediction model to detect  
6 undiagnosed HFpEF, AIM-HFpEF.

7 Methods: We applied natural language processing (NLP) and machine learning methods to routinely  
8 collected electronic health record (EHR) data from a tertiary centre hospital trust in London, UK, to  
9 derive the AIM-HFpEF model. We then externally validated the model and performed benchmarking  
10 against existing HFpEF prediction models (H2FPEF and HFpEF-ABA) for diagnostic power in  
11 patients of non-white ethnicity and patients from areas of increased socioeconomic deprivation.

12 Results: An XGBoost model combining demographic, clinical and echocardiogram data showed  
13 strong diagnostic performance in the derivation dataset (n=3170, AUC=0.88, [95% CI, 0.86-0.91])  
14 and validation cohort (n=5383, AUC: 0.88 [95% CI, 0.87-0.89]). Diagnostic performance was  
15 maintained in patients of non-White ethnicity (AUC=0.88 [95% CI, 0.84-0.93]) and patients from  
16 areas of high socioeconomic deprivation (AUC=0.89 [95% CI, 0.84-0.94]). and AIM-HFpEF  
17 performed favourably in comparison to H2FPEF and HFpEF-ABA models. AIM-HFpEF model  
18 probabilities were associated with an increased risk of death, hospitalisation and stroke in the external  
19 validation cohort (P<0.001, P=0.01, P<0.001 respectively for highest versus middle tertile).

20 Conclusion: AIM-HFpEF represents a validated equitable diagnostic model for HFpEF, which can be  
21 embedded within an EHR to allow for fully automated HFpEF detection.

22

23 Keywords: Heart Failure with Preserved Ejection Fraction; Prediction Model; Electronic Health  
24 Records

25

## 1 **Introduction**

2 Heart Failure with Preserved Ejection Fraction (HFpEF) accounts for approximately half of all heart  
3 failure (HF) cases and is associated with significant healthcare costs, morbidity, and mortality. Early  
4 diagnosis is important and has prognostic benefit <sup>1</sup>. However, despite the high prevalence of HFpEF,  
5 most cases remain undiagnosed <sup>2</sup> and therefore a high proportion of patients do not benefit from  
6 specialist cardiology care, evidence-based therapies, and eventually improved outcomes. <sup>3,4</sup>

7 Strategies have been developed to improve HFpEF detection, including the H2FPEF scoring system <sup>5</sup>.  
8 The H2FPEF score is validated for the diagnosis of HFpEF, but requires an *a priori* suspicion of  
9 HFpEF. More recently, the HFpEF-ABA score, based on clinical features alone with no cardiac  
10 imaging features, has been developed as a screening tool to identify possible HFpEF cases and guide  
11 the need for specialist cardiac imaging and clinical evaluation <sup>6</sup>.

12 HFpEF is disproportionately under-diagnosed particularly in patients of non-white race and ethnicity <sup>7</sup>.  
13 Indeed, Black and Asian patients with HFpEF have different patterns of comorbidity to White patients,  
14 including features used in current HFpEF diagnostic systems such as atrial fibrillation and body mass  
15 index.

16 The widespread deployment of electronic health record (EHR) platforms provides the potential to  
17 enable access to a wide range of routinely collected clinical data in a fraction of the time taken to  
18 perform manual case record completion. EHR-based diagnostic approaches lend themselves to  
19 automation, removing the need for clinician-initiated suspicion of disease and therefore potentially  
20 decreasing the risk of bias. Moreover, advances in artificial intelligence (AI) methods allow for the  
21 capture of both structured and unstructured data, including AI-based detection of clinical concepts  
22 from free text via natural language processing (NLP) <sup>8-11</sup> with potentially less data missingness in  
23 populations less engaged with health services. We have previously used these methods to detect  
24 undiagnosed HFpEF from the EHR, finding that less than 10% of all cases of HFpEF have a clinician-  
25 assigned diagnosis, while the remaining 90% are undiagnosed. <sup>2</sup>

1 The aim of this study was to develop and externally validate a prediction model for detecting HFpEF  
2 by applying artificial intelligence methods to routinely collected data from the EHR of two  
3 independent, ethnically diverse cohorts of HFpEF patients, including a representative distribution of  
4 cases of Confirmed and Undiagnosed HFpEF. A further aim was to assess the performance of the  
5 prediction model across racial and ethnicity groups and in patients from socially deprived areas.

1 **Methods**

2 *Ethical Considerations*

3 This project operated under London South-East Research Ethics Committee approval (18/LO/2048)

4 granted to the King's Electronic Records Research Interface (KERRI) and London Dulwich Research

5 Ethics Committee approval (19/LO/1957), which did not require written informed patient consent.

6 The study complies with the Declaration of Helsinki. Patients were consulted on the study via a

7 dedicated patient and public involvement (PPI) meeting held during the study design phase. A formal

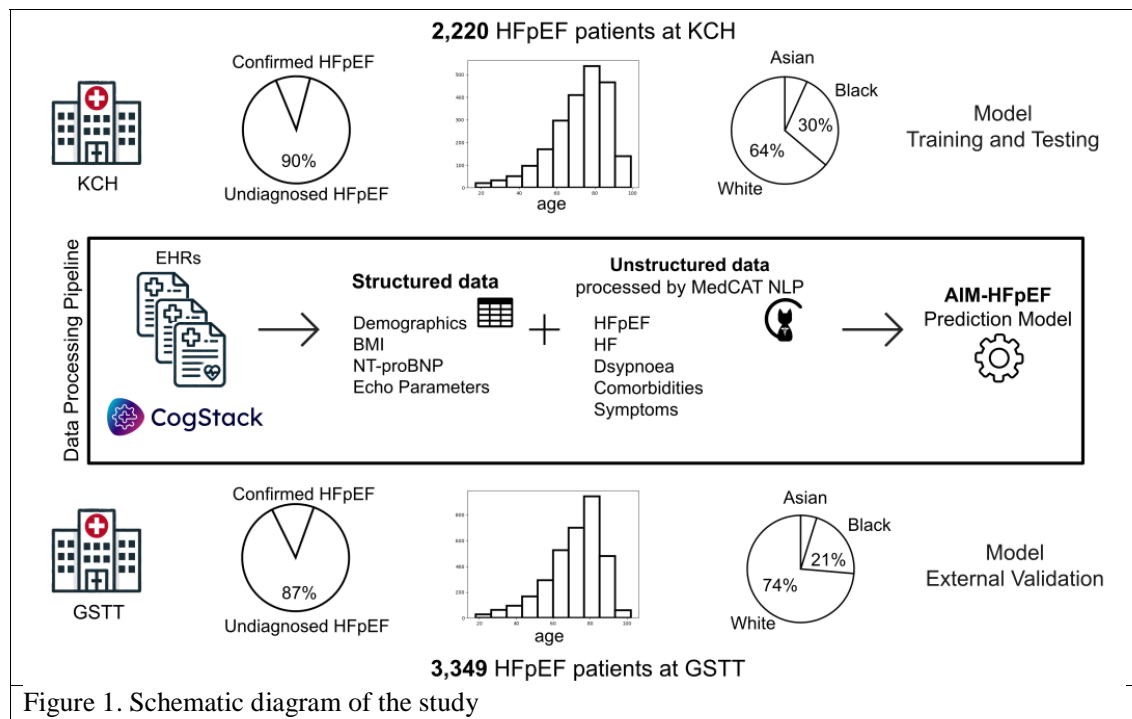
8 protocol was not published and the study was not registered.

9 *Participating Centres*

10 King's College Hospital NHS Foundation Trust (KCH) and Guy's and St Thomas' NHS Foundation

11 Trust (GSTT) are two large, multi-site tertiary hospitals in London, UK, providing specialist

12 cardiology services and a dedicated heart failure service open to referral by any physician.



13

14

1 *Study Design and inclusion criteria*

2 At each participating centre, we established a registry comprising a retrospective anonymised  
3 database of adult patients with a clinical diagnosis of HF documented within the EHR between 2010-  
4 2022. **Figure 1** shows the schematic diagram of the study. Patients were included if they had two or  
5 more mentions of a “heart failure” (HF) diagnosis in the clinical text as determined by a well-  
6 validated NLP pipeline<sup>12-15</sup>. Specifically, a random sample of 100 HFpEF patients identified by the  
7 NLP pipeline were manually validated for this study and 100% of them were true positive. Patients  
8 were included regardless of nature of clinical episode (inpatient or outpatient). Both structured and  
9 unstructured portions of the echocardiogram report were used to extract LVEF data of patients and  
10 other relevant echocardiographic parameters. Patients with a clinical diagnosis of HF and LVEF  
11  $\geq 50\%$  were categorised into one of the following 2 groups:

- 12 1. Confirmed HFpEF: Clinician-assigned diagnosis of HFpEF
- 13 2. Undiagnosed HFpEF (meeting the ESC diagnostic criteria): Patients with HF, LVEF $\geq 50\%$ ,  
14 and imaging/biochemical evidence of diastolic dysfunction meeting the ESC diagnostic  
15 criteria<sup>16</sup> but who have not received a HFpEF diagnosis.

16 Patients were excluded if at any point they had an echocardiogram with LVEF  $< 50\%$ . Patients with a  
17 clinical diagnosis of HF and LVEF $\geq 50\%$  on echocardiography, but not meeting ESC diagnostic  
18 criteria and who have not received a HFpEF diagnosis were excluded, as were patients with an  
19 alternative diagnosis (severe valvular heart disease, hypertrophic cardiomyopathy, restrictive  
20 cardiomyopathy, constrictive pericarditis, and cardiac amyloidosis). **Supplementary Table 1**  
21 contains the SNOMED terms used in the classification of included and excluded patients. Patients  
22 needed to have the echocardiogram report within one year of clinical diagnosis of HF, otherwise they  
23 were excluded.

24 At each site we also established a registry of non-cardiac dyspnoea patients to act as control groups.  
25 Patients in the control group needed (1) two or more mentions of dyspnoea in the clinical text as  
26 determined by NLP; (2) an echocardiogram report (LVEF $\geq 50\%$ ) within one year of the dyspnoea

1 mentions; and (3) no mentions of HF diagnosis in the clinical text. Patients with severe valvular heart  
2 disease, hypertrophic cardiomyopathy, restrictive cardiomyopathy, constrictive pericarditis, or cardiac  
3 amyloidosis were also excluded from the control group.

#### 4 *Prediction Model Development*

5 The dataset collected from KCH was used as the derivation cohort, with 80% of the data used for  
6 training and 20% for testing, while the GSTT dataset was used for external validation. All model  
7 development and evaluation were performed using Python version 3.10.12.

8 We derived prediction models using four different methods: Logistic Regression (LR), Support Vector  
9 Machine (SVM), Random Forest (RF) and XGBoost (XGB). The machine learning models were built  
10 and evaluated using the ‘scikit-learn’ (for LR, SVM and RF) and ‘xgboost’ (for XGB) python  
11 packages. Hyperparameter tuning for each model was performed through grid search and 10-fold  
12 cross-validation using the training dataset only. The ‘StandardScaler’ from ‘scikit-learn.preprocessing’  
13 was applied to the data before model training for consistent scaling across echocardiograph parameter  
14 features for LR, SVM and RF models. Mean imputation was applied to the missing values for the  
15 three (LR, SVM and RF) machine learning models, missForest imputation was also tested and the  
16 correlation of the predicted probabilities from both methods were greater than 98%. XGB can handle  
17 scaling and missing values internally. Model evaluation was based on metrics including accuracy,  
18 precision, recall, F1 score, and AUC-ROC, calculated using functions from ‘scikit-learn.metrics’. The  
19 comparison of AUCs was performed using a python implementation of the DeLong test<sup>17</sup>. We  
20 selected the model which had the best overall performance for the Results section while the  
21 performances for non-selected models are reported in **Supplementary Table 2**.

#### 22 *Feature Engineering*

23 We aimed to produce a model based on the parameters included in the ESC criteria for diagnosing  
24 HFpEF. Features representing key demographic, echocardiographic, comorbid, and symptomatic  
25 factors commonly associated with HFpEF are used to construct the model and they are extracted from  
26 both structured and unstructured data. Features from structured data include age, sex, BMI,



1 NTproBNP and echocardiographic parameters (E/e', LA volume, LA volume indexed, IVSd, LVPWD,  
2 LVEDD, LV V1 max, LV V1 max PG, LV mass, LV mass indexed, and PASP), while comorbidities  
3 (diabetes mellitus type 1, diabetes mellitus type 2, ischemic heart disease, myocardial infarction,  
4 cerebrovascular accident, hypertensive disorder, transient ischemic attack, atrial fibrillation,  
5 pulmonary hypertension, kidney disease, and angina) and symptoms (dyspnoea at rest, chest pain,  
6 dizziness, and syncope) were extracted from unstructured clinical notes through natural language  
7 processing (NLP) using MedCAT<sup>13</sup> within the CogStack platform<sup>18</sup>. Echocardiographic parameters  
8 included in the ESC criteria were prioritized in the model, while other parameters with more than 30%  
9 missingness were excluded. Comorbidities and symptoms are represented as binary features, with a  
10 positive value indicating the presence of the comorbidity or symptom before the first mention of HF  
11 (for the HFpEF group) or dyspnoea (for the control group) in the EHR while a negative value  
12 indicates the absence of the condition. Dyspnoea was excluded as a feature in the model since it was  
13 part of the inclusion criteria for control patients, resulting in 100% of the control group presenting  
14 with dyspnoea. The full model includes 30 features in total and can be executed automatically on the  
15 CogStack platform using routinely collected EHR data.

16 A generalised linear model (GLM) from the 'statsmodels' package was used to simplify the full  
17 prediction model through feature selection using the KCH training dataset, retaining the top 10  
18 features ranked by significance based on their p-values. These selected features were then used to  
19 construct a simplified prediction model. In the simplified model, diabetes mellitus type 1 and type 2  
20 were combined into a single feature. The simplified model can be useful for manually inputting  
21 feature values in the form of an online application.

## 22 *Model explainability*

23 A Shapley Additive Explanations (SHAP) graph was plotted for the full model using the 'shap' python  
24 library (version 0.45.1)<sup>19</sup> to show the importance and values of each feature contributing to the  
25 prediction outcome.

## 26 *Comparison with H2FPEF and HFpEF-ABA*

1 For comparison with the H2FPEF score, we followed the more precise version of the H2FPEF  
2 probability using the formula provided in Reddy et al. (2018)<sup>5</sup>, specifically the online calculator from  
3 the supplemental material. The formula requires five key variables: BMI, atrial fibrillation (AF),  
4 pulmonary artery systolic pressure (PASP), age, and filling pressure (E/e'). We also compared the  
5 results with the point-based version of the H2FPEF score (which requires six variables with  
6 antihypertensive drugs added) in **Supplementary Table 2**. Patients with any missing values for these  
7 key variables were excluded from the comparison to ensure accurate probability estimation.

8 For comparison with the HFpEF-ABA score, we used the formula provided in Reddy et al. (2024)<sup>6</sup>,  
9 specifically from “Extended Data Table 4 Regression equations for clinical variable models”. This  
10 model requires three variables: age, BMI, and atrial fibrillation (AF). Similar to the comparison with  
11 H2FPEF score, patients with missing values for any of these three variables were excluded from the  
12 comparison.

### 13 *Subgroup analysis*

14 In the subgroup analysis, we examined two patient subgroups: (1) non-White individuals based on  
15 self-ascribed ethnicity<sup>2</sup> and (2) those with low socioeconomic status, as assessed by the English  
16 Indices of Multiple Deprivation 2019 (IMD). The IMD was determined using postcodes of the  
17 patients. Patients were classified as having low IMD if their postcodes fell within the most deprived  
18 quintiles according to the national index. The model’s performance was evaluated separately in the  
19 two subgroups to assess potential variations in predictive accuracy for HFpEF based on ethnicity and  
20 socioeconomic status.

### 21 *Prediction Model Output and Calibration*

22 The machine learning models output the probability of HFpEF for each patient. For consistency, the  
23 cut-off value for a positive HFpEF prediction is set at 0.5 when computing the accuracy, precision and  
24 recall values in **Supplementary Table 2**. The same threshold is used when comparing those metrics  
25 with the H2FPEF and HFpEF-ABA scores. In practice, the threshold can be adjusted according to  
26 individual treatment goals and preferences<sup>6</sup>. Calibration of the prediction probabilities were assessed

1 graphically using calibration curves produced by the 'calibration\_curve' function in 'scikit-learn'  
2 package.

3

4 *Outcome data*

5 Mortality data were obtained from death notification letters in the EHR system and the master  
6 demographic patient indices of the hospitals (synchronised with NHS Spine for demographics).

7 Hospitalisations were estimated by the number of discharge notifications in the EHR in the study  
8 timeframe i.e. 2010-2022. Diagnoses of myocardial infarction and stroke were recorded as outcome

9 data if they occurred after the first mention of HF (for the HFpEF group) or dyspnoea (for the control

10 group) in the EHR.

1 **Results**

2 We identified 2,220 HFpEF patients (231 [10%] confirmed HFpEF and 1,989 [90%] undiagnosed  
3 HFpEF) at KCH and 3,349 HFpEF patients (430 [13%] confirmed HFpEF and 2,919 [87%]  
4 undiagnosed HFpEF) at GSTT. There were 950 and 2,034 non-HF patients with dyspnoea at KCH and  
5 GSTT, respectively, for the control groups. **Table 1** shows the baseline characteristics of the patients.

6 We randomly divided the KCH patients into 80% for training (a total of 2,536 patients: 177 confirmed  
7 HFpEF, 1,599 undiagnosed HFpEF, and 760 control) and 20% for testing (a total of 634 patients: 54  
8 confirmed HFpEF, 390 undiagnosed HFpEF, and 190 control), stratified by HFpEF and non-HF  
9 groups. The GSTT patients were used for external validation.

10 Among the four machine learning models: logistic regression (LR), random forest (RF), support  
11 vector machine (SVM), and XGBoost (XGB), the XGB model achieved the highest performance,  
12 with an AUC-ROC of 0.8910 (95% CI, 0.8665-0.9154) in the KCH testing cohort, outperforming the  
13 other models (LR:0.8204 [95% CI, 0.7868-0.8540], RF: 0.8472 [95% CI, 0.8161-0.8784],  
14 SVM:0.8386 [95% CI, 0.8063-0.8709]). Therefore, we focused on the XGBoost model for further  
15 analysis and validation. The full report of the performances of all the models is shown in  
16 **Supplementary Table 2.**

17

18

19

20

21

22

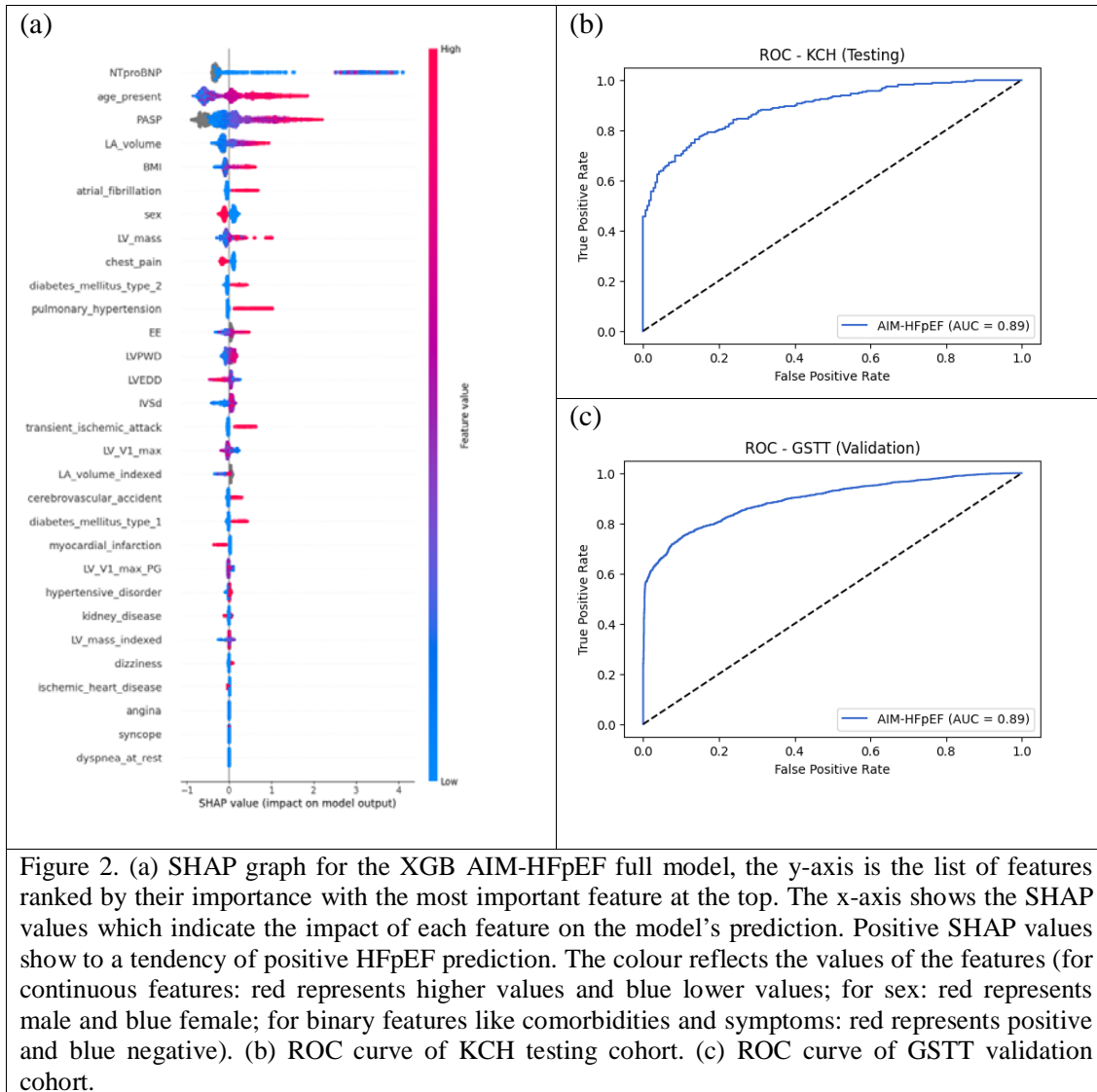


Figure 2. (a) SHAP graph for the XGB AIM-HFpEF full model, the y-axis is the list of features ranked by their importance with the most important feature at the top. The x-axis shows the SHAP values which indicate the impact of each feature on the model's prediction. Positive SHAP values show to a tendency of positive HFpEF prediction. The colour reflects the values of the features (for continuous features: red represents higher values and blue lower values; for sex: red represents male and blue female; for binary features like comorbidities and symptoms: red represents positive and blue negative). (b) ROC curve of KCH testing cohort. (c) ROC curve of GSTT validation cohort.

1

2 We employed SHAP analysis to understand the contribution of individual features to the model's  
3 predictions from XGB as shown in **Figure 2 (a)**. The top contributing features include NTproBNP,  
4 age, PASP, LA volume and BMI. Age, PASP and BMI are also included in the H2FPEF formula but  
5 the XGB model has no prior information of this.

6 **Figure 2(b) and (c)** shows the Receiver Operating Characteristic (ROC) curves for the XGB  
7 prediction model on the KCH testing dataset and GSTT validation dataset, respectively. The AUC for  
8 the GSTT validation dataset was 0.8934 (95% CI, 0.8852-0.9016) for the XGB model.

9

1 *Simplified Model*

2 While the SHAP graph can help us to understand how the XGB model predicts the patients, in order  
3 to identify features with the strongest statistical relationship with the prediction outcome, a  
4 Generalized Linear Model (GLM) analysis was performed to select the top 10 features for a simplified  
5 model. **Table 2** shows the results of the GLM with the top selected features and their coefficients and  
6 p-values associated with HFpEF prediction using the KCH training dataset. All the top 10 features  
7 have a p-value of less than 0.02 indicating that they are statistically significant predictors.

Table 2. GLM selected features.		
Feature	Coefficient	P-Value
Age at presentation	0.5640	<0.0001
PASP	0.5713	<0.0001
Sex	-0.2801	<0.0001
NT-proBNP	1.0217	<0.0001
Atrial fibrillation	0.2816	0.0002
LVEDD	-0.3566	0.0024
Transient ischemic attack	0.2012	0.0033
Diabetes mellitus	0.1804	0.0046
BMI	0.1790	0.0113
LA volume	0.4207	0.0139

8

9 A simplified version of the machine learning models was next built using the 10 features identified by  
10 GLM. The XGB model again performed the best (AUC-ROC of 0.8845 [95% CI, 0.8592-0.9097])  
11 among the four models (LR: 0.8025 [95% CI, 0.7678-0.8373], RF: 0.8379 [95% CI, 0.8051-0.8707],  
12 SVM: 0.8205 [95% CI, 0.7852-0.8559]) for the KCH testing cohort. The AUC of the simplified  
13 model is slightly less than that in the full model (AUC of 0.8809 [95% CI, 0.8721-0.8897] for the  
14 GSTT validation cohort), but it is more usable in practice. The individual performances of the model  
15 on the confirmed HFpEF and undiagnosed HFpEF groups are shown in **Supplementary Figure 1**.  
16 The calibration curves for the two HFpEF groups in the GSTT validation cohort is shown in  
17 **Supplementary Figure 2**. The simplified XGB model is used for subsequent comparison and  
18 subgroup analysis.

19

1 *Comparison with H2FPEF and HFpEF-ABA scores*

2 To evaluate the performance of the simplified XGB model in comparison with established clinical  
3 scoring systems, we compared it to the widely used H2FPEF score and the more recently published  
4 HFpEF-ABA score. In the KCH testing cohort (n=634), 182 (29%) patients had complete data for all  
5 five variables (age, BMI, PASP, E/e', and AF) required to calculate the H2FPEF probability (AUC,  
6 AIM-HFpEF: 0.8751 [95% CI, 0.8225-0.9276], H2FPEF: 0.7873 [95% CI, 0.7075-0.8672],  
7 p=0.0064), while 483 (76%) patients had all three variables (age, BMI, and AF) necessary for the  
8 HFpEF-ABA score calculation (AUC, AIM-HFpEF: 0.8792 [95% CI, 0.8493-0.9091], HFpEF-ABA:  
9 0.7425 [95% CI, 0.6949-0.7901], p<0.0001). In the GSTT validation cohort (n=5,621), 967 (17%)  
10 patients had the full set of variables to compute the H2FPEF probability (AUC, AIM-HFpEF: 0.8746  
11 [95% CI, 0.8510-0.8982]), H2FPEF: 0.7805 [95% CI, 0.7505-0.8105], p<0.0001) and 3,735 (66%)  
12 patients had the variables required for the HFpEF-ABA score (AUC, AIM-HFpEF: 0.8788 [95% CI,  
13 0.8665-0.8912], HFpEF-ABA: 0.7624 [95% CI, 0.7462-0.7787], p<0.0001). **Figure 3 (a)** presents the  
14 ROC curves comparing the performance of the simplified XGB model, H2FPEF score, and HFpEF-  
15 ABA score in the GSTT validation cohort. Comparison for the KCH testing cohort is shown in  
16 **Supplementary Figure 3(a)**.

17

18

19

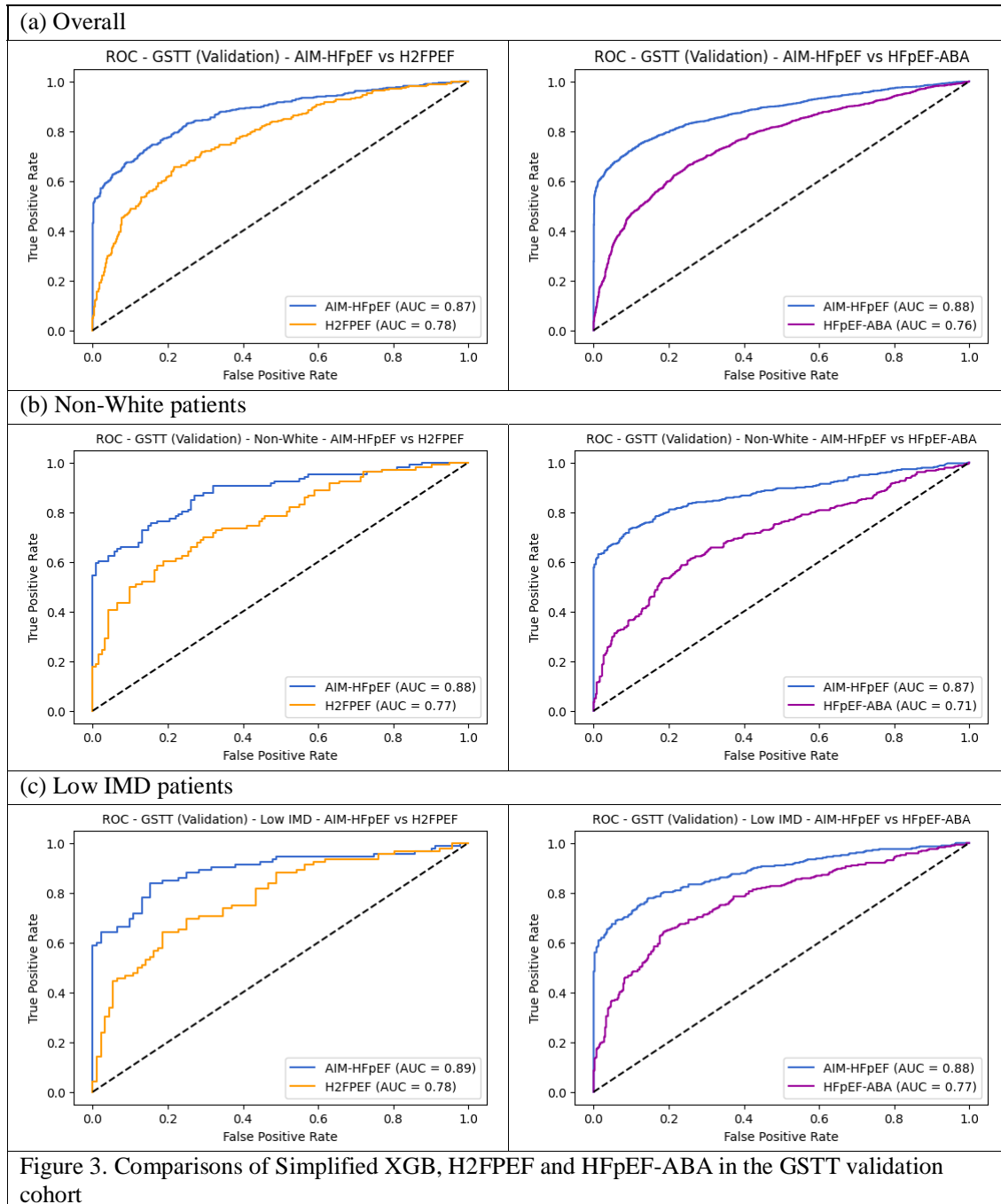
20

21

22

23

24



1

2 *Subgroup analysis*

3 In GSTT validation cohort (n=5,621), there were 1,282 (23%) non-White patients and 1,136 (20%)

4 patients from the lowest quintile of IMD.

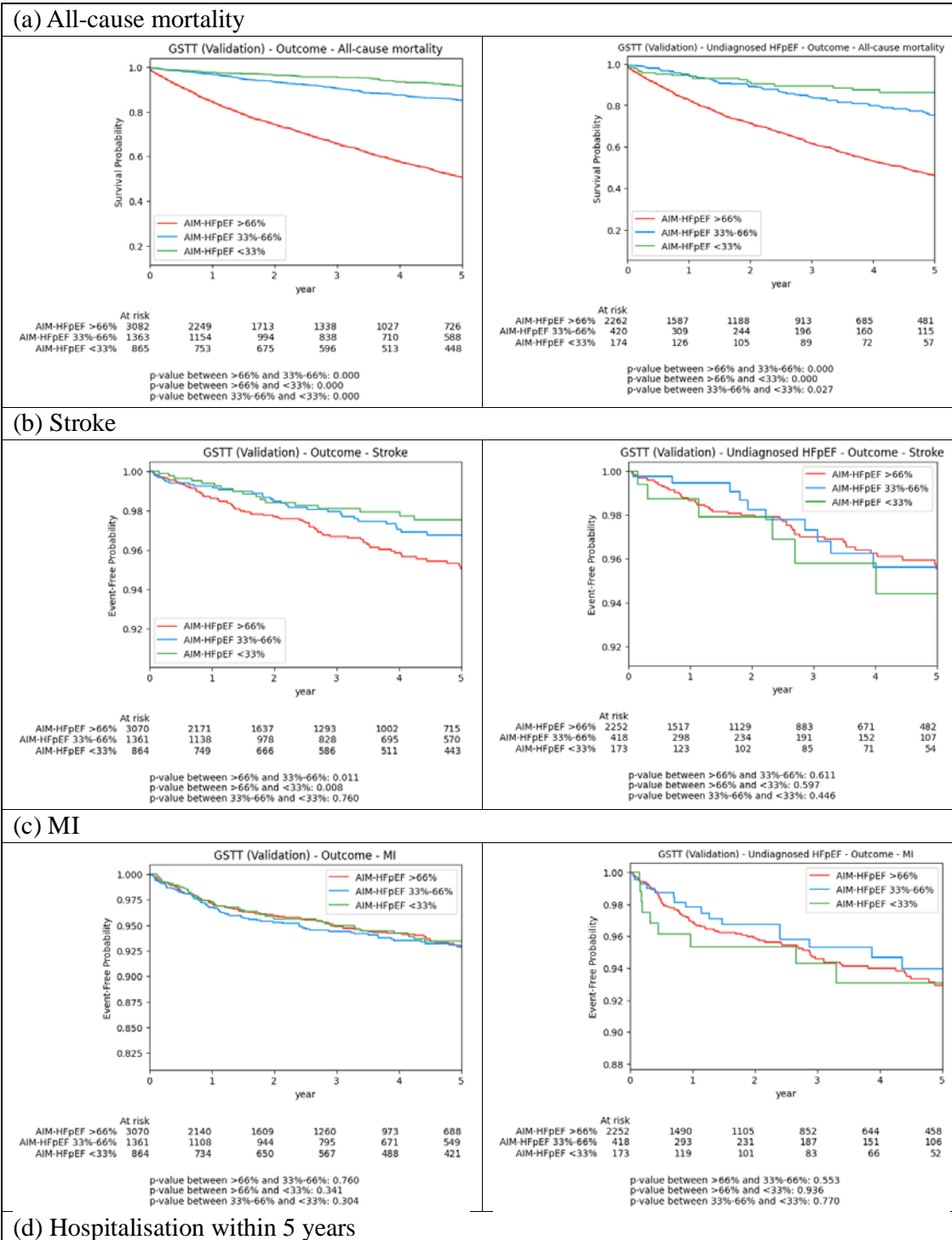


1 Non-White patients: In the GSTT validation cohort, the number of patients included in the  
2 comparison for the H2FPEF and HFpEF-ABA scores were 249 (AUC, AIM-HFpEF: 0.8812 [95% CI,  
3 0.8365-0.9259], H2FPEF: 0.7681 [95% CI, 0.7071-0.8291],  $p < 0.0001$ ) and 923 (AUC, AIM-HFpEF:  
4 0.8744 [95% CI, 0.8477-0.9012], HFpEF-ABA: 0.7101 [95% CI, 0.6734-0.7468],  $p < 0.0001$ ),  
5 respectively. **Figure 3 (b)** shows the ROC curves for the GSTT validation cohort. Comparisons for  
6 the KCH testing cohort is shown in **Supplementary Figure 3(b)**.

7 Low IMD patients: In the GSTT validation cohort, the corresponding numbers were 195 (AUC, AIM-  
8 HFpEF: 0.8914 [95% CI, 0.8423-0.9406], H2FPEF: 0.7793 [95% CI, 0.7127-0.8459],  $p = 0.0003$ ) and  
9 726 (AUC, AIM-HFpEF: 0.8829 [95% CI, 0.8557-0.9100], HFpEF-ABA: 0.7745 [95% CI, 0.7383-  
10 0.8107],  $p < 0.0001$ ), respectively. **Figure 3 (c)** shows the ROC curves for the GSTT validation cohort.  
11 Comparisons for the KCH testing cohort is shown in **Supplementary Figure 3(c)**.

## 12 Outcome analysis

13 In the outcome analysis, we selected patients with a predicted probability  $\geq 90\%$  of having HFpEF  
14 from the models and investigated their all-cause mortality over a 5-year period. In the KCH testing  
15 cohort, the AIM-HFpEF model identified 223 patients and 89 (40%) of them died within 5 years of  
16 HF diagnosis. H2FPEF identified 32 patients and 13 (33%) died within 5 years, while HFpEF-ABA  
17 identified 87 and 29 (33%) of them died. In the GSTT validation cohort, the numbers are AIM-HFpEF:  
18 2272 and 957 (42%); H2FPEF: 105 and 53 (50%); HFpEF-ABA: 606 and 193 (32%). Kaplan-Meier  
19 (KM) curves of the outcome analysis for all-cause mortality, MI and stroke are shown in  
20 **Supplementary Figure 4**. We also investigated whether the AIM-HFpEF model can predict outcome  
21 by dividing the GSTT validation cohort into three groups based on AIM-HFpEF predicted probability  
22 tertiles in **Figure 4**. In the overall cohort, AIM-HFpEF produced probabilities that were associated  
23 with an increased risk of death ( $p < 0.0001$ ), stroke ( $p = 0.01$ ) and hospitalisation ( $p < 0.0001$ ) when  
24 comparing the highest tertile to the middle tertile. In the undiagnosed cohort, AIM-HFpEF  
25 probabilities were associated with an increased risk of death ( $p < 0.0001$ ) and hospitalisation  
26 ( $p < 0.0001$ ).



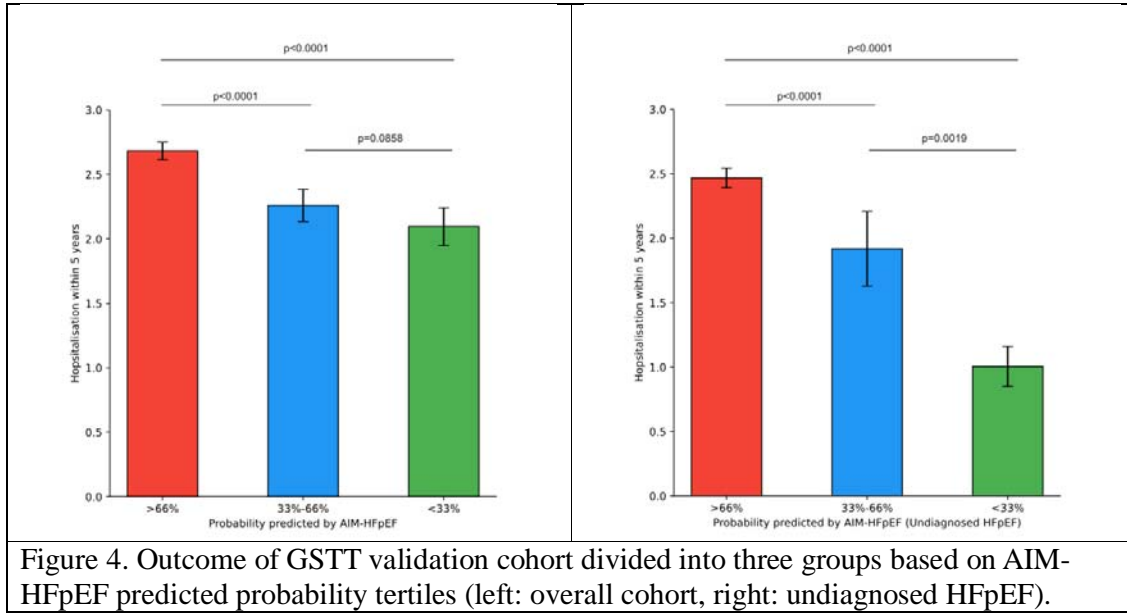


Figure 4. Outcome of GSTT validation cohort divided into three groups based on AIM-HFpEF predicted probability tertiles (left: overall cohort, right: undiagnosed HFpEF).

1

2

3

4

## 1 Discussion

2 Using well-validated NLP and machine learning techniques applied to routinely collected data from  
3 the EHR, we have developed the AIM-HFpEF predictive model to accurately detect patients with  
4 HFpEF. Importantly from a health equity point of view and to address algorithmic bias <sup>20</sup>, it  
5 performs well in patients of non-White ethnicity and in patients from areas of higher socioeconomic  
6 deprivation.

7

8 AIM-HFpEF is designed to be fully automated and integrated into EHR platforms, has been externally  
9 validated and performs favourably to current diagnostic and screening models. It is anticipated that  
10 patients with a high likelihood of HFpEF as ascertained by the AIM-HFpEF model could be identified  
11 to the clinician by way of an electronic pop-up prompt within the EHR, with subsequent referral to a  
12 cardiologist for specialist assessment and initiation of treatment if appropriate.

13

14 In constructing AIM-HFpEF, we have taken the novel step of including data from patients with  
15 undiagnosed HFpEF in the datasets. We see this as a potential key approach in addressing the issue of  
16 underdiagnosis in HFpEF. Conceptually, it can be considered that the characteristics of the  
17 undiagnosed HFpEF patients are the clinically most important predictors, not captured through  
18 analysis solely of diagnosed HFpEF. One possible reason is that these patients not yet diagnosed have  
19 information missing in their structured data and only present in unstructured form. Patients within the  
20 Confirmed HFpEF group have *already been diagnosed* and therefore are in less of a need of a  
21 predictive model, whereas the patients within the Undiagnosed HFpEF group are those who are being  
22 missed by current diagnostic methods.

23

24 A key concern of AI-based disease prediction tools is the risk of potentiating any biases contained  
25 within the training dataset. A key aim of this study was to ensure that AIM-HFpEF retained good

1 performance in diagnosing the significant minority of HFpEF patients of non-White race and ethnicity,  
2 to ensure alignment with future frameworks for assessing algorithmic bias <sup>20</sup>. We note that the  
3 derivation and validation populations of the HFPEF-ABA score were overwhelming of White  
4 ethnicity and therefore its generalisability to non-White populations has not previously been  
5 ascertained. Furthermore, several studies have identified a lack of generalisability in the H2FPEF  
6 score *viz a viz* its performance in non-White populations, likely at least in part due to the heavy  
7 weighting afforded to a diagnosis of AF in the H2FPEF score (only a small proportion of Black  
8 patients with HFpEF have a diagnosis of AF) <sup>7</sup>. Our results show that AIM-HFpEF performs better  
9 than H2FpEF and HFpEF-ABA in non-White patients in the UK, although we do note that both  
10 existing scores have reasonable performance in this patient group and our study can also be  
11 considered as additional external validation of these models.

12

13 The predictors identified in the full model can be related to HFpEF either in terms of direct  
14 pathophysiological mechanisms or by their relation to clinical features associated with the syndrome.  
15 In contrast to HFpEF-ABA, in our GLM model we found several echocardiogram variables to be  
16 significant predictors of HFpEF and therefore our model includes a number of echocardiogram  
17 measures i.e. Pulmonary artery systolic pressure (PASP), left atrial volume and LVEDD. The least  
18 explainable feature in our simplified model is TIA: we consider that it is likely to represent a  
19 composite of age, atrial fibrillation and cardiovascular disease, each of which are known to contribute  
20 to HFpEF pathophysiology. As expected, there is significant overlap between the variables included in  
21 AIM-HFpEF and those in other HFpEF predictive models. A key difference is the inclusion of  
22 NTproBNP in our model. Given that not all patients will have NTproBNP results available, we have  
23 confirmed the acceptable model performance even when natriuretic peptide results are not available  
24 **(Supplementary Document 1)**.

25

26 *Strengths and Limitations*

1 A key strength of an EHR-based approach is that it lends itself to automation i.e. there is no  
2 requirement for an *a priori* suspicion of HFpEF as is the case with diagnostic scores such as H2FPEF  
3 and HFA-PEFF. Conversely, the use of routinely collected, retrospective data does have key  
4 limitations, including non-standardised reporting of clinical features across the two participating  
5 centres, therefore requiring prospective validation.

6 As discussed above, the inclusion of patients with undiagnosed HFpEF in both the derivation and  
7 validation datasets is a strength of our approach, potentially allowing the final model to be much more  
8 generalisable and less prone to bias compared to diagnostic models derived from smaller, more highly  
9 preselected groups of confirmed HFpEF patients.

10 Another limitation of our model is the reliance on an advanced data extraction platform to employ  
11 NLP methods and retrieve clinical data from the EHR. Hospital systems with informatics capabilities  
12 to employ our model are in the minority globally, particularly in low- and lower-middle income  
13 countries, despite the CogStack technology being low cost and light weight, and available open source.  
14 We have sought to mitigate this limitation by producing an alternative prediction model that does not  
15 require advanced EHR data analytic capabilities. This **Simplified Model** could potentially be  
16 accessed via a smartphone app to enable clinicians to define the likelihood of a HFpEF diagnosis.

17 A further limitation is that although both the derivation and validation cohorts come from separate  
18 large multi-hospital NHS trusts, they are both within the same large urban metropolis i.e. London.  
19 Further external validation in different settings is therefore required to ensure generalisability of our  
20 findings across broader geographic areas. Further work will therefore involve assessment of wider  
21 generalisability both in larger UK datasets and in international datasets. Additional future avenues  
22 including prospective validation will be a key step toward assessing the ability of AIM-HFpEF to  
23 affect patient outcomes through improved diagnosis via the model. Finally, incorporation of primary  
24 care data will be important to ensure accurate diagnosis in the unknown proportion of undiagnosed  
25 HFpEF patients without clinical data within secondary care.

26

1

2

### 3 **Conclusion**

4 In this study we describe the use of AI methods to develop an automated, EHR based diagnostic  
5 prediction model for HFpEF. The AIM-HFpEF model has been externally validated and is seen to  
6 perform favourably to existing diagnostic and screening models and is accurate in non-White patients  
7 and in those from areas of high socio-economic deprivation.

8

9

### 10 **Funding**

11 This work was supported by the Adrian Beecroft British Heart Foundation Cardiovascular Catalyst  
12 Award (CC/22/250022 to AMS, KOG, RJD and JT) with further support from the British Heart  
13 Foundation (CH/1999001/11735, RG/20/3/34823 and RE/18/2/34213 to AMS) and King's College  
14 Hospital Charity (D/3003/122022/Shah/1188 to AMS). KOG and DIB are each supported by MRC  
15 Clinician Scientist Fellowships (MR/Y001311/1 to KOG, MR/X001881/1 to DB).

16

### 17 **Conflicts of interest**

18 TAM has received speaker's fees and advisory board fees from Abbott, Edwards, Boehringer  
19 Ingelheim, and AstraZeneca. AMS serves as an advisor to Forcefield Therapeutics and CYTE- Global  
20 Network for Clinical Research. All other authors have nothing to disclose. TFL does not accept any  
21 honoraria from industry, but has received research and educational grants from Abbott, Amgen  
22 AstraZeneca, Boehringer Ingelheim, Daich-Sankyo, Eliy Lilly, Novartis, Novo Nordisk, Sanofi and  
23 Vifor.

24

### 25 **Data availability statement**

- 1 The data and code underlying this article will be shared on reasonable request to the corresponding
- 2 author.



## 1   **References**

- 2   1. Saito Y, Obokata M, Harada T, et al. Prognostic benefit of early diagnosis with exercise stress  
3   testing in heart failure with preserved ejection fraction. *Eur J Prev Cardiol.* 2023;30(9):902–911. doi:  
4   10.1093/eurjpc/zwad127.
- 5   2. Wu J, Biswas D, Ryan M, et al. Artificial intelligence methods for improved detection of  
6   undiagnosed heart failure with preserved ejection fraction. *Eur J Heart Fail.* 2024;26(2):302–310. doi:  
7   10.1002/ejhf.3115.
- 8   3. Anker SD, Butler J, Filippatos G, et al. Empagliflozin in heart failure with a preserved ejection  
9   fraction. *N Engl J Med.* 2021;385(16). doi: 10.1056/nejmoa2107038.
- 10  4. Solomon SD, McMurray JJV, Vaduganathan M, et al. Finerenone in heart failure with mildly  
11  reduced or preserved ejection fraction. *N Engl J Med.* 2024.
- 12  5. Reddy YNV, Carter RE, Obokata M, Redfield MM, Borlaug BA. A simple, evidence-based  
13  approach to help guide diagnosis of heart failure with preserved ejection fraction. *Circulation.*  
14  2018;138(9). doi: 10.1161/CIRCULATIONAHA.118.034646.
- 15  6. Reddy YNV, Carter RE, Sundaram V, et al. An evidence-based screening tool for heart failure with  
16  preserved ejection fraction: The HFpEF-ABA score. *Nat Med.* 2024;30(8):2258–2264.
- 17  7. Brown S, Biswas D, Wu J, et al. Race-and ethnicity-related differences in heart failure with  
18  preserved ejection fraction using natural language processing. *JACC: Advances.* 2024;3(8):101064.
- 19  8. Bean DM, Kraljevic Z, Shek A, Teo J, Dobson RJB. Hospital-wide natural language processing  
20  summarising the health data of 1 million patients. *PLOS Digital Health.* 2023;2(5). doi:  
21  10.1371/journal.pdig.0000218.
- 22  9. Searle T, Ibrahim Z, Teo J, Dobson RJB. Discharge summary hospital course summarisation of in  
23  patient electronic health record text with clinical concept guided deep pre-trained transformer models.  
24  *J Biomed Inform.* 2023;141. doi: 10.1016/j.jbi.2023.104358.
- 25  10. Kraljevic Z, Bean D, Shek A, et al. Foresight—a generative pretrained transformer for modelling  
26  of patient timelines using electronic health records: A retrospective modelling study. *The Lancet*  
27  *Digital Health.* 2024;6(4). doi: 10.1016/S2589-7500(24)00025-6.
- 28  11. Williams SC, Noor K, Sinha S, et al. Concept recognition and characterization of patients  
29  undergoing resection of vestibular schwannoma using natural language processing. *Journal of*  
30  *Neurological Surgery Part B: Skull Base.* 2024.
- 31  12. Bean DM, Kraljevic Z, Searle T, et al. Angiotensin-converting enzyme inhibitors and angiotensin  
32  II receptor blockers are not associated with severe COVID-19 infection in a multi-site UK acute  
33  hospital trust. *European Journal of Heart Failure.* 2020;22(6). doi: 10.1002/ejhf.1924.
- 34  13. Kraljevic Z, Searle T, Shek A, et al. Multi-domain clinical natural language processing with  
35  MedCAT: The medical concept annotation toolkit. *Artif Intell Med.* 2021;117. doi:  
36  10.1016/j.artmed.2021.102083.

- 1 14. Wu H, Wang M, Wu J, et al. A survey on clinical natural language processing in the united  
2 kingdom from 2007 to 2022. *npj Digital Medicine*. 2022;5(1). doi: 10.1038/s41746-022-00730-6.
- 3 15. Farajidavar N, O’Gallagher K, Bean D, et al. Diagnostic signature for heart failure with preserved  
4 ejection fraction (HFpEF): A machine learning approach using multi-modality electronic health record  
5 data. *BMC Cardiovascular Disorders*. 2022;22(1). doi: 10.1186/s12872-022-03005-w.
- 6 16. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC guidelines for the diagnosis and treatment of  
7 acute and chronic heart failure: Developed by the task force for the diagnosis and treatment of acute  
8 and chronic heart failure of the european society of cardiology (ESC). with the special contribution of  
9 the heart failure association (HFA) of the ESC. *European Journal of Heart Failure*. 2022;24(1). doi:  
10 10.1002/ejhf.2333.
- 11 17. Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under  
12 correlated receiver operating characteristic curves. *IEEE Signal Process Lett*. 2014;21(11). doi:  
13 10.1109/LSP.2014.2337313.
- 14 18. Jackson R, Kartoglu I, Stringer C, et al. CogStack - experiences of deploying integrated  
15 information retrieval and extraction services in a large national health service foundation trust hospital.  
16 *BMC Medical Informatics and Decision Making*. 2018;18(1). doi: 10.1186/s12911-018-0623-9.
- 17 19. Scott M, Su-In L. A unified approach to interpreting model predictions. *Advances in neural  
18 information processing systems*. 2017;30:4765–4774.
- 19 20. Collins GS, Dhiman P, Andaur Navarro C,L., et al. Protocol for development of a reporting  
20 guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction  
21 model studies based on artificial intelligence. *BMJ Open*. 2021;11(7):e048008.  
22 <http://bmjopen.bmj.com/content/11/7/e048008.abstract>. doi: 10.1136/bmjopen-2020-048008.

23

## 1 **Figure Legends**

2 Figure 1. Schematic diagram of the study

3 Figure 2. (a) SHAP graph for the XGB AIM-HFpEF full model, the y-axis is the list of features  
4 ranked by their importance with the most important feature at the top. The x-axis shows the SHAP  
5 values which indicate the impact of each feature on the model's prediction. Positive SHAP values  
6 show to a tendency of positive HFpEF prediction. The colour reflects the values of the features (for  
7 continuous features: red represents higher values and blue lower values; for sex: red represents male  
8 and blue female; for binary features like comorbidities and symptoms: red represents positive and  
9 blue negative). (b) ROC curve of KCH testing cohort. (c) ROC curve of GSTT validation cohort.

10 Figure 3. Comparisons of Simplified XGB, H2FPEF and HFpEF-ABA in the GSTT validation cohort

11 Figure 4. Outcome of GSTT validation cohort divided into three groups based on AIM-HFpEF  
12 predicted probability tertiles (left: overall cohort, right: undiagnosed HFpEF).

13 Supplementary Figure 1. (a) Performance of the AIM-HFpEF prediction model on Confirmed HFpEF  
14 (left) and Undiagnosed HFpEF (right) patients in GSTT validation cohort. (b) Comparison with  
15 H2FPEF. (c) Comparison with HFpEF-ABA.

16 Supplementary Figure 2. Calibration curves of predicted probabilities produced by the AIM-HFpEF  
17 simplified model. The undiagnosed HFpEF group is more calibrated, while the confirmed HFpEF  
18 group tends to have over-estimated probabilities when predicted probabilities are low.

19 Supplementary Figure 3. ROC curves for comparisons with H2FPEF and HFpEF-ABA scores in KCH  
20 testing cohort, (a) Overall (b) Non-White patients (c) low IMD patients.

21 Supplementary Figure 4. Kaplan-Meier (KM) curves for patients having a predicted probability <sup>3</sup> 90%  
22 of having HFpEF by the models with (a) all-cause mortality (b) MI and (c) Stroke within 5 years.

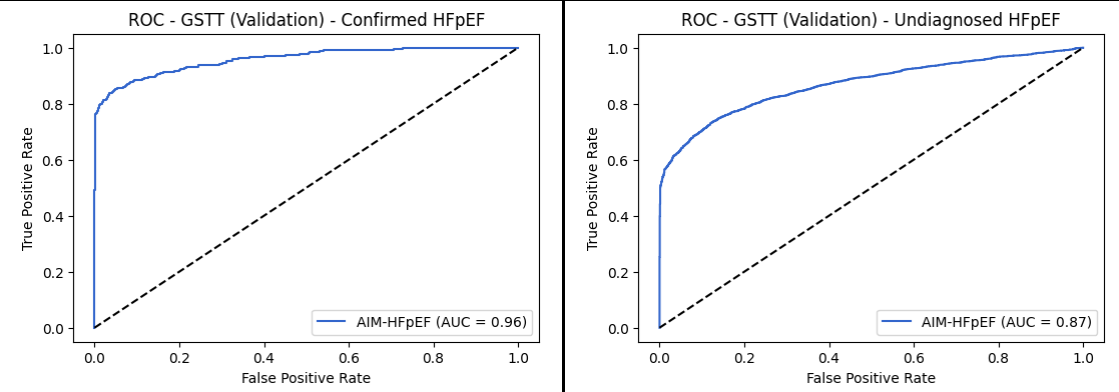
23

Table 1. Baseline characteristics (demographics, comorbidities, symptoms, echocardiograph parameters) of patients						
	KCH			GSTT		
	HFpEF (n=2220)		non-HF Dyspnoea	HFpEF (n=3349)		non-HF Dyspnoea
	Confirmed HFpEF	Undiagnosed HFpEF	Control	Confirmed HFpEF	Undiagnosed HFpEF	Control
n	231	1989	950	430	2919	2034
age, mean (SD)	78.8 (10.3)	71.1 (15.5)	57.5 (15.7)	78.1 (10.3)	70.7 (15.1)	56.5 (15.5)
sex, n (%) Female	153 (66.2)	1176 (59.1)	438 (46.1)	257 (59.8)	1561 (53.5)	1080 (53.1)
ethnicity, n (%) White	122 (52.8)	1136 (57.1)	531 (55.9)	249 (57.9)	1701 (58.3)	1170 (57.5)
Black	65 (28.1)	513 (25.8)	196 (20.6)	85 (19.8)	485 (16.6)	350 (17.2)
Asian	17 (7.4)	117 (5.9)	57 (6.0)	29 (6.7)	99 (3.4)	138 (6.8)
diabetes mellitus type 1, n (%)	26 (11.3)	171 (8.6)	30 (3.2)	49 (11.4)	214 (7.3)	123 (6.0)
diabetes mellitus type 2, n (%)	85 (36.8)	493 (24.8)	118 (12.4)	230 (53.5)	962 (33.0)	449 (22.1)
ischemic heart disease, n (%)	93 (40.3)	455 (22.9)	242 (25.5)	152 (35.3)	843 (28.9)	378 (18.6)
myocardial infarction, n (%)	36 (15.6)	147 (7.4)	95 (10.0)	72 (16.7)	409 (14.0)	205 (10.1)
cerebrovascular accident, n (%)	81 (35.1)	379 (19.1)	95 (10.0)	38 (8.8)	113 (3.9)	61 (3.0)
hypertensive disorder, n (%)	214 (92.6)	1223 (61.5)	423 (44.5)	379 (88.1)	1932 (66.2)	998 (49.1)
transient ischemic attack, n (%)	47 (20.3)	200 (10.1)	35 (3.7)	40 (9.3)	194 (6.6)	94 (4.6)
atrial fibrillation, n (%)	119 (51.5)	409 (20.6)	60 (6.3)	247 (57.4)	1151 (39.4)	202 (9.9)
pulmonary hypertension, n (%)	40 (17.3)	180 (9.0)	11 (1.2)	51 (11.9)	281 (9.6)	61 (3.0)
kidney disease, n (%)	132 (57.1)	601 (30.2)	156 (16.4)	250 (58.1)	1184 (40.6)	783 (38.5)
angina, n (%)	58 (25.1)	221 (11.1)	101 (10.6)	55 (12.8)	243 (8.3)	104 (5.1)
dyspnoea, n (%)	212 (91.8)	1169 (58.8)	950 (100.0)	353 (82.1)	1784 (61.1)	2034 (100.0)
dyspnoea at rest, n (%)	10 (4.3)	22 (1.1)	NA	34 (7.9)	18 (0.6)	3 (0.1)
chest pain, n (%)	141 (61.0)	651 (32.7)	413 (43.5)	171 (39.8)	787 (27.0)	643 (31.6)
dizziness, n (%)	86 (37.2)	346 (17.4)	125 (13.2)	143 (33.3)	558 (19.1)	341 (16.8)
presyncope, n (%)	5 (2.2)	27 (1.4)	8 (0.8)	8 (1.9)	44 (1.5)	23 (1.1)
syncope, n (%)	33 (14.3)	120 (6.0)	49 (5.2)	49 (11.4)	209 (7.2)	156 (7.7)

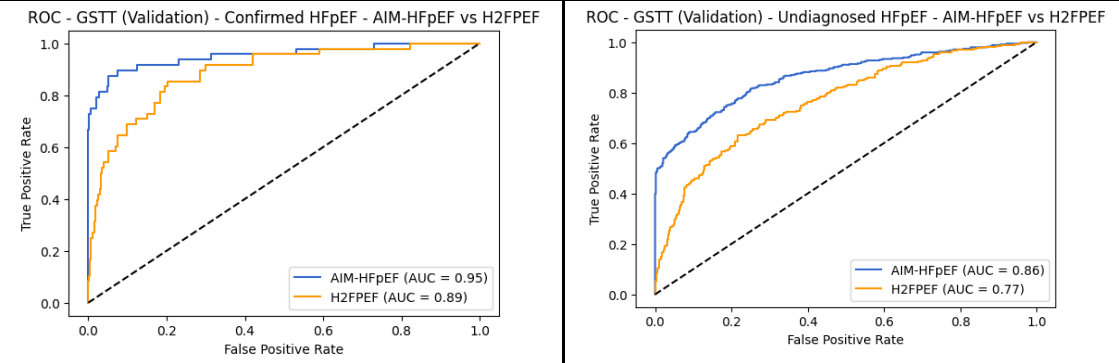
NTproBNP (pg/ml), mean (SD)	2714.3 (4412.3)	3086.0 (5519.3)	111.4 (83.5)	3938.9 (7518.0)	3533.3 (6905.5)	122.6 (97.9)
BMI, mean (SD)	30.6 (6.7)	29.2 (7.0)	28.0 (6.1)	31.3 (5.7)	28.8 (5.7)	28.0 (5.3)
IVSd (mm), mean (SD)	1.2 (0.2)	1.2 (0.3)	1.1 (0.2)	1.2 (0.3)	1.2 (0.3)	1.1 (0.3)
LVEDD (mm), mean (SD)	4.5 (0.6)	4.4 (0.6)	4.5 (0.6)	4.5 (0.7)	4.5 (0.7)	4.5 (0.6)
LVEF (%), mean (SD)	59.8 (5.8)	60.5 (6.1)	60.1 (6.3)	59.7 (5.4)	59.3 (5.6)	60.6 (4.9)
LVPWD (mm), mean (SD)	1.1 (0.2)	1.1 (0.2)	1.0 (0.2)	1.1 (0.2)	1.1 (0.2)	1.0 (0.2)
E/e', mean (SD)	13.7 (6.0)	11.6 (5.9)	8.5 (3.3)	13.0 (6.1)	11.1 (5.3)	8.1 (3.8)
LA volume (ml), mean (SD)	80.2 (34.3)	69.0 (38.4)	50.7 (19.3)	70.2 (40.1)	66.7 (38.9)	44.5 (24.0)
LA volume indexed (ml/m <sup>2</sup> ), mean (SD)	41.7 (14.7)	35.7 (17.8)	27.1 (8.9)	42.8 (22.0)	36.7 (16.1)	24.0 (12.5)
LV mass (g), mean (SD)	191.4 (59.6)	185.5 (66.6)	167.2 (55.8)	188.7 (61.1)	190.1 (65.4)	163.6 (59.6)
LV mass indexed (g/m <sup>2</sup> ), mean (SD)	101.3 (28.4)	95.8 (33.4)	84.1 (24.2)	104.0 (31.5)	100.7 (29.7)	86.1 (27.9)
PASP (mmHg), mean (SD)	31.6 (12.4)	32.5 (14.3)	23.2 (8.1)	33.1 (13.7)	32.0 (15.3)	23.5 (9.9)
TR max vel (cm/s), mean (SD)	276.1 (52.3)	278.6 (60.2)	237.3 (40.5)	273.7 (55.5)	267.3 (64.9)	227.0 (50.2)
H2FPEF, mean (SD)	5.8 (1.8)	4.3 (2.1)	2.2 (1.6)	4.5 (1.9)	3.5 (2.0)	2.2 (1.7)
IMD, mean (SD)	4.4 (2.2)	4.4 (2.3)	4.7 (2.4)	3.9 (1.9)	4.6 (2.4)	4.7 (2.4)

Supplementary Figure 1

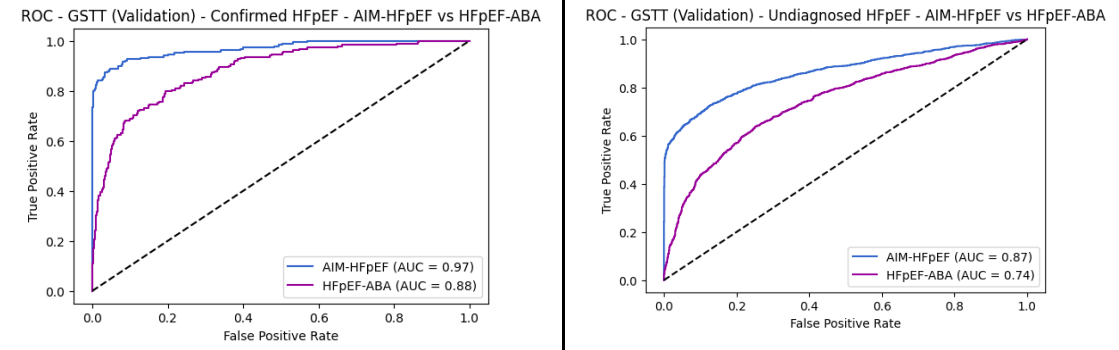
(a) AIM-HFpEF simplified model on Confirmed HFpEF patients (AUC: 0.9586 [95% CI, 0.9474-0.9699]) and Undiagnosed HFpEF patients (AUC: 0.8694 [95% CI, 0.8597-0.8792])



(b) Comparison with H2FPEF in GSTT validation cohort. Left, AIM-HFpEF: 0.9534 [0.9141-0.9927], H2FPEF: 0.8875 [0.8383-0.9368],  $p=0.0030$ . Right, AIM-HFpEF: 0.8637 [0.8378-0.8895], H2FPEF: 0.7656 [0.7335-0.7978],  $p<0.001$ .

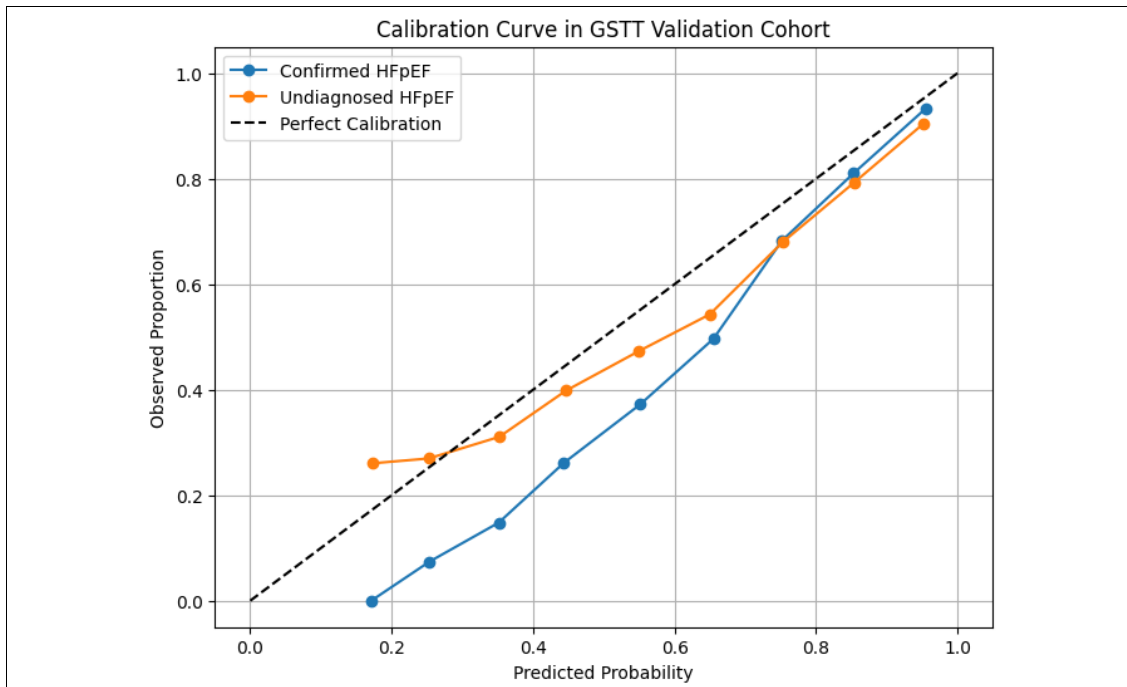


(c) Comparison with HFpEF-ABA in GSTT validation cohort. Left, AIM-HFpEF: 0.9691 [0.9548-0.9834], HFpEF-ABA: 0.8827 [0.8572-0.9082],  $p<0.0001$ . Right, AIM-HFpEF: 0.8654 [0.8516-0.8793], HFpEF-ABA: 0.7446 [0.7270-0.7622],  $p<0.0001$



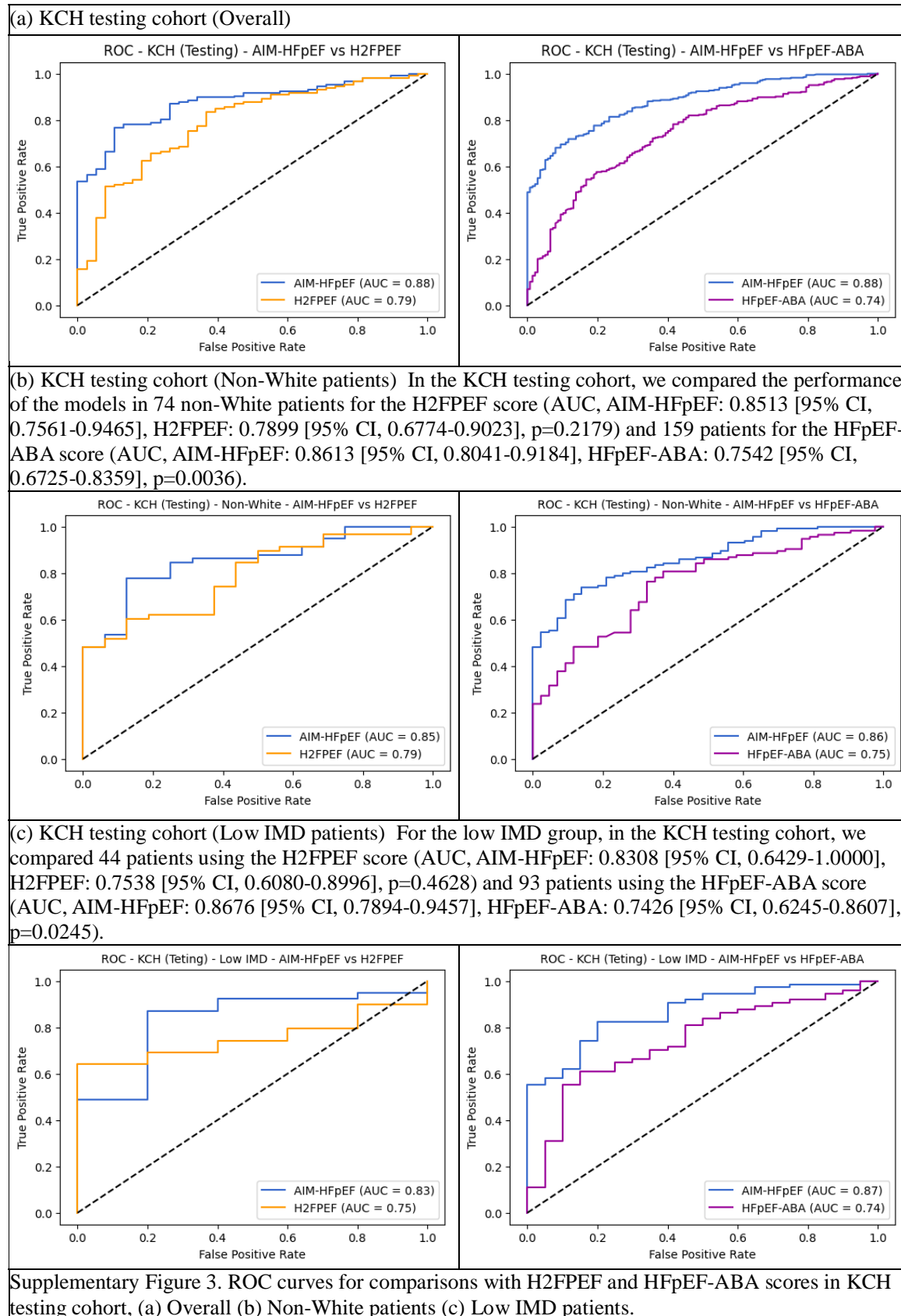
Supplementary Figure 1. (a) Performance of the AIM-HFpEF prediction model on Confirmed HFpEF (left) and Undiagnosed HFpEF (right) patients in GSTT validation cohort. (b) Comparison with H2FPEF. (c) Comparison with HFpEF-ABA.

## Supplementary Figure 2



Supplementary Figure 2. Calibration curves of predicted probabilities produced by the AIM-HFpEF simplified model. The undiagnosed HFpEF group is more calibrated, while the confirmed HFpEF group tends to have over-estimated probabilities when predicted probabilities are low.

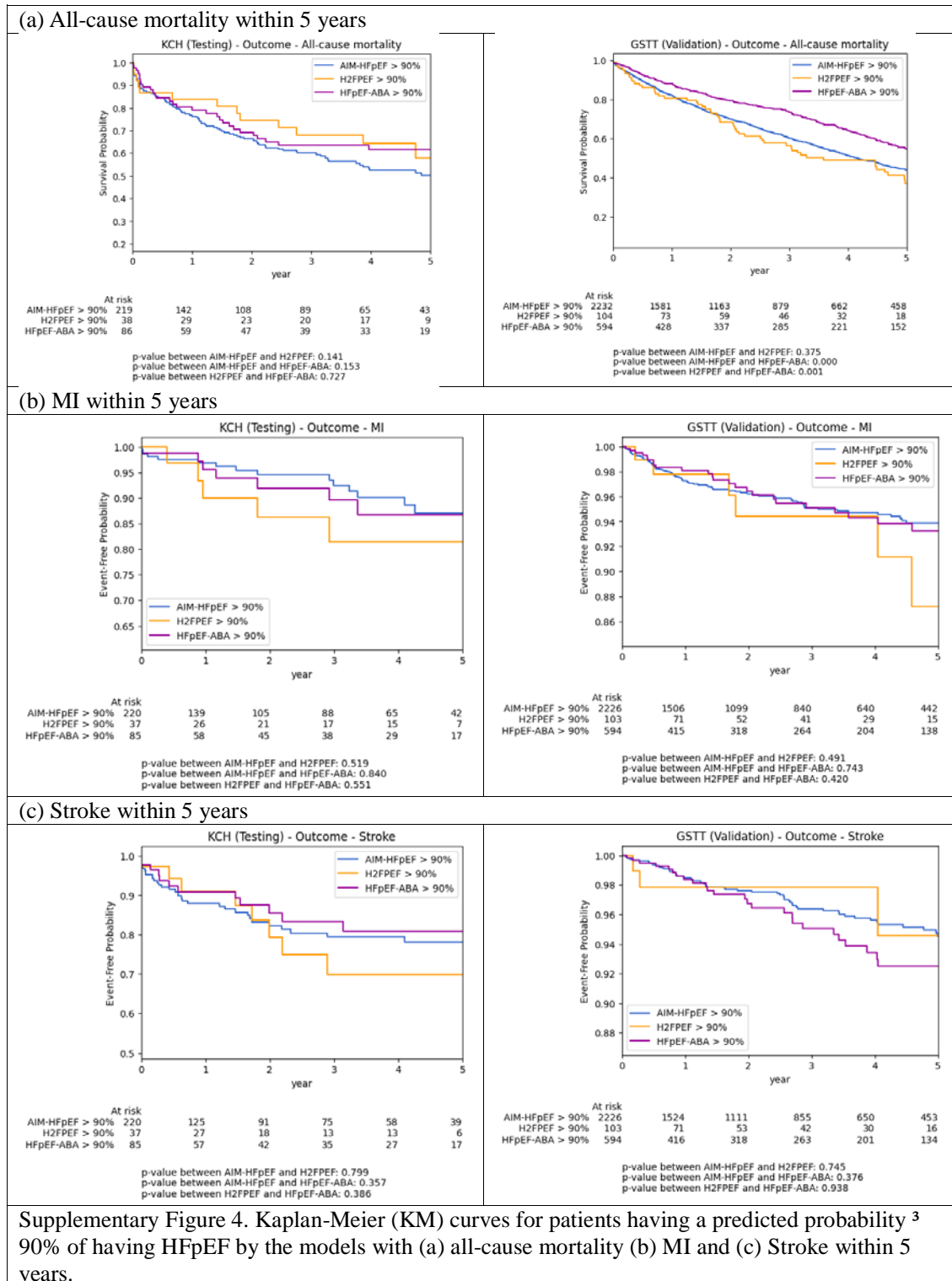
Supplementary Figure 3.







Supplementary Figure 4



Supplementary Table 1. SNOMED-CT terms used in the study for the inclusion and exclusion criteria. All child terms are included by default. Inclusion criteria include HF, HFpEF and Dyspnoea. Exclusion criteria include severe valvular heart disease, hypertrophic cardiomyopathy, restrictive cardiomyopathy, constrictive pericarditis, and cardiac amyloidosis

Criteria	SNOMED-CT Term	ID
Inclusion	Heart failure (disorder)	84114007
	Heart failure with normal ejection fraction (disorder)	446221000
	Dyspnoea (finding)	267036007
Exclusion	Aortic valve stenosis (disorder)*	60573004
	Aortic valve regurgitation (disorder)*	60234000
	Mitral valve stenosis (disorder)*	79619009
	Mitral valve regurgitation (disorder)*	48724000
	Pulmonic valve stenosis (disorder)*	56786000
	Pulmonic valve regurgitation (disorder)*	91434003
	Hypertrophic cardiomyopathy (disorder)	233873004
	Senile cardiac amyloidosis (disorder)	16573007
	Restrictive cardiomyopathy (disorder)	415295002
	Constrictive pericarditis (disorder)	85598007

\* - the **severe** valvular heart disease patients are identified with the 'severe' keyword before the mentions of the SNOMED-CT terms.

Supplementary Table 2. Performances of all machine learning models

F – Full Model  
S – Simplified Model

LR – Logistic Regression  
RF – Random Forest  
SVM – Support Vector Machine  
XGB – XGBoost

H2FPEF-Point – Point-based H2FPEF score

KCH Testing Cohort (n=634)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]
F-LR	0.7760	0.8094	0.8896	0.8476	0.8204 [0.7868-0.8540]
F-RF	0.7382	0.7473	0.9459	0.8350	0.8472 [0.8161-0.8784]
F-SVM	0.7918	0.8250	0.8919	0.8571	0.8386 [0.8063-0.8709]
F-XGB	<b>0.8186</b>	0.8600	0.8851	<b>0.8724</b>	<b>0.8910 [0.8665-0.9154]</b>
S-LR	0.7539	0.7975	0.8694	0.8319	0.8025 [0.7678-0.8373]
S-RF	0.7240	0.7307	<b>0.9595</b>	0.8296	0.8379 [0.8051-0.8707]
S-SVM	0.7918	0.8223	0.8964	0.8578	0.8205 [0.7852-0.8559]
S-XGB	0.8060	<b>0.8607</b>	0.8626	0.8616	0.8845 [0.8592-0.9097]

KCH Testing Cohort – Comparison with H2FPEF (n=182)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.7935	0.8649	0.8767	0.8707	0.8398 [0.7764-0.9031]	0.0753
F-RF	0.8043	0.8354	0.9384	0.8839	0.8600 [0.8060-0.9141]	0.0318
F-SVM	0.7989	0.8658	0.8836	0.8746	0.8506 [0.7844-0.9168]	0.0447
F-XGB	0.8261	0.8800	0.9041	0.8919	<b>0.8756 [0.8245-0.9267]</b>	0.0062
S-LR	0.7935	0.8699	0.8699	0.8699	0.8327 [0.7684-0.8971]	0.0490
S-RF	0.8043	0.8274	<b>0.9521</b>	0.8854	0.8634 [0.8034-0.9233]	0.0351
S-SVM	0.8152	0.8636	0.9110	0.8867	0.8131 [0.7350-0.8912]	0.4362
S-XGB	<b>0.8315</b>	0.8912	0.8973	<b>0.8942</b>	0.8751 [0.8225-0.9276]	0.0064
H2FPEF	0.6848	<b>0.9231</b>	0.6575	0.7680	0.7873 [0.7075-0.8672]	-
H2FPEF-Point	0.8152	0.8544	0.9247	0.8882	0.7222 [0.6259-0.8184]	0.1663

KCH Testing Cohort – Comparison with HFpEF-ABA (n=483)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.7734	0.8150	0.8837	0.8480	0.8206 [0.7821-0.8591]	<0.0001
F-RF	0.7443	0.7552	0.9506	0.8417	0.8454 [0.8096-0.8811]	<0.0001
F-SVM	0.7879	0.8288	0.8866	0.8567	0.8353 [0.7973-0.8732]	<0.0001
F-XGB	<b>0.8170</b>	0.8657	0.8808	<b>0.8732</b>	<b>0.8873 [0.8587-0.9159]</b>	<0.0001
S-LR	0.7588	0.8081	0.8692	0.8375	0.7983 [0.7573-0.8392]	0.0013
S-RF	0.7297	0.7388	<b>0.9622</b>	0.8359	0.8275 [0.7880-0.8671]	<0.0001
S-SVM	0.7963	0.8254	0.9070	0.8643	0.8148 [0.7724-0.8572]	0.0007
S-XGB	0.8046	<b>0.8676</b>	0.8576	0.8626	0.8792 [0.8493-0.9091]	<0.0001
HFpEF-ABA	0.7173	0.8230	0.7703	0.7958	0.7425 [0.6949-0.7901]	-

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

KCH Testing Cohort – Non-White (n=202)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]
F-LR	0.7824	0.8108	0.8955	0.8511	0.8327 [0.7715-0.8938]
F-RF	0.7668	0.7633	<b>0.9627</b>	0.8515	0.8576 [0.8044-0.9108]
F-SVM	0.7927	0.8264	0.8881	0.8561	0.8443 [0.7852-0.9034]
F-XGB	<b>0.8135</b>	<b>0.8451</b>	0.8955	<b>0.8696</b>	<b>0.8848 [0.8390-0.9305]</b>
S-LR	0.7513	0.7945	0.8657	0.8286	0.8034 [0.7397-0.8672]
S-RF	0.7513	0.7529	0.9552	0.8421	0.8578 [0.8014-0.9143]
S-SVM	0.7720	0.8041	0.8881	0.8440	0.8252 [0.7645-0.8859]
S-XGB	0.7927	0.8406	0.8657	0.8529	0.8823 [0.8356-0.9290]

KCH Testing Cohort – Non-White – Comparison with H2FPEF (n=74)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.7973	0.8909	0.8448	0.8673	0.8513 [0.7472-0.9554]	0.1586
F-RF	<b>0.8243</b>	0.8571	<b>0.9310</b>	<b>0.8926</b>	0.8594 [0.7721-0.9467]	0.1175
F-SVM	0.8108	0.9074	0.8448	0.8750	0.8524 [0.7477-0.9571]	0.1786
F-XGB	0.8108	0.8793	0.8793	0.8793	<b>0.8739 [0.7899-0.9580]</b>	0.0614
S-LR	0.7703	0.8727	0.8276	0.8496	0.8222 [0.7160-0.9284]	0.4025
S-RF	0.8108	0.8438	<b>0.9310</b>	0.8852	0.8815 [0.7954-0.9676]	0.0559
S-SVM	0.7703	0.8475	0.8621	0.8547	0.8060 [0.6980-0.9141]	0.7285
S-XGB	0.8108	0.8929	0.8621	0.8772	0.8513 [0.7561-0.9465]	0.2179
H2FPEF	0.6622	<b>0.9459</b>	0.6034	0.7368	0.7899 [0.6774-0.9023]	-
H2FPEF-Point	0.7703	0.8154	0.9138	0.8618	0.6546 [0.4886-0.8207]	0.1400

KCH Testing Cohort – Non-White – Comparison with HFpEF-ABA (n=159)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.7707	0.8197	0.8772	0.8475	0.8027 [0.7268-0.8787]	0.2133
F-RF	0.7707	0.7786	<b>0.9561</b>	0.8583	0.8355 [0.7721-0.8988]	0.0420
F-SVM	0.7834	0.8390	0.8684	0.8534	0.8154 [0.7408-0.8900]	0.1153
F-XGB	<b>0.8089</b>	0.8559	0.8860	<b>0.8707</b>	<b>0.8694 [0.8142-0.9247]</b>	0.0020
S-LR	0.7580	0.8167	0.8596	0.8376	0.7752 [0.6976-0.8528]	0.5610
S-RF	0.7580	0.7676	<b>0.9561</b>	0.8516	0.8308 [0.7610-0.9006]	0.0279
S-SVM	0.7707	0.8095	0.8947	0.8500	0.7997 [0.7260-0.8734]	0.2027
S-XGB	0.7834	0.8448	0.8596	0.8522	0.8613 [0.8041-0.9184]	0.0036
HFpEF-ABA	0.7197	<b>0.8571</b>	0.7368	0.7925	0.7542 [0.6725-0.8359]	-

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

KCH Testing Cohort – Low IMD (n=118)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]
F-LR	0.7812	0.8447	0.8788	0.8614	0.8213 [0.7416-0.9010]
F-RF	0.8125	0.8319	0.9495	0.8868	0.8467 [0.7745-0.9190]
F-SVM	0.8047	0.8700	0.8788	0.8744	0.8746 [0.8151-0.9341]
F-XGB	<b>0.8438</b>	0.899	0.8990	<b>0.8990</b>	<b>0.9011 [0.8460-0.9562]</b>
S-LR	0.7656	0.8350	0.8687	0.8515	0.8001 [0.7177-0.8824]
S-RF	0.7891	0.8000	<b>0.9697</b>	0.8767	0.8558 [0.7846-0.9270]
S-SVM	0.7969	0.8763	0.8586	0.8673	0.8335 [0.7550-0.9120]
S-XGB	0.8281	<b>0.9053</b>	0.8687	0.8866	0.8812 [0.8198-0.9427]

KCH Testing Cohort – Low IMD – Comparison with H2FPEF (n=44)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.7727	0.9143	0.8205	0.8649	0.8359 [0.7065-0.9653]	0.3105
F-RF	0.8409	0.8810	0.9487	0.9136	0.8077 [0.6738-0.9416]	0.5139
F-SVM	0.7727	0.9143	0.8205	0.8649	0.8513 [0.7404-0.9622]	0.1805
F-XGB	0.8409	0.9000	0.9231	0.9114	0.8615 [0.7282-0.9948]	0.2082
S-LR	0.7955	0.9167	0.8462	0.8800	0.8513 [0.7320-0.9705]	0.1345
S-RF	0.8636	0.8837	<b>0.9744</b>	<b>0.9268</b>	<b>0.8923 [0.7955-0.9891]</b>	0.0973
S-SVM	0.7955	0.9167	0.8462	0.8800	0.7795 [0.6015-0.9575]	0.7758
S-XGB	<b>0.8636</b>	0.9459	0.8974	0.9211	0.8308 [0.6429-1.0000]	0.4628
H2FPEF	0.6818	<b>1.0000</b>	0.6410	0.7812	0.7538 [0.6080-0.8996]	-
H2FPEF-Point	0.7955	0.9167	0.8462	0.8800	0.6821 [0.3958-0.9683]	0.5871

KCH Testing Cohort – Low IMD – Comparison with HFpEF-ABA (n=93)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.7660	0.8421	0.8649	0.8533	0.8034 [0.7023-0.9044]	0.3102
F-RF	0.7979	0.8161	<b>0.9595</b>	0.8820	0.8007 [0.7034-0.8980]	0.2763
F-SVM	0.7872	0.8553	0.8784	0.8667	0.8500 [0.7728-0.9272]	0.0319
F-XGB	<b>0.8404</b>	0.8831	0.9189	<b>0.9007</b>	0.8892 [0.8185-0.9599]	0.0098
S-LR	0.7660	0.8333	0.8784	0.8553	0.7682 [0.6585-0.8779]	0.5925
S-RF	0.7979	0.8022	0.9865	0.8848	0.8324 [0.7381-0.9268]	0.0784
S-SVM	0.7872	0.8462	0.8919	0.8684	0.8142 [0.7131-0.9152]	0.1978
S-XGB	0.8298	<b>0.8919</b>	0.8919	0.8919	<b>0.8676 [0.7894-0.9457]</b>	0.0245
HFpEF-ABA	0.7234	0.8636	0.7703	0.8143	0.7426 [0.6245-0.8607]	-

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

GSTT Validation Cohort (n=5,621)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]
F-LR	0.7251	0.7364	0.8692	0.7973	0.7989 [0.7870-0.8108]
F-RF	0.7485	0.7340	0.9343	0.8221	0.8348 [0.8237-0.8459]
F-SVM	0.7377	0.7481	0.8719	0.8053	0.8086 [0.7970-0.8203]
F-XGB	<b>0.7986</b>	<b>0.8112</b>	0.8815	<b>0.8449</b>	<b>0.8934 [0.8852-0.9016]</b>
S-LR	0.7269	0.7315	0.8865	0.8016	0.8053 [0.7936-0.8170]
S-RF	0.7111	0.6998	<b>0.9382</b>	0.8016	0.8315 [0.8204-0.8426]
S-SVM	0.7215	0.7288	0.8797	0.7972	0.7952 [0.7830-0.8073]
S-XGB	0.7826	0.7986	0.8701	0.8328	0.8809 [0.8721-0.8897]

GSTT Validation Cohort – Comparison with H2FPEF (n=967)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.6685	0.5817	0.8223	0.6814	0.7849 [0.7550-0.8149]	0.6847
F-RF	0.6411	0.5493	0.9340	0.6917	0.8242 [0.7972-0.8513]	0.0001
F-SVM	0.6783	0.5871	0.8553	0.6963	0.8016 [0.7726-0.8306]	0.0623
F-XGB	0.7123	0.6105	<b>0.9188</b>	<b>0.7335</b>	<b>0.8821 [0.8596-0.9047]</b>	<0.0001
S-LR	0.6783	0.5890	0.8401	0.6925	0.7815 [0.7514-0.8117]	0.8918
S-RF	0.5722	0.5020	0.9569	0.6585	0.8071 [0.7788-0.8355]	0.0199
S-SVM	0.6422	0.5548	0.8604	0.6746	0.7821 [0.7515-0.8128]	0.8819
S-XGB	0.6860	0.5876	0.9112	0.7144	0.8746 [0.8510-0.8982]	<0.0001
H2FPEF	<b>0.7144</b>	<b>0.6692</b>	0.6675	0.6684	0.7805 [0.7505-0.8105]	-
H2FPEF-Point	0.6772	0.5980	0.7665	0.6719	0.7550 [0.7242-0.7858]	0.0072

GSTT Validation Cohort – Comparison with HFpEF-ABA (n=3,735)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.6391	0.5440	0.8496	0.6633	0.7800 [0.7644-0.7956]	0.0119
F-RF	0.6446	0.5442	0.9255	0.6854	0.8236 [0.8098-0.8375]	<0.0001
F-SVM	0.6597	0.5610	0.8585	0.6786	0.7949 [0.7798-0.8101]	<0.0001
F-XGB	<b>0.7518</b>	<b>0.6511</b>	0.8763	<b>0.7471</b>	<b>0.8898 [0.8784-0.9012]</b>	<0.0001
S-LR	0.6357	0.5399	0.8742	0.6675	0.7915 [0.7762-0.8068]	<0.0001
S-RF	0.5882	0.5042	<b>0.9371</b>	0.6557	0.8220 [0.8079-0.8362]	<0.0001
S-SVM	0.6340	0.5385	0.8742	0.6665	0.7886 [0.7731-0.8041]	<0.0001
S-XGB	0.7341	0.6331	0.8667	0.7317	0.8788 [0.8665-0.8912]	<0.0001
HFpEF-ABA	0.6623	0.5687	0.7977	0.6640	0.7624 [0.7462-0.7787]	-

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

GSTT Validation Cohort – Non-White (n=1,182)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]
F-LR	0.7125	0.7112	0.8610	0.7790	0.7956 [0.7703-0.8210]
F-RF	0.7302	0.7019	<b>0.9413</b>	0.8042	0.8301 [0.8066-0.8536]
F-SVM	0.7218	0.7201	0.8625	0.7849	0.8013 [0.7764-0.8261]
F-XGB	<b>0.7850</b>	<b>0.7814</b>	0.8811	<b>0.8283</b>	<b>0.8893 [0.8717-0.9070]</b>
S-LR	0.7015	0.7014	0.8582	0.7719	0.7794 [0.7534-0.8053]
S-RF	0.6855	0.6694	0.9198	0.7749	0.8166 [0.7926-0.8406]
S-SVM	0.6813	0.6878	0.8395	0.7561	0.7616 [0.7343-0.7888]
S-XGB	0.7572	0.7615	0.8553	0.8057	0.8641 [0.8437-0.8844]

GSTT Validation Cohort – Non-White – Comparison with H2FPEF (n=249)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.7018	0.6439	0.8019	0.7143	0.7877 [0.7284-0.8469]	0.4246
F-RF	0.6798	0.5976	<b>0.9528</b>	0.7345	0.8304 [0.7786-0.8823]	0.0119
F-SVM	0.7149	0.6414	0.8774	0.7410	0.8017 [0.7447-0.8588]	0.1444
F-XGB	<b>0.7412</b>	0.6556	0.9340	<b>0.7704</b>	<b>0.9041 [0.8666-0.9416]</b>	<0.0001
S-LR	0.6886	0.6259	0.8208	0.7102	0.7674 [0.7062-0.8286]	0.9653
S-RF	0.5965	0.5376	0.9434	0.6849	0.7856 [0.7265-0.8447]	0.4863
S-SVM	0.6535	0.5931	0.8113	0.6853	0.7543 [0.6888-0.8199]	0.5595
S-XGB	0.7018	0.6218	0.9151	0.7405	0.8812 [0.8365-0.9259]	<0.0001
H2FPEF	0.6974	<b>0.7033</b>	0.6038	0.6497	0.7681 [0.7071-0.8291]	-
H2FPEF-Point	0.6711	0.6260	0.7264	0.6725	0.7394 [0.6760-0.8028]	0.1688

GSTT Validation Cohort – Non-White – Comparison with HFpEF-ABA (n=923)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.6312	0.5326	0.8201	0.6458	0.7678 [0.7347-0.8010]	0.0003
F-RF	0.6409	0.5350	<b>0.9469</b>	0.6837	0.8219 [0.7937-0.8500]	<0.0001
F-SVM	0.6590	0.5543	0.8584	0.6736	0.7898 [0.7581-0.8216]	<0.0001
F-XGB	<b>0.7497</b>	<b>0.6387</b>	0.8968	<b>0.7460</b>	<b>0.8990 [0.8768-0.9212]</b>	<0.0001
S-LR	0.6215	0.5243	0.8289	0.6423	0.7569 [0.7231-0.7907]	<0.0001
S-RF	0.5816	0.4944	0.9145	0.6418	0.8031 [0.7725-0.8337]	<0.0001
S-SVM	0.6058	0.5119	0.8230	0.6312	0.7494 [0.7146-0.7841]	0.0045
S-XGB	0.7170	0.6096	0.8614	0.7139	0.8744 [0.8477-0.9012]	<0.0001
HFpEF-ABA	0.6518	0.5609	0.6932	0.6201	0.7101 [0.6734-0.7468]	-



It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

GSTT Validation Cohort – Low IMD (n=1,136)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]
F-LR	0.7324	0.7528	0.8670	0.8059	0.8074 [0.7816-0.8332]
F-RF	0.7736	0.7580	<b>0.9499</b>	0.8432	0.8513 [0.8276-0.8751]
F-SVM	0.7544	0.7731	0.8727	0.8199	0.8202 [0.7949-0.8454]
F-XGB	<b>0.8130</b>	<b>0.8322</b>	0.8870	<b>0.8587</b>	<b>0.8969 [0.8789-0.9148]</b>
S-LR	0.7369	0.7470	0.8913	0.8128	0.8170 [0.7918-0.8422]
S-RF	0.7287	0.7212	0.9399	0.8161	0.8376 [0.8137-0.8615]
S-SVM	0.7324	0.7455	0.8841	0.8089	0.7993 [0.7724-0.8262]
S-XGB	0.7837	0.8099	0.8655	0.8368	0.8811 [0.8615-0.9006]

GSTT Validation Cohort – Low IMD – Comparison with H2FPEF (n=195)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.6902	0.6667	0.7609	0.7107	0.7811 [0.7133-0.8488]	0.9449
F-RF	0.6957	0.6364	0.9130	0.7500	0.8355 [0.7751-0.8959]	0.0365
F-SVM	0.7609	0.7264	0.8370	0.7778	0.8358 [0.7762-0.8954]	0.0323
F-XGB	<b>0.7935</b>	<b>0.7411</b>	0.9022	<b>0.8137</b>	<b>0.8982 [0.8505-0.9459]</b>	0.0001
S-LR	0.7120	0.6726	0.8261	0.7415	0.7883 [0.7222-0.8544]	0.6006
S-RF	0.6359	0.5817	<b>0.9674</b>	0.7265	0.8159 [0.7545-0.8773]	0.1997
S-SVM	0.6793	0.6364	0.8370	0.7230	0.8029 [0.7366-0.8693]	0.3887
S-XGB	0.7337	0.6720	0.9130	0.7742	0.8914 [0.8423-0.9406]	0.0003
H2FPEF	0.7065	0.7317	0.6522	0.6897	0.7793 [0.7127-0.8459]	-
H2FPEF-Point	0.7065	0.6759	0.7935	0.7300	0.7517 [0.6819-0.8214]	0.1754

GSTT Validation Cohort – Low IMD – Comparison with HFpEF-ABA (n=726)

Model	Accuracy	Precision	Recall	F1-Score	AUC [95% CI]	P-value
F-LR	0.6388	0.5488	0.8374	0.6630	0.7761 [0.7399-0.8123]	0.9205
F-RF	0.6608	0.5602	0.9343	0.7004	0.8365 [0.8064-0.8665]	<0.0001
F-SVM	0.6799	0.5828	0.8651	0.6964	0.8058 [0.7721-0.8395]	0.0459
F-XGB	<b>0.7709</b>	<b>0.6736</b>	0.8927	<b>0.7679</b>	<b>0.8993 [0.8749-0.9237]</b>	<0.0001
S-LR	0.6344	0.5433	0.8685	0.6684	0.7949 [0.7599-0.8298]	0.0832
S-RF	0.6050	0.5189	<b>0.9481</b>	0.6707	0.8240 [0.7927-0.8552]	0.0013
S-SVM	0.6358	0.5443	0.8720	0.6702	0.7871 [0.7518-0.8224]	0.3572
S-XGB	0.7342	0.6378	0.8651	0.7342	0.8829 [0.8557-0.9100]	<0.0001
HFpEF-ABA	0.6711	0.5799	0.8166	0.6782	0.7745	-