Scientific Reports

https://doi.org/10.1038/s41598-026-56217-x

Article in Press

Meta-heuristic and machine learning based functional capacity prediction using gait parameters in patients with heart failure

Received: 27 December 2025

Accepted: 29 May 2026

Published online: 03 June 2026

Cite this article as: Tanriverdi Eyolcu A., Doğru A., Buyrukoğlu S. *et al.* Meta-heuristic and machine learning based functional capacity prediction using gait parameters in patients with heart failure. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-56217-x

Aylin Tanriverdi Eyolcu, Ayşe Doğru, Selim Buyrukoğlu, Sema Savci, Buse Ozcan Kahraman & Mehmet Birhan Yilmaz

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Full title: Meta-heuristic and machine learning based functional capacity prediction using gait parameters in patients with heart failure

Aylin Tanriverdi Eyolcu[1*], Ayşe Doğru[2], Selim Buyrukoğlu[3], Sema Savci[4], Buse Ozcan Kahraman[5], Mehmet Birhan Yilmaz[6]

[1] Department of Physiotherapy and Rehabilitation, Çankırı Karatekin University, Çankırı, Türkiye

[2] Department of Information Technology, Çankırı Karatekin University, Çankırı, Türkiye

[3] Department of Computer Engineering, Çankırı Karatekin University, Çankırı, Türkiye

[4] Department of Physiotherapy and Rehabilitation, Acıbadem Mehmet Ali Aydınlar University, Istanbul, Türkiye

[5] Faculty of Physical Therapy and Rehabilitation, Dokuz Eylül University, Izmir, Türkiye

[6] Department of Cardiology, Dokuz Eylül University, Izmir, Türkiye

ARTICLE IN PRESS

Short running head: Machine learning based prediction of functional capacity

***Corresponding Author:** Aylin Tanriverdi Eyolcu

Telephone number: +905352359989

E-mail address: aylintanriverdi@karatekin.edu.tr

Postal address: Uluyazı Campus of Çankırı Karatekin University, Faculty of Health Science, 18100, Çankırı, Turkey

ARTICLE IN PRESS

Meta-heuristic and machine learning based functional capacity prediction using gait parameters in patients with heart failure

This study aimed to predict functional capacity using wearable sensor-based spatiotemporal and kinematic gait parameters, and meta-heuristic-based machine learning approaches in patients with heart failure (HF). This cross-sectional study included 70 patients with HF and assessed functional capacity and gait parameters using the six-minute walk test and a wearable inertial sensor, respectively. The Synthetic Minority Over-Sampling Technique for Regression was employed to increase the number of instances. The optimisation of four machine learning models (XGBoost, LightGBM, Random Forest, CatBoost) was implemented using Simulated Annealing, Genetic Algorithms, Particle Swarm Optimisation (PSO) and Bayesian Optimisation. The model performance was evaluated via 10-fold cross-validation through the R-squared coefficient (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean squared error (MSE). The interpretability of the optimal model was investigated using SHapley Additive exPlanations (SHAP) to explain individual feature contributions. The hybrid PSO-CatBoost model provided better predictive accuracy (R^2=0.9456; RMSE=17.0063; MAE=10.2239; MSE=289.2154) as compared to other hybrid optimisation-model configurations. SHAP ranked right stride length and gait speed as the most important features. The hybrid PSO-CatBoost demonstrated high predictive accuracy in estimating functional capacity in patients with HF using gait parameters, but its potential as a clinical decision-support tool requires further validation.

Keywords: CatBoost, Functional capacity, Gait analysis, Heart failure, Particle swarm optimisation

ARTICLE IN PRESS

Introduction

A universal definition of heart failure (HF) is that 'HF is a clinical syndrome with symptoms and/or signs caused by a structural and/or functional cardiac abnormality and corroborated by elevated natriuretic peptide levels and/or objective evidence of pulmonary or systemic congestion'[1]. The prevalence of HF is 56 million people on a global scale, and rises with age in both women and men, with high mortality and morbidity rates[2,3]. This condition is progressive if properly unmanaged, as the inability of the heart to pump efficiently causes inadequate blood flow to organs and tissues, stimulating compensatory mechanisms that can worsen the condition over time[4]. Orthopnoea, paroxysmal nocturnal dyspnea, fatigue, ankle swelling, reduced exercise tolerance, and inability to exercise are symptoms frequently encountered in clinical trials and used in risk scoring in patients with HF[1].

Patients with HF experience poor functional capacity caused by the coupling of reduced cardiac and pulmonary reserve, skeletal muscle and peripheral dysfunction, impaired autonomic regulation and existing comorbidities[5]. The 6-minute walk test (6MWT) is a widely applied tool to evaluate the functional capacity of HF patients[6]. 6MWT has been determined to be a safe, simple and reliable tool in this population[7-9]. 6MWT predicts mortality and morbidity in patients with HF[10]. Additionally, functional impairment assessed using 6MWT provides additional prognostic information over recognised risk markers in patients with HF[11]. Furthermore, the 6MWT distance (6MWD) provides prognostic information similar to that of peak oxygen consumption and the ventilatory efficiency obtained through cardiopulmonary exercise testing, a gold standard for measuring exercise capacity, in this population[12].

Gait speed has an incremental value in predicting mortality and hospitalisation in older patients with HF[13]. Gait parameters obtained from wearable sensors provide a large amount of data on human movement, such as gait speed, stride length, and cadence, regardless of whether those assessments are performed in laboratory or clinical settings[14,15]. Nowadays, machine learning

ARTICLE IN PRESS

(ML) approaches using parameters of gait analysis are widely investigated[16-18]. ML is one subfield of artificial intelligence that focuses on developing algorithms that allow computers to automatically learn patterns and make predictions or decisions based on data[19]. The applications of ML in patients with HF are becoming widespread. ML algorithms can be utilised for predicting HF risk, diagnosing HF, classifying HF phenotypes, and predicting outcomes following HF diagnosis[20]. Meta-heuristic algorithms are essential in enhancing the performance of ML models, as they effectively search a complex and multidimensional search space and avoid local minima. Their adaptive optimisation potential allows for optimising model hyperparameters with fine accuracy, which results in improved predictive accuracy, robustness and extrapolation to a wide variety of datasets[21]. By determining the feasibility of ML applications that estimate functional capacity using wearable sensor-based gait parameters, these models would have the potential to provide practical and accurate assessment and to optimise exercise prescription in cardiac rehabilitation, leading to better disease management in patients with HF. Therefore, this study aimed to predict functional capacity using wearable sensor-based spatiotemporal and kinematic gait parameters and meta-heuristic-based ML approaches in patients with HF and to determine the optimal gait parameters.

ARTICLE IN PRESS

Methods

Study design and participants

This was a single-centre, retrospective and cross-sectional study. Ethical approval of this this study was provided by Çankırı Karatekin University Ethics Commission (decision code: f4e68ddf5d3f451f, decision date: 09.07.2025). All participants in this study provided written informed consent. The study was conducted in accordance with the Declaration of Helsinki.

Patients with heart failure and reduced ejection fraction (HFrEF), defined according to international guideline, comprised the study population[22]. Inclusion criteria consisted of being

ARTICLE IN PRESS

clinically stable, being between 35 and 85 years, and being in New York Heart Association (NYHA) functional class II or III. Exclusion criteria consisted of neurological or orthopaedical disease affecting mobility, significant chronic respiratory or kidney disease or diabetic neuropathy.

Functional capacity assessment

The 6MWT was performed to evaluate functional capacity following the international statement[23]. Participants were instructed to walk as quickly as possible along a 20-meter flat corridor for 6 minutes. The standard phrases of encouragement were used every minute during the test. The distance covered during 6MWT was noted as 6MWD.

Gait assessment

Gait assessment was performed using the reliable G-WALK® gait analysis system (BTS Bioengineering S.p.A., Italy)[24]. A wearable inertial sensor (G-SENSOR®) comprising a triaxial accelerometer (16 bit/axes; 8g), a triaxial magnetometer (13 bit; 1200 μT), a triaxial gyroscope (16 bit/axes; 250°/s), and a global positioning system receiver was positioned at the S1 and S2 vertebrae and fastened with a semi-elastic strap[25]. Each participant was asked to walk along an 8-meter path at their preferred walking speed. Obtained data, sampled at 100 Hz, were wirelessly sent to a computer via Bluetooth, then processed using software (BTS G-STUDIO, version 3.5.25.0)[25]. In total, 37 gait parameters, automatically analysed by the software, were used for model development (see Supplementary, Table S1 online).

Model development

Figure 1 illustrates the framework for generating a meta-heuristic-based ML model construction and validation.

Data Augmentation and Preprocessing

ARTICLE IN PRESS

The dataset used in this research was expanded via the Synthetic Minority Over-sampling Technique for Regression (SMOGN), which boosted the number of total observations to 410[26-28].

In contrast to standard oversampling, SMOGN uses a relevance function to determine which areas of the continuous target variable are underrepresented and increases the number of observations of the training sample to 410[26]. The strategy aims at increasing the capacity of the model to acquire the uncommon pathological gait patterns without affecting the general data distribution. The final sampling size chosen at 410 observations was calculated assessing that they will give the meta-heuristic optimization process enough statistical power and reduce the risk of overfitting that commonly comes with defining too much synthetic data creation.

The modeling pipeline was designed in such a way that it all possibilities of data leakage and ensure methodological rigor. First, 80% training set and a 20% independent testing set were created out of the original data (n=70). More importantly, all preprocessing operations such as data augmentation (SMOGN) and Recursive Feature Elimination (RFE) were carried out in the training folds. The test set, made up of 14 original patients, was completely untouched by any data generation or feature selection processes used to create a synthetic data set. This will guarantee that the model test is based on real clinical generalizability on unseen data (see Figure 1 to see the detailed pipeline).

In order to make the performance estimates of the small dataset (10-fold cross-validation) stable and reliable, a 10-fold cross-validation framework was applied on the development set (80% of the original data). This strategy, together with an independent hold-out test set (20%), is a strong evaluation of how the model can be applied to unobserved clinical cases.

Recursive Feature Elimination

ARTICLE IN PRESS

Recursive Feature Elimination (RFE) is a feature selection method which removes and keeps the most relevant features in a recursive manner, training a model and removing the least significant features[29]. The model is fitted to the data at every iteration, and features are ranked by their importance or contribution to the predictive performance. The most insignificant feature(s) are eliminated, and the process continues until the number of features required is achieved.

RFE was used to reduce the noise and increase the interpretability of the models. It was necessary to optimize the feature space, which was achieved by selecting the 25 important gait parameters out of the initial 37. Besides, the SMOGN algorithm was incorporated into the pipeline to take care of data imbalance. SMOGN was effective in enhancing the learning ability of the model because of complex biomechanical interactions by having a relevance function that concentrated on underrepresented target values.

Meta-Heuristic Algorithms for Hyperparameter Optimisation

The optimisation of hyperparameters is an essential stage in ensuring the overall performance of an ML model is optimised. In the present research, the traditional ML methods were supplemented by meta-heuristic and Bayesian algorithms, which provide more advanced and efficient optimisation strategies[21]. These algorithms are characterised by the possibility to explore a wide parameter space, locate global optima and minimise the probability of local optima entrapment. Simulated Annealing (ASA) automatically changes the cooling schedule and the sampling strategy in the parameter space based on the nature of the problem[30]. The adaptive mechanism allows the algorithm to find the global optimum faster and more efficiently and decreases the chances of local optima trapping. It is especially useful in high-dimensional, nonlinear optimisation problems. Genetic algorithm (GA) is powerful, nature-inspired optimization tool that use evolutionary principles to solve complex problems across diverse domains. Their adaptability and ability to search large, complex spaces make them

ARTICLE IN PRESS

valuable for both theoretical and practical applications[31]. In a GA, mutation and crossover operations are used on the entire new generation. Hybrid Particle Swarm Optimisation (HPSO) is a hybrid algorithm that is built upon the Particle Swarm Optimisation (PSO)[32]. HPSO aims to address the issue of premature convergence and enhance the process of global solution discovery by keeping the quick convergence property of PSO and exploiting the advantage of other algorithms. HPSO usually uses the global search power of PSO with the local search or exploration powers of other complementary algorithms. This combination allows the algorithm to both perform a wide-ranging search and perform more detailed searches in promising areas. Bayesian Optimisation is based on the development of a probabilistic model of the objective function, typically a Gaussian Process[33]. The model has already been assessed to create a pre-belief regarding the behaviour of the function. A function of acquisition is then used to decide the next most promising point to consider. The acquisition function is a trade-off between exploration (searching areas where the model is uncertain) and exploitation (searching in areas that are likely to give good results). This is done iteratively until a predetermined number of iterations has been completed or until a convergence criterion has been reached.

ARTICLE IN PRESS

Machine Learning Algorithms

In the present work, a collection of ML algorithms that are typically used in regression problems and which exemplify various learning paradigms was used. These algorithms are famous in terms of performance and effectiveness in different spheres of application[34]. Extreme Gradient Boosting **(**XGBoost) is a faster and more efficient type of the gradient boosting model with a specific optimised implementation[35]. XGBoost works by repeatedly building the weak learners, usually decision trees, by gradient of the loss function, with each new tree designed to reduce the errors made by the previous learners. Regularisation techniques are also used in this method to control overfitting, thus turning XGBoost into an effective and robust modelling tool. Light Gradient Boosting Machine (LightGBM), unlike the classic gradient boosting

ARTICLE IN PRESS

algorithms, builds trees in a leaf-wise manner instead of level-wise[36]. The approach enables more complex trees to be built and thus becomes more accurate but can also run the risk of overfitting. LightGBM also offers rapid training times, histogram-based algorithms and parallel learning features. Random forest (RF) works through the pooling of several decision trees, each of which has been trained on a randomly chosen subset of the dataset[37]. Moreover, the random subset of features is taken into account at every split. This added randomness also minimises overfitting and increases the generalizability of the model. RF is an excellent option when it comes to classification and regression tasks because it can be highly effective and interpreted. CatBoost uses ordered boosting to avoid target leakage, symmetric (oblivious) trees to create balanced trees, and boosting to build computationally efficient trees[38]. Other sampling techniques, such as minimal variance sampling, stabilise the gradient estimation and enhance convergence and generalisation. These properties allow CatBoost to enjoy good predictive accuracy, high efficiency in training on a large sample, and high performance on heterogeneous and categorical attributes.

Hyperparameter Tuning

The process of hyperparameter optimisation is an essential part of the process of ML optimisation[39]. In this work, a wide hyperparameter range was specified in relation to each of the ML models. The hyperparameters set aside in the individual models are provided in Supplementary, Table S2 online.

The best hyperparameter settings were found within such ranges using meta-heuristic algorithms like SA, GA, PSO and Bayesian Optimisation.

In order to fully replicate the results, the particular optimization parameters, such as the iterations, the population sizes, and convergence criterion of each algorithm, are clearly written in Supplementary Table S3. The last optimized settings are given in the same table as well.

ARTICLE IN PRESS

Computational experiments were performed all on Google Colab platform with a Tesla T4 to provide a standard of processing power and disclosure of computational cost.

Explainability via SHapley Additive exPlanations

Lundberg and Lee proposed a game theory-based interpretability tool called SHapley Additive exPlanations (SHAP), which is designed to explain the black-box behaviour of ML models[34]. This technique measures the amount that each feature adds to the prediction of the model in a just way, according to the concept of Shapley values; that is, it measures the effect of a feature on the model prediction relative to the baseline.

Evaluation Metrics

All the models were analysed based on the following metrics, as compared to their performance [40].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \underline{y})^2} \quad (1)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (2)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (3)$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (4)$$

In these equations, n is the total observations, y_i is the actual value of an observation, $\hat{y}_i$ is the prediction given by the model of the same observation and $\underline{y}$ is the average of all the actual values.

Statistical Analysis

Continuous variables are presented as mean ± standard deviation values, whereas those with alternative distributions are presented as median and interquartile range values. Categorical variables are presented as the number (percentage) of patients with the corresponding attribute. Moreover, comparisons between the training and testing datasets were performed using the

ARTICLE IN PRESS

Student t-test (Welch t-test) with continuous variables that are normally distributed, the Mann-Whitney U test with non-normally distributed continuous variables, and the Fisher exact test with categorical variables. p-value < 0.05 was considered statistically significant. Python (version 3.12) was used for ML and statistical analyses.

Results

Baseline characteristics of 56 patients from the training dataset and 14 patients from the testing dataset are presented in Table 1. The age, gender, ejection fraction, NYHA functional class, medications, implantable cardioverter defibrillator, and 6MWD were similar between the training and testing datasets ($p > 0.05$). The body mass index was significantly higher in the training dataset compared to the testing datasets ($p = 0.03$).

Base model results with non-augmented and augmented data

In RFE, 25 important gait features were revealed. Table 2 shows the relative performance of the various combinations of metaheuristic algorithms and ML models with non-augmented and augmented data**.** Many metaheuristic algorithms have been developed and tested, but among them, SA, GA, PSO, Bayesian Optimisation, and Recursive Feature RFE with the use of ML algorithms XGBoost, LightGBM, RF, and CatBoost, we can see clear differences in predictive accuracy. When basing model performance is evaluated without data augmentation, the results indicate significant variation in base model performance depending on which metaheuristic algorithm or RFE is used together with the ML models. RF demonstrates the best predictive performance across all tested settings compared to XGBoost, CatBoost, and LightGBM. RFE with RF showed the most overall performance, with the lowest RMSE (71.1277), MAE (40.2006) and MSE (5059.1552). When basing model performance is evaluated with data augmentation, CatBoost is always the best-performing model in the different metaheuristic algorithms, mostly in PSO, where the model has the lowest RMSE (17.0445), MAE (11.4098),

ARTICLE IN PRESS

and MSE (290.5135). This finding highlights the better predictive ability of the hybrid PSO-CatBoost in the process of modeling thorny relationships in the augmented data.

Figure 2 displays the bias-variance decomposition for the top-performing models from each algorithm family on the non-augmented and augmented datasets. The results indicate that all models suffer from high Mean Squared Error (MSE), with values ranging from 6269 for SA-XGBoost to 7512 for PSO-CatBoost. A significant portion of this error is attributable to high bias, suggesting that the models were underfitting and unable to capture the underlying complexity of the original dataset. Recognizing the limitations of the original dataset, data augmentation was employed to enhance model training. There is a dramatic reduction in MSE for all models; for example, the MSE for the PSO-CatBoost model decreased from 7512 to 598.26. This improvement is largely due to a significant reduction in the bias component, indicating that the augmented data enabled the models to learn the target function more effectively and overcome the initial underfitting problem.

Figure 3 summarizes a comprehensive comparison of the coefficient of determination (R^2) across four ML models (XGBoost, LightGBM, RF, and CatBoost) and five optimization strategies based on the non-augmented and the augmented data sets. On the non-augmented dataset, the RF model demonstrated superior performance, particularly with its default hyperparameter settings, achieving the highest R^2 score of 0.614. On the augmented dataset, in a notable shift from the initial results, the CatBoost model emerged as the clear top performer, achieving the highest R^2 scores with all optimization algorithms. The combination of CatBoost with PSO yielded the peak performance with an R^2 of 0.9456. XGBoost also performed exceptionally well, with scores consistently above 0.919, while the relative performance of RF declined.

ARTICLE IN PRESS

To understand feature influence on the model predictions, a global SHAP decision plot was generated for the models trained on the non-augmented data and augmented data, as shown in Figures 4. Features such as left pelvic obliquity range and left pelvic tilt range appear to be among the most influential, as indicated by the wide range their decision paths cover across the model's output values for non-augmented data. After data augmentation, features such as the right stride length and gait speed were recognised as the most dominant predictors.

ARTICLE IN PRESS

ARTICLE IN PRESS

Discussion

The main findings of the present study were (1) the RFE-RF model, when applied to inertial sensor-based gait parameters, had the best potential to predict functional capacity measured using 6MWT without data augmentation, (2) when data were augmented, the PSO-CatBoost model was superior compared to other ML models for the prediction of functional capacity in patients with HF, (3) left pelvic obliquity and pelvic tilt ranges were the most important features for predicting functional capacity without data augmentation, in contrast, right stride length and gait speed were the most important features for predicting functional capacity with data augmentation. This analysis provides initial insights into the key biomechanical parameters driving the predictions.

Our findings demonstrate that data augmentation was a critical and transformative step in this modelling task. While the combination of the RFE-RF was the strongest performer on the limited original dataset, its advantage disappeared when the models were trained on the richer, augmented data. The overall high performance of CatBoost, regardless of the optimisation strategy used, is an indication of its strength and flexibility, which makes it a valid option in predictive modelling in augmented, complex datasets. Additionally, the augmentation process effectively mitigated the high bias (underfitting) observed initially, allowing the gradient boosting models, particularly CatBoost, to excel. Data augmentation led not only to more accurate models but also to more stable and reliable ones. Furthermore, the combination of the PSO-CatBoost for hyperparameter tuning yielded the best predictive performance, achieving an outstanding R^2 score of 0.9456. This demonstrates a high degree of accuracy and reliability for the model's predictions on unseen data. The present study also underscores the importance of a sufficiently large and diverse dataset for training robust ML models and highlights the effectiveness of CatBoost for this particular regression problem.

ARTICLE IN PRESS

A variety of related studies have been proposed in the literature, and they are presented in Table 3. De Cannière et al. found that a support vector machine (SVM) model using chronotropic response and effort as input features predicted 6MWD with an overestimation of 42.8 m in the cardiac rehabilitation population[41]. The present study demonstrated that hybrid PSO-CatBoost was the best model predicting 6MWD with an 11.4 m overestimation in patients with HF. This variation in distance is less than 32 m and 25 m, which is considered a clinically meaningful change in overall disease status in patients with HF and in assessing the effect of cardiac rehabilitation, respectively[42,43]. These findings indicate that hybrid PSO-CatBoost applied to gait parameters achieves a more accurate prediction for functional capacity in patients with HF.

In the literature, several studies identified different ML models using gait parameters for classification or prediction[16-18]. Zhou et al. investigated three different ML methods, including Kernel principal component Analysis (KPCA**)** in combination with Support Vector Machine (SVM)**,** RF and Artificial Neural Network (ANN), to classify healthy young-middle-aged adults, healthy old adults, and geriatric patients based on gait outcomes[16]. Authors reported that KPCA in combination with SVM, RF, and ANN, when applied to dynamic gait outcomes, are capable of discriminating between age groups with the accuracy of 89%, 73.6% and 90%, respectively. Additionally, cross-entropy between medial-vertical and vertical acceleration, vertical direction, maximal Lyapunov exponent of the vertical acceleration, gait speed and the variability of the accelerations in anterior-posterior and vertical direction have been found to be the most important gait outcomes for classification[16]. Lim et al. reported that the LightGBM applied to sensor-based gait parameters is superior to predict falls with an accuracy of 96%[17]. Mizugucci et al. aimed to develop a ML–based automatic rating model of the Clinical Frailty Scale using gait analysis from smartphone video recordings for patients with HF[18]. The author demonstrated that the LightGBM automatic rating model achieved a good performance in estimating Clinical Frailty Scale, with Shapley analysis identifying 15 primary gait parameters,

ARTICLE IN PRESS

including gait speed, total gait time, stance phase time, and ankle swing speed, as key contributors to the Clinical Frailty Scale automatic rating system[18]. In the present study, global SHAP analysis revealed that left pelvic obliquity and pelvic tilt ranges were the most influential; however, when data were augmented, right stride length and gait speed emerged as the strongest features of functional capacity. This suggests that the augmented data allowed the model to uncover more robust and significant relationships between the spatiotemporal and kinematic gait parameters and functional capacity. Gait speed appears to be a key contributor across the studies, underpinning its prominence in prognosis in patients with HF[13].

ML model (PSO-CatBoost) and wearable sensor-based gait parameters may offer a promising approach for estimating functional status over time, optimising exercise prescription and assessing the effectiveness of cardiac rehabilitation to increase functional capacity in patients with HF. Gait speed and stride length emerged as the most influential variables in the best model, indicating their role in estimating functional capacity in patients with HF. Identifying declines in these gait parameters may reflect reduced functional capacity in patients with HF[15]. From a clinical perspective, monitoring changes in gait speed and stride length may support clinical decision-making, particularly when 6MWT is not feasible.

While data limitations remain a challenge, our findings provide a methodological improvement by combining augmentation, feature elimination, and optimization in a unified pipeline. Our study was based on a single-center, cross-sectional design, which may limit the generalizability and clinical applicability of the findings, given the absence of external validation, longitudinal validation, and evaluation of clinical decision impact. Therefore, longitudinal and multicentre studies are required to confirm the generalizability of the model and to better define its role in clinical decision support.

ARTICLE IN PRESS

The present study advances the application of ML in clinical prediction tasks by systematically integrating feature selection, data augmentation, and optimisation techniques. Compared to earlier studies, our study applied recursive feature elimination and advanced meta-heuristic algorithms, alongside SMOGN-based data augmentation, to build robust predictive models using state-of-the-art learners. The method is in line with the recent literature, including the study by Trabassi et al. that stated that data balancing needs to be combined with generative AI to discover rare pathological patterns in small clinical gait datasets[44]. By leveraging PSO-CatBoost as the optimal model, our study achieved competitive predictive efficiency, outperforming several previous studies that lacked either feature selection or optimisation strategies[16-18,41,45-46]. This contribution demonstrates that the integration of meta-heuristic algorithms significantly enhances predictive accuracy and robustness compared to conventional approaches.

Conclusion

This study demonstrates that meta-heuristic-based ML models are useful for predicting functional capacity based on gait parameters in patients with HF. Augmentation through the use of SMOGN worked to enhance data representativeness and lower underfitting, whereas PSO was successful in hyperparameter fine-tuning and the resulting eventual capacity in prediction. CatBoost turned out to be significantly more accurate and more stable with PSO, which proves the power of such a combination to capture the intricate relationships between gait parameters. These results highlight the strength of CatBoost that is PSO-enhanced as an effective and feasible model of predictive analytics. Further studies should evaluate the findings on a much larger dataset in different HF trajectories.

ARTICLE IN PRESS

Disclosures

Conflict of Interest: The authors declare no competing interests.

Funding: No funding was received for this study.

Data availability statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

All authors are responsible for conseptualization, methodology, interpretation, and writing-review & editing. A.T.E., B.O.K., M.B.Y., and S.S. performed the investigation. A.D., S.B. performed formal analysis, software, pre-processing and visualization. ATE, A.D., and S.B. are responsible for writing the original draft.

ARTICLE IN PRESS

ARTICLE IN PRESS

References

1. Bozkurt B, et al. Universal definition and classification of heart failure: a report of the heart failure society of America, heart failure association of the European society of cardiology, Japanese heart failure society and writing committee of the universal definition of heart failure. J Card Fail. 27, 387-413 (2021).

2. Yan T, et al. Burden, trends, and inequalities of heart failure globally, 1990 to 2019: a secondary analysis based on the global burden of disease 2019 study. J Am Heart Assoc. 12, e027852 (2023).

3. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GM, Coats AJ. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res. 118, 3272-3287 (2022).

4. Tanai E, Frantz S. Pathophysiology of heart failure. Compr Physiol. 6, 187-214 (2016).

5. Del Buono MG, et al. Exercise intolerance in patients with heart failure: JACC state-of-the-art review. J Am Coll Cardiol. 73, 2209-2225 (2019).

6. Giannitsi S, Bougiakli M, Bechlioulis A, Kotsia A, Michalis LK, Naka KK. 6-minute walking test: a useful tool in the management of heart failure patients. Ther Adv Cardiovasc Dis. 13, 1753944719870084 (2019).

7. Pollentier B, et al. Examination of the six minute walk test to determine functional capacity in people with chronic heart failure: a systematic review. Cardiopulm Phys Ther J. 21, 13-21 (2010).

8. Du H, Wonggom P, Tongpeth J, Clark RA. Six-minute walk test for assessing physical functional capacity in chronic heart failure. Curr Heart Fail Rep. 14, 158-166 (2017).

9. Uszko-Lencer NH, et al. Reliability, construct validity and determinants of 6-minute walk test performance in patients with chronic heart failure. Int J Cardiol. 240, 285-290 (2017).

ARTICLE IN PRESS

10. Bittner V, et al. Prediction of mortality and morbidity with a 6-minute walk test in patients with left ventricular dysfunction. JAMA. 270, 1702-1707 (1993).

11. Scrutinio D, et al. Incremental prognostic value of functional impairment assessed by 6-min walking test for the prediction of mortality in heart failure. Sci Rep. 14, 3089 (2024).

12. Forman DE, et al. 6-min walk test provides prognostic utility comparable to cardiopulmonary exercise testing in ambulatory outpatients with systolic heart failure. J Am Coll Cardiol. 60, 2653-2661 (2012).

13. Pulignano G, Del Sindaco D, Di Lenarda A, Alunni G, Senni M, Tarantini L, et al. Incremental value of gait speed in predicting prognosis of older adults with heart failure: insights from the IMAGE-HF study. JACC Heart Fail. 4, 289-298 (2016).

14. Muro-De-La-Herran A, Garcia-Zapirain B, Mendez-Zorrilla A. Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. Sensors (Basel). 14, 3362-3394 (2014).

15. Eyolcu AT, Kahraman BO, Yilmaz MB, Savci S. Wearable sensor-based gait parameters in patients with heart failure with reduced ejection fraction: associations with clinical and physical outcomes. Heart Lung. 75, 277-284 (2026).

16. Zhou Y, et al. The detection of age groups by dynamic gait outcomes using machine learning approaches. Sci Rep.10, 4426 (2020).

17. Lim ZK, Connie T, Goh MKO, Saedon NIB. Fall risk prediction using temporal gait features and machine learning approaches. Front Artif Intell. 7, 1425713 (2024).

18. Mizuguchi Y, et al. Machine learning–based gait analysis to predict clinical frailty scale in elderly patients with heart failure. Eur Heart J Digit Health. 5, 152-162 (2024).

19. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol. 9, 14-14 (2020).

ARTICLE IN PRESS

ARTICLE IN PRESS

20. Averbuch T, et al. Applications of artificial intelligence and machine learning in heart failure. Eur Heart J Digit Health. 3, 311-322 (2022).

21. da Costa Oliveira AL, Britto A, Gusmão R. Machine learning enhancing metaheuristics: a systematic review. Soft Comput. 27, 15971-15998 (2023).

22. McDonagh TA, 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, 3599-3726 (2021).

23. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med. 166, 111-117 (2002).

24. De Ridder R, Lebleu J, Willems T, De Blaiser C, Detrembleur C, Roosen P. Concurrent validity of a commercial wireless trunk triaxial accelerometer system for gait analysis. J Sport Rehabil. 28 (2019).

25. BTS. G Walk User Manual. English Version 7.0.0 (2016).

26. Torgo L, Ribeiro RP, Pfahringer B, Branco P. SMOTE for regression. In: Correia L, Reis LP, Cascalho J. (eds) Progress in Artificial Intelligence. EPIA 2013. Lecture Notes in Computer Science. 8154 (2013). Springer, Berlin, Heidelberg.

27. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 49, 1373-1379 (1996).

28. Harrell FE, Levy DG. Regression modeling strategies. R package version. 6.3-0 (2022).

ARTICLE IN PRESS

ARTICLE IN PRESS

29. Bulut O, Tan B, Mazzullo E, Syed A. Benchmarking variants of recursive feature elimination: Insights from predictive tasks in education and healthcare. Information. 16, 476 (2025).

30. Ingber L. Adaptive simulated annealing (ASA): Lessons learned. arXiv preprint cs/0001018 (2000).

31. Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput. 6, 182-197 (2002).

32. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95-international conference on neural networks. Abstract 4, p.1942-1948. ieee (1995).

33. Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N. Taking the human out of the loop: A review of Bayesian optimization. Proc IEEE. 104, 148-175 (2015).

34. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 30 (2017).

35. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 785-794 (2016).

36. Ke G, et al. Lightgbm: A highly efficient gradient boosting decision tree. dv Neural Inf Process Syst. 30 (2017).

37. Breiman L. Random forests. Machine learning. 45, 5-32 (2001).

38. Pothupogu P, Santhakumar Y, Yenmanagandla A, Sriram A. Health monitoring system based on LightGBM, CatBoost and Explainable Boost Classifier. In: 2024 2nd World Conference on Communication & Computing (WCONF). 1-7. IEEE (2024).

39. Alwazy ASH, Buyrukoğlu G, Buyrukoğlu S, Baker MR. Evaluating machine learning and statistical learning techniques for cancer classification and diagnosis. Iran J Comput Sci. 8, 471-490 (2025).

ARTICLE IN PRESS

ARTICLE IN PRESS

40. Nooruldeen O, Baker MR, Aleesa A, Ghareeb A, Shaker EH. Strategies for predictive power: Machine learning models in city-scale load forecasting. e-Prime-Adv Electr Eng Electron Energy. 6, 100392 (2023).

41. De Cannière H, et al. Wearable monitoring and interpretable machine learning can objectively track progression in patients during cardiac rehabilitation. Sensors (Basel). 20, 3601 (2020).

42. Shoemaker MJ, Curtis AB, Vangsnes E, Dickinson MG. Clinically meaningful change estimates for the six-minute walk test and daily activity in individuals with chronic heart failure. Cardiopulm Phys Ther J. 24, 21 (2013).

43. Gremeaux V, et al. Determining the minimal clinically important difference for the six-minute walk test and the 200-meter fast-walk test during cardiac rehabilitation program in coronary artery disease patients after acute coronary syndrome. Arch Phys Med Rehabil. 92, 611-619 (2011).

44. Trabassi D, et al. Optimizing rare disease gait classification through data balancing and generative AI: Insights from hereditary cerebellar ataxia. Sensors (Basel). 24, 3613 (2024).

45. Noh B, et al. XGBoost based machine learning approach to predict the risk of fall in older adults using gait outcomes. Sci Rep. 11, 12183 (2021).

46. Wenzel C, et al. Machine learning predicts peak oxygen uptake and peak power output for customizing cardiopulmonary exercise testing using non-exercise features. Eur J Appl Physiol.124, 3421-3431 (2024).

ARTICLE IN PRESS

ARTICLE IN PRESS

Figure Legends

Figure 1. The framework for generating the meta-heuristic-based ML model construction and validation. Abbreviation: ML: machine learning; RFE, recursive feature elimination; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; RMSE, root-mean-square error; MAE, mean absolute error; MSE, mean squared error.

Figure 2. a) Bias-variance decomposition for best models of non-augmented data, and b) Bias-variance decomposition for best models of augmented data. Abbreviation: RFE, recursive feature elimination; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine

Figure 3. a) Heatmap of R^2 performance of ML models based on optimization algorithms for non-augmented data, and b) Heatmap of R^2 performance of ML models based on optimization algorithms for augmented data. Abbreviation: ML: machine learning; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; PSO, particle swarm optimization; RFE, recursive feature elimination.

Figure 4. a) Global SHAP decision plot of the most important features of the PSO-CatBoost model for Non-Augmented Data, and b) Global SHAP decision plot of the most important features of the PSO-CatBoost model for Augmented Data. Abbreviation: PSO, particle swarm optimization.

ARTICLE IN PRESS