Non-invasive measurement of intracardiac pressures using machine learning techniques

A.E. van Ravensberg

Master Thesis
Technical Medicine
April 2023

This page was intentionally left blank

NON-INVASIVE MEASUREMENT OF INTRACARDIAC PRESSURES IN HEART FAILURE PATIENTS USING MACHINE LEARNING TECHNIQUES

Annemiek E. van Ravensberg

Student number: 4545346

18th of April 2023

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University; Delft University of Technology; Erasmus University Rotterdam

Master thesis project (TM30004; 35 ECTS)

Dept. of Biomechanical Engineering, TUDELFT

September 2022 – March 2023

Supervisor(s):

Dr. Robert M.A. van der Boon MD

Dr. Ing. Nico Bruining

Niels T.B. Scholte MD

Thesis committee members:

Dr. Ing. Nico Bruining, Erasmus MC

Dr. Robert M.A. van der Boon MD, Erasmus MC

Dr. Ir. Enno T. van der Velde, LUMC

Dr. Arend W. van Deutekom MD, Erasmus MC

Niels T.B. Scholte MD, Erasmus MC

1 Table of contents

2 Abstract

Background: Solutions targeting early recognition of congestion in heart failure (HF) patients have the potential to prevent readmissions and can thus significantly reduce the burden on HF care. The gold standard measure of congestion is invasively measured pulmonary capillary wedge pressure (PCWP). However, the invasive nature and accessibility of this measurement limits its clinical use. Non-invasive approximation of the PCWP using biosensing wearables could be a promising replacement for HF monitoring.

Purpose: The primary aim of this retrospective study was to create a model that estimates the PCWP based on non-invasive measurements of vital signs using both traditional statistics and machine learning (ML) techniques.

Methods: The study cohort comprised right-sided heart catheterisations between 23/6/2017 and 19/8/2022 performed in the Erasmus MC, Rotterdam, The Netherlands. The following models were used: linear regression or classification, k-nearest neighbours, random forest, gradient boosting, and multilayer perceptron. The outcome measure for the regression models was the continuous PCWP as measured during the catheterisation. The two outcome classes for the classification models were low (<12 mmHg) and high (≥12 mmHg) PCWP. Non-invasive mean arterial blood pressure (MAP), saturation, heart rate, weight and temperature measured at most 72 hours before or after the catheterisation were collected as the features for the models, as well as the age and gender of the patient. Additionally, ECG-signals acquired during the catheterisation were used to calculate the heart rate variability (HRV). The data was split into a validation (20%) and training (80%) data set. The models were built based on the training set and then applied on the validation set to determine the coefficients of determination (R2) for the regression models and the area under the curve (AUC) for the classification models.

Results: A total of 853 catheterisation patients were included of which 31.3% had HF as primary diagnosis and 48.7% had a PCWP of 12 mmHg or higher. The average age of the cohort was 58.9 ± 13.7 years and 52.1% were male. The HRV had the highest correlation with the PCWP with a correlation of 0.16. All the regression models resulted in low R2 values of up to 0.11 and the classification models in AUC values of up to 0.59.

Conclusion: In the current study, PCWP could not be approximated with non-invasive measurements using traditional statistics and ML techniques. These findings support the notion that traditional measures for monitoring HF are poorly correlated with hemodynamic parameters. Perhaps repeated measurements over time, e.g. trends, or continuously measured signals such as photoplethysmography (PPG) could overcome the shortcomings of a single vital signs measurement for congestion evaluation. Therefore, future prospective research is needed to evaluate the potential of wearable devices measuring trends based on PPG signals in predicting hemodynamics.

3 List of Abbreviations

AF	Atrial Fibrillation
AUC	Area Under the Curve
CT	Crest Time
DA	Diastolic Amplitude
DBP	Diastolic Blood Pressure
E/e’	Early mitral inflow velocity to early diastolic mitral annulus velocity ratio
ECG	Electrocardiography
eGFR	Estimated glomerular filtration rate
EHR	Electronic Health Record
HF	Heart Failure
HRV	Heart Rate Variability
LASI	Large Artery Stiffness Index
LDA	Linear Discriminant Analysis
LVEF	Left Ventricular Ejection Fraction
MAP	Mean Arterial Pressure
ML	Machine Learning
NT-proBNP	N-Terminal pro B-type natriuretic peptide
PA	Pulmonary Artery
PCWP	Pulmonary Capillary Wedge Pressure
PPG	Photoplethysmography
R2	Coefficient of determination
ROC	Receiver Operating Characteristic
SA	Systolic Amplitude
SD	Standard Deviation
SBP	Systolic Blood Pressure
TAVI	Transcatheter Aortic Valve Implantation
WHF	Worsening Heart Failure

4 Introduction

Heart failure (HF) is a global pandemic affecting approximately 26 million patients worldwide.(1) Both its prevalence and incidence are expected to further rise in the future because of an aging population, increase in patients with co-morbidities and higher survival rates after myocardial infarctions.(2,3) Despite advances in medical and device therapy, HF remains associated with a poor prognosis with a 30 day mortality of 10%.(4) Additionally, HF places a significant burden on the healthcare system due to frequent outpatient follow-up and recurrent hospitalization for worsening HF (WHF). (5)

The main reason for heart failure hospitalisation is congestion, which is defined as fluid accumulation in the intravascular compartment and the interstitial space.(7) Solutions targeting early recognition of congestion prevent readmissions, which significantly reduces the burden on HF care. The gold standard measure of congestion is the pulmonary capillary wedge pressure (PCWP) measured in the pulmonary artery (PA) by a Swan-Ganz catheter (Figure 1).(7,8) In recent years, several studies have shown a reduction in hospital admissions in patients with implantable sensors measuring the diastolic PA pressure as a surrogate measurement.(9–12) Although the use of such implantable sensors is an effective solution, these systems are invasive and expensive and therefore not available for every HF patient.

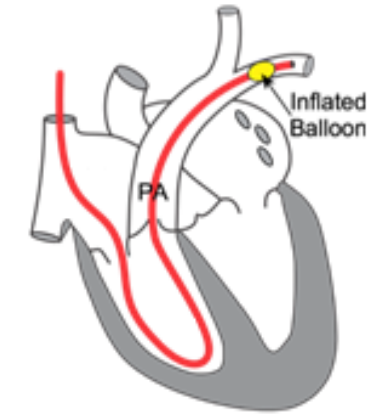


Figure 1. Pulmonary capillary wedge pressure measurement with a Swan-Ganz catheter in the Pulmonary Artery (PA).(6)

Non-invasive measurements (e.g. heart rate, blood pressure, saturation) are currently commonly used to monitor patients with HF because of their predictive value for worsening heart failure (WHF), either on an outpatient basis or remotely.(13–17) Especially during COVID-19, remote monitoring has increased (18) and new emerging technologies create more and more opportunities to use wearable devices (wearables) to collect non-invasive measurements.(19–21) The accessibility and convenience of wearables could make them ideal replacements for the invasively measured PCWP's.

Therefore, the primary aim of this thesis was to create a model that can use non-invasive measurements of vital signs to estimate the PCWP using both traditional statistics and machine learning (ML) techniques. The secondary aim of this study was to evaluate the additional predictive power of echocardiographic and laboratory measurements when added to the models and to evaluate whether the performances of the models change for specific populations.

5 Method

5.1 Study design and population

In this retrospective study cohort, all patients older than 18 years without a left ventricular assist device who underwent right-sided heart catheterisation between 23/6/2017 and 19/8/2022 at the Erasmus MC, were included. The study was approved by the ethics committee (MEC-2022-0822). All data was pseudo-anonymised/encoded and data traceable to individual patients was only accessible to the members of the research team, health care inspection, and the Medical Research Ethics Committee of the Erasmus MC.
The complete study method is summarised and visualised in Figure 2, containing the feature extraction, train-validation split, pre-processing, hyperparameter selection and model fitting and eventually, the model validation.

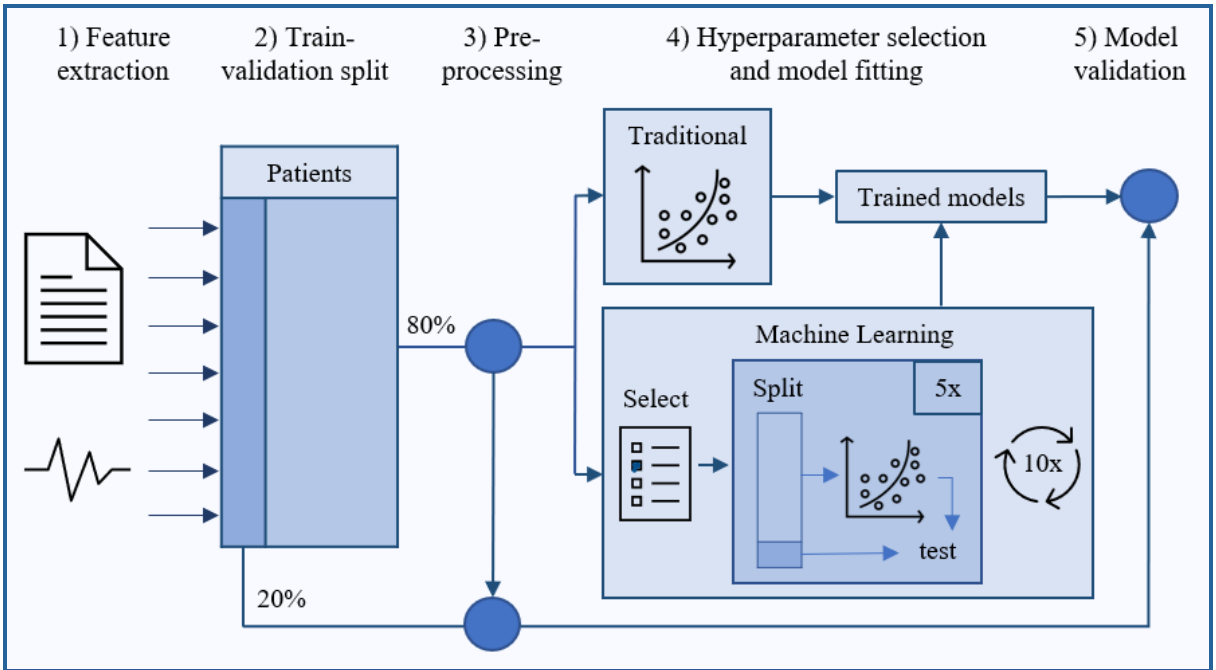


Figure 2. Schematic representation of the pulmonary capillary wedge pressure estimation pipeline. **1)** Retrospective feature extraction from the electronic health records and electrocardiograms. **2)** Splitting the data in 80% train data and 20% validation data. **3)** Pre-processing the train and validation data based on the train data. **4)** Fitting the models based on the train data. For the machine learning models, the best combination of hyperparameters was found by selecting 10 combinations and determining their performance with 5-fold cross validation. **5)** The resulting models were validated with the validation data.

5.2 Data extraction

Patient data was retrospectively extracted from the electronic health record (EHR) system. The data can be divided into the outcome value, non-invasive measurements, additional features and patient characteristics for a subgroup analysis (Table 1). Data was only included when measured at most 72 hours before or after the heart catheterisation.

Table 1. The collected data, divided into different categories. The non-invasive features were used for the models of the primary analysis to estimate the outcome and the additional features were only added for a secondary analysis. The patient characteristics were used for a secondary analysis to divide the population in subgroups.

Type	Predictor	Primary analysis	Secondary analyses
Outcome	PCWP	X	X
Non-invasive	Age	X	X
	Sex	X	X
	Weight	X	X
	Height	X	X
	SBP	X	X
	DBP	X	X
	Heart rate	X	X
	HRV	X	X
	Saturation	X	X
	Temperature	X	X
Patient characteristics	HF diagnosis		X
	Valve disease diagnosis		X
Additional features	LVEF		X
	E/e’		X
	Haemoglobin		X
	Sodium		X
	eGFR		X
	NT-proBNP		X
	Creatinine		X

Abbreviations: Pulmonary Capillary Wedge Pressure (PCWP), Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Heart Rate Variability (HRV), Heart Failure (HF), Left Ventricular Ejection Fraction (LVEF), early mitral inflow velocity to early diastolic mitral annulus velocity ratio (E/e’), estimated Glomerular Filtration Rate (eGFR), N-Terminal pro B-type Natriuretic Peptide (NT-proBNP)

5.2.1 Outcome measure: PCWP

The outcome measure of all the models was the PCWP as measured during the Swan-Ganz catheterisation. Since this pressure can be measured multiple times during a catheterisation, only the last recorded value was used under the assumption that this was the accepted PCWP measurement.

5.2.2 Non-invasive measurements

The first non-invasive features for the primary analysis were the gender and the age of the patients at the time of the catheterisation. Additionally, non-invasive systolic and diastolic blood pressure (SBP and DBP respectively), saturation, heart rate, weight, height and body temperature measured at most 72 hours before or after the catheterisation were collected. When multiple measurements were present, the last measurement before the catheterisation was selected and otherwise the first measurement afterwards. The SBP and DBP were used to calculate the mean arterial pressure (MAP) according to formula 1.

$$\mathrm{MAP} = \frac{1}{3}(\mathrm{SBP} + 2 * \mathrm{DBP}) \qquad (1)$$

Lastly, a MATLAB algorithm was built to calculate the heart rate variability (HRV) based on ECG-signals from during the catheterisation. The first step of the algorithm was signal pre-processing by filtering out clipping (values that exceed the minimum or maximum thresholds of the sensor) and periods with flat lines. Additionally, a 2nd order Butterworth band-pass filter with lower threshold 10 Hz and upper threshold 30 Hz was used to filter out the wandering baseline (low frequencies) and noise (high frequencies). Then, peak detection was used to detect the R-peaks and subsequently determine the time between following R-peaks in milliseconds as the RR-interval. The HRV was calculated as the average of the moving variance per 300 RR values. By taking the variance of small parts of the ECG recording, the HRV was defined as a feature of short-term variance and longer recordings with a slow increase or decrease of heart rate did not automatically result in high HRV values. Figure 3 shows an ECG signal with a regular HR and therefore a low HRV value. In contrast, figure 4 shows an ECG signal with an irregular HR and a high HRV value.

5.2.3 Additional data

The secondary analyses consisted of an analysis with patient subgroups and an analysis with additional features. For the subgroup analysis, it was determined for every patient whether or not the patient was diagnosed with HF or heart valve disease.

Echocardiographic and laboratory predictors of WHF were collected for the analysis with additional features based on the earlier literature study. The echocardiographic predictors were the left ventricular ejection fraction (LVEF) and early mitral inflow velocity to early diastolic mitral annulus velocity ratio (E/e'). The laboratory values were the haemoglobin, sodium, estimated glomerular filtration rate (eGFR), N-Terminal pro B-type natriuretic peptide (NT-proBNP) and creatinine. Equally to the non-invasive measurements, the last documented value before the catheterisation was selected and otherwise the first documented value afterwards within 72 hours before or after the procedure.

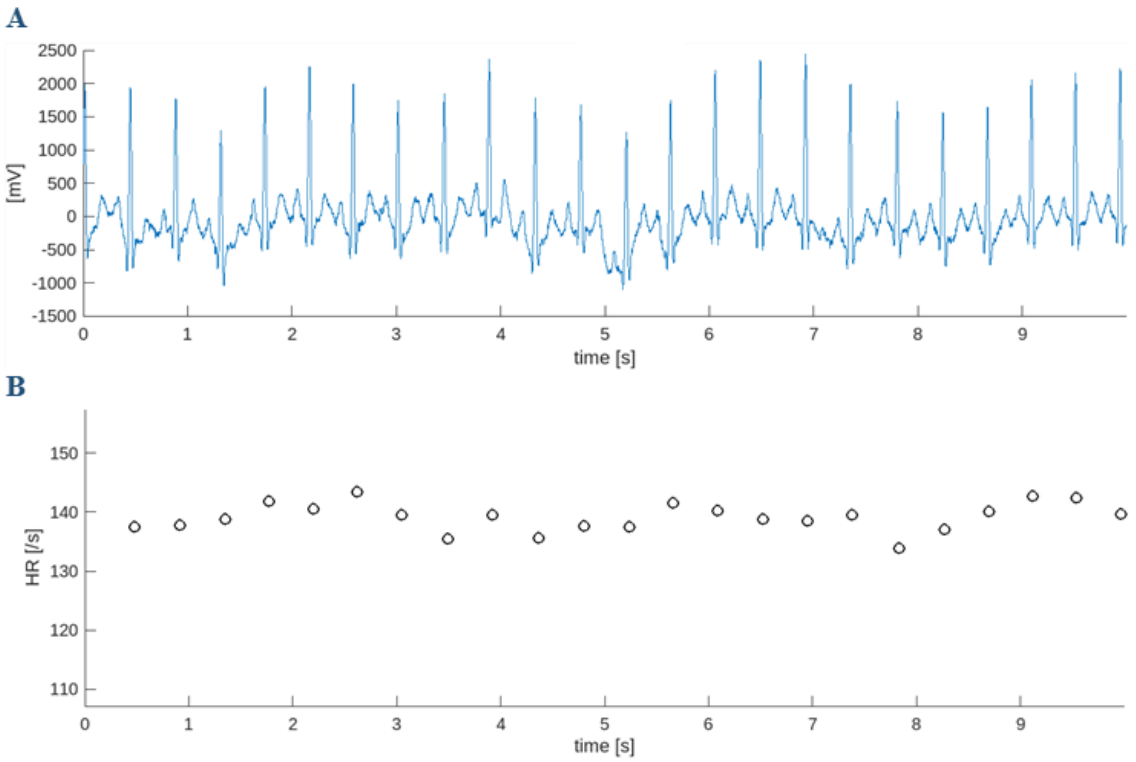


Figure 3. Regular electrocardiogram (A) and the corresponding heart rate (B) with a heart rate variability of 8 ms.

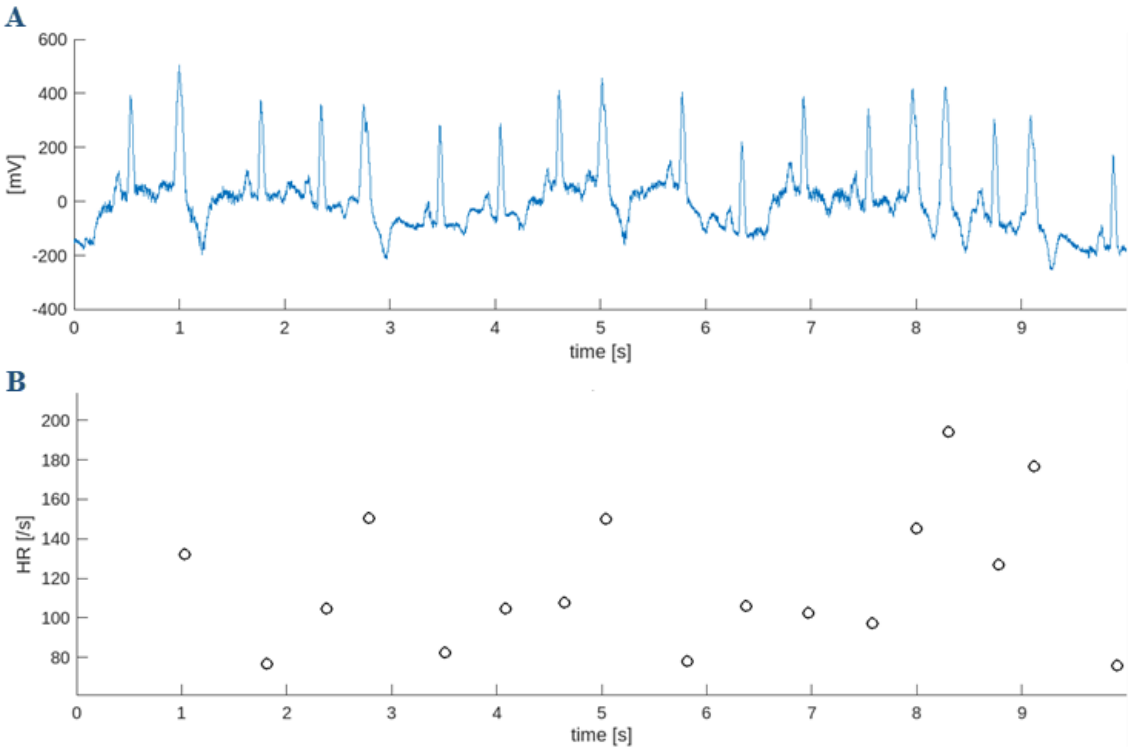


Figure 4. Irregular electrocardiogram (A) and the corresponding heart rate (B) with a heart rate variability of 154 ms.

5.3 Data pre-processing

The data pre-processing was performed in Python. First, patients without the outcome measure (PCWP) were removed and the non-invasive features were selected as the feature set for the primary analysis. Then, the data was split into a 20% validation and 80% train data set. Features with over 20% missing values or patients with over 50% missing features were removed. Then, all features were scaled based on the train data by removing the mean and dividing it by its variance. All remaining missing values were replaced with a value of 0. The correlations between the features were determined and for feature sets with a correlation higher than 0.5, only the feature with the highest correlation with the PCWP was selected, the other features were removed. As a result, for the blood pressure values (SBP, DBP and MAP), only the MAP was included and the height was excluded because of its correlation with gender.

5.4 Model development

To approximate the PCWP based on the non-invasive features, both regression and classification models were used. The models were all trained with the same train-validation split as during pre-processing. As the outcome measure of the regression models, the continuous PCWP values were used. For the classification models, the PCWP values were divided in two outcome classes: low (<12 mmHg) PCWP and high (≥12 mmHg) PCWP. This cut-off value was selected since normal wedge pressures range from 4 to 12 mmHg.(22)

Five different techniques (described in more detail later on) were used to build the models:

- Traditional statistical models
- K-nearest neighbours
- Random forest
- Multilayer perceptron (Deep learning)
- Gradient boosting

The Python library scikit-learn (23) was used for the first four techniques, while the latter used the Python package XGBoost (24). Every technique was used to build both a regression model and a classification model. The traditional models were directly fitted on the train data set. In contrast, for the ML algorithms, the train data was further split into 5-folds to optimise the model hyperparameters. Hyperparameters are algorithm settings that are assigned before the model training. In this study, 10 randomly selected value combinations of hyperparameters were tried per ML algorithm. The predefined options for the hyperparameter values can be found in the Supplementary Materials (Table S1). The combination of hyperparameter values with the best performance after 5-fold cross validation was selected and then fitted on the complete train data set.

5.4.1 Traditional statistical models

For the traditional regression models, a linear function (Figure 5A) and a second degree polynomial (Figure 5B) were fitted on the train data to estimate the continuous PCWP values. Similarly for the classification, a linear and second degree curve were fitted to divide the train data in low and high PCWP. The scikit-learn methods that were used were Linear Regression for the regression curves and Linear Discriminant Analysis (LDA) for the classifications. To create the second degree curves, the data set was expanded by the square of the features.

5.4.2 K-nearest neighbours

A k-nearest neighbours algorithm (Figure 5C) predicts the outcome values of a new datapoint (either categorical for classification or continuous for regression) based on the outcome values of a predefined number (k) of patients from the train data that have the closest feature values. The features must be scaled to make the distance comparison between the features correct.

5.4.3 Random forest

A random forest (Figure 5D) is based on the average outcome of multiple decision trees. A decision tree is a structure where each node splits the data set in smaller subgroups based on high or low values of a feature. Eventually, each patient ends in a terminal node with an estimate for the outcome value, either continuous for regression or categorical for classification.

5.4.1 Gradient boosting

A gradient boosting model (Figure 5E) is an ensemble of other prediction models, in this case decision trees. In contrast to a random forest, the training proceeds iteratively, adding new trees that take into account the errors of prior trees. Therefore, it usually outperforms a random forest.

5.4.1 Multilayer perceptron

A multilayer perceptron is a deep learning technique that comprises a neural network with at least three layers of nodes: an input layer, one or more hidden layers, and an output layer (Figure 5F). In every hidden layer, the neurons have a mathematical function that calculates the output based on the input of the neuron.

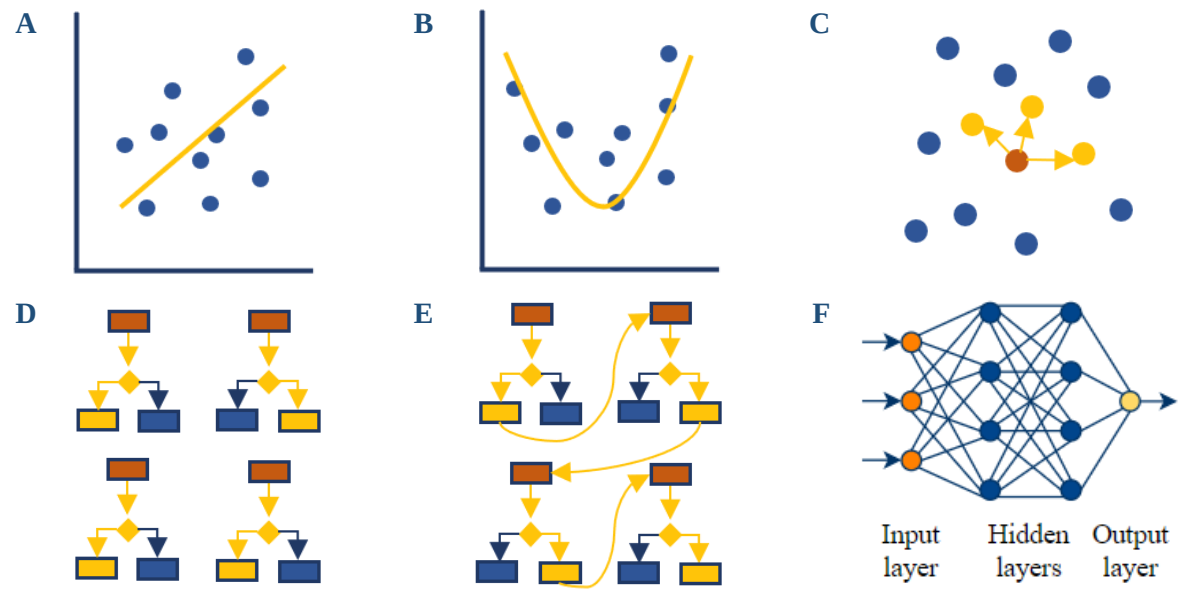


Figure 5. Schematic representation of the regression and classification techniques. **A.** Linear regressor or Linear Discriminant Analysis (LDA); **B.** Second degree regressor or LDA; **C.** Random forest; **D.** K-nearest neighbour; **E.** Gradient boosting. **F.** Multilayer perceptron.

5.5 Model evaluation

After the regression and classification models were built based on the train data, the resulting models were run on the validation data to determine the model performances.

The performance measure for the regression models was the coefficient of determination (R2). The R2 is the proportion of the variation in the outcome that is predicted by the model, calculated according to formula 2.

$$R2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \qquad (2)$$

Where y_i are the true PCWP values and f_i the predicted PCWP values of the validation set. The $\bar{y}$ is the mean of the validation data. Therefore, the numerator is the total sum of the squared errors of the estimated values and the denominator is the total sum of squares, which is proportional to the variance of the data. When a model is a perfect fit, the R2 is equal to 1 and when a constant function that always predicts the mean of the data is a better fit to the outcome of the model, the R2 becomes negative.

To determine the performances of the classification models, the area under the curve (AUC) of the Receiver Operator Characteristics (ROC) curves were determined. A value higher than 0.5 translates to a higher performance than random change when choosing an outcome class.

5.6 Secondary analyses

To test whether the performances of the models change with more specific and homogenous populations, the primary regression analysis was repeated with only HF patients and thereafter with all patients without valve disease.

Then, the regression models were trained with different sets of features. The first subset of features comprised the echocardiographic and laboratory features (additional features, Table 1) and the second subset comprised both the additional and the non-invasive features to determine the added value of echocardiographic and laboratory measurements for PCWP prediction.

6 Results

A total of 853 right-sided heart catheterisations from 791 unique patients met the inclusion criteria. The mean age of the patients was 58.9 years (SD ± 13.7), 47.9% were female, 31.3% had HF and 16.9% had any form of moderate/severe valve disease. The mean PCWP value was 13.1 mmHg (SD ± 8.6) and 48.7% of the values were 12 mmHg or higher. The train and validation data set had similar characteristics (Table 2).

Table 2. Population characteristics of the total data set, as well as the characteristics of the train data set and validation data set

	Total **N (%)**	**Train data** **N (%)**	**Validation data** **N (%)**
Total catheterisations	853	682	171
Male	444 (52.1%)	361 (52.9%)	83 (48.5%)
Female	409 (47.9%)	321 (47.1%)	88 (51.5%)
Heart failure	267 (31.3%)	213 (31.2%)	54 (31.6%)
Heart valve disease	144 (16.9%)	115 (16.9%)	29 (17.0%)
PCWP ≥ 12 mmHg	415 (48.7%)	338 (49.6%)	77 (45.0%)
	Mean ± SD	**Mean ± SD**	**Mean ± SD**
PCWP (mmHg)	13.1 ± 8.6	13.0 ± 8.4	13.2 ± 9.1
Age (years)	58.9 ± 13.7	59.1 ± 13.4	58.2 ± 15.2
LVEF (%)	48.8 ± 15.6	48.5 ± 15.1	50.1 ± 17.4
NT-proBNP (pmol/L)	455.8 ± 1086.8	386.5 ± 745.5	847.8 ± 2157.1
E/e'	11.4 ± 8.2	11.4 ± 8.6	11.3 ± 6.4
eGFR (mL/min)	69.6 ± 24.5	69.3 ± 24.1	71.1 ± 26.3
HRV (ms)	80.7 ± 65.4	80.9 ± 66.8	80.1 ± 59.3
Hemoglobin (mmol/L)	8.4 ± 1.32	8.4 ± 1.3	8.4 ± 1.3
MAP (mmHg)	88.3 ± 13.9	88.1 ± 13.5	89.3 ± 15.4
Sodium (mmol/L)	139.7 ± 3.5	139.7 ± 3.5	140.0 ± 3.6
Weight (kg)	78.3 ± 17.5	78.4 ± 17.3	78.1 ± 18.0
Heart rate (/min)	77.5 ± 14.4	77.4 ± 14.1	78.1 ± 15.7
Temperature (ºC)	36.5 ± 0.5	36.5 ± 0.4	36.6 ± 0.5
Saturation (%)	95.4 ± 2.9	95.4 ± 2.9	95.3 ± 2.8

Abbreviations: Number of patients (N), Standard Deviation (SD), Left Ventricular Ejection Fraction (LVEF), estimated Glomerular Filtration Rate (eGFR), N-Terminal pro-BNP (NT-proBNP), early mitral inflow velocity to early diastolic mitral annulus velocity ratio (E/e'), Heart Rate Variability (HRV), Mean Arterial Pressure (MAP)

Table 3 shows the correlations between the features and the PCWP. Scatterplots visualising the PCWP distributions per feature can be found in the supplementary materials (Figure S1). In the complete population, the LVEF had the strongest correlation with the PCWP of -0.32 and had even stronger negative correlations (-0.44) in the subpopulations without valve disease. This feature was followed by the NT-proBNP and E/e' with correlations in the complete population of 0.22 and higher values in most subpopulations. The HRV had the highest correlation with the PCWP out of the non-invasive features, with a correlation of 0.16 in the complete population.

Table 3. Correlations between the features and the pulmonary capillary wedge pressure for different subpopulations sorted by absolute correlation in the total population (All).

	All (N = 853)	HF (N = 267)	No valve disease (N=609)	HF without valve disease (N=178)
LVEF	-0.32	-0.26	-0.44	-0.44
NT-proBNP	0.22	0.19	0.26	0.26
E/e'	0.22	0.27	0.16	0.36
eGFR	-0.18	-0.05	-0.14	-0.02
HRV	0.16	0.11	0.16	0.21
Haemoglobin	-0.15	-0.10	-0.13	-0.15
MAP	-0.10	-0.09	-0.09	-0.08
Sodium	-0.10	-0.14	-0.07	-0.10
Weight	0.09	0.07	0.10	0.04
Gender	0.09	0.11	0.11	0.10
Age	0.04	0.01	0.03	0.07
Heart rate	-0.04	-0.02	-0.03	-0.01
Temperature	0.01	0.05	0.01	0.07
Saturation	0.00	-0.07	-0.02	-0.07

Abbreviations: Number of patients (N), Heart Failure (HF), Left Ventricular Ejection Fraction (LVEF), N-Terminal pro B-type Natriuretic Peptide (NT-proBNP), early mitral inflow velocity to early diastolic mitral annulus velocity ratio (E/e'), estimated Glomerular Filtration Rate (eGFR), Heart Rate Variability (HRV), Mean Arterial Pressure (MAP)

6.1 Primary analysis

The highest performance of the regression models was an R2 value of 0.11 for the second degree regression (Table 4). The classification algorithms for low (<12 mmHg) and high PCWP (≥12 mmHg) showed AUC of the ROC values of at most 0.59 (Table 4). All ROC curves are visualised in Figure 6.

Table 4. Performance results of the primary analysis. For every technique, a regression model and classification model were built and validated, resulting in an R2 value for every regression model and an AUC value for every classification model.

Technique	Regression R2	Classification AUC
Linear regression/classification	0.08	0.59
Second degree regression/classification	0.11	0.57
Random forest	0.10	0.55
K-nearest neighbours	0.05	0.58
Gradient boosting	0.09	0.54
Multilayer perceptron	0.05	0.57

Abbreviations: coefficient of determination (R2), Area Under the Curve (AUC)

6.2 Secondary analysis

Using different subgroups of the population (only HF patients or only patients without heart valve disease) or using additional features (the echocardiographic and laboratory features) did not increase the R2 values of the regression models (Table 5) compared to the models from the primary analysis (Table 4).

Table 5. Performance results of the secondary analyses. For every regression technique, a model was built and validated per data or feature subgroup. The four subgroups were: all HF patients with the non-invasive features; all patients without valve disease with the non-invasive features; all patients with the only the additional features; and all patients with all features (non-invasive and additional).

Technique	HF R2 (N = 267)	No valve disease R2 (N = 609)	Additional features R2 (N = 853)	All features R2 (N = 853)
Linear regression	-0.11	-0.02	0.06	-0.04
Second degree regression	-0.34	-1.12	0.00	-0.10
Random forest	-0.14	-0.06	0.01	0.04
K-nearest neighbours	-0.03	0.05	0.04	-0.01
Gradient boosting	-0.08	0.03	0.02	-0.12
Multilayer perceptron	-0.17	-0.56	-0.07	0.00

Abbreviations: Heart Failure (HF), Number of patients (N), coefficient of determination (R2)

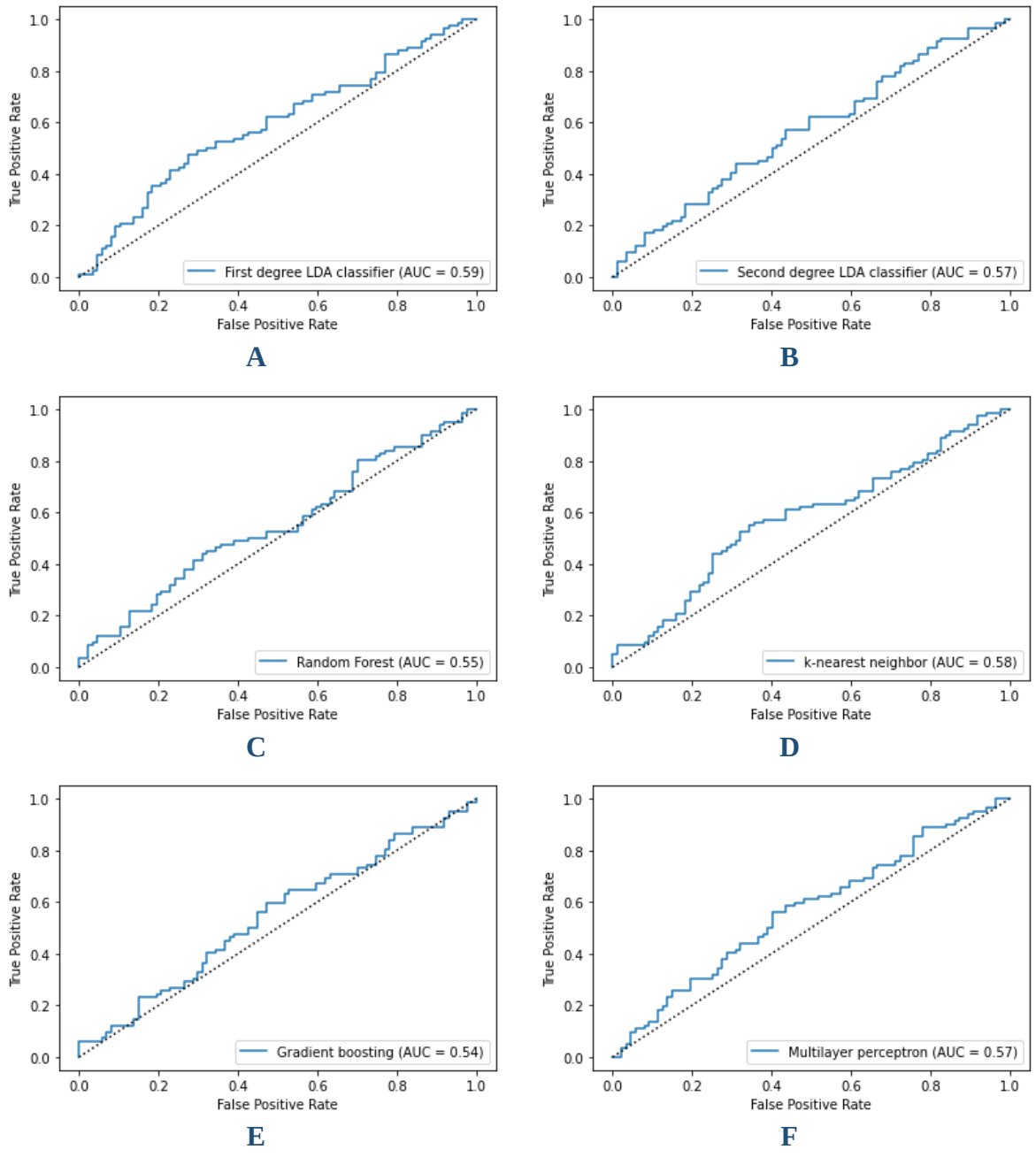


Figure 6. ROC-plots of the classification models. **A.** First degree Linear Discriminant Analysis (LDA) classifier; **B.** Second degree LDA classifier; **C.** Random Forest; **D.** k-Nearest neighbour; **E.** Gradient boosting. **F.** Multilayer perceptron.

7 PPG pilot

7.1 Introduction

As an additional pilot study, photoplethysmography (PPG) data of a prospective study was analysed to investigate the potential of PPG features for PCWP approximation. PPG sensors measure blood volume changes non-invasively by sending a continuous pulse of photons through the skin. The variable intensity of reflected photons is a representation of the heart rhythm.(20) Studies show that PPG signals could approximate the non-invasive blood pressure as there is a physiological relation between the pulse amplitude of the PPG signal and the blood pressure.(25,26) Moreover, the time between the systolic and diastolic peaks can be used as a measure for the arterial stiffness.(26–28) The aim of this additional analysis was to describe the relationship between those PPG features and the PCWP.

7.2 Methods

7.2.1 Study design and population

This pilot analysis was performed on prospectively collected PPG data from patients who underwent right catheterisation before Transcatheter Aortic Valve Implantation (TAVI) between 21/11/2022 and 07/02/2023 at the Erasmus MC. The study cohort comprised patients older than 18 years without a left ventricular assist device. The PPG data was collected with a Corsano Cardiowatch Bracelet (29) during the right catheterisation.

7.2.2 Data collection

The Corsano watch continuously measured PPG signals generated using green, red and infrared light. For this analysis, the green light signal with a wavelength of 525 nm was used, as it contains less artifacts.(30) The data had a sample frequency of 128 Hz.

In addition, it was determined whether the patients had valve disease, HF and atrial fibrillation (AF) at the time of the procedure. AF is a relevant diagnosis for this PPG analysis, since the morphology of PPG signals is affected by irregular rhythms when waveforms close together merge into each other. The diagnosis of AF was made by visual assessment of the perioperative ECGs for any sign of AF.

7.2.3 Signal pre-processing and feature extraction

For every raw PPG signal, a 70 seconds window right before the PCWP measurement was selected. In case of visually identifiable artefacts, the start time of the window was advanced by a maximum of 5 minutes. Then, a MATLAB algorithm was created to pre-process the PPG signal and extract the features. A 2^{nd} order Butterworth band-pass filter with a lower threshold of 1 Hz and an upper threshold of 5 Hz was used to filter out the wandering baseline (low frequencies) and noise (high frequencies). Afterwards, the first ten seconds of the signals were removed.

Six unique features were extracted (Figure 7): the HR, HRV, Systolic Amplitude (SA), Diastolic Amplitude (DA), Crest Time (CT) and Large Artery Stiffness Index (LASI), based on proposed features from literature.(26–28) To calculate these features, three points of the PPG morphology were used: the foot of a wave, the systolic peak and lastly the diastolic peak. Peak detection was used to detect the positive peaks of the signal (systolic peak) and the negative peaks (foot). The systolic peaks were used to determine the time interval between following heart beats in milliseconds. The HRV was calculated as the standard deviation of the interval values and the HR as 60 divided by the mean of the intervals. The systolic peaks and foots were also used to calculate the SA as the average vertical distance between them. The CT was defined as the average time between a foot and the following systolic peak. Finally, the diastolic peaks were detected as the first peak between the systolic peak and following foot. The DA was defined as the average vertical distance between the diastolic peaks and the directly following foot and the LASI was calculated as the average time between every systolic peak and the following diastolic peak.

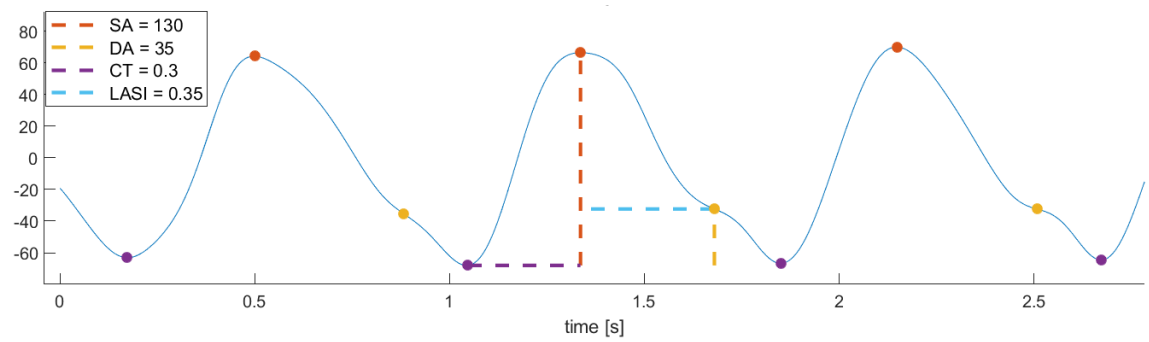


Figure 7. Features extracted from the photoplethysmography (PPG) signal: Systolic Amplitude (SA), Diastolic Amplitude (DA), Crest Time (CT) and Large Artery Stiffness Index (LASI). Additionally, the distance between detected systolic peaks was used to calculate the heart rate and heart rate variability.

7.2.4 Model development and validation

The features were used to estimate the PCWP values with a simplified version of the pipeline that was used before. Because of the small sample size of this pilot study, only the traditional statistical techniques (linear regression and LDA classification) were used to estimate the PCWP. The models were based on a single feature at a time and validated with 3-fold cross-validation. The second degree regression and LDA models were not applied to avoid over-fitting of the models since this would require an expansion of the feature set. The regression models were evaluated with the resulting R2 value and the classification models with the AUC of the ROC curve. Additionally, for features that resulted in AUC values above 0.50, the two PCWP classes were statistically compared using a Mann-Whitney U test. The significance level was set to 0.05, corrected for multiple testing using Bonferroni.

7.3 Results

A total of 14 patients with available PCWP values and PPG signals were included. Since the measurements were performed just before a TAVI, all patients had significant aortic stenosis. Additionally, 28.6% of the patients had heart failure, 14.3% had atrial fibrillation and 78.6% was male. The median PCWP value was 11.5 mmHg and 50.0% of the values were 12 mmHg or higher. All population characteristics are presented in Table 6. The features did not have a normal distribution due to the small population and therefore, the median and the lower and upper quartile were used to describe the feature distributions.

Table 6. Population characteristics and the correlations between the PCWP and the PPG features.

	N (%)		
Total catheterisations	14		
Male	11 (78.6%)		
Female	3 (21.4%)		
Heart failure	4 (28.6%)		
Heart valve disease	14 (100.0%)		
Atrial fibrillation	2 (14.3%)		
PCWP ≥ 12 mmHg	7 (50.0%)		
	Median	**Q1**	**Q2**
PCWP (mmHg)	11.5	5.5	13.5
HRV (ms)	50.9	33.2	93.4
Diastolic Amplitude	55.4	40.1	78.2
Systolic Amplitude	181.9	153.4	395.2
Heart rate (/s)	72.1	67.5	73.9
LASI (s)	0.32	0.27	0.37
Crest time (s)	0.30	0.27	0.33

Abbreviations: Pulmonary Capillary Wedge Pressure (PCWP), Photoplethysmography (PPG), Number of patients (N), lower Quartile (Q1), upper Quartile (Q3), Heart Rate Variability (HRV), Large Artery Stiffness Index (LASI)

All PPG features resulted in regression models with negative R2 values, which is in line with their low correlations with the PCWP (Table 7). Scatterplots visualising the PCWP distributions per feature can be found in the supplementary materials (Figure S2). The classification models resulted in higher performances: the model based on the SA and the model based on the LASI both resulted in an AUC of 0.86 and the model based on the CT resulted in an AUC of 0.72 (Table 7). The ROC-curves of these features are shown in figure 8.

Table 7. Regression and classification performance results (R2 and AUC respectively) of the models based on the individual PPG features.

Feature	Correlation with the PCWP	Regression R2	Classification AUC
Heart rate (/s)	0.13	-1.42	0.36
HRV (ms)	0.09	-2.47	0.50
SA	0.08	-0.79	0.86
DA	0.07	-0.93	0.44
CT (s)	0.06	-2.74	0.72
LASI (s)	0.05	-1.14	0.86

Abbreviations: coefficient of determination (R2), Area Under the Curve (AUC), Photoplethysmography (PPG), Pulmonary Capillary Wedge Pressure (PCWP), Heart Rate Variability (HRV), Systolic Amplitude (SA), Diastolic Amplitude (DA), Crest Time (CT) and Large Artery Stiffness Index (LASI)

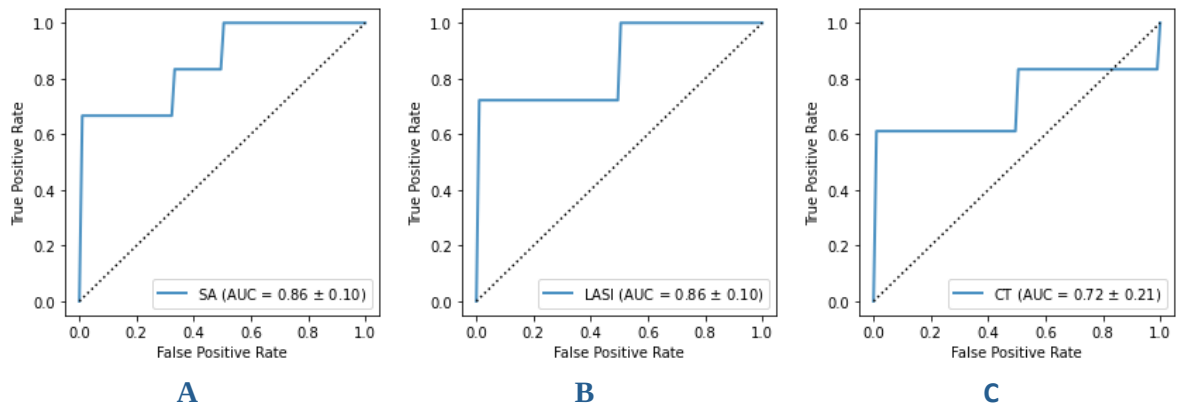


Figure 8. ROC-curves of the best performing classification models. **A**. Model based on the Systolic Amplitude (SA) with an AUC of 0.86. **B**. Model based on the Large Artery Stiffness Index (LASI) with an AUC of 0.86. **C.** Model based on the Crest Time (CT) with an AUC of 0.72.

The distribution of the SA and the LASI per PCWP class (Figure 9) showed that a reasonable distinction can be made between the two classes based on these individual features, in line with the high AUC scores of the models based on the SA or LASI. Notably, the two patients with AF are regularly at the edge of their class spread, illustrating the difference in morphology. Despite the visible difference between the classes for every feature, the SA, LASI and CT did not significantly differ between the PCWP classes according to the Mann-Whitney tests after Bonferroni correction (significance level $\alpha = 0.0167$, Table 8).

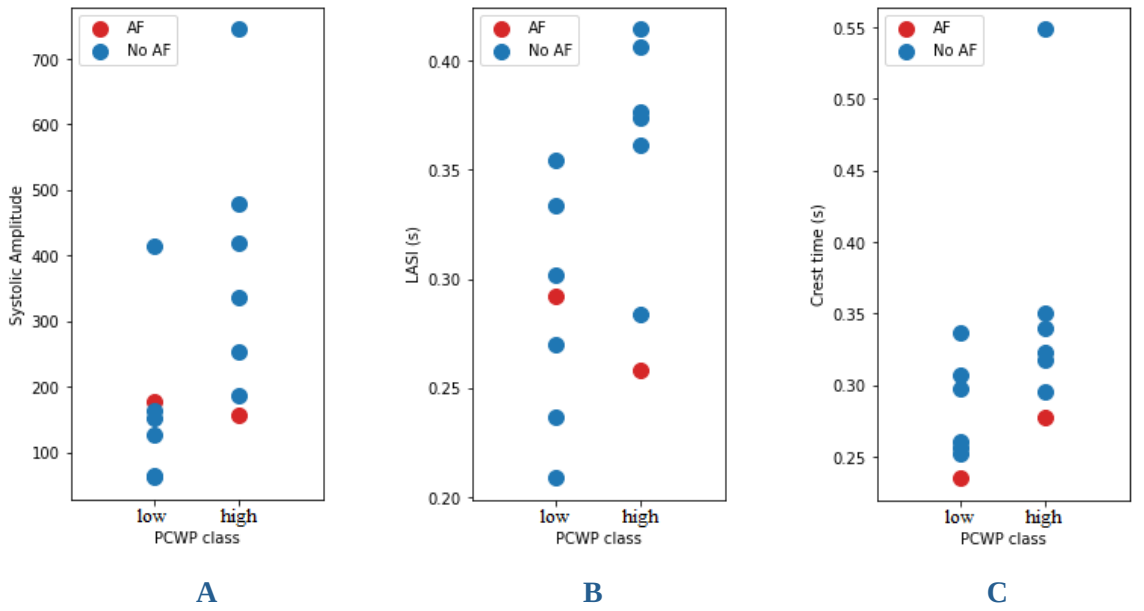


Figure 9. Distribution of the systolic amplitude (**A**), Large Artery Stiffness Index (LASI) (**B**) and crest time (**C**) for the two Pulmonary Capillary Wedge Pressure (PCWP) classes (low (<12 mmHg) and high (≥12 mmHg)). The patients with Atrial Fibrillation (AF) are marked red, patients without AF are marked blue.

Table 8. Median and quartile values of the SA, LASI and CT values per PCWP class and the p-value of the Mann-Whitney U comparison between the classes, with the significance level set to $\alpha = 0.0167$ after Bonferroni correction for mutliple testing.

	Low PCWP			High PCWP			
Feature	**Median**	**Q1**	**Q2**	**Median**	**Q1**	**Q2**	**P-value**
SA	152.7	96.2	170.7	336.3	220.0	449.6	0.02
LASI (s)	0.29	0.25	0.32	0.37	0.32	0.39	0.05
CT (s)	0.26	0.25	0.30	0.32	0.31	0.35	0.04

Abbreviations: Systolic Amplitude (SA), Crest Time (CT) and Large Artery Stiffness Index (LASI), Pulmonary Capillary Wedge Pressure (PCWP), lower Quartile (Q1), upper Quartile (Q3)

8 Discussion

In this retrospective study, the main research aim to approximate the invasively measured PCWP with the non-invasive MAP, saturation, heart rate, HRV, temperature, weight, age and gender was unsuccessful. However, in an additional prospective pilot analysis, different features of PPG signals showed discriminatory power between patients with low and high PCWP values.

The low performances of the regression and classification models based on the retrospectively collected non-invasive measurements suggest that vital parameters of a single point in time do not provide sufficient information about the patient's volume status. Yet the traditional model of HF care primarily focuses on monitoring vital signs and symptoms of the patient, as outlined in the guidelines.(16) Estimated jugular venous pressure, graded oedema and shortness of breath appeared to be unreliable in determining the PCWP as well in a similar study by Polcz et al.,(31) further reflecting the shortcomings of signs and symptoms for HF monitoring. However, a recent meta-analysis concluded that remote monitoring of vital signs and symptoms in the HF population reduces all-cause mortality and HF related hospitalisations, with structured telephone support based on regular assessments outperforming other categories of remote monitoring.(32) Therefore, the clinical relevance of the monitoring programs using non-invasive measurements may lie in detected changes over time. For instance, Churpek et al. showed that including the trends of vital signs can increase the accuracy of models designed to detect critical illness.(33) A limitation of our study was the inability to perform a trend analysis due to the data heterogeneity. The time intervals between measurements varied and a substantial number of patients had only one vital sign measurement in the 72-hour time interval before or after the procedure. A future prospective study with several structured vital signs measurements in an appropriate time interval – at least multiple days – in advance of the procedure is needed to determine the clinical relevance of more frequently measured non-invasive parameters. Moreover, another benefit of a measurement protocol is that it can improve the overall data quality by limiting the variation in measurement execution and notation.

Similarly, the secondary analysis showed that echocardiographic and laboratory values also failed to predict the PCWP, despite the slightly higher correlations with the PCWP. Since the clinical relevance of these parameters may lie primarily in changes over time, future prospective studies should include several echocardiographic and laboratory measures in the protocol when assessing the potential of these measures.

Another possible direction for further research into the non-invasive estimation of PCWP is the use of PPG signals. Although the population was small and therefore sensitive to overfitting, the results of the included prospective pilot study showed the potential of parameters extracted from PPG signals to classify patients based on low or high PCWP values. The SA and the LASI of the signals both had high AUC scores of 0.86, despite the lack of a significant difference

between the classes. Therefore, the PPG signals seem to contain relevant information which could be further substantiated in future studies with an expansion of the population. In addition, the PPG sensors were incorporated in easy-to-wear wristbands, which makes it possible for future research to measure the signals over a longer period and thus to include trend analyses. A challenge that will arise when measuring the PPG signals for a longer period is the varying signal quality due to motion artifacts or temporarily reduced contact with the skin. To improve signal quality, averaging the green signal measured by the two available sensors could reduce noise, although averaging signals can lead to information loss. For the current analysis, the varying signal quality was not a substantial problem because of the selection of a short time window in the signals and therefore, averaging was not needed.

Although the current study already included a feature derived from ECG signals (HRV) without any power to estimate the PCWP values, there might still be useful information in the ECG signals. Raghu et al. (34) identified patients with a PCWP higher than 18 mmHg with an AUC of 0.82 based on ECG signals. This higher cut-off value for the low or high PCWP determination did not improve our classification results based on the non-invasive features, including HRV. Therefore, the promising results of Raghu et al. suggest that expanding the non-invasive feature set with additional ECG features can be beneficial. Although they used a deep learning model and therefore could not identify specific ECG features that resulted in the classification, the classification seemed to depend on the P-wave and QRS-complex amplitudes and barely on the T-wave amplitude.

8.1 Future perspective

This study showed that a single measurement moment for vital signs does not provide sufficient information about the patient's volume status. Therefore, there is a need for large prospective studies to evaluate the clinical value of trends in non-invasive measurements. The accessibility and convenience of wearable devices could enable more frequent measurements of vital signs for remote monitoring of HF patients based on PPG or ECG signals. For example, the Apple watch (35) and Fitbit (36) can measure the HR or HRV using PPG sensors and the VitalPatch described by Stehlik et al. (37) is FDA approved for ECG monitoring. The latter already built a ML algorithm to predict HF readmission and achieved an accuracy comparable to implanted devices that measure the PA pressures.

8.2 Conclusion

To conclude, this large retrospective study showed the shortcomings of sporadic measurements of non-invasive MAP, saturation, heart rate, HRV, temperature and weight for volume status evaluation. Both traditional statistical and ML models based on single non-invasive measurements could not approximate the PCWP. However, the SA and LASI of PPG signals showed discriminatory power between patients with low and high PCWP values. Therefore, future prospective research is needed to evaluate the clinical benefit of wearable devices measuring trends based on PPG signals.

9 References

1. Kumar P, Clark M. Clinical Medicine. Elsevier; 2017. 980–991 p.
2. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. Nat Rev Cardiol. 2011 Jan;8(1):30.
3. Savarese G, Lund LH. Global Public Health Burden of Heart Failure. Card Fail Rev. 2017;3(1):7.
4. Krumholz HM, Merrill AR, Schone EM, Schreiner GC, Chen J, Bradley EH, et al. Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission. Circ Cardiovasc Qual Outcomes. 2009 Sep;2(5):407–13.
5. Dharmarajan K, Hsieh AF, Lin Z, Bueno H, Ross JS, Horwitz LI, et al. Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. JAMA. 2013 Jan 23;309(4):355–63.
6. Klabunde RE. Normal and Abnormal Blood Pressure (Physiology, Pathophysiology & Treatment) [Internet]. 2013 [cited 2023 Feb 28]. Available from: https://www.cvphysiology.com/Heart%20Failure/HF008
7. Boorsma EM, ter Maaten JM, Damman K, Dinh W, Gustafsson F, Goldsmith S, et al. Congestion in heart failure: a contemporary look at physiology, diagnosis and treatment. Nature Reviews Cardiology 2020 17:10. 2020 May 15;17(10):641–55.
8. Mooney DM, Fung E, Doshi RN, Shavelle DM. Evolution from electrophysiologic to hemodynamic monitoring: the story of left atrial and pulmonary artery pressure monitors. Front Physiol. 2015;6(OCT):271.
9. Abraham WT. The Role of Implantable Hemodynamic Monitors to Manage Heart Failure. Heart Fail Clin. 2015 Apr 1;11(2):183–9.
10. Hawkins NM, Virani SA, Sperrin M, Buchan IE, McMurray JJV, Krahn AD. Predicting heart failure decompensation using cardiac implantable electronic devices: a review of practices and challenges. Eur J Heart Fail. 2016 Aug 1;18(8):977–86.
11. Dickinson MG, Allen LA, Albert NA, DiSalvo T, Ewald GA, Vest AR, et al. Remote Monitoring of Patients With Heart Failure: A White Paper From the Heart Failure Society of America Scientific Statements Committee. J Card Fail. 2018 Oct 1;24(10):682–94.
12. Adamson PB, Abraham WT, Stevenson LW, Desai AS, Lindenfeld J, Bourge RC, et al. Pulmonary Artery Pressure-Guided Heart Failure Management Reduces 30-Day Readmissions. Circ Heart Fail. 2016 Jun 1;9(6).
13. Fox K, Ford I, Steg PG, Tendera M, Robertson M, Ferrari R. Heart rate as a prognostic risk factor in patients with coronary artery disease and left-ventricular systolic

dysfunction (BEAUTIFUL): a subgroup analysis of a randomised controlled trial. Lancet. 2008;372(9641):817–21.

14. Sydó N, Sydó T, Gonzalez Carta KA, Hussain N, Farooq S, Murphy JG, et al. Prognostic Performance of Heart Rate Recovery on an Exercise Test in a Primary Prevention Population. J Am Heart Assoc. 2018 Apr 1;7(7).

15. Singh N, Moneghetti KJ, Christle JW, Hadley D, Froelicher V, Plews D. Heart Rate Variability: An Old Metric with New Meaning in the Era of Using mHealth technologies for Health and Exercise Training Guidance. Part Two: Prognosis and Training. Arrhythm Electrophysiol Rev. 2018 Dec 1;7(4):247–55.

16. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failureThe Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J. 2016 Jul 14;37(27):2129–200.

17. Lip GYH, Skjøth F, Overvad K, Rasmussen LH, Larsen TB. Blood pressure and prognosis in patients with incident heart failure: the Diet, Cancer and Health (DCH) cohort study. Clin Res Cardiol. 2015 Dec 1;104(12):1088–96.

18. Varma N, Marrouche NF, Aguinaga L, Albert CM, Arbelo E, Choi JI, et al. HRS/EHRA/APHRS/LAHRS/ACC/AHA worldwide practice update for telehealth and arrhythmia monitoring during and after a pandemic. J Arrhythm. 2020 Oct 1;36(5):813–26.

19. Samol A, Bischof K, Luani B, Pascut D, Wiemer M, Kaese S. Single-Lead ECG Recordings Including Einthoven and Wilson Leads by a Smartwatch: A New Era of Patient Directed Early ECG Differential Diagnosis of Cardiac Diseases? Sensors (Basel). 2019 Oct 2;19(20):4377.

20. Kamiŝalić A, Fister I, Turkanović M, Karakatiĉ S. Sensors and Functionalities of Non-Invasive Wrist-Wearable Devices: A Review. Sensors 2018, Vol 18, Page 1714. 2018 May 25;18(6):1714.

21. Dagher L, Shi H, Zhao Y, Marrouche NF. Wearables in cardiology: Here to stay. Heart Rhythm. 2020 May 1;17(5 Pt B):889–95.

22. Nair R, Lamaa N. Pulmonary Capillary Wedge Pressure. StatPearls [Internet]. 2022 Apr 21 [cited 2023 Mar 14]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK557748/

23. scikit-learn: machine learning in Python — scikit-learn 1.2.2 documentation [Internet]. [cited 2023 Mar 19]. Available from: https://scikit-learn.org/stable/index.html

24. Python Package Introduction — xgboost 1.7.4 documentation [Internet]. [cited 2023 Mar 19]. Available from: https://xgboost.readthedocs.io/en/stable/python/python_intro.html

25. Riaz F, Azad MA, Arshad J, Imran M, Hassan A, Rehman S. Pervasive blood pressure monitoring using Photoplethysmogram (PPG) sensor. Future Generation Computer Systems. 2019 Sep 1;98:120–30.

26. Kachuee M, Mahdi Kiani M, Mohammadzade H, Shabany M. Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time. 2015 IEEE International Symposium on Circuits and Systems (ISCAS). 2015;1006–9.

27. Mok Ahn J. New aging index using signal features of both photoplethysmograms and acceleration plethysmograms. Healthc Inform Res. 2017 Jan 1;23(1):53–9.

28. Mehrabbeik M, Rashidi S. Estimation of Cuffless Systolic and Diastolic Blood Pressure Using Pulse Transient Time. 5th Iranian Conference on Signal Processing and Intelligent Systems, ICSPIS 2019. 2019 Dec 1.

29. Bracelet - Corsano Health [Internet]. [cited 2023 Mar 19]. Available from: https://corsano.com/products/bracelet-2/

30. Lee J, Matsumura K, Yamakoshi KI, Rolfe P, Tanaka S, Yamakoshi T. Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion. Annu Int Conf IEEE Eng Med Biol Soc. 2013;2013:1724–7.

31. Polcz M, Huston J, Breed M, Case M, Leisy P, Schmeckpeper J, et al. Comparison of clinical symptoms and bioimpedance to pulmonary capillary wedge pressure in heart failure. American Heart Journal Plus: Cardiology Research and Practice. 2022 Mar 1;15:100133.

32. Scholte NTB, Gürgöze MT, Aydin D, Theuns DAMJ, Manintveld OC, Ronner E et al. Efficacy of Telemonitoring Modalities in Heart Failure: A State-of-the-Art Systematic Review and Meta-Analysis. Eur Heart J. Ahead of publication.

33. Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. Resuscitation. 2016 May 1;102:1–5.

34. Raghu A, Schlesinger D, Pomerantsev E, Devireddy S, Shah P, Garasic J, et al. ECG-guided non-invasive estimation of pulmonary congestion in patients with heart failure. Scientific Reports 2023 13:1. 2023 Mar 9;13(1):1–10.

35. Monitor your heart rate with Apple Watch - Apple Support [Internet]. [cited 2023 Feb 17]. Available from: https://support.apple.com/en-us/HT204666

36. Irregular Rhythm; Check for atrial fibrillation with Fibit [Internet]. [cited 2023 Feb 17]. Available from: https://www.fitbit.com/global/us/technology/irregular-rhythm

37. Stehlik J, Schmalfuss C, Bozkurt B, Nativi-Nicolau J, Wohlfahrt P, Wegerich S, et al. Continuous Wearable Monitoring Analytics Predict Heart Failure Hospitalization: The LINK-HF Multicenter Study. Circ Heart Fail. 2020;13(3):e006513.

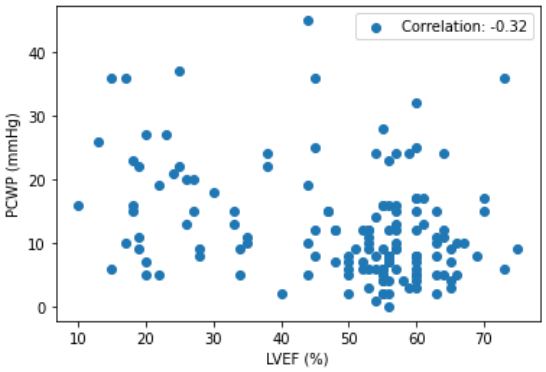
10 Supplementary Materials

10.1 Hyperparameters Machine learning models

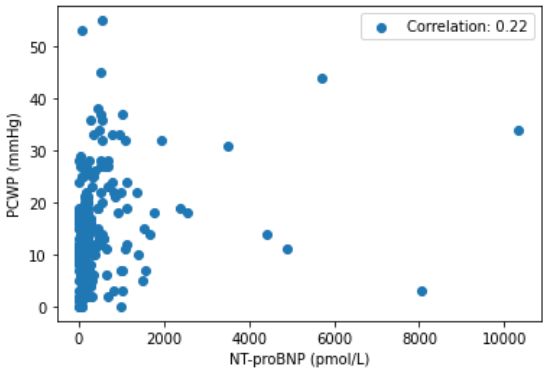
Table S1. Used Python package and hyperparameter options for the Machine Learning models

Technique	Python package	Python module	Hyperparameter	Predefined range
Random forest	Sklearn.ensemble	RandomForestRegressor / RandomForestClassifier	'n_estimators' 'bootstrap' 'max_depth'	list(range(100,301)) [True, False] [4, 5, 6, 7, 8, 9, 10, None]
K-nearest neighbor	Sklearn.neighbors	KNeighborsRegressor / KNeighborsClassifier	'n_neighbors' 'weights'	list(range(50,200,2)) ['uniform', 'distance']
Multilayer perceptron	Sklearn. neural_network	MLPRegressor / MLPClassifier	'hidden_layer_sizes' 'solver' 'max_iter'	[(10,), (20,), (50,), (100,)] ['lbfgs', 'sgd', 'adam'] list(range(1000,3000,200))
Gradient boosting	xgboost	XGBRegressor /XGBClassifier	'max_depth' 'nthread'	list(range(1,10)) list(range(1,10))

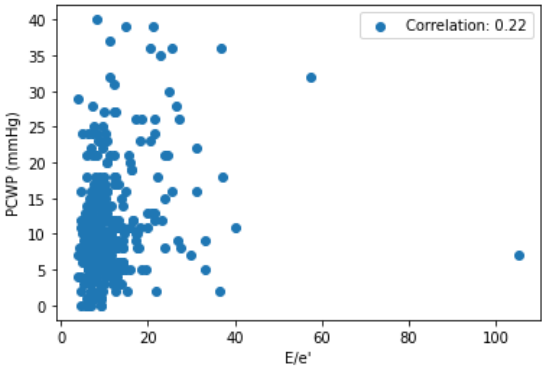
10.2 Scatterplots retrospective features and PCWP



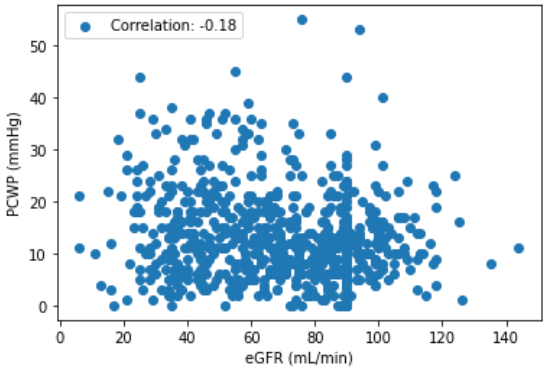
A LVEF



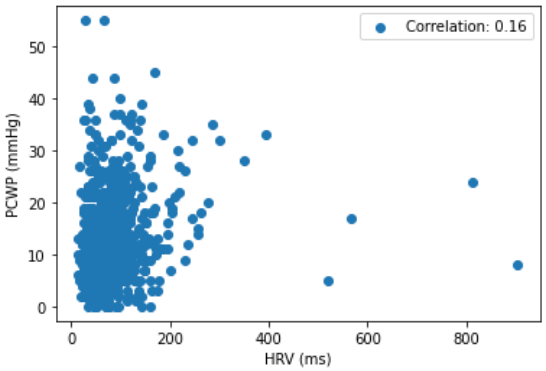
B NT-proBNP



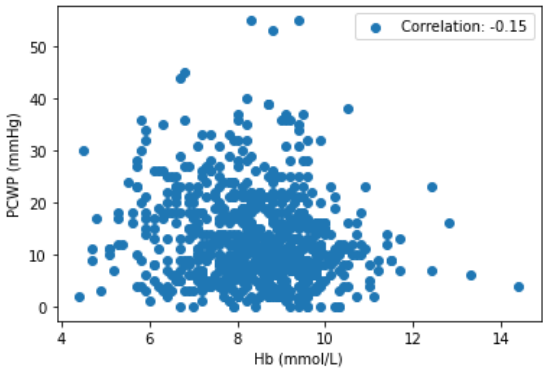
C E/e'



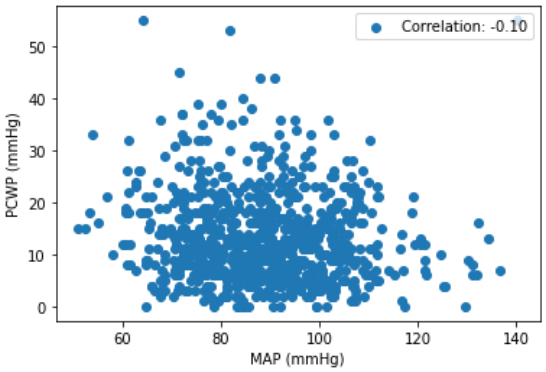
D eGFR



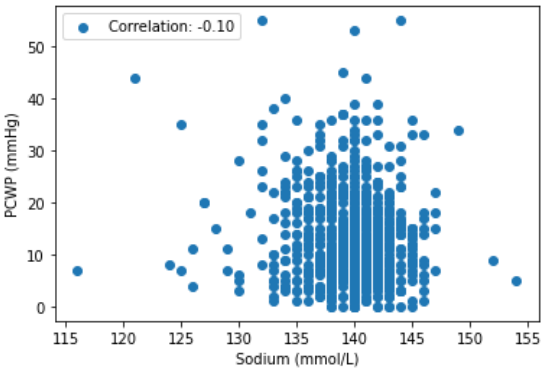
E Heart rate variability



F Hemoglobin



G Mean Arterial Pressure



H Sodium

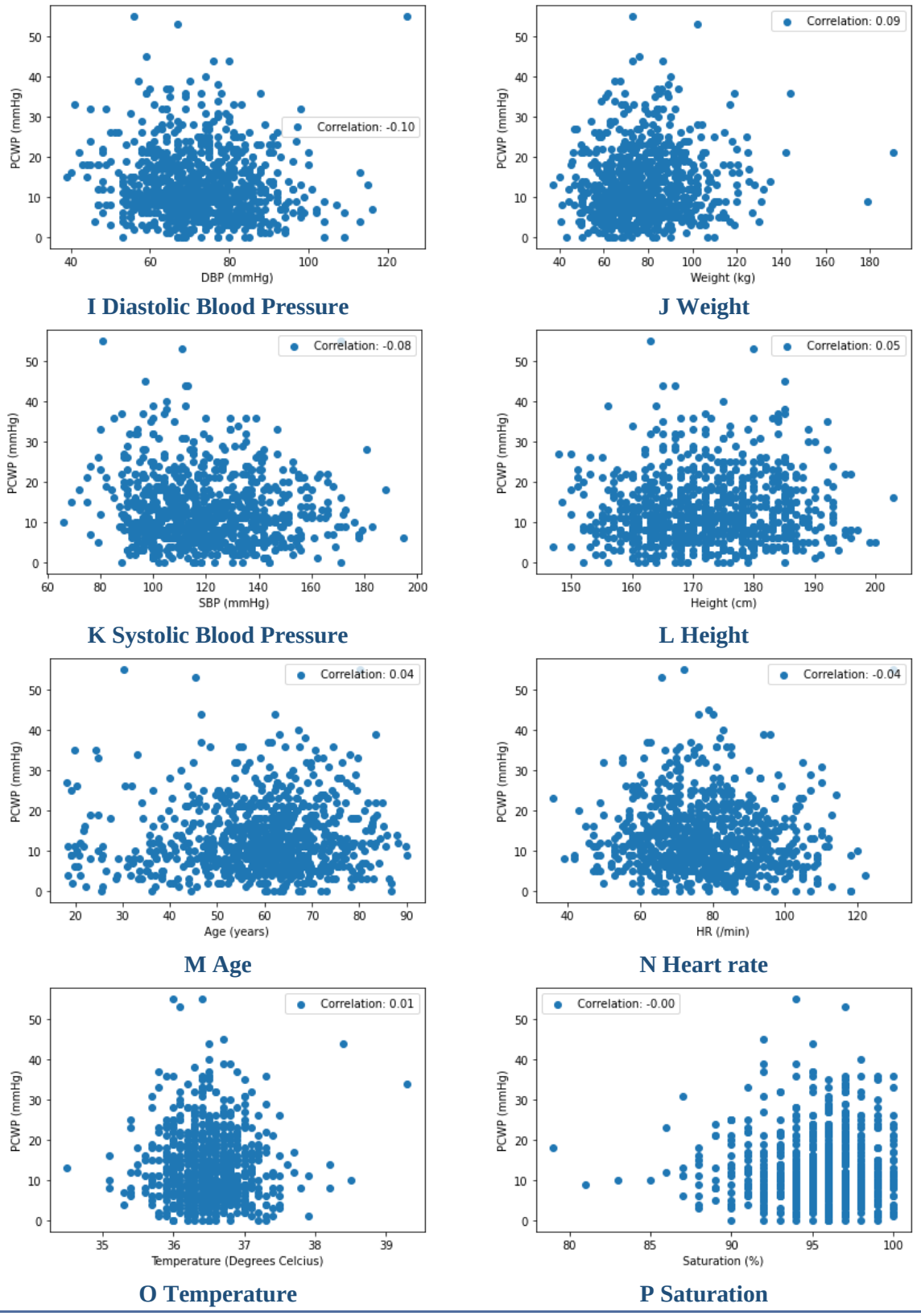


Figure S1. Scatterplots of the retrospectively collected features versus pulmonary capillary wedge pressure (PCWP), sorted by the absolute value of the corresponding correlation coefficient. **Abbreviations:** Left Ventricular Ejection Fraction (LVEF), estimated Glomerular Filtration Rate (eGFR), N-Terminal pro B-type Natriuretic Peptide (NT-proBNP), early mitral inflow velocity to early diastolic mitral annulus velocity ratio (E/e')

10.3 Scatterplots PPG features and PCWP

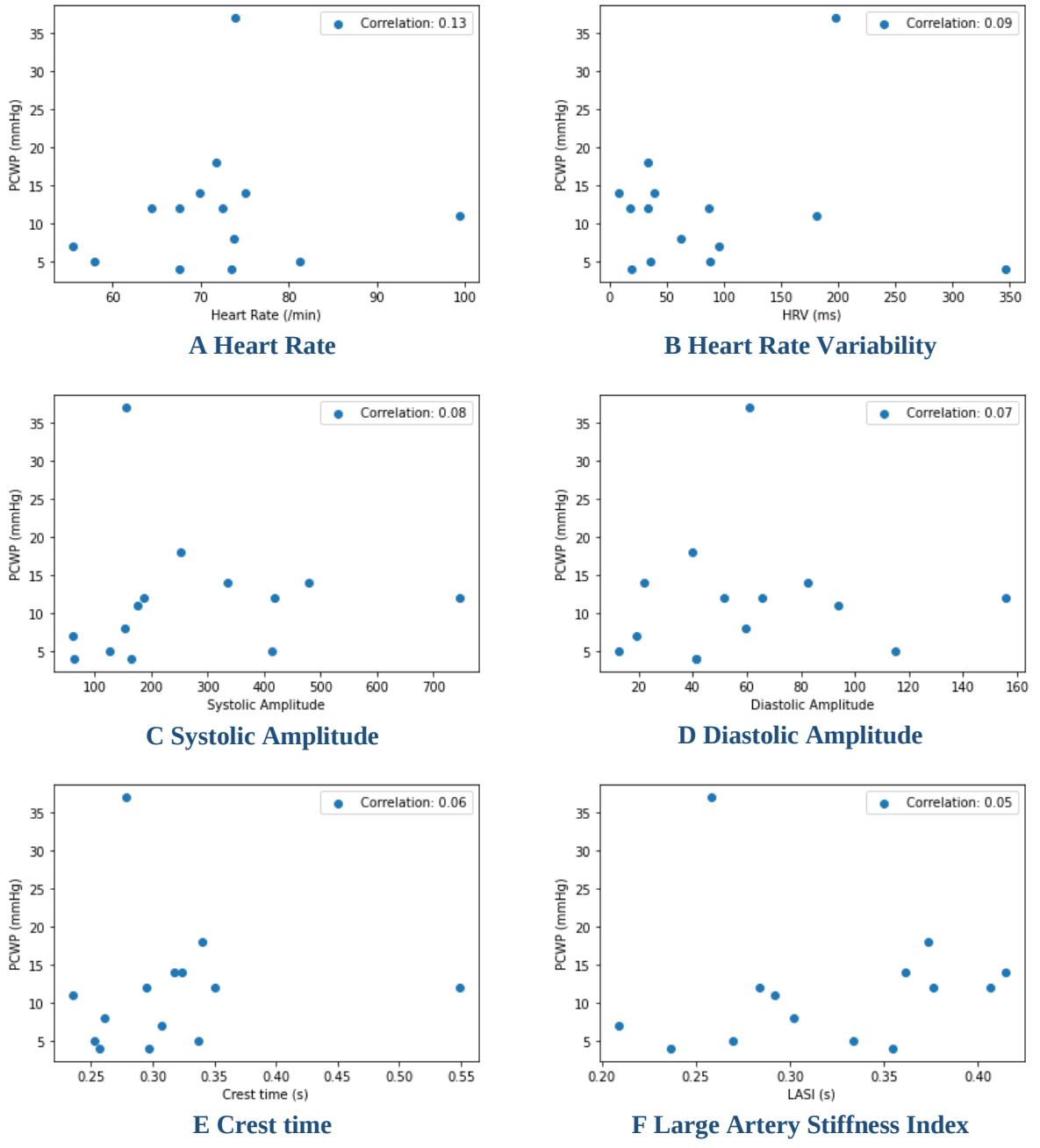


Figure S2. Scatterplots of the photoplethysmography features versus pulmonary capillary wedge pressure (PCWP), sorted by the absolute value of the corresponding correlation coefficient.