Clinical utility of an AI-based model for detecting Heart Failure with preserved Ejection Fraction (HFpEF) through decision curve analysis

Akerman, A.P*., Porumb, M*., Scott, C. G#., Beqiri, A*., Chartsias, A*., Ryu, A.J#., Hawkes, W*., Huntley, G.D#., Arystan, A.Z#., Kane, G.C#., Pislaru, S.V#., Lopez-Jiminez, F#., Sarwar, R*.^., O'Driscoll, J*, $., Leeson, P*, ^., Upton, R*., Woodward, G*., Pellikka, P.A#.

* Ultromics Ltd, Oxford, UK; # Mayo Clinic, MN, USA; ^ University of Oxford, UK; $ Canterbury Christ Church University, UK

EchoGo
MAYO CLINIC
St George's University Hospitals NHS Foundation Trust

PURPOSE / OBJECTIVES

Heart failure with preserved ejection fraction (HFpEF) is a clinical syndrome with increasing prevalence, poor 5-year survival rates, high re-admission rates, and substantial morbidity. Echocardiography is critical in the HFpEF diagnostic pathway, but algorithms for echocardiographic interpretation, and the integration into broader clinical decision making are limited by discordant or incomplete data. This leads to variable diagnostic capacity, increasing requirements for further confirmatory testing or incorrect patient management.

MATERIAL & METHODS

A three-dimensional convolutional neural network was developed to automatically detect HFpEF using only the apical four-chamber videoclip (EchoGo Heart Failure; Ultromics Ltd). Model development utilized retrospective, multi-site, and multi-national cohort data (Mayo Clinic, USA; NHS, UK). Echocardiogram databases and electronic medical records were used to identify patients with preserved ejection fraction (≥50%), and evidence of increased intra-cardiac filling pressure, and a diagnosis of heart failure (ICD-9/10) within one year of the echocardiogram or lack thereof (cases and controls, respectively). In an independent testing dataset comprised of multi-site retrospective data from Mayo Clinic Health System (USA), the AI model was compared to clinically validated algorithms (HFA-PEFF Score[1] and H2FPEF Score[2]) with respect to classification performance (sensitivity and specificity) and the impact on clinical decision making (decision curve analysis).

RESULTS

Patient demographics for the 2971 cases and 3785 controls utilized for training and validation of the AI model, and 646 cases and 638 controls utilized for independent testing are presented in **Table 1**. The AI model demonstrated excellent discrimination performance in all datasets, with AUROC between 0.91 and 0.97 (**Figure 1**), and very good sensitivity (mean: 87.8% [95% CI: 84.5, 90.9]) and specificity (81.9% [78.2, 85.6]) on 1190/1284 patients in the training dataset (uncertain in 7.3%). The HFA-PEFF and H2FPEF scores also demonstrated very good sensitivity (84.1% [78.1, 91.4] and 98.2% [96.3, 99.8]) and specificity (99.7% [98.8, 100] and 74.0% [66.9, 79.0]), but were indeterminate in 820 (63.9%) and 776 (60.4%) patients, respectively. When indeterminate patients according to the HFA-PEFF score or H2FPEF score were assessed by the AI model, 610 (74.4%) and 571 (73.6%) of patients were correctly reclassified (respectively; **Figure 2**). In the testing dataset, modelling patient management decisions (e.g., prescription of SGLT2i) based on the combined diagnostic capacity of the AI HFpEF model and the HFA-PEFF or H2FPEF score, compared with the clinical score alone, resulted in more true positives being identified per 100 in the target population (**Figure 3**).

SUMMARY / CONCLUSION

A novel AI model demonstrated excellent discrimination between patients with HFpEF compared to patients with common risk factors, but no clinical diagnosis. Comparing the AI model and current clinical algorithms supports a use-case for in a screening paradigm or to support uncertain diagnoses.

DISCLOSURES

REFERENCES

1. Pieske et al., 2019. https://doi.org/10.1093/eurheartj/ehz641;
2. 2. Reddy et al., 2018, . https://doi.org/10.1161/CIRCULATIONAHA.118.034646

A novel AI model to detect HFpEF provided **more diagnostic outputs** than clinically validated algorithms.

Patients with **indeterminate** outputs by clinical algorithms were often **correctly re-classified** by the AI model.

Use of such a model in a **screening** paradigm or to **support uncertain diagnoses** could facilitate correct patient management.

DEMOGRAPHICS

Table 1. Demographics for cases and controls in training, validation, and testing

Patient Characteristics	Controls			Cases		
	Training	Validation	Testing	Training	Validation	Testing
Age, years	55.8 (15.7)	57.5 (15.8)	64.6 (17.4)	73.2 (11.5)	73.7 (11.5)	72.4 (13.3)
Women	1632 (52.2)	344 (52.4)	326 (51.1)	1277 (50.5)	237 (53.7)	337 (52.2)
Obesity	1943 (62.1)	427 (65.1)	451 (70.7)	1924 (76.1)	332 (75.3)	504 (78.0)
Hypertension	1120 (35.8)	272 (41.5)	301 (47.2)	1994 (78.8)	349 (79.1)	559 (86.5)
Diabetes mellitus	364 (11.6)	92 (14.0)	107 (16.8)	1003 (39.6)	171 (38.8)	326 (50.5)
LV mass index, g/m2	84 (17)	83 (17)	85 (21)	112 (32)	111 (29)	121 (36)
LA volume index, mL/m2	28.0 (7.5)	27.5 (7.2)	28.8 (35.4)	43.5 (12.0)	43.9 (12.1)	44.8 (15.2)
Ejection fraction, %	63 (5)	63 (5)	63 (6)	62 (6)	62 (6)	61 (6)
E velocity, cm/s	73 (16)	73 (16)	73 (17.0)	97 (23)	99 (24)	102 (26)
E: A, ratio	1.24 (0.46)	1.22 (0.47)	1.08 (0.4)	1.45 (0.84)	1.46 (0.75)	1.52 (0.87)
Average E/e', ratio	7.6 (1.9)	7.8 (2.1)	9.4 (2.7)	17.7 (6.3)	18.0 (6.3)	19.9 (6.8)
TR velocity, m/sec	2.4 (0.3)	2.4 (0.3)	2.4 (0.3)	2.9 (0.5)	2.9 (0.5)	3.0 (0.5)

DISCRIMINATION

Figure 1. Receiver operating characteristic curves for training, validation, and testing datasets
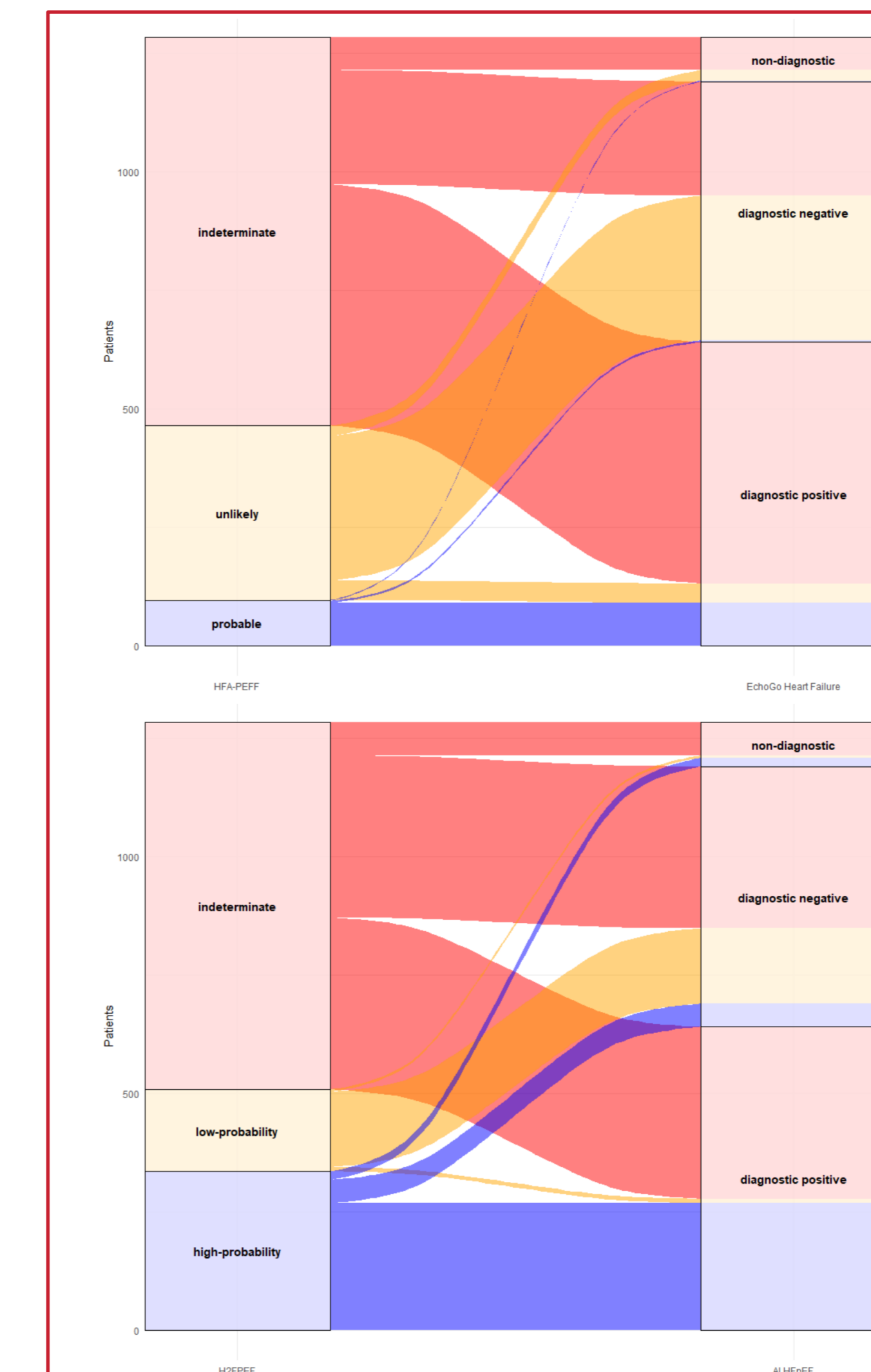
RECLASSIFICATION

Figure 2. Alluvial plots demonstrating classification of patients according to HFA-PEFF (top panel) and H2FPEF (bottom panel) compared with the AI model

DECISION MAKING
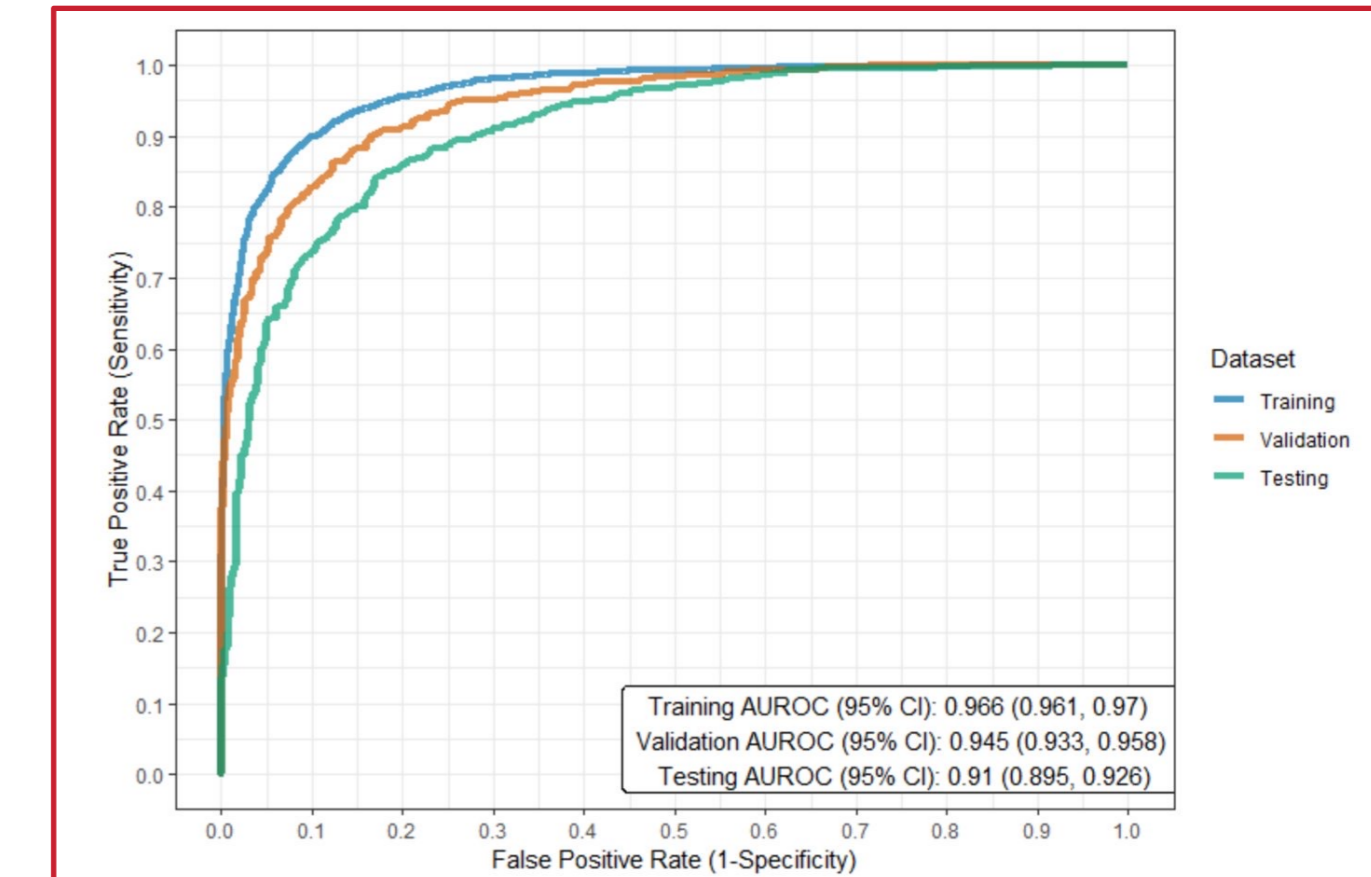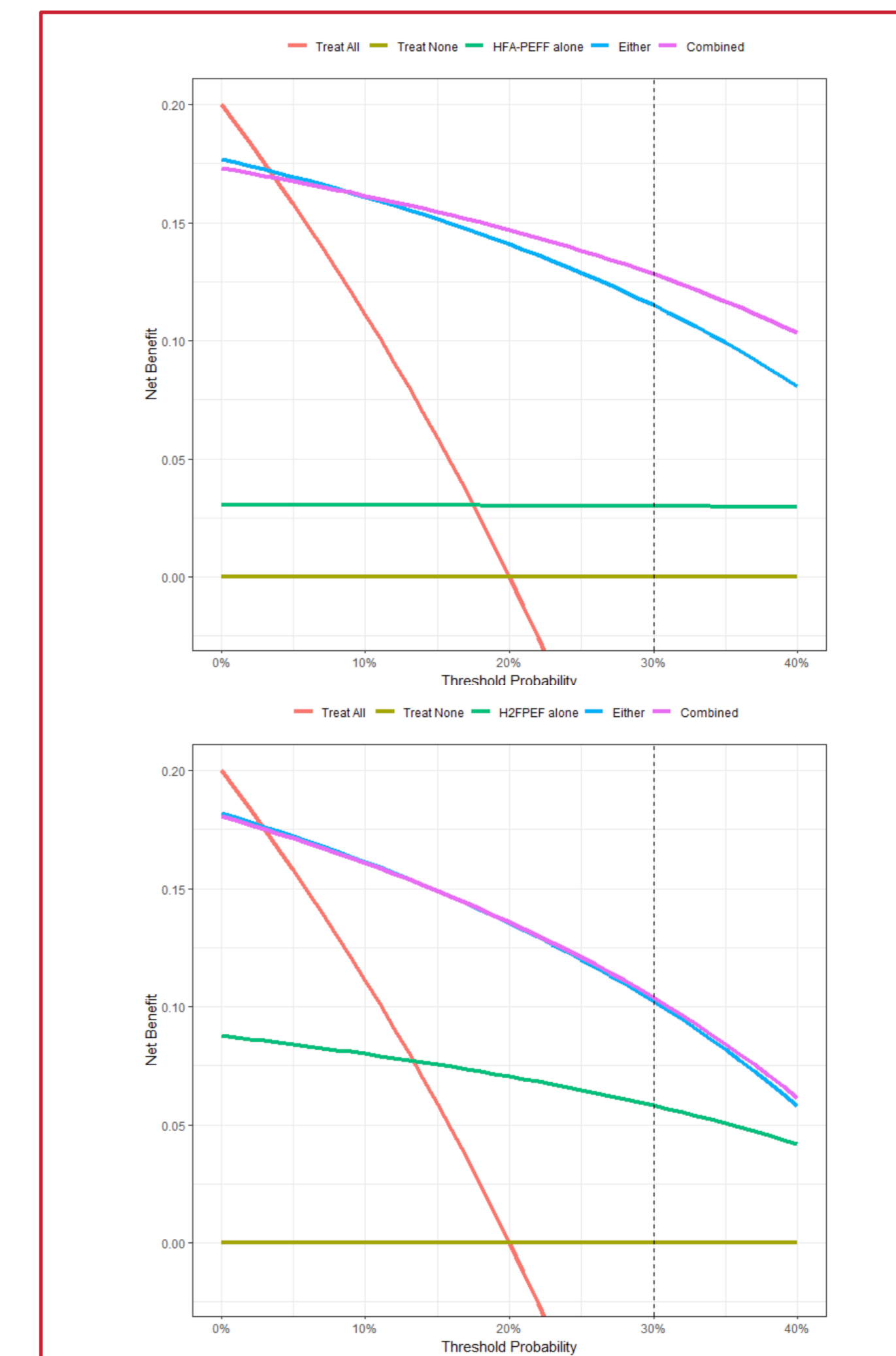

Figure 3. Decision curve analysis of patients management according to the use of HFA-PEFF (top panel) and H2FPEF (bottom panel) compared with the AI model