Computer Methods and Programs in Biomedicine 248 (2024) 108107

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Circadian assessment of heart failure using explainable deep learning and novel multi-parameter polar images

Mohanad Alkhodari [a,b,*], Ahsan H. Khandoker [a], Herbert F. Jelinek [a,c], Angelos Karlas [d,e,f,g], Stergios Soulaidopoulos [h], Petros Arsenos [h], Ioannis Doundoulakis [h], Konstantinos A. Gatzoulis [h], Konstantinos Tsioufis [h], Leontios J. Hadjileontiadis [a,i,*]

[a] Healthcare Engineering Innovation Center (HEIC), Department of Biomedical Engineering and Biotechnology, Khalifa University, Abu Dhabi, United Arab Emirates
[b] Cardiovascular Clinical Research Facility, Radcliffe Department of Medicine, University of Oxford, Oxford, UK
[c] Biotechnology Center (BTC), Department of Biomedical Engineering, Khalifa University, Abu Dhabi, United Arab Emirates
[d] Chair of Biological Imaging, Central Institute for Translational Cancer Research (TranslaTUM), Technical University of Munich, Munich, Germany
[e] Helmholtz Zentrum München, Institute of Biological and Medical Imaging, Neuherberg, Germany
[f] Clinic for Vascular and Endovascular Surgery, Technical University of Munich, Klinikum rechts der Isar, Munich, Germany
[g] DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany
[h] First Cardiology Department, 'Hippokration' General Hospital, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece
[i] Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

ABSTRACT

Background and objective: Heart failure (HF) is a multi-faceted and life-threatening syndrome that affects more than 64.3 million people worldwide. Current gold-standard screening technique, echocardiography, neglects cardiovascular information regulated by the circadian rhythm and does not incorporate knowledge from patient profiles. In this study, we propose a novel multi-parameter approach to assess heart failure using heart rate variability (HRV) and patient clinical information.
Methods: In this approach, features from 24-hour HRV and clinical information were combined as a single polar image and fed to a 2D deep learning model to infer the HF condition. The edges of the polar image correspond to the timely variation of different features, each of which carries information on the function of the heart, and internal illustrates color-coded patient clinical information.
Results: Under a leave-one-subject-out cross-validation scheme and using 7,575 polar images from a multi-center cohort (American and Greek) of 303 coronary artery disease patients (median age: 58 years [50–65], median body mass index (BMI): 27.28 kg/m^2 [24.91–29.41]), the model yielded mean values for the area under the receiver operating characteristics curve (AUC), sensitivity, specificity, normalized Matthews correlation coefficient (NMCC), and accuracy of 0.883, 90.68%, 95.19%, 0.93, and 92.62%, respectively. Moreover, interpretation of the model showed proper attention to key hourly intervals and clinical information for each HF stage.
Conclusions: The proposed approach could be a powerful early HF screening tool and a supplemental circadian enhancement to echocardiography which sets the basis for next-generation personalized healthcare.

1. Introduction

Heart failure (HF) is a chronic pathological state that prevents the heart from pumping regularly to meet the body's need for oxygenated blood [1]. It is mostly caused by the presence of coronary artery disease (CAD), which is characterized by an accumulation of plagues in the arteries feeding the heart, leading them to become narrow or blocked [2]. Globally, 64.3 million people are living with HF with an estimated 7.2 million deaths every year [3,4]. HF patients suffer from a significant deterioration in the systolic function that is evaluated based on the left ventricular ejection fraction (LVEF), which is the amount of blood pumped at each contraction of the left ventricle [5]. HF stages based on LVEF are variable and even though several guidelines have set certain thresholds to classify them, *i.e.*, European Society of Cardiology (ESC) [6], there are still no strict rules to decide due to the etiology of HF, treatment procedures, and overall clinical presentation of patients [7].

The most preferred tool for screening LVEF-based HF is echocardiography [8]. Although reliable, it requires expensive equipment, which decreases its availability in public healthcare sectors, especially in less

* Corresponding author at: Healthcare Engineering Innovation Center (HEIC), Department of Biomedical Engineering and Biotechnology, Khalifa University, Abu Dhabi, United Arab Emirates.
E-mail addresses: mohanad.alkhodari@ku.ac.ae (M. Alkhodari), leontios.hadjileontiadis@ku.ac.ae (L.J. Hadjileontiadis).

https://doi.org/10.1016/j.cmpb.2024.108107
Received 4 July 2023; Received in revised form 28 February 2024; Accepted 29 February 2024
Available online 6 March 2024

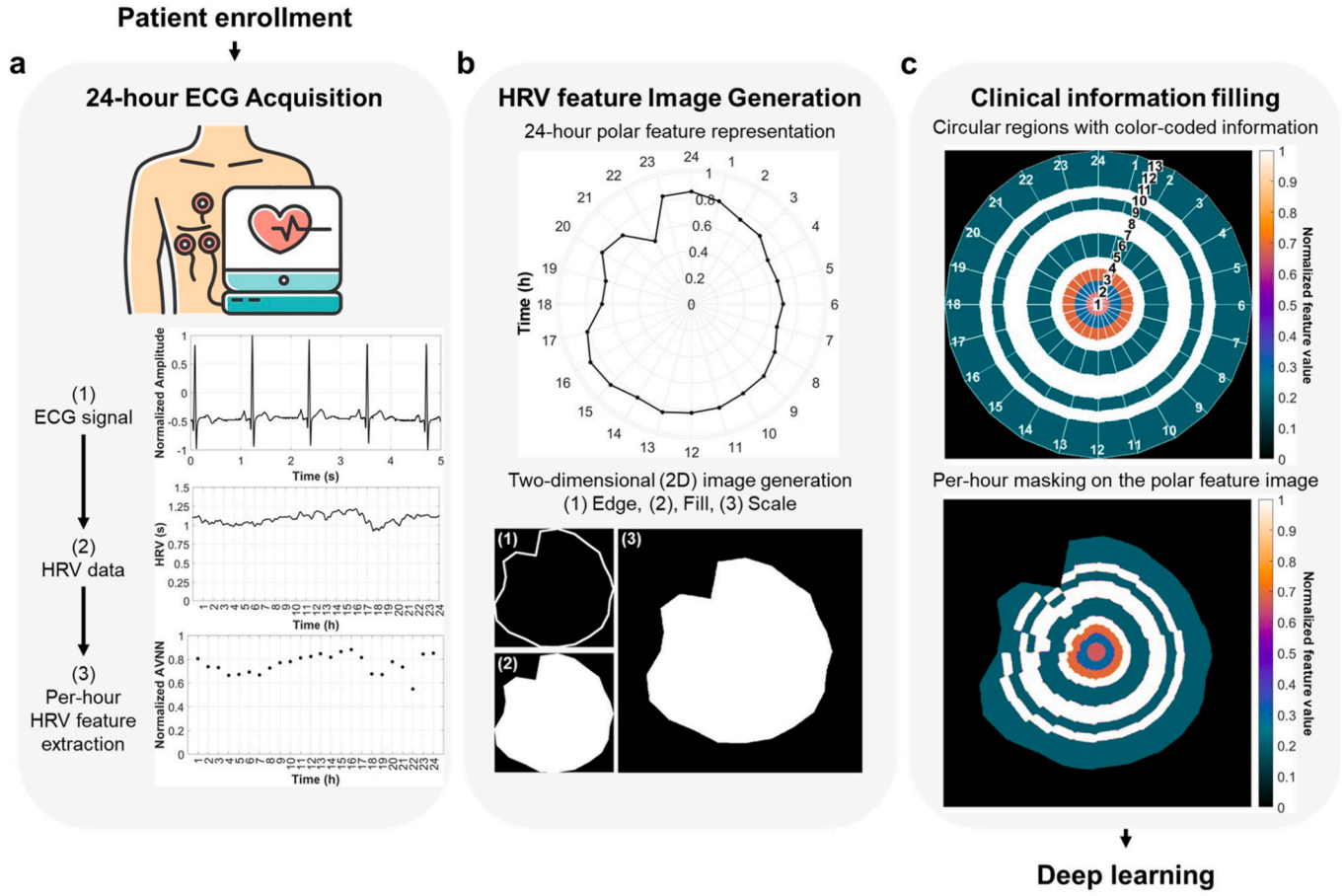


Fig. 1. Overview of the study including the generation of the filled-in heart rate variability (HRV) polar image. a, The approach starts by taking raw 24-hour electrocardiography (ECG) signal using a Holter device from a heart failure (HF) patient (showing only a 5-second segment as an example). The approach proceeds by extracting the corresponding HRV data and representing them in the form of extracted features on an hourly basis (showing the average normal-to-normal (AVNN) feature in (3). **b,** The second step is to depict these per-hour feature values on a polar map plot and then transform it into a two-dimensional (2D) binary image. The algorithm transforms the plot from polar coordinates to pixels to create (1) edges, (2) a filled image, and (3) a scaled region relative to the original plot dimensions. **c,** The last step is to fill in patient profile information within the generated 2D binary image of the HRV feature. The algorithm uses pre-defined color coding rules for the 13 patient information (arranged from 1 to 13, see Fig. 2c). The final filled-in HRV polar image is used as an input to the deep learning framework for the prediction of HF.

developed countries. In addition, it cannot provide cardiovascular information regulated by the circadian rhythm, as it is usually done for a short period [9], or account for any personalized information from the patient's profile. Developing other, more holistic indicators is thus an essential clinical aim. One such option is electrocardiography (ECG) and its corresponding heart rate variability (HRV) that is usually associated with the endocrine, autonomic nervous system (ANS), and intrinsic modulation of the cardiac electrophysiological rhythm [10]. Due to the presence of CAD in HF patients, the autonomic regulatory balance gets interrupted, and such behavior has been usually observed in literature through HRV analysis [11,12]. However, a deep understanding of the relationship between HRV and HF is still not well defined. In addition, the conventional diagnostic procedures of HF are highly dependent on medical experts, which poses difficulties, especially in the presence of big patient data in the form of images, signals, or clinical profiles.

The use of deep learning may be a promising approach to resolving the heavy dependency on medical expertise. Recent advances in deep learning have facilitated the growth of computerized algorithms in diagnosing HF. There have been many efforts to develop trained deep learning tools for detecting HF in ECG signals [13–15]. The use of ECG signals may be highly affected by the quality of the recordings, and training models on long ECG signals may require high computational demands. Others have tried simplifying ECG signals into corresponding short-term HRV data, a short sequence of consecutive R-peaks distances [16–18]. Although short-term HRV is less complex, it does not include additional knowledge about cardiac variations throughout the circadian rhythm of the heart. Several works have also reported the use of deep learning in HF diagnostics using patient profiles [19–21]. However, much demographic and clinical patient information could be highly overlapped among HF stages, especially when a narrower threshold is used to determine each stage.

We propose a novel approach that simplifies long-term 24-hour HRV data (in the form of extracted features) and combines it with patient profile information in a single two-dimensional (2D) image (Fig. 1). The generated image is a polar image of the patient's HRV data with edges corresponding to per-hour feature variations and the interior of filled-in color-coded demographic and clinical information. This novel multi-parameter polar image presents, in a holistic way, an overview 24 hours of the patient's cardiovascular status, considering personalized information from his/her clinical profile as well [22]. These polar images are inputted to a deep learning model to infer for the HF stage, *i.e.*, preserved LVEF, mid-range LVEF, and reduced LVEF. We validated the reliability of the model on predicting HF stages and discovering circadian/medical characteristics of patients at each stage using 7,575 polar images from a cohort of 303 coronary artery disease patients, achieving mean accuracy greater than 92%.

Table 1
Demographic and clinical information of patients included in this study (American and Greek cohorts).

Variables	Total patients (n = 303)	p-value (datasets)	HF stages						p-value (stages)
			HFpEF (n = 129)	p-value (datasets)	HFmEF (n = 92)	p-value (datasets)	HFrEF (n = 82)	p-value (datasets)	
LVEF, %	55 (46.5-63)	**<0.001**	63 (60-70)	**<0.001**	52.5 (50-55)	**<0.001**	45 (40-47)	0.099	**<0.001**
Cardiovascular risk factors									
Age, yrs	58 (50-65)	0.929	57 (38-64.5)	0.300	58.5 (52-68)	0.277	60.5 (50-66)	0.282	0.110
Male	258 (85.15)	0.566	108 (83.72)	0.189	76 (82.61)	1.000	74 (90.24)	1.000	0.237
BMI, kg/m^2	27.28 (24.91-29.41)	**<0.001**	27.12 (24.39-28.95)	0.121	27.22 (25.35-29.92)	**0.045**	27.68 (25.31-29.74)	0.197	0.133
Smoker	203 (67.00)	0.146	87 (67.44)	0.328	62 (67.39)	0.225	54 (65.85)	0.010	0.757
Diabetic	43 (14.19)	**0.002**	10 (7.75)	0.083	13 (14.13)	0.785	20 (24.39)	0.009	**<0.001**
Hypertensive	154 (50.83)	0.885	64 (49.61)	**0.001**	46 (50.00)	1.000	44 (53.66)	0.070	0.587
Cardiovascular history									
AP	186 (61.39)	**<0.001**	89 (68.99)	**<0.001**	46 (50.00)	**<0.001**	51 (62.20)	0.292	0.189
VT	21 (6.93)	0.408	8 (6.20)	0.162	2 (2.17)	0.160	11 (13.42)	0.744	0.088
Prior MI	223 (73.60)	**<0.001**	77 (59.69)	**<0.001**	76 (82.61)	**<0.001**	70 (85.37)	0.023	**<0.001**
Cardiovascular medication									
Beta-blockers	245 (80.86)	**0.005**	101 (78.30)	0.162	77 (83.70)	**0.002**	67 (81.71)	0.201	0.478
ACE-inhibitors	113 (37.29)	**<0.001**	47 (36.43)	**<0.001**	33 (35.87)	**0.002**	33 (40.24)	0.037	0.611
Anti-arrhythmics	12 (3.96)	0.083	3 (2.33)	0.173	4 (4.35)	**0.044**	5 (6.10)	0.044	0.166
diuretics	114 (37.62)	**<0.001**	24 (18.61)	**<0.001**	51 (55.44)	**<0.001**	39 (47.56)	<0.001	**<0.001**

All values are represented as median (inter-quartile range) or n (%). Bold p-values show statistically significant differences ($p < 0.050$) amongst the three stages using the linear regression fitting testing. HFpEF = Heart failure with preserved ejection fraction; HFmEF = Heart failure with mid-range ejection fraction; HFrEF = Heart failure with reduced ejection fraction; LVEF = Left ventricular ejection fraction; BMI = Body mass index; AP = Angina pectoris; VT = Ventricular tachycardia; MI = Myocardial infarction; ACE = Angiotensin-converting enzyme.

2. Methods

2.1. Study design and dataset

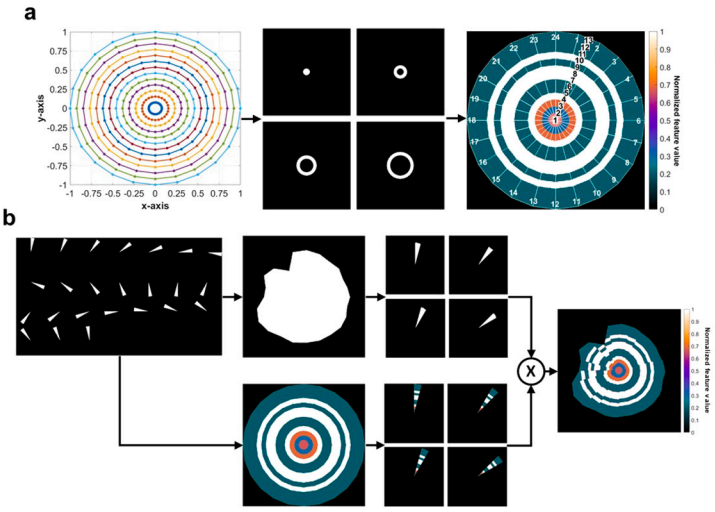
In this multivariable discovery and modelling study, we trained and tested a deep learning model for the prediction of HF stages from multiple clinical data sources, *i.e.*, ECG, HRV, and patient medical records, which were obtained from 303 coronary artery disease (CAD) patients suffering from HF (Table 1). These patients were classified, based on their corresponding LVEF levels, as HF with preserved LVEF (HFpEF, LVEF > 55%), HF with mid-range LVEF (HFmEF, 55% ≥ LVEF ≥ 50%), and HF with reduced LVEF (HFrEF, LVEF < 50%) with a narrower range than regular guidelines for the HFmEF stage, following the recommendations of the American Society of Echocardiography and the European Association of Cardiovascular Imaging (ASE/EACVI) [23–25]. Thus, the dataset consisted of 129 patients in the HFpEF stage, 92 in HFmEF, and 82 in HFrEF. In addition, the dataset used in this study consisted of two patient cohorts; one from an American database and the other from a Greek database.

The American patient cohort was obtained from the archives of the Intercity Digital Electrocardiography Alliance (IDEAL) study of the University of Rochester Medical Center Telemetric and Holter ECG Warehouse (THEW) [26,27]. The enrolment protocol was conducted following the Declaration of Helsinki and in accordance to Title 45, U.S. Code of Federal Regulations, Part 46, protection of human subjects (revised: November 13, 2001 – Effective: December 13, 2001). The study was further approved by the research subject review board at the University of Rochester and all enrolled subjects provided a signed consent before participating in the study. Only subjects with evidence of previous myocardial infarction, exercise-induced ischemia, and stable phase of ischemic heart disease at least two months after their last event were allowed to enrol. In addition, the eligibility criteria included not being diagnosed with a congenital HF and being under a stable sinus rhythm. On the other hand, the exclusion criteria included having dilated cardiomyopathy with a left ventricular diameter (LVD) > 60 mm and LVEF < 40%), coronary artery bypass grafting (CABG) surgery, non-sinus rhythm, and any cerebral, severe hepatic, or malignancy diseases. From the IDEAL study, a total of 199 patients were obtained and divided into 106 HFpEF, 46 HFmEF, and 47 HFrEF (Supplementary Material).

On the other hand, the Greek patient cohort was obtained from the PRESERVE EF study [28]. Patients were enrolled in seven actively participating cardiology departments in Greece. The ethics committee at each department approved the protocol in the study endorsed by the Hellenic Society of Cardiology, Greece [29]. All patients provided consent before participating in the study at each cardiology department. To be eligible for enrollment in the study, patients should have a post-angiographically proven myocardial infarction (at least 40 days after the event or 90 days after any CABG surgeries). In addition, they must be revascularized, or if not revascularized, they should be without evidence of any active ischemia (within 6 months). All patients should be under optimal and tolerate medical therapy. The exclusion criteria included having a secondary prevention indication for implantable cardioverter defibrillator (ICD) implantation, permanent pacemaker, persistent, long-standing persistent, and permanent atrial fibrillation, and any neurological symptoms of syncope or pre-syncope (within 6 months). In addition, patients with systemic illnesses, such as liver failure, renal diseases, rheumatic diseases, thyroid dysfunction, and cancer, were excluded. From the PRESERVE EF study, 104 patients were obtained and divided as 23 HFpEF, 46 HFmEF, and 35 HFrEF (Supplementary Material).

2.2. Patient profiles and statistical analysis

Patient profiles were obtained from two different patient cohorts. Therefore, some information found in one patient cohort was missing in the other. To address this, we have created a set of common demographic and clinical information between the two datasets and used it for further analysis (Table 1). A total of 13 patient information items were shared between the two cohorts that were grouped as cardiovascular risk factors: age (years), gender (male/female), BMI (kg/m^2), smoking (yes/no), diabetes (yes/no), and hypertension (yes/no), cardiovascular history: angina pectoris, ventricular tachycardia, and prior myocardial infarction, and cardiovascular medication: beta-blockers, ACE-inhibitors, antiarrhythmics, and diuretics. The additional infor-



c

Circle label	Demographic/clinical information	Color codes	Additional normalization notes
1	Age, yrs	Blue: < 35 Purple: 35-65 Orange: 65-100 White: > 100	Patient's age is divided by 100 to normalize as [0 - 1] Patients older than 100 are fixed to 1
2	Sex	Blue: Male Pink: Female	Male patients are set to 0.3 Female patients are set to 0.6
3	BMI, kg/m^2	Blue: < 25 Purple: 25-30 Orange: 30-35 White: > 35	Normalize using a sigmoid function with: Mean = 27.28 Standard deviation = 3.45
4	Smoking	Blue: No White: Yes	No is set to 0.2 Yes is set to 1
5	Diabetes		
6	Hypertension		
7	AP		
8	VT		
9	Prior MI		
10	Beta-blockers		
11	ACE-inhibitors		
12	Anti-arrhythmics		
13	diuretics		

BMI = Body mass index; AP = Angina pectoris; VT = Ventricular tachycardia; MI = Myocardial infarction; ACE = Angiotensin-converting enzyme

Fig. 2. Methodology followed to combine the heart rate variability (HRV) feature image with patient information to generate a color-coded image. a, Generating circular rings of different diameters (maximum: 2) and converting them to binarized images to be able to color coding each with clinical information of patients. **b,** The complete process of filling in the per-hour (pie-shaped segment) in the 24-hour polar image with clinical information. The featured image is multiplied by the pie-shaped circular rings after size adjustments to form the final image. **c,** Color-coding rules used to fill the final image in the order suggested by cardiologists.

mation for each patient cohort dataset is provided in Supplementary Material.

We performed a Student's t-test based on linear regression fitting [30] to analyze the significance of patient information in discriminating between the three LVEF-based HF stages. We also checked this significance between the two patient cohorts for the common information only. A significant difference was obtained whenever the *p*-value reached less than 0.05.

2.3. Color-coded patient information

To convert patient information into more visual-friendly images, we have applied color coding rules for each demographic and clinical information (Fig. 2). All information was normalized according to these rules, in a way that forms unique color codes for each variable. For age (in years), we divided the value by 100 to get a value between 0 and 1. If the patient's age was more than 100, we set it to 1. Accordingly, the color images for age can be described as blue < 35, purple 35-65, orange 65-100, and white > 100. For sex, we fixed male patients to 0.3 (blue) and female patients to 0.6 (pink) to be easily discriminated against through visual inspection. For BMI (in kg/m^2), we normalized all values using a regular sigmoid function with the mean (27.28) and standard deviation (3.45) of the current dataset. Thus, BMI was represented in colors as blue < 25, purple 25-30, orange 30-35, and white > 35. For categorical variables (yes/no answers), including smoking, diabetes, hypertension, angina pectoris, ventricular tachycardia, prior myocardial infarction, beta-blockers, ACE-inhibitors, antiarrhythmics, and diuretics, a value of 0.2 was given to the no answer (blue) and a value of 1 was given to the yes answer (white).

We compensate for color-blind people using suitable palettes with their vision, including protanopia, deuteranopia, and tritanopia (Supplementary Material); the color codes for every patient information are provided in the figure alongside the original color palette.

2.4. HRV data and features extraction

All patients in this study performed 24-hour Holter ECG recordings (Fig. 1a) using a three pseudo-orthogonal lead configuration. The corresponding HRV data (distance between R-peaks) for each patient's recordings were obtained through expert annotations or the Pan-Tompkins algorithm applied to the third ECG lead due to having a high amplitude with clear ECG waveform. We ensured that all recordings are fixed to start by hour 01:00 either through a pre-determined starting hour in the annotation file or through the Cosinor analysis fitting algorithm [31].

We extracted HRV features on hourly basis from time-domain [32]: average normal-to-normal (NN) interval (AVNN, ms), standard deviation of the NN intervals (SDNN, ms), square root of the mean of the sum of squares of differences between adjacent NN intervals (RMSSD, ms), percentage of NN intervals more than 50 ms (pNN50, %), and standard error of the average NN interval (SEM, ms), frequency-domain [32]: slope of the linear interpolation of the spectrum for frequencies less than the very-low frequency (VLF) band upper bound (BETA), high frequency normalized power (HFNorm, %), peak frequency of the high frequency band (HFPeak, Hz), power in the high frequency band (HFPower, ms^2), low frequency normalized power (LFNorm, %), peak frequency of the low frequency band (LFPeak, Hz), power in the low frequency band (LFPower, ms^2), ratio between the LFPower and the HFPower (LF/HF), total power in both frequency bands (Total Power, ms^2), VLF normalized power (VLFNorm, %), and power in the VLF band (VLFPower, ms^2), non-linear metrics [33,34]: standard deviation of the NN intervals along the perpendicular to the line-of-identity (SD1, ms), standard deviation of the NN intervals along the line-of-identity (SD2, ms), de-trended fluctuations analysis for the low-scale slope (alpha1), de-trended fluctuations analysis for the high-scale slope (alpha2), and complexity of physiological time-series signals (Sample Entropy), and fragmentation metrics [35]: percentage of inflection points in the NN intervals (PIP, %), acceleration and deceleration segments inverse average length (IALS), percentage of short segments (PSS, %), and percentage of alternation segments (PAS, %).

2.5. Filled-in HRV feature image generation

The generation of the HRV feature image consists of two components: the transformation of HRV features values to a 2D polar image and the filling-in with color coded patient information (Fig. 1b-c, Fig. 2 and Algorithm 1). Initially, each HRV feature value was normalized and drawn as a polar map representing a 24-hour clock. Each per-hour feature value was converted from polar to Cartesian (x, y) coordinates. Then, the Cartesian points were mapped to a 2D space with dimensions of 512x512 and connected to form the outline of the feature image. After filling the outline with a binary value of 1 (background with 0),

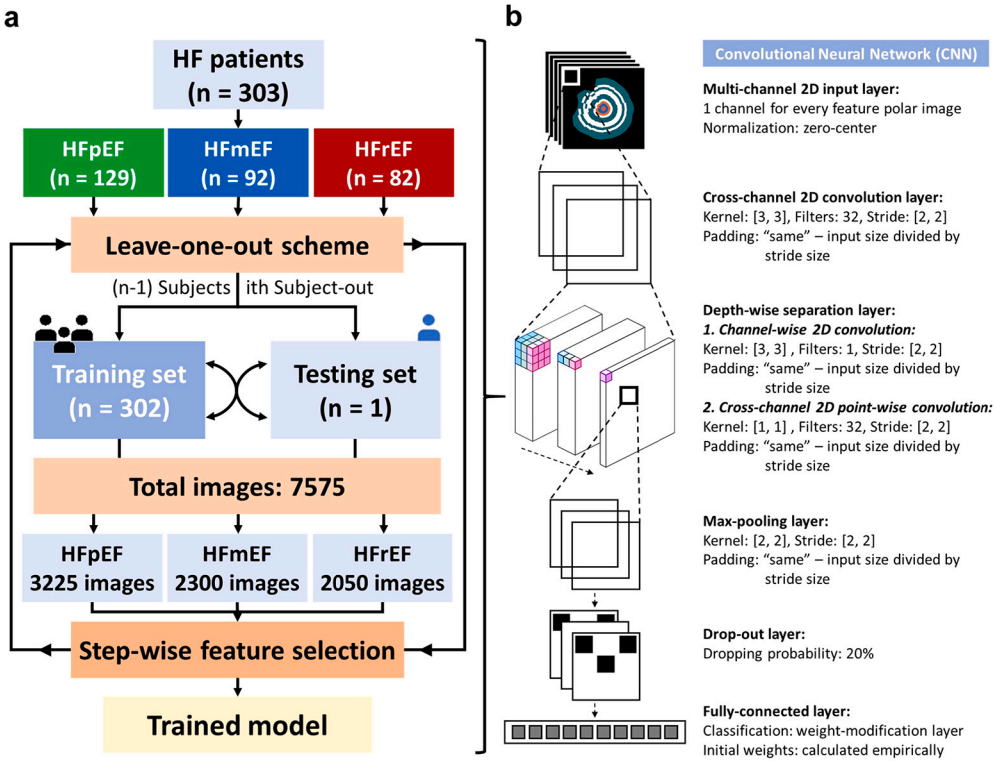


Fig. 3. Deep learning training mechanism and model's architecture. a, The training included a total of 303 patients and followed a leave-one-subject-out (LOSO) cross-validation scheme. Out of the overall 7575 HRV feature images, the training ensures a maximized performance through a step-wise feature selection approach. **b,** The architecture of the model was designed to allow it to learn from HRV features altogether as well as individually. The multi-channel two-dimensional (2D) input layer passes the best-selected images first to a cross-channel convolutional layer and then to a channel-wise convolutional layer. To reduce complexity, the network uses a max-pooling layer followed by a drop-out layer (20% probability) to prevent over-fitting. The last layer in the network (fully connected) includes a weight-modified layer with initial weights that were calculated empirically to handle the slight imbalance in the dataset.

Algorithm 1 The complete procedure used to generate color-coded HRV feature images.

for $feature = 1, 2, \ldots, 25$ **do**
  Calculate the angle (theta) for 24 points of 24 hours
  Define R as the 24 hourly values of the feature
  Convert polar dimensions (theta, R) to Cartesian coordinates (x, y)
  Calculate scaling factor as a mean division of x and y between the feature and a full circle
  Use Cartesian coordinates to create a binary edge image filled at each (x, y) point pairs
  Fill the space inside the edges to create the feature image (1's and 0's)
  Rescale the image based on the calculated scaling value of step 5
  Generate per-hour pie-shaped binarized segments using steps from 2 to 8 with R as the maximum value of the circle, *i.e.*, 1
  Generate 13 binarized circular rings each multiplied by one of the color-coded clinical information
  Multiply each pie-shaped segment by the full circular rings image
  Multiply each pie-shaped segment by the generated feature image in step 8
  Combine all resulting segments to form the color-coded feature image
end for

we re-scaled the generated image to its original dimensions relative to the polar map plot. The re-scaling factor (a value between 0 and 1) was calculated by measuring the difference between every point on the Cartesian coordinates and a circular base plot that represents the maximum possible value of every feature after normalization, *i.e.*, 1 in a scale from 0 to 1.

To fill in color-coded patient information, we generated 13 circular rings (corresponding to 13 patient information items) that extend from the center to the edges of the image. Each circular ring was filled with a color-coded demographic or clinical variable in a pre-defined order (described in the next section). For each hourly pie-shaped segment, the binary HRV feature segment and the circular rings segment were extracted. To be able of masking the hourly circular rings segment on the binary HRV feature segment, we scaled them to make them the same size. Then, they were multiplied to form the corresponding mask. To generate the final filled-in HRV feature image, we combined all hourly scaled segments and smoothed their edges to form a unique connected shape without sharp edges.

We decided on the optimum order of patient information circular rings within the HRV feature image through a questionnaire spread among eight cardiologists at Greek and United Arab Emirates hospitals and medical centers. In the questionnaire, we requested each cardiologist to order each patient information category (cardiovascular risk factors, history, and medication) as center, middle, and edges. The current order of categories (Fig. 1c and Fig. 2a, c) had a *perfect* Fleiss kappa [36] agreement of 0.88 (95% confidence interval: 0.84-0.91, p-value: < 0.001) among cardiologists.

2.6. Deep learning model training

We trained the deep learning convolutional neural network (CNN) model (Fig. 3a) using a leave-one-subject-out (LOSO) cross-validation scheme, where on each iteration, an ith subject was used for testing and the remaining $n-1$ subjects were used for training. The convolutional operation is defined as follows,

$$C_i^{lj} = h(b_j + \sum_{m=1}^{M} w_m^j x_{i+m-1}^j) \quad (1)$$

Table 2
Detailed description of the structure of the used convolutional neural network (CNN) in this study.

#	Layers	Kernel size	Input dimension	Output dimension	Additional Notes
1	2D Input	-	512x512x7	-	Zero-center normalization
2	2D cross-channel convolution	[3, 3]	512x512x7	256x256x32	Stride: [2, 2], Padding: 'same'
3	Batch normalization	-	256x256x32	256x256x32	-
4	ReLU	-	256x256x32	256x256x32	-
5	2D channel-wise convolution	[3, 3]	256x256x32	128x128x32	Stride: [2, 2], Padding: 'same'
6	Batch normalization	-	128x128x32	128x128x32	-
7	ReLU	-	128x128x32	128x128x32	-
8	2D point-wise cross-channel convolution	[1, 1]	128x128x32	64x64x32	Stride: [2, 2], Padding: 'same'
9	Batch normalization	-	64x64x32	64x64x32	-
10	ReLU	-	64x64x32	64x64x32	-
11	2D max-pool	[2, 2]	64x64x32	32x32x32	Stride: [2, 2], Padding: 'same'
12	Dropout	-	32x32x32	32x32x32	Percentage: 20%
13	Fully-connected	-	32x32x32	1x1x3	Weight-modification: initial weights calculated empirically
14	Soft-max	-	1x1x3	1x1x3	Probabilities for the three heart failure stages

where $x_i = [x_1, x_2, ..., x_n]$ is the input, n is the total number of points, l is the layer index, h is the activation function, b is the bias of the j_{th} feature map, M is the kernel size, w_m^j is the weight of the j^{th} feature map and m_{th} filter index.

Instead of training the network on the whole 25 color-coded HRV features (7575 HRV images) all at once, features that were of high discrimination ability in exhibiting differences between the three LVEF-based HF stages had to be determined. To achieve this, a step-wise feature selection approach was followed. In this approach, 25 iterations were used based on the total number of HRV features covered in the study. At the first iteration, each feature was used to train the model and output the corresponding classification accuracy. After picking the highest-performing feature, it was added to the best-features set that will be used in all the following iterations. On the second iteration, the process continues by adding HRV features one by one (step-by-step) to the best feature set and observing the classification performance accordingly. This iterative process continued until all features were ranked based on their impact on the overall classification process. In the end, the iteration that gave the highest accuracy was selected, and all features included in the best features set up to this iteration were selected as optimal features for the training and classification process.

The structure of the deep learning network (Fig. 3b and Table 2) was designed to start with a multi-channel 2D input layer that accepts a variable number of color-coded HRV features. Each feature was assigned to a channel and a zero-center normalization was applied accordingly. Then, the network applied an initial cross-channel 2D convolutions with a kernel size of [3, 3], 32 filters, and stride of [2, 2] with padded output similar to the input. Another depth-wise convolution was applied to the outputs of the first convolution to ensure a channel-wise 2D feature extraction mechanism. The layer had a single filter with a kernel size of [3, 3] and stride of [2, 2]. The padding was similar to the previous convolutional step. The depth-wise convolution proceeded with cross-channel 2D point-wise convolution with a kernel of [1, 1], 32 filters, and a stride of [2, 2]. After applying convolutions, a max-pooling mechanism was applied to reduce the dimensionality and complexity in the network with a kernel size of [2, 2] and stride of [2, 2]. To prevent the network from over-fitting, a 20% drop-out layer was added. For classification, a fully-connected layer was used with soft-max and weight-modified layers that calculated class weights empirically. All parameters used in the design of the deep learning network were optimized based on the iterative hyperparameter optimization approach [37].

3. Results

3.1. Patients characteristics analysis

The demographic and clinical variables recorded for patients' were categorized as cardiovascular risk factors, history, and medication. Most patients were males (85.15%) with a median age of 58. They had an interquartile range for the body mass index (BMI) of 24.91-29.41 kg/m^2, which shows a range for the overweight patient category. In addition, most of these patients had a history of angina pectoris (61.39%) and myocardial infractions (73.60%). A total of 245 patients were on beta-blockers medication, which accounts for 80.86% of the whole patient population.

We have further assessed the ability to distinguish between patients across the three LVEF-based HF stages, as well as the two patient cohorts (Table 1). A statistically significant difference was found among the three HF stages in diabetes status, having a prior myocardial infarction, and being on diuretics medication ($p < 0.001$). The same variables were found to be significantly different between the two patient cohorts in addition to BMI, having an angina pectoris condition, and being on beta-blockers and ACE-inhibitor medications. Additional analysis on each patient cohort separately is provided in Supplementary Material. It is worth noting that much information about patients was not common between these two cohorts.

3.2. HF stages classification performance

A selection procedure of the best color-coded HRV features, based on the maximization of the deep learning model accuracy, resulted in seven color-coded HRV features (Fig. 4a), namely RMSSD and SEM from the time-domain, HFNorm, HFPower, and LFNorm from the frequency domain, SD1 from non-linear metrics, and IALS from fragmentation metrics. Breaking down the performance, the model resulted in average sensitivity, specificity, and accuracy of 90.68%, 95.19%, and 92.62%, respectively, in predicting the three HF stages (Fig. 4b, c). In addition, the most precise predictions were achieved with the HFpEF stage with 93.02% followed by HFmEF (90.22%) and HFrEF (86.59%), with an F1-score of 91.95%, 88.3%, and 90.45% for HFpEF, HFmEF and HFrEF, respectively. In terms of the normalized Matthews correlation coefficient (NMCC), HFrEF had the highest percentage of 93.64%, followed by HFpEF and HFmEF with 92.94% and 91.54%, respectively.

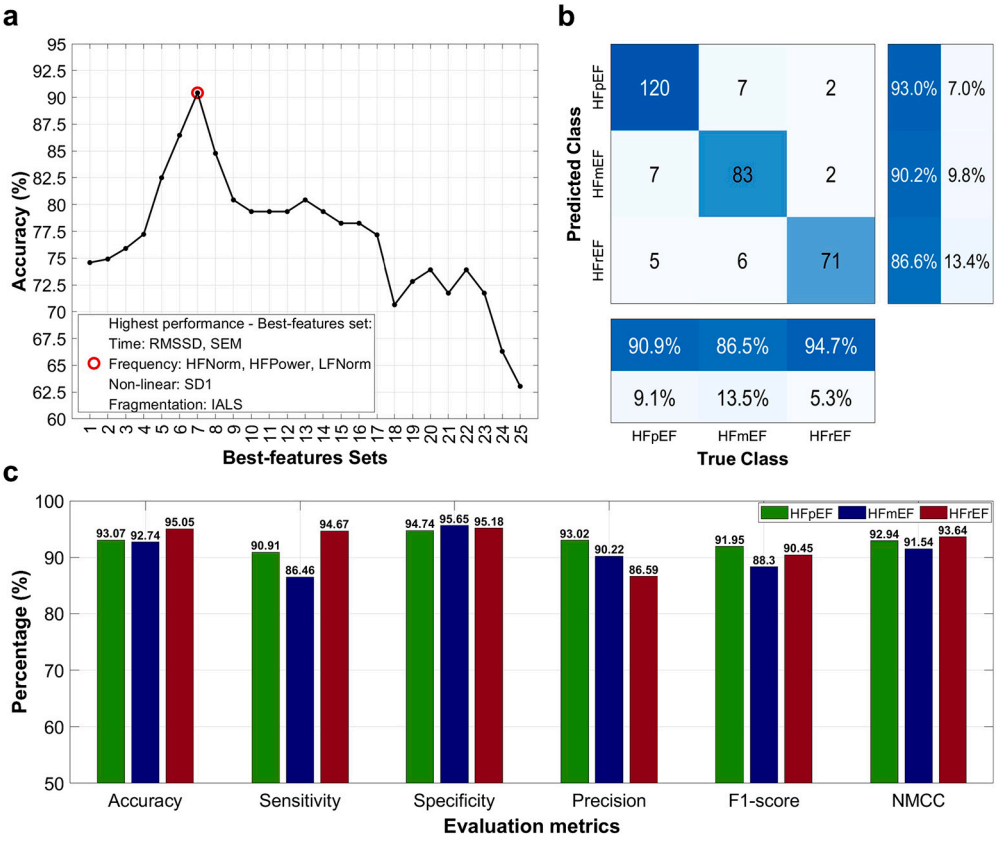


Fig. 4. Performance of the trained deep learning model with a leave-one-subject-out (LOSO) cross-validation. a, Accuracy variations using different combinations of features based on the step-wise feature selection. Best-features sets (25 sets for 25 features) were determined by having the maximum achievable accuracy (red circle) when adding features step-by-step to the combination set (see Methods for definitions of features). **b,** The confusion matrix for the predictions of heart failure with preserved ejection fraction (HFpEF), failure with mid-range ejection fraction (HFmEF), and failure with reduced ejection fraction (HFrEF) following the LOSO cross-validation scheme. **c,** The performance metrics are calculated from the confusion matrix for the accuracy, sensitivity, specificity, precision, F1-score, and normalized Matthews correlation coefficient (NMCC). The analysis of the receiver operating characteristic (ROC) and precision-recall (PR) curves is provided in Fig. 5.

We further analyzed the performance of the model through the receiver operating characteristics (ROC) and precision-recall (PR) curves (Fig. 5a, b). The highest area under the ROC curve (AUROC) and the PR curve (AUPR) was achieved in predicting the HFpEF stage (AUROC: 0.915, 95% Confidence Interval: 0.828-1.000). In predicting the HFmEF stage, the model achieved an AUROC of 0.886±0.081 and AUPR of 0.730±0.055 (the lowest among the three HF stages). Lastly, the predictions of the HFrEF yielded an overall AUROC of 0.850 (95% Confidence Interval (CI): 0.781-0.919) and AUPR of 0.778 (95% CI: 0.707-0.849). In all three HF stages, the sensitivity and precision values were sustained high, exhibiting a drop only for specificity and recall threshold values > 0.90 (Fig. 5c).

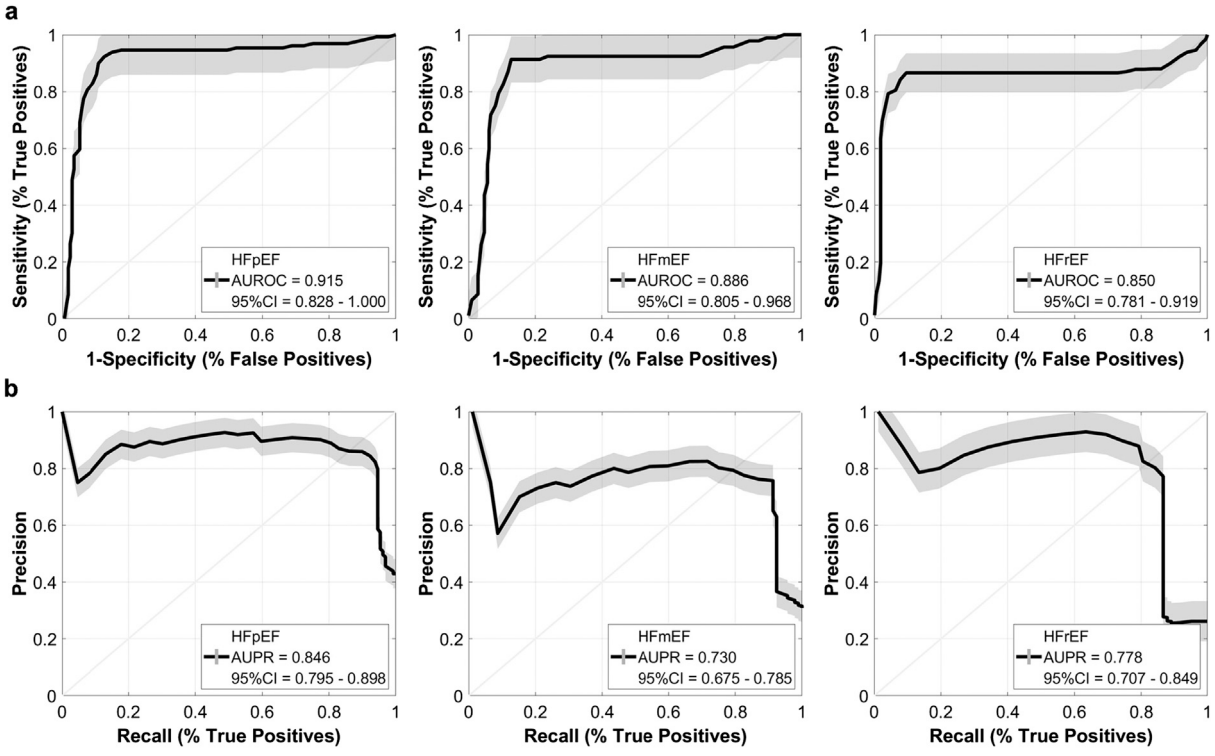
3.3. Interpretable deep learning decisions

To be able to interpret the decisions made by the model, we extracted heatmaps from the last features layer (max-pooling) of the trained deep learning model (Fig. 6). These heatmaps are defined as an attention-based mechanism that mimics human perception, and are considered as selective attention for identifying regions with the highest impacts on the decisions [38]. Using the seven-feature images (Fig. 6a), the extracted heatmap is a color image ranging from blue to red, referring to the lowest and highest impacts, respectively (Fig. 6b). After extracting a heatmap for every patient's images, we generated a unique averaged heatmap per HF stage (Fig. 7). Through these unique heatmaps, we performed deeper analysis on the impact of each hourly segment, as well as each demographic and clinical information ring on the predicted HF stage (Fig. 8).

Through visual inspection, the unique heatmaps showed that the predictions were mostly affected by the evening and early-morning time changes of the HRV features for HFpEF patients, late-night and afternoon for HFmEF patients, and afternoon to evening for HFrEF patients (Fig. 8a). In addition, by transforming these heatmaps into line plots, we observed time regions of high importance (≥ 0.70) for each HF stage (Fig. 8b). For HFpEF, hours 01:00-02:00, 07:00-10:00, and 17:00-00:00 were found highly important in deriving the model's decisions, whereas for HFmEF, hours 02:00-07:00 and 11:00-15:00 had high importance values. For HFrEF, only a one-time region (15:00-19:00) was found to be of high importance in the predictions. We further transformed demographic and clinical information rings at these important time regions into line plots as shown in Fig. 8c. Age, sex, and BMI were found important almost across all important hours for the predictions of the HFpEF stage. Moreover, smoking, diabetes, hypertension, and angina pectoris were important information only during evening hours (21:00-00:00). For HFrEF predictions, smoking, diabetes, and hypertension were the only important information for the trained model. The predictions for the HFmEF stage were highly impacted by smoking, diabetes, hypertension, and angina pectoris during hours 02:00-07:00, whereas sex and BMI slightly affected the decisions during afternoon hours (13:00-15:00).

3.4. Additional experiments

Additional experimentational scenarios were followed to validate the modelling approach (Fig. 9). First, the performance was investigated when varying the order of clinical information groups when filling in-



c

Thresholds (of specificity to find sensitivity) (of recall to find precision)	Performance per HF category					
	HFpEF		HFmEF		HFrEF	
	Sensitivity	Precision	Sensitivity	Precision	Sensitivity	Precision
0.70	0.946	0.909	0.924	0.821	0.866	0.919
0.75	0.946	0.907	0.924	0.793	0.866	0.897
0.80	0.946	0.896	0.913	0.779	0.866	0.825
0.85	0.930	0.859	0.913	0.780	0.866	0.787
0.90	0.899	0.847	0.859	0.769	0.866	0.256
0.95	0.705	0.508	0.467	0.357	0.793	0.262
AUROC (95% Confidence Interval)	0.915 (0.828 - 1.000)		0.866 (0.805 - 0.968)		0.850 (0.781 - 0.919)	
AUPR (95% Confidence Interval)	0.846 (0.795 - 0.898)		0.730 (0.675 - 0.785)		0.778 (0.707 - 0.849)	

Fig. 5. Receiver operating characteristics (ROC) and precision-recall (PR) curves analysis. a, ROC curves and area under the ROC (AUROC) values for heart failure with preserved ejection fraction (HFpEF), heart failure with mid-range ejection fraction (HFmEF), and heart failure with reduced ejection fraction (HFrEF) stages. **b,** PR curves and area under PR (AUPR) values for each heart failure (HF) stage. The shaded area in **a-b** represents a 95% confidence interval. **c,** tabulated data extracted from the ROC and PR curves of each HF stage. The data shows sensitivity and precision variations according to different specificity and recall thresholds (from 0.70 to 0.95). It includes the values of AUROC and AUPR with a 95% confidence interval.

side the HRV feature images (Fig. 9a). The six scenarios for clinical information sets were as follows,

- Set 1: cardiovascular risk factors, cardiovascular history, cardiovascular medication
- Set 2: cardiovascular risk factors, cardiovascular medication, cardiovascular history
- Set 3: cardiovascular medication, cardiovascular risk factors, cardiovascular history
- Set 4: cardiovascular medication, cardiovascular history, cardiovascular risk factors
- Set 5: cardiovascular history, cardiovascular medication, cardiovascular risk factors
- Set 6: cardiovascular history, cardiovascular risk factors, cardiovascular medication

The performance in general had minor differences compared with the performance using the order decided by clinicians, *i.e.*, set 1. For the 5 additional sets with varying orders, the overall accuracies were 90.51%, 88.42%, 86.90%, 89.33%, and 90.05%, respectively. The closest accuracy, *i.e.*, set 2, had the order of cardiovascular risk factors, medication, and history; which was close to the highest and chosen set.

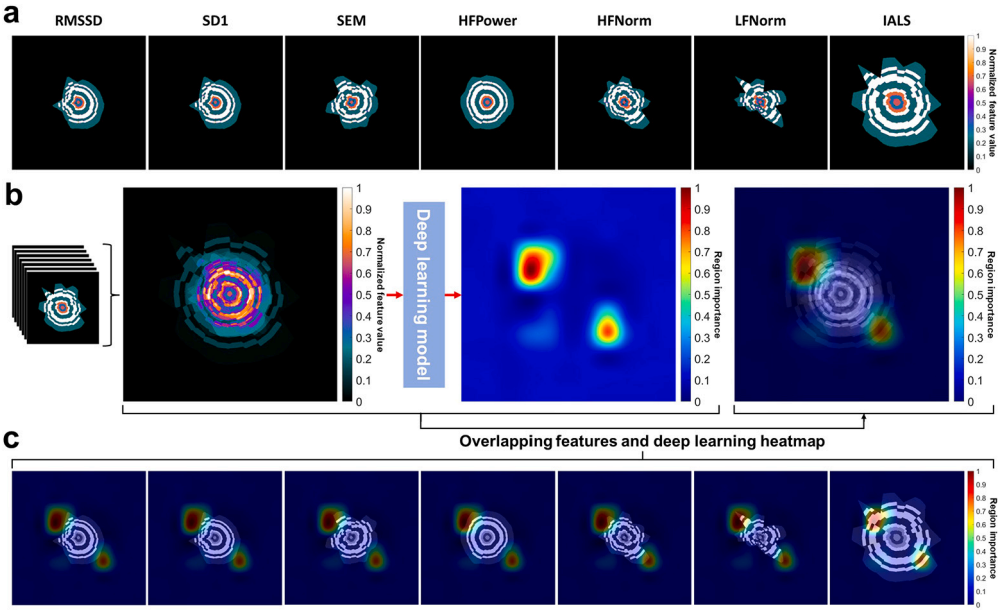


Fig. 6. Example for the extraction of attention-based heatmaps from the last features layer of the trained network (max-pooling). a, Example of a patient's set of HRV features images. The seven features shown are determined through the step-wise feature selection for a maximized performance. **a,** Stacking all seven features, inputting them to the trained deep learning model, and extracting the corresponding heatmap (attention map). **c,** Overlapping features images with the extracted heatmap to visualize the effect of different regions on the decision of the model.

The lowest accuracy levels were obtained when having cardiovascular medication information as the start of sets 3 and 4.

Moreover, we performed the modelling and cross-validation testing to evaluate the performance using binary HRV feature images without the incorporation of clinical information (Fig. 9b). In other words, the best set of HRV features were used as input images without being color-coded, *i.e.*, black and white images (1's correspond to feature points); which was performed by excluding steps shown in Fig. 2. Without clinical information, the average accuracy was 80.67%; which shows a reduction in accuracy by nearly 10%; which suggest the importance of clinical information when color-coded to the overall modelling and discrimination between heart failure stages with narrower ranges.

Lastly, we evaluated the information gain before and after the transformation of HRV features and clinical information into a 2D polar image using Shannon entropy analysis [39,40]. This allowed for understanding the impact of capturing multi-parametric data as a single source of information; a color-coded polar image. Here, the average entropy of the best performing HRV features was nearly 3.35 across the three heart failure groups. In addition, the clinical information was assessed similarly and had an average entropy of 3.97. On the other hand, the entropy dropped heavily when combining both as a single image to 1.08. The drop in entropy suggests proper capturing of useful information within both data sources. Moreover, it suggests less randomness in the polar image compared to 1D signal or feature inputs; which had eventually caused an increase in the overall performance of the model, inclusion of high information content within images, and better characterization of multiparameter patient information.

4. Discussion

This study provides evidence that AI could facilitate cardiovascular disease diagnostics [41,42] by accurately assessing the HF stages in novel clinical multi-parameter polar images. Under a LOSO experiment, we validated our results across two different datasets of CAD patients, exhibiting high accuracy in the HF stage predictions. In addition, the interpretation of deep learning decisions that the model provided enables crucial discoveries regarding the disease that may not be well-known yet to clinicians, owing to its ability to analyze big data. Relative to recent studies in the field (Table 3), the proposed study had higher performance with a novel approach for the integration of multiple parameters, *i.e.*, HRV and clinical information.

Our results have several implications. First, knowledge about the extent of HF is of significant importance in reducing morbidity and mortality rates among CAD patients, especially for the heterogeneous preserved and mid-range groups, where individualized treatment options are most important. Using an accurate, simple, and cost-effective technique for medical diagnosis and management is always the main target for medical practitioners as a first-stage screening tool before requesting additional tests that are more expensive and not universally available. Therefore, the proposed study could pave the way towards applying chronopharcology, by suggesting 24-hour ECG as a suitable assistive tool for circadian and personalized medicine in HF treatment. In addition, it could be repeated for specific patients more often to monitor disease progression with or without treatment procedures as a supplementary tool to echocardiography.

Second, the ability of the model to integrate multiple sources of medical information (HRV and patient profiles) in a single 2D color polar image has a significant impact on reducing the stress applied on cardiologists diagnosing HF cases or other cardiac diseases. Moreover, it allows them to establish a connection between 24-hour HRV variations and demographic/clinical information for the patient undergoing diagnosis. The simplicity of the proposed approach could suggest color-coded HRV images as a new protocol in assessing cardiovascular diseases where a single image can carry multiple patient-specific characteristics. This can be considered crucial for providing timely diagnostic and treatment procedures, especially when there are big patient data that require prompt intervention.

Third, analyzing the influence of HRV features in HF progression with deep learning allowed for a better understanding of the role of the autonomic nervous system. For example, the RMSSD feature was found highly important, as it reflects the modulation of the parasympathetic activity. Its variations are usually linked to HF in addition to sympathetic upregulation, where lower RMSSD has been linked to an increase in all-cause mortality [43–45]. In addition, frequency-domain features in the low and high-frequency bands were more important than other features, as they reflect cardiac autonomic outflows that are mainly under the control of baroreflex activity [46]. It is worth not-

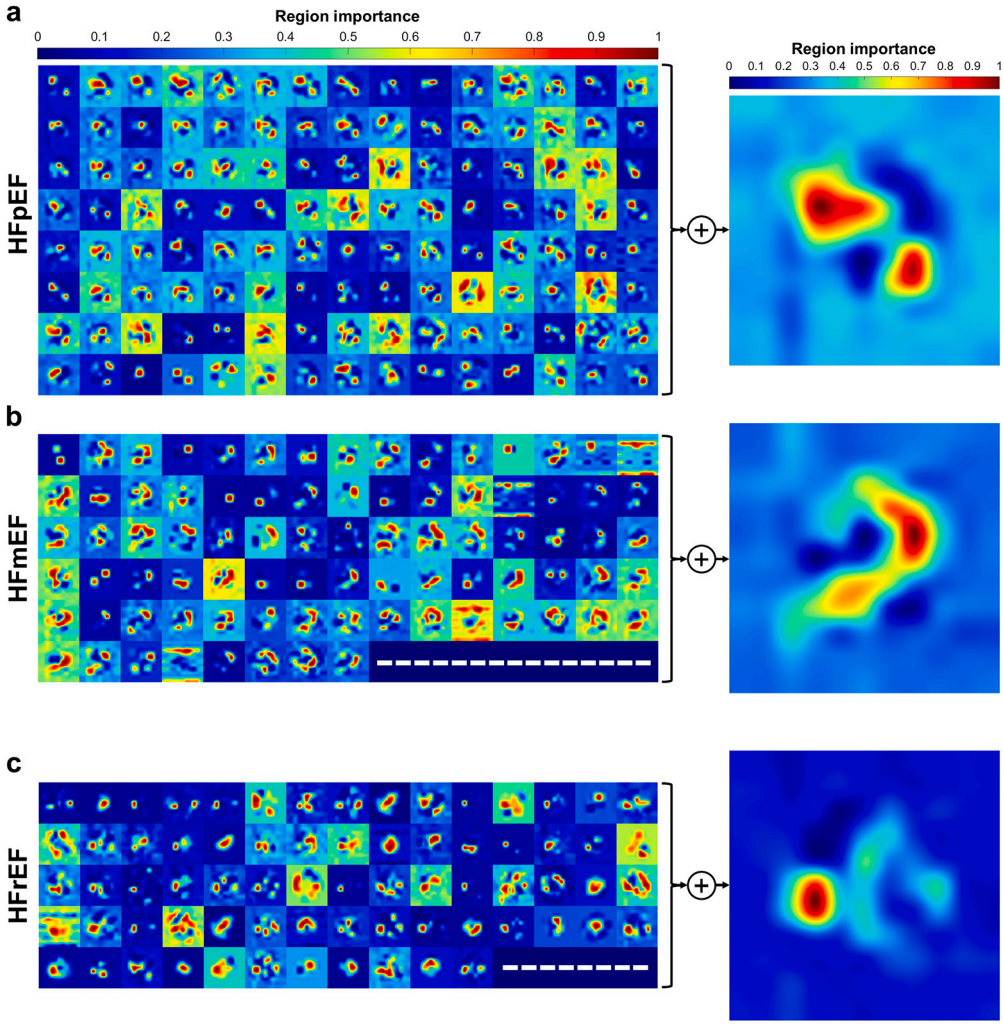


Fig. 7. All heatmaps extracted from the network for every heart failure (HF) patient. a, Heart failure with preserved ejection fraction (HFpEF) patients' images with the corresponding unique heatmap generated by averaging all images. **b,** Heart failure with mid-range ejection fraction (HFmEF). **c,** Heart failure with reduced ejection fraction (HFrEF).

ing that SD1 and IALS were found important as measures to reflect the parasympathetic modulation and R-peak sequence fragments, respectively.

Fourth, our study transformed the conventional black-box deep learning train-predict mechanism into an easily interpretable and interpretable approach. This could be a huge advance in HF discoveries, especially when differentiating between three stages concerning their LVEF ranges. The analysis of the 24-hour HRV was motivated by the knowledge that cardiac arrhythmias and heart attacks appear more often at certain times during the 24-hour time interval, but may be a function of HF progression. Inspection of HRV through deep learning attention-based heatmaps allowed for discovering useful information on the circadian changes in the cardiac rhythm and how this differed between HF stages. Our study indicated that time intervals in the early morning and late evening are important for characterizing HFpEF patients, which are well-known high-risk times for cardiac arrests [47]. Furthermore, hours in the late afternoon were found to be highly correlated with the HFrEF stage, which could be linked to the higher mortality rates often observed at this time interval [48]. The model has discovered that late-night to early-morning hours, as well as early-afternoon hours, are the best times for HFmEF categorization, as can be seen from the corresponding heatmaps (Fig. 7). These HFmEF-related periods precede the significant hours found for both HFpEF and HFrEF, suggesting them as a unique focal point at which HF may (or may not) progress further (or recover) [49].

Fifth, via the generated heatmaps analysis, the discovery of unique patterns relative to the demographic and clinical information embedded as part of the input images is feasible. In literature, previous studies have demonstrated that HFpEF patients tend to be way much older than patients with HFmEF or HFrEF [54], which has also been observed in Fig. 7c with high importance across all significant hours. Of interest is the BMI that was important for HFpEF and to a lesser extent for HFmEF, which were previously shown to have a strong correlation that results in developing further adverse cardiac events [55]. Moreover, the high importance of other comorbidities (diabetes and hypertension) was distributed across all evening hours for HFpEF and HFrEF and late-night hours for HFmEF. This could be correlated with the fact that all-cause mortality has been increasingly reported in diabetic HFrEF patients [56]. In addition, several studies have shown a strong association between hypertension and HF, especially for patients with preserved LVEF [57,58]. Concerning cardiovascular history, only angina pectoris was found important for the HFmEF patients' group, which mainly shows that these patients had high activation of the myocardial ischemia caused by the myocardial blood supply and oxygen demand imbalance [59,60]. It is worth noting that medication intake was not significant for the decisions of the model, despite the importance of antiarrhythmics and diuretics in characterizing HFpEF patients [54,61].

Moreover, when changing the order of clinical information and having medication in the center of input images, the performance was

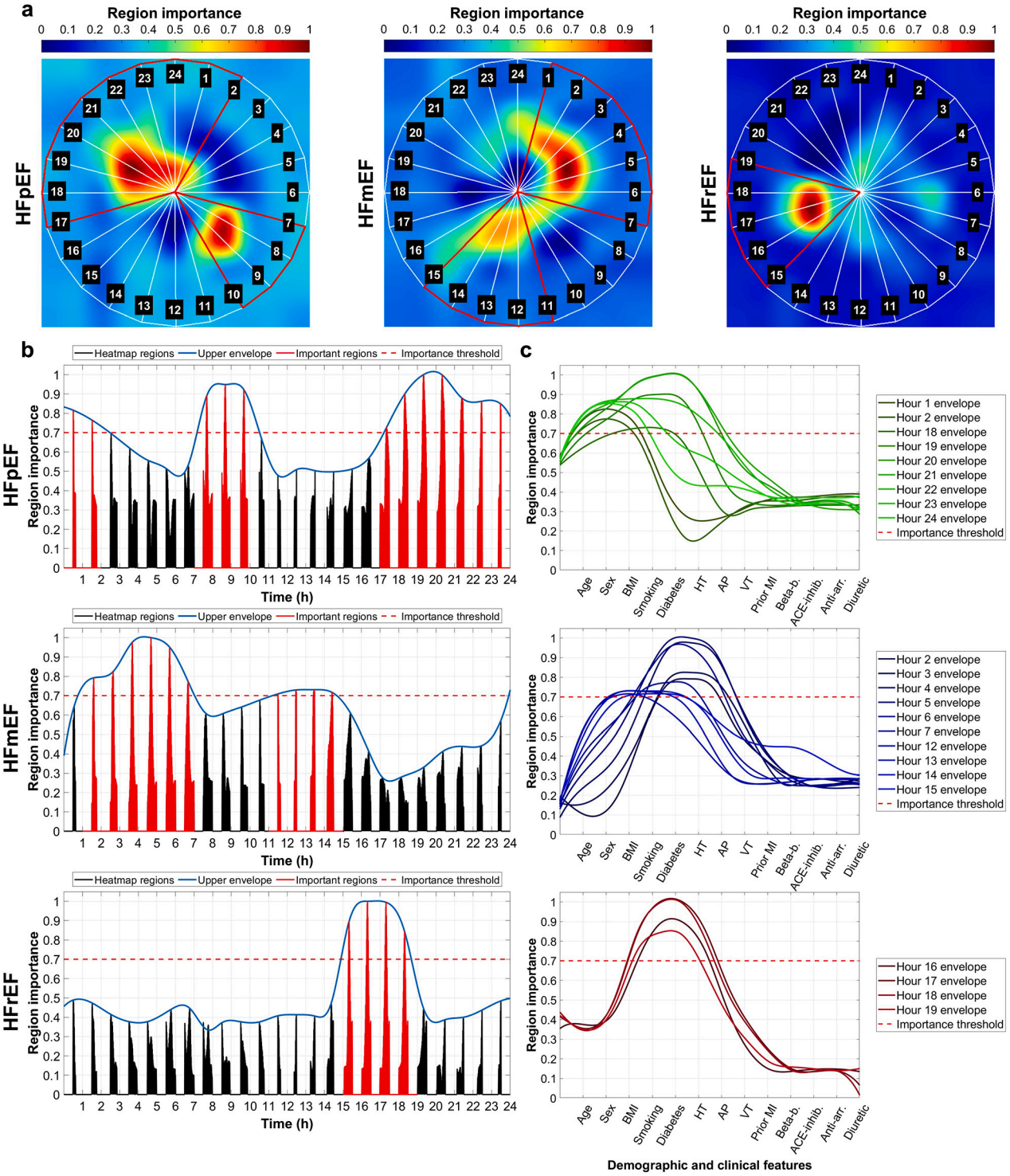


Fig. 8. Analysis of the attention-based unique heatmaps for the three heart failure (HF) stages. a, The unique heatmaps for heart failure with preserved ejection fraction (HFpEF), heart failure with mid-range ejection fraction (HFmEF), and heart failure with reduced ejection fraction (HFrEF) stages. Each image shows the 24-hour time regions plus the most important regions for the decisions of the stage (red cover). **b,** Time analysis of the heatmaps throughout the 24-hour cycle. Each hourly region was extracted from the pie-shaped hour segment shown in **a**. The blue envelope shows smoother variations in the heatmap across the hours. The red highlighted regions represent regions that were most important for the network decisions (≥ 0.70). **c,** Demographic and clinical information analysis after extracting them from the heatmaps using the ring-shaped segments shown in Fig. 1. The colors start with dark color at hour 01:00 and get brighter going up to hour 00:00.

slightly decreased. This could be related with how the model extracts features from the images based on their importance. Accordingly, having the center of the images with less effective variables in the discrimination between heart failure stages leads to unoptimized learned features. In addition, statistical analysis has revealed that no significant differences were observed between the three heart failure stages in the selected cohort. This could be due to having similar percentages in medication intake; excluding diuretics which was significantly lower in the HFpEF group only. Compared with well-known risk factors put in the center of images, which had the highest accuracy levels,

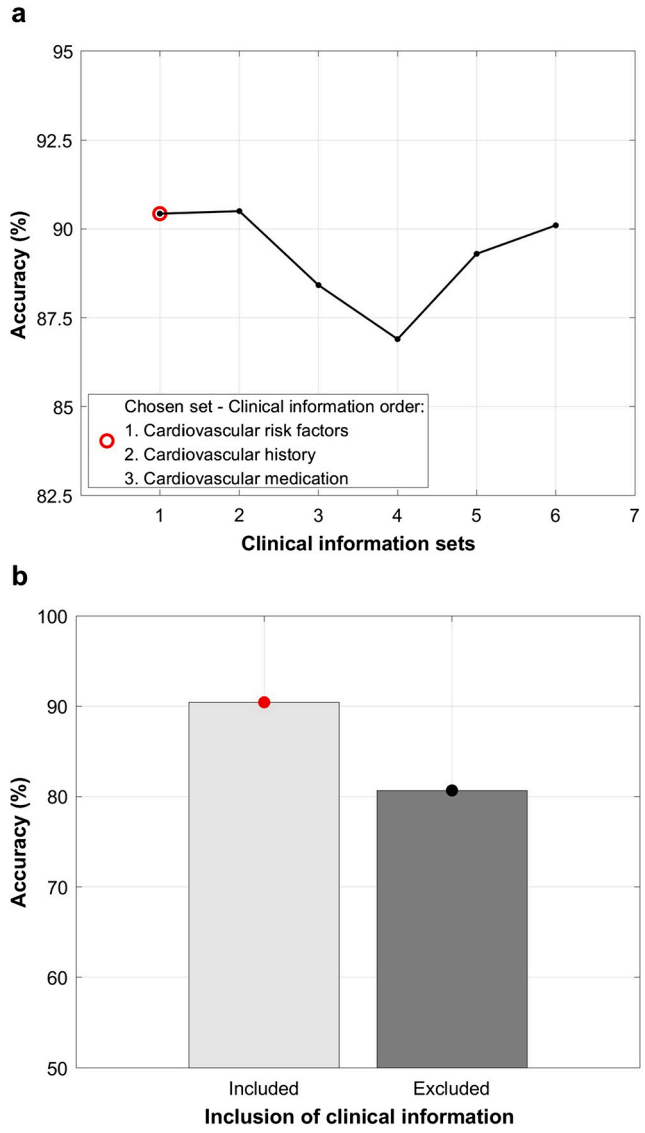


Fig. 9. Experimentational scenarios to validate the modelling approach. a, Variation of the order of the clinical information groups (risk factors, history, and medication) in multiple sets. Each set had clinical information filled within the HRV feature images at a different order. **b,** Performance when excluding clinical information from the HRV feature images, *i.e.*, using binary HRV feature images.

the model tends to extract better feature patterns that allow discriminating between the stages and thus, the more these important variables are towards the center, the higher the performance. This comes in line with the decisions made by clinicians on the optimal order of variables within images.

Sixth, in our study, a narrower range was adopted for the HFmEF stage, which differs from the current European gold-standard range [6]. The variable criteria for this stage come because the ranges are based on the etiology of HF. Thus, there are no strict rules on the decision for a fixed threshold due to the loose association between the treatment for this group and clinical presentation. In addition, in literature, 90% of the patients under the regular 41% to 49% LVEF threshold had either deteriorated or recovered from this range [7]. Our results show the potential of distinguishing patients under the narrow HFmEF range, which always created confusion in the diagnosis [24].

Despite the presented study's potential, some limitations could be considered. We classified HF patients with LVEF thresholds that differ from the current gold-standard guidelines. It will be interesting to explore different LVEF-based HF ranges in future studies to explore the reliability of the developed approach under various scenarios. Also, while we have confirmed that the model successfully could separate the three HF stages, we did not investigate the ability of our model to be applied to a broader range of cardiovascular diseases. In addition, the model's performance was validated in a sufficient number of participants; nevertheless, future studies with extended populations should be employed to further confirm those results. Additionally, while we have tested the model across two institutions, further studies can expand the diversity of datasets and institutions with more data and less imbalance between the classes. Moreover, the performance of the models, although accurate, it showed a slight decline in the PR curves. This tanking phenomena was mostly due to having few true positives at higher thresholds despite having false positives, *i.e.*, negative examples, that were incorrectly assigned with a very high probability of being positive. Therefore, more calibration for the models could suggest slight improvement in the performance. However, the trade-off between the performance and the overall training costs should be taken into consideration. Finally, our empirical results highlight a strong connection between HF staging and circadian rhythm-based cardiovascular regulation, in compliance with past work on the topic; however, both, the etiology and phenotype of HF largely differ across CAD patients; hence, the inclusion of circadian rhythm effect and patient's profile in the multi-parameter polar image assists the model to account for such difference in the HF staging. More studies, yet, are needed to better understand this relation, further exploring the role of personalization, both in the HF staging and initiation of personalized therapeutic options.

5. Conclusions

Overall, we have developed the approach based on echocardiography-free clinical quantitative and qualitative features (HRV and patient's demographics and clinical profile) through novel multi-parameter polar images. The proposed model is the first step towards implementing a simple, cost-effective, and frequent-use screening tool to evaluate the extent of HF based on LVEF levels. Polar map images of HRV features with filled-in patient information integrated within a deep learning-based trained model promise to enhance the diagnosis of HF at various degrees of LVEF, as they allow for an initial screening and evaluation of the circadian left ventricle function before echocardiography tests. Further research should evaluate the clinical cost-effectiveness of the model to provide, via polar plots - heat map images, a personalized HF management over the 24 hours.

CRediT authorship contribution statement

Mohanad Alkhodari: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Ahsan H. Khandoker:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Herbert F. Jelinek:** Writing – review & editing, Supervision, Investigation, Formal analysis, Conceptualization. **Angelos Karlas:** Writing – review & editing, Resources, Data curation. **Stergios Soulaidopoulos:** Writing – review & editing, Resources, Data curation. **Petros Arsenos:** Writing – review & editing, Resources, Data curation. **Ioannis Doundoulakis:** Writing – review & editing, Resources, Data curation. **Konstantinos A. Gatzoulis:** Writing – review & editing, Resources, Data curation. **Konstantinos Tsioufis:** Writing – review & editing, Resources, Data curation. **Leontios J. Hadjileontiadis:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Table 3
Summary of recent studies that utilized heart rate variability (HRV) and machine learning for heart failure assessment.

Study	Year	Dataset	Type	Technique	Model	Performance
Hussain et al. [50]	2020	PhysioNet databases (n = 98)	24-hour Holter HRV	HRV features	SVM Decision trees	Accuracy: 90.50% AUC: 0.90
Agliari et al. [51]	2020	Ascoli-Piceno hospital (n = 232)	24-hour Holter HRV	HRV features	Feed-forward ANN	Accuracy: 85.00%
Alkhodari et al. [45]	2021	THEW database (n = 92)	24-hour Holter HRV	HRV features	SVR	RMSE: 10.4 Accuracy: 89.60%
Alkhodari et al. [21]	2021	Multi-center cohorts THEW and PRESERVE-EF (n = 303)	Clinical information	NA	CNN SVM GLM	RMSE: 4.13 Accuracy: 90.10%
Aggarwal et al. [52]	2022	PhysioNet databases (n = 419)	24-hour Holter HRV	HRV features	SVM Decision trees	Accuracy: 86.35% AUC: 0.85
Tian et al. [53]	2022	Jinan 4th People's hospital (n = 179)	24-hour Holter HRV	HRV features	XGBoost Decision trees	Accuracy: 90.00%
This study	**2024**	**Multi-center cohorts THEW and PRESERVE-EF (n = 303)**	**24-hour Holter HRV Clinical information**	**Multi-parameter HRV feature images**	**Deep CNN**	**Accuracy: 92.62% AUC: 0.883**

HRV = Heart rate variability; SVM = Support vector machine; AUC = Area under the ROC curve; ANN = Artificial neural network; RMSE = Root mean square error; SVR = Support vector regression; GLM = Generalized linear model; CNN = Convolutional neural network.

Declaration of competing interest

The authors declare no competing interests about the content or interpretation of the data in the present study.

Data and code availability

The data used in this study can be obtained by directly contacting the corresponding author or the owners of the datasets (American cohort: Dr. Jean-Philippe Couderc, Greek cohort: Dr. Stergios Soulaidopoulos). The datasets can not be shared publicly due to embargo for disclosure by the owners.

All code requests can be made by contacting the corresponding author. Codes related to the generation of polar images and deep learning modelling/predictions are available partially with an example on a GitHub repository here: https://github.com/malkhodari/Alkhodari_2023_CMPB.git.

Acknowledgement

The authors would like to acknowledge Khwaja Y. Hasan, MBBS, from the Cardiology Department at Cleveland Clinic, Abu Dhabi, UAE, for his advice on diagnosing LVEF in heart failure patients. Funding for this study was provided by a grant (award number: 8474000132) from the Healthcare Engineering Innovation Center (HEIC) at Khalifa University, Abu Dhabi, UAE. The funders were not involved in any part of the study, including study design, data collection, data analysis, data interpretation, and report writing. All authors had final responsibility for the decision to submit for publication. The corresponding authors fully accessed data used in this study.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cmpb.2024.108107.

References

[1] Véronique L. Roger, Epidemiology of heart failure, Circ. Res. 113 (2013) 646–659.
[2] Mihai Gheorghiade, George Sopko, Leonardo De Luca, Eric J. Velazquez, John D. Parker, Philip F. Binkley, Zygmunt Sadowski, Krzysztof S. Golba, David L. Prior, Jeanl L. Rouleau, et al., Navigating the crossroads of coronary artery disease and heart failure, Circulation 114 (2006) 1202–1213.
[3] Amy Groenewegen, Frans H. Rutten, Arend Mosterd, Arno W. Hoes, Epidemiology of heart failure, Eur. J. Heart Fail. 22 (2020) 1342–1356.
[4] Judith Mackay, George A. Mensah, Kurt Greenlund, The Atlas of Heart Disease and Stroke, World Health Organization, 2004.
[5] Jeptha P. Curtis, Seth I. Sokol, Yongfei Wang, Saif S. Rathore, Dennis T. Ko, Farid Jadbabaie, Edward L. Portnay, Stephen J. Marshalko, Martha J. Radford, Harlan M. Krumholz, The association of left ventricular ejection fraction, mortality, and cause of death in stable outpatients with heart failure, J. Am. Coll. Cardiol. 42 (2003) 736–742.
[6] Theresa A. McDonagh, Marco Metra, Marianna Adamo, Roy S. Gardner, Andreas Baumbach, Michael Böhm, Haran Burri, Javed Butler, Jelena Čelutkienė, Ovidiu Chioncel, et al., 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: developed by the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) with the special contribution of the Heart Failure Association (HFA) of the ESC, Eur. Heart J. 42 (2021) 3599–3726.
[7] Hong-Mi Choi, Myung-Soo Park, Jong-Chan Youn, Update on heart failure management and future directions, Korean J. Intern. Med. 34 (2019) 11.
[8] Q. Ciampi, B. Villari, Role of echocardiography in diagnosis and risk stratification in heart failure with left ventricular systolic dysfunction, Cardiovasc. Ultrasound 5 (2007) 1–12.
[9] T. Foley, S. Mankad, N. Anavekar, C. Bonnichsen, M. Morris, T. Miller, P. Araoz, Measuring left ventricular ejection fraction–techniques and potential pitfalls, Eur. Cardiol. Rev. 8 (2012) 108–114.
[10] F. Shaffer, J.P. Ginsberg, An overview of heart rate variability metrics and norms, Front. Public Health 5 (2017) 258.
[11] Ken Kiyono, Junichiro Hayano, Eiichi Watanabe, Zbigniew R. Struzik, Yoshiharu Yamamoto, Non-Gaussian heart rate as an independent predictor of mortality in patients with chronic heart failure, Heart Rhythm 5 (2008) 261–268.
[12] Heikki V. Huikuri, M.J. Pekka Raatikainen, Rikke Moerch-Joergensen, Juha Hartikainen, Vesa Virtanen, Jean Boland, Olli Anttonen, Nis Hoest, Lucas V.A. Boersma, Eivind S. Platou, et al., Prediction of fatal or near-fatal cardiac arrhythmia events in patients with depressed left ventricular function after an acute myocardial infarction, Eur. Heart J. 30 (2009) 689–698.
[13] U. Rajendra Acharya, Hamido Fujita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, Ru San Tan, Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals, Appl. Intell. 49 (2019) 16–27.
[14] K. Kim, J. Kwon, Deep learning for diagnosing heart failure from ECG signals, J. Heart Lung Transplant. 38 (2019) S375.
[15] Matthias Unterhuber, Karl-Philipp Rommel, Karl-Patrik Kresoja, Julia Lurz, Jelena Kornej, Gerhard Hindricks, Markus Scholz, Holger Thiele, Philipp Lurz, Deep learning detects heart failure with preserved ejection fraction using a baseline electrocardiogram, Eur. Heart J. Digit. Health (2021).
[16] Joon-myoung Kwon, Kyung-Hee Kim, Ki-Hyun Jeon, Hyue Mee Kim, Min Jeong Kim, Sung-Min Lim, Pil Sang Song, Jinsik Park, Rak Kyeong Choi, Byung-Hee Oh, Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification, Korean Circ. J. 49 (2019) 629–639.
[17] Ludi Wang, Wei Zhou, Qing Chang, Jiangen Chen, Xiaoguang Zhou, Deep ensemble detection of congestive heart failure using short-term RR intervals, IEEE Access 7 (2019) 69559–69574.
[18] Ludi Wang, Xiaoguang Zhou, Detection of congestive heart failure based on LSTM-based deep network via short-term RR intervals, Sensors 19 (2019) 1502.
[19] Edward Choi, Andy Schuetz, Walter F. Stewart, Jimeng Sun, Using recurrent neural network models for early detection of heart failure onset, J. Am. Med. Inform. Assoc. 24 (2017) 361–370.
[20] Robert Chen, Walter F. Stewart, Jimeng Sun, Kenney Ng, Xiaowei Yan, Recurrent neural networks for early detection of heart failure from longitudinal electronic health record data: implications for temporal modeling with respect to time before

diagnosis, data density, data quantity, and data type, Circ. Cardiovasc. Qual. Outcomes 12 (2019) e005114.
[21] Mohanad Alkhodari, Herbert F. Jelinek, Angelos Karlas, Stergios Soulaidopoulos, Petros Arsenos, Ioannis Doundoulakis, Konstantinos A. Gatzoulis, Konstantinos Tsioufis, Leontios J. Hadjileontiadis, Ahsan H. Khandoker, Deep learning predicts heart failure with preserved, mid-range, and reduced left ventricular ejection fraction from patient clinical profiles, Front. Cardiovasc. Med. 8 (2021) 1604.
[22] Mohanad Alkhodari, Leontios J. Hadjileontiadis, Herbert F. Jelinek, Ahsan H. Khandoker, Heart failure assessment using multiparameter polar representations and deep learning, in: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2023, pp. 1–4.
[23] Roberto M. Lang, Michelle Bierig, Richard B. Devereux, Frank A. Flachskampf, Elyse Foster, Patricia A. Pellikka, Michael H. Picard, Mary J. Roman, James Seward, Jack Shanewise, et al., Recommendations for chamber quantification, Eur. J. Echocardiogr. 7 (2006) 79–108.
[24] Gregg C. Fonarow, Jeffrey J. Hsu, Left ventricular ejection fraction: what is "normal"?, JACC Heart Fail. 4 (2016) 511–513.
[25] Connie W. Tsao, Asya Lyass, Martin G. Larson, Susan Cheng, Carolyn S.P. Lam, Jayashri R. Aragam, Emelia J. Benjamin, Ramachandran S. Vasan, Prognosis of adults with borderline left ventricular ejection fraction, JACC Heart Fail. 4 (2016) 502–510.
[26] Jean-Philippe Couderc, The telemetric and Holter ECG warehouse initiative (THEW): a data repository for the design, implementation and validation of ECG-related technologies, in: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE, 2010, pp. 6252–6255.
[27] University of Rochester Medical Center. Telemetric and Holter ECG Warehouse.
[28] Konstantinos A. Gatzoulis, Dimitrios Tsiachris, Petros Arsenos, Christos-Konstantinos Antoniou, Polychronis Dilaveris, Skevos Sideris, Emmanuel Kanoupakis, Emmanouil Simantirakis, Panagiotis Korantzopoulos, Ioannis Goudevenos, et al., Arrhythmic risk stratification in post-myocardial infarction patients with preserved ejection fraction: the PRESERVE EF study, Eur. Heart J. 40 (2019) 2940–2949.
[29] Konstantinos A. Gatzoulis, Dimitris Tsiachris, Petros Arsenos, Polychronis Dilaveris, Skevos Sideris, Emmanouil Simantirakis, Michalis Efremidis, Nikolaos Dagres, Panagiotis Korantzopoulos, Nikolaos Fragkakis, et al., Post myocardial infarction risk stratification for sudden cardiac death in patients with preserved ejection fraction: PRESERVE-EF study design, Hellenic J. Cardiol. (Ell. Kardiol. Epitheōr.) 55 (2014) 361–368.
[30] Samprit Chatterjee, D.L. McLeish, Fitting linear regression models to censored data by least squares and maximum likelihood methods, Commun. Stat., Theory Methods 15 (1986) 3227–3243.
[31] Margaret M. Doyle, Terrence E. Murphy, Margaret A. Pisani, Henry K. Yaggi, Sangchoon Jeon, Nancy S. Redeker, Melissa P. Knauert, A SAS macro for modelling periodic data using cosinor analysis, Comput. Methods Programs Biomed. 209 (2021) 106292.
[32] Marek Malik, Heart rate variability: standards of measurement, physiological interpretation, and clinical use: task force of the European society of cardiology and the North American society for pacing and electrophysiology, Ann. Noninvasive Electrocardiol. 1 (1996) 151–181.
[33] Chung-Kang Peng, Jeffrey M. Hausdorff, A.L. Goldberger, Fractal mechanisms in neuronal control: human heartbeat and gait dynamics in health and disease, in: Self-Organized Biological Dynamics and Nonlinear Control, 2000, pp. 66–96.
[34] Madalena Costa, Ary L. Goldberger, C-K. Peng, Multiscale entropy analysis of biological signals, Phys. Rev. E 71 (2005) 021906.
[35] Madalena D. Costa, Roger B. Davis, Ary L. Goldberger, Heart rate fragmentation: a new approach to the analysis of cardiac interbeat interval dynamics, Front. Physiol. 8 (2017) 255.
[36] Jonathan Moeyersons, Elena Smets, John Morales, Amalia Villa, Walter De Raedt, Dries Testelmans, Bertien Buyse, Chris Van Hoof, Rik Willems, Sabine Van Huffel, et al., Artefact detection and quality assessment of ambulatory ECG signals, Comput. Methods Programs Biomed. 182 (2019) 105050.
[37] Matthias Feurer, Frank Hutter, Hyperparameter Optimization, Springer, Cham, 2019.
[38] Yang Liu, Lixin Ji, Ruiyang Huang, Tuosiyu Ming, Chao Gao, Jianpeng Zhang, An attention-gated convolutional neural network for sentence classification, Intell. Data Anal. 23 (2019) 1091–1107.
[39] G.I. Redford, R.M. Clegg, Polar plot representation for frequency-domain analysis of fluorescence lifetimes, J. Fluoresc. 15 (2005) 805–815.
[40] Y. Wu, et al., Local Shannon entropy measure with statistical tests for image randomness, Inf. Sci. 222 (2013) 323–342.
[41] F. Shahbazi, B.M. Asl, Generalized discriminant analysis for congestive heart failure risk assessment based on long-term heart rate variability, Comput. Methods Programs Biomed. 122 (2015) 191–198.
[42] Z. Masetic, A. Subasi, Congestive heart failure detection using random forest classifier, Comput. Methods Programs Biomed. 130 (2016) 54–64.
[43] Brian Olshansky, Hani N. Sabbah, Paul J. Hauptman, Wilson S. Colucci, Parasympathetic nervous system and heart failure: pathophysiology and potential implications for therapy, Circulation 118 (2008) 863–871.
[44] Daiyi Luo, Weifeng Pan, Yifan Li, Kaicheng Feng, Guanzheng Liu, The interaction analysis between the sympathetic and parasympathetic systems in CHF by using transfer entropy method, Entropy 20 (2018) 795.
[45] Mohanad Alkhodari, Herbert F. Jelinek, Naoufel Werghi, Leontios J. Hadjileontiadis, Ahsan H. Khandoker, Estimating left ventricle ejection fraction levels using circadian heart rate variability features and support vector regression models, IEEE J. Biomed. Health Inform. 25 (2020) 746–754.
[46] David S. Goldstein, Oladi Bentho, Mee-Yeong Park, Yehonatan Sharabi, Low-frequency power of heart rate variability is not a measure of cardiac sympathetic tone but may be a measure of modulation of cardiac autonomic outflows by baroreflexes, Exp. Physiol. 96 (2011) 1255–1261.
[47] Heikki V. Huikuri, Markku K. Linnaluoto, Tapio Seppänen, K.E. Juhani Airaksinen, Kenneth M. Kessler, Juha T. Takkunen, Robert J. Myerburg, Circadian rhythm of heart rate variability in survivors of cardiac arrest, Am. J. Cardiol. 70 (1992) 610–615.
[48] Bryan G. Schwartz, Guy S. Mayeda, Steven Burstein, Christina Economides, Robert A. Kloner, When and why do heart attacks occur? Cardiovascular triggers and their potential role, Hosp. Pract. 38 (2010) 144–152.
[49] Qing Zhou, Peixin Li, Hengli Zhao, Xingbo Xu, Shaoping Li, Jing Zhao, Dingli Xu, Qingchun Zeng, Heart failure with mid-range ejection fraction: a distinctive subtype or a transitional stage?, Front. Cardiovasc. Med. 8 (2021) 495.
[50] L. Hussain, et al., Detecting congestive heart failure by extracting multimodal features and employing machine learning techniques, BioMed Res. Int. 2020 (2020).
[51] E. Agliari, et al., Detecting cardiac pathologies via machine learning on heart-rate variability time series and related markers, Sci. Rep. 10 (2020) 8845.
[52] R. Aggarwal, S. Kumar, HRV based feature selection for congestive heart failure and normal sinus rhythm for meticulous presaging of heart disease using machine learning, Meas.: Sens. 24 (2022) 100573.
[53] X. Tian, et al., Prediction of the left ventricular ejection fraction by machine learning algorithms based on heart rate variability parameters in patients with ischemic heart disease, 2022.
[54] Wilson Nadruz Jr., Erin West, Mário Santos, Hicham Skali, John D. Groarke, Daniel E. Forman, Amil M. Shah, Heart failure and midrange ejection fraction: implications of recovered ejection fraction for exercise tolerance and outcomes, Circ. Heart Fail. 9 (2016) e002826.
[55] Tiffany M. Powell-Wiley, Julius Ngwa, Selomie Kebede, Di Lu, Phillip J. Schulte, Deepak L. Bhatt, Clyde Yancy, Gregg C. Fonarow, Michelle A. Albert, Impact of body mass index on heart failure by race/ethnicity from the Get With The Guidelines–Heart Failure (GWTG–HF) Registry, JACC Heart Fail. 6 (2018) 233–242.
[56] Mohammed Al-Jarallah, Rajesh Rajan, Ibrahim Al-Zakwani, Raja Dashti, Bassam Bulbanat, Mustafa Ridha, Kadhim Sulaiman, Alawi A. Alsheikh-Ali, Prashanth Panduranga, Khalid F. AlHabib, et al., Mortality and morbidity in HFrEF, HFmrEF, and HFpEF patients with diabetes in the Middle East, Oman Med. J. 35 (2020) e99.
[57] Marc A. Pfeffer, Amil M. Shah, Barry A. Borlaug, Heart failure with preserved ejection fraction in perspective, Circ. Res. 124 (2019) 1598–1617.
[58] Marco Guazzi, Stefano Ghio, Yochai Adir, Pulmonary hypertension in HFpEF and HFrEF: JACC review topic of the week, J. Am. Coll. Cardiol. 76 (2020) 1102–1111.
[59] R.J. Mentz, S. Broderick, L. Shaw, M. Fiuzat, C. O'Connor, Association between angina and outcomes in heart failure patients with preserved ejection fraction: analysis from the duke databank for cardiovascular disease, J. Am. Coll. Cardiol. 61 (2013) E682.
[60] A.A. Badar, et al., Relationship between angina pectoris and outcomes in patients with heart failure and reduced ejection fraction: an analysis of the controlled rosuvastatin multinational trial in heart failure (corona), Eur. Heart J. 35 (2014) 3426–3433.
[61] Takeshi Tsuda, Kristi K. Fitzgerald, Dystrophic cardiomyopathy: complex pathobiological processes to generate clinical phenotype, J. Cardiovasc. Dev. Dis. 4 (2017) 14.