AI-based Echocardiographic Assessment Is Associated With HF Hospitalization and Cardiac Mortality in HFpEF

C. CASSIANNI [1], G. D. HUNTLEY [1], M. CASTRICHINI [1], A. P. AKERMAN [2], M. PORUMB [2], C. G. SCOTT [1], H.N. DAVISON [1], W. HAWKES [2], G. WOODWARD [2], B. BORLAUG [1], R. UPTON [2], P. A. PELLIKKA [1]

1 Mayo Clinic, MN, United States of America; 2 Ultromics Ltd, United Kingdom

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. Advances in development and implementation of artificial intelligence (AI) in the HFpEF diagnostic pathway show promise but require progressive development and validation to ensure clinical utility and meaningful impact. We therefore examined the association between version 2 of an FDA-approved HFpEF diagnostic aid with patient outcomes.

MATERIAL & METHODS

A three-dimensional convolutional neural network was developed using retrospective, multi-site, and multi-national cohort data (Mayo Clinic, USA, and NHS, UK) to automatically detect HFpEF (EchoGo Heart Failure; Ultromics Ltd [1]). HFpEF cases represented patients with EF ≥50%, evidence of increased intra-cardiac filling pressure, and a diagnosis of heart failure (ICD-9/10) within one year of the echocardiogram. Controls represented patients with EF ≥50%, but no evidence of increased intra-cardiac filling pressure or diagnosis of HF. Version 2 of the AI model provided a continuous probability of HFpEF to support the existing binary classification for high and low likelihood of HFpEF, and an uncertain output as risk mitigation. Version 2 also implemented more rigorous augmentation of the model in pre-processing based on real-world experience of the model in clinical implementation. Model performance and association with outcomes was examined in a multi-site retrospective dataset[1], consisting of 646 patients with HFpEF and 638 patients without HFpEF. Incident HF hospitalization was obtained from electronic health record chart review. Mortality was obtained from the National Death Index, and causes of cardiac deaths were manually reviewed. Cardiac mortality and HF hospitalization were plotted accounting for death as a competing risk, and Fine and Gray method was used to estimate the hazard ratios (HRs) adjusted for differences in age and sex between groups. Integration of AI model and existing clinical scores (H2FPEF[2]) are examined.

RESULTS

Table 1 presents model performance between version 1 and version 2 in a previously described independent test population (**Table 2**). Version 2 of the model produced 95 uncertain outputs, compared to 94 in version 1 (both 7.4%). Discrimination, classification, and calibration were all improved in the 2nd version of the model. Among the 1,284 patients followed for a median of 3.4 years (interquartile range, 1.7-6.5 years), there were 252 HF hospitalizations and 540 deaths. **Figure 1** demonstrates the risk for HF hospitalization according to AI model categorical output (top) and quartiles of continuous probability output (bottom). Positive AI output was associated with a higher risk for HF hospitalization than negative output (HR, 3.76; 95% CI, 2.71-5.21; P < .001) and likewise for uncertain output (HR, 2.79; 95% CI, 1.60-4.62; P < .001). Similarly, cardiac mortality (n=135) was higher in patients with positive output (HR, 5.55; 95% CI, 3.28-9.37; P < .001); patients with an uncertain output tended to have a higher mortality (HR, 2.22; 95% CI, 0.94-5.24; P = .07). Patients with higher continuous probability outputs demonstrated incrementally higher risk for cardiac mortality (fourth quartile vs first quartile: HR, 11.65; 95% CI, 4.65-29.20; P < .0001). **Figure 2** demonstrates the application of the AI model to patients with nondiagnostic H2FPEF outputs for stratification of risk of HF hospitalization. This sequential approach allowed the classification of all but 68 of the 776 patients (8.8%) and demonstrated similar associations with patient outcomes.

RESULTS

Table 1. Discrimination, classification, and calibration of version 1 and version 2 of the AI model

Version	AUROC	Sensitivity	Specificity	NPV	PPV	Intercept	Slope
v1 (2023)	0.91 (0.90-0.93)	87.8% (84.5-90.9%)	81.9% (78.2-85.6%)	86.5% (83.0-90.0%)	83.6% (80.2-87.0%)	-0.31 (-0.49- -0.13)	0.71 (0.63-0.78)
v2 (2024)	0.95 (0.94-0.96)	89.8% (87.5-92.5%)	86.3% (83.6-89.7%)	89.0% (87.0-91.7%)	87.2% (84.7-89.8%)	-0.31 (-0.49- -0.14)	1.03 (0.92-1.14)

Assessing only a **single echocardiographic video**, an AI model demonstrated **high diagnostic performance**.

Further development of the model based on **learnings from real-world clinical implementation** improved performance and generalizability.
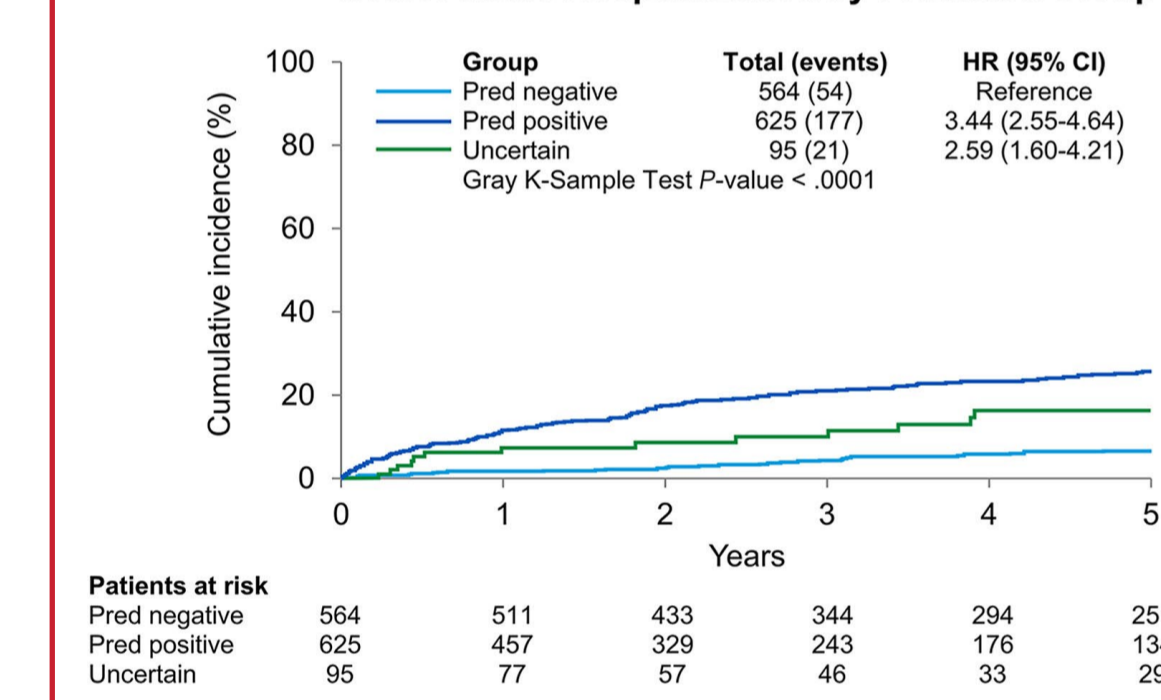
Categorical and continuous AI model outputs were associated **with incremental risk of HF hospitalization and cardiac mortality** and clinical utility when integrated into the existing diagnostic pathway.
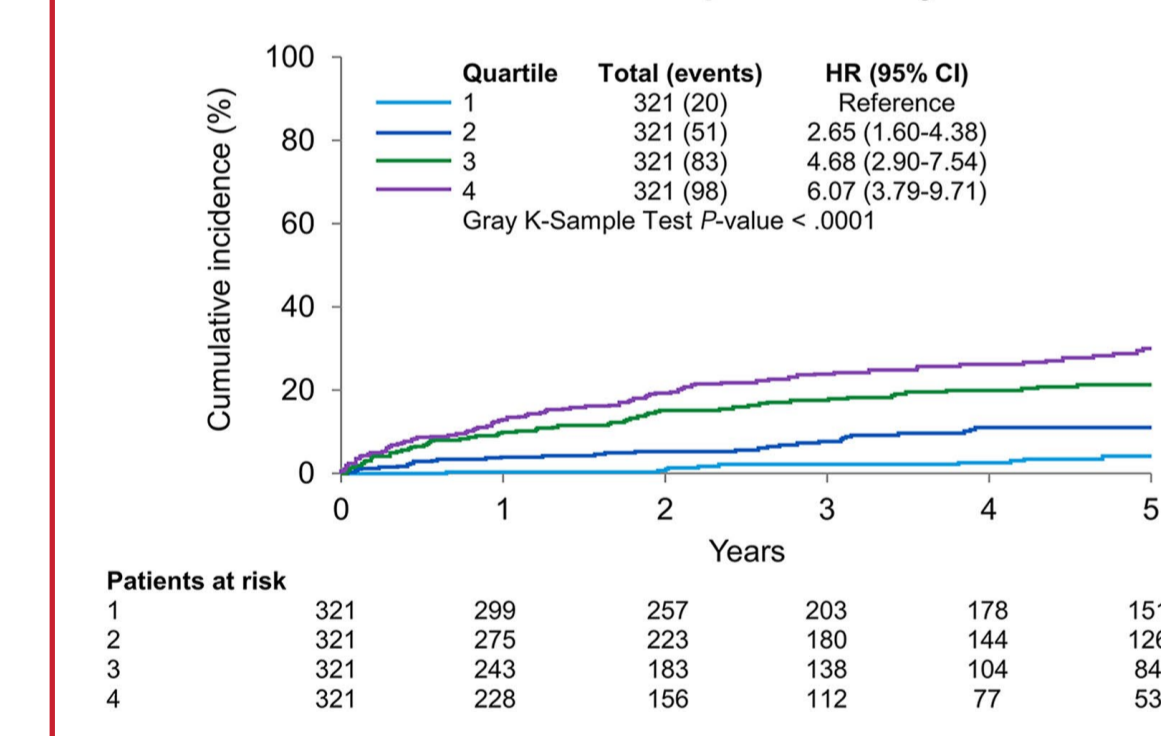

Figure 1. Association of AI model binary (A) and continuous (B) outputs with HF hospitalization

Table 2. Patient demographics for development 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)	324 (73.5)	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)
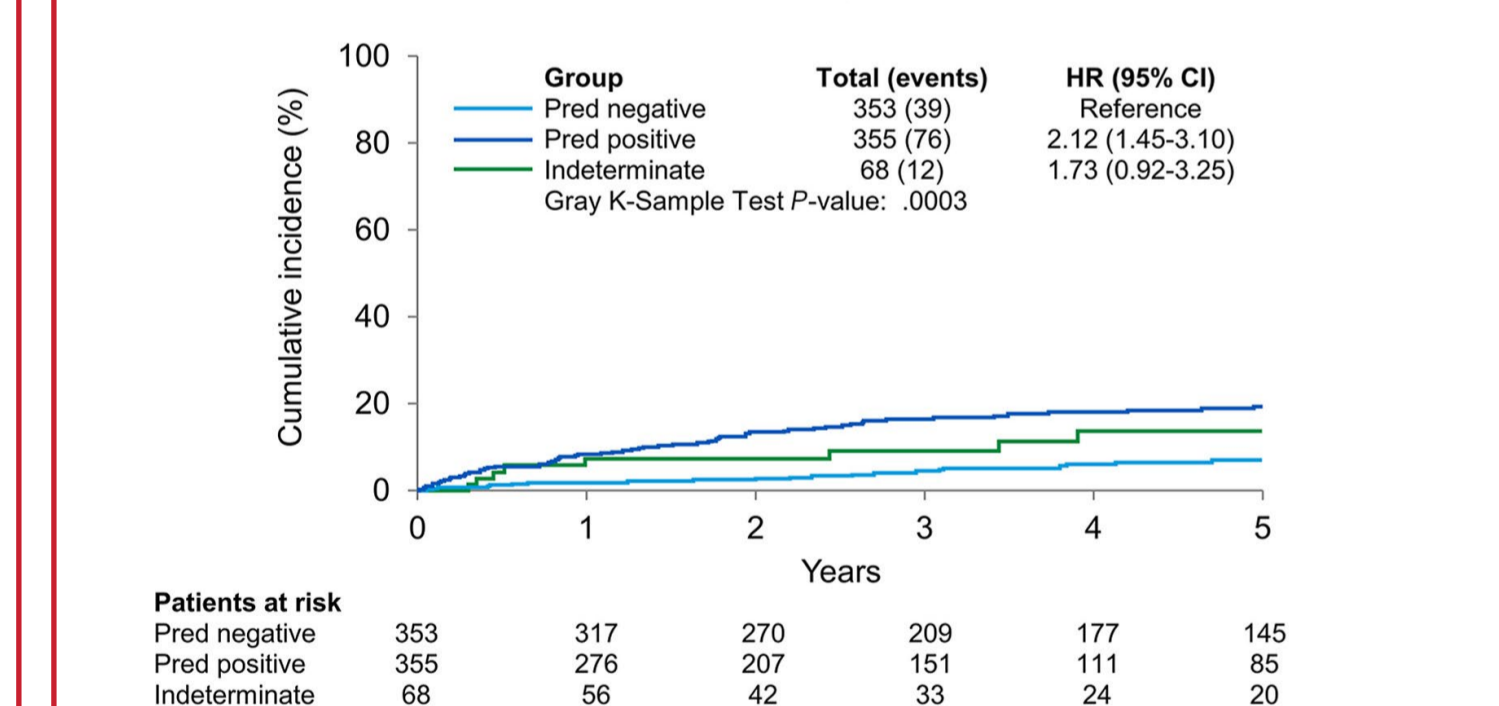
Figure 2. Integration of AI model into existing clinical score diagnostic pathway

SUMMARY / CONCLUSION

The AI model was associated with higher risk of HF hospitalization and cardiac mortality, with incremental risk according to categorical and continuous probability outputs. Integration of the AI model into the diagnostic pathway permitted reclassification of previously indeterminate patient risk, and identification of patients at higher risk of hospitalization.

REFERENCES

1. Akerman et al., 2023, https://doi/10.1016/j.jacadv.2023.100452

2. Reddy et al., 2018, . https://doi.org/10.1161/CIRCULATIONAHA.118.034646