Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine learning prediction of spirometric parameters from clinical data and video-based walk test features

Burçin Çelik
Ondokuz Mayıs University
Sema Gül
Ondokuz Mayıs University
Muammer Türkoğlu
Samsun University
Öykü Çelik
Duke University School of Medicine
Melda İşevi
Ondokuz Mayıs University
Caner İşevi
drcanerisevi@gmail.com

Ondokuz Mayıs University

Research Article

Keywords: Chronic obstructive pulmonary disease, Spirometry, Machine learning, Gait analysis, Six-minute walk test, Explainable artificial intelligence

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11706261/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Machine learning prediction of spirometric parameters from clinical data and video-based walk test features

Burçin Çelik[1], Sema Gül[2], Muammer Türkoğlu[3], Öykü Çelik[4], Melda İşevi[5], Caner İşevi[1*]

[1] Department of Thoracic Surgery, Faculty of Medicine, Ondokuz Mayıs University, Samsun, Türkiye

[2] Division of Neonatology, Department of Pediatrics, Faculty of Medicine, Ondokuz Mayıs University, Samsun, Türkiye; ADAPHA Artificial Intelligence R&D and Software Inc., Samsun Technopark, Samsun, Türkiye

[3] Department of Software Engineering, Faculty of Computer and Information Sciences, Samsun University, Samsun, Türkiye

[4] Duke University School of Medicine, Durham, NC, USA; Department of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst, MA, USA

[5] Division of Intensive Care, Department of Anaesthesiology and Reanimation, Faculty of Medicine, Ondokuz Mayıs University, Samsun, Türkiye

***Correspondence:** Caner İşevi, Department of Thoracic Surgery, Faculty of Medicine, Ondokuz Mayıs University, Kurupelit, Samsun, Türkiye. Tel: +90 362 312 19 19 (ext. 4240); e-mail: drcanerisevi@gmail.com

Email addresses: Burçin Çelik: cburcin@hotmail.com; Sema Gül: fztsmgl2011@gmail.com; Muammer Türkoğlu: muammer.turkoglu@samsun.edu.tr; Öykü Çelik: oykucelik@umass.edu; Melda İşevi: drmeldaisevi@gmail.com; Caner İşevi: drcanerisevi@gmail.com

Abstract

Background: Spirometry is essential for diagnosing chronic obstructive pulmonary disease and for assessment before lung resection but requires equipment, trained staff and patient

cooperation. We evaluated whether machine learning models combining clinical data with video-based gait features from the six-minute walk test can predict spirometric parameters.

Methods: In this prospective single-centre study, 100 participants (33 with chronic obstructive pulmonary disease, 67 controls) underwent spirometry and a video-recorded six-minute walk test. Inputs were 14 clinical variables, the manually measured walk distance and three video-derived gait features. Thirteen regression models predicted six spirometric targets separately over five repeated stratified 90%/10% train–test splits. Variable contributions in the best model per target were assessed with Shapley additive explanations.

Results: Twenty-four of 33 patients (72.7%) were newly diagnosed by study spirometry. The highest mean coefficients of determination were 0.782 ± 0.049 for forced vital capacity in litres, 0.730 ± 0.078 for mid-expiratory flow, 0.701 ± 0.135 for the ratio of forced expiratory volume in 1 s to forced vital capacity, 0.701 ± 0.047 for forced expiratory volume in 1 s in litres, and 0.555–0.595 for percent-predicted values. Mean absolute errors for both volumes were 0.24 L. Linear and regularized models were the most consistent. Sex and age dominated volume predictions; smoking pack-years ranked in the top five for all targets, and gait features for selected targets.

Conclusions: Clinical data combined with video-based walk test features predicted several spirometric parameters with moderate-to-good accuracy. These exploratory findings support a complementary pre-spirometry triage tool, pending external validation.

Trial registration: ClinicalTrials.gov, NCT07010211. Registered 30 May 2025.

Keywords: Chronic obstructive pulmonary disease; Spirometry; Machine learning; Gait analysis; Six-minute walk test; Explainable artificial intelligence

Background

Chronic obstructive pulmonary disease (COPD) is a preventable and treatable but heterogeneous disease characterized by persistent respiratory symptoms and airflow limitation, and it remains a leading cause of morbidity and mortality worldwide. Its diagnosis

and severity assessment rely on the spirometric demonstration of airflow limitation [1], and spirometry is also central to the assessment of fitness for lung resection [2]. However, spirometry requires a calibrated device, trained personnel, quality control and maximal patient effort. It is therefore not always available in primary care and resource-limited settings, which contributes to substantial underdiagnosis [3], and access to spirometry in primary care is unequal [4].

Artificial intelligence (AI) is increasingly used to support clinical decisions in COPD. In a blinded primary-care diagnostic accuracy study, software using raw spirometry and basic demographic data identified COPD in 1113 symptomatic patients with 84.0% sensitivity and 86.8% specificity [4]. Deep learning models have also predicted forced expiratory volume in 1 s (FEV1) and forced vital capacity (FVC) from computed tomography images with strong correlation [5]. Most of these approaches, however, depend either on spirometry itself or on costly imaging involving radiation.

The six-minute walk test (6MWT) is a simple and inexpensive test of functional capacity in chronic respiratory disease, and the six-minute walk distance (6MWD) is a component of the BODE (body mass index, airflow obstruction, dyspnoea and exercise capacity) index, which predicts mortality in COPD [6]. In routine practice, the output of the 6MWT is reduced to the total distance walked, and temporal–motor information such as step count, fatigue-related decline in performance and gait variability is not recorded. Markerless pose estimation frameworks such as MediaPipe Pose extract body landmarks in real time from standard camera or smartphone video [7, 8], offering a contactless, low-cost opportunity to quantify gait during the 6MWT without additional devices.

Most AI models in COPD remain at an early developmental stage, relying on single-centre retrospective data with limited external validation [9], and complex machine learning models do not always outperform established clinical scores [10]. Transparent comparison of model families on the same data and clinician-interpretable outputs are therefore important. To our knowledge, no study has analysed standard 6MWT recordings with markerless pose

estimation, combined the resulting gait features with clinical variables to predict multiple spirometric parameters, and interpreted these predictions with explainable AI.

We aimed to predict FEV1 (L), FEV1 (%), FVC (L), FVC (%), FEV1/FVC (%) and forced expiratory flow between 25% and 75% of FVC (FEF25–75, L/s) from demographic and clinical variables and video-derived 6MWT features in a prospective sample of patients with COPD and controls, by comparing 13 regression models and interpreting the best model for each target with SHapley Additive exPlanations (SHAP).

Methods

Study design and ethics

This prospective, single-centre model development study was conducted at the Department of Thoracic Surgery, Ondokuz Mayıs University Faculty of Medicine, Samsun, Türkiye. It was approved by the Ondokuz Mayıs University Clinical Research Ethics Committee (decision no. OMÜ KAEK 2025/164, 26 March 2025) and registered at ClinicalTrials.gov (NCT07010211). In accordance with the ethics committee decision, data collection began in June 2025, after project funding had been secured, and continued until July 2026. The study was conducted in accordance with the Declaration of Helsinki. Written informed consent covering video recording and the scientific use and publication of the data was obtained from all participants. Reporting follows the TRIPOD+AI statement [11], and the completed checklist is provided in Additional file 2. The study workflow is summarized in Fig. 1.

Participants

Individuals aged 20–80 years who presented to the clinic and were physically able to perform spirometry and the 6MWT were enrolled. Participants were classified by spirometry as having COPD (FEV1/FVC < 0.70; n = 33) [1] or as controls (FEV1/FVC ≥ 0.70, with or without a smoking history; n = 67). Exclusion criteria were acute respiratory or other active infection; severe heart failure, advanced arrhythmia or other serious cardiovascular disease; neurological or orthopaedic disease markedly limiting walking; serious lung disease other than COPD (e.g. interstitial lung disease, active tuberculosis); pregnancy or breastfeeding; inability

to complete spirometry or the 6MWT; and refusal of video recording. Of 114 individuals screened, 14 were excluded, leaving 100 participants for analysis. There were no missing data.

Sample size was calculated with Minitab 16 (Minitab Inc., State College, PA, USA) for the classification aim of the original protocol, based on a similar study [12]: with a type I error of 5% and power of 80%, at least 24 participants per group were required, giving a target of approximately 53 participants after allowing for data loss. This calculation does not reflect the requirements of regression models with continuous outcomes. According to published criteria [13], approximately 120–190 participants could be needed for 18 candidate predictors and an expected R^2 of 0.55–0.70; the present findings should therefore be considered exploratory.

Clinical assessment

Age, sex and body mass index (BMI) were recorded. Smoking history was recorded as pack-years, current smoking and time since cessation (weeks). Prior physician-diagnosed COPD and its duration, comorbidity and allergy were noted. Resting pulse and respiratory rates were measured, and oxygen saturation was determined by pulse oximetry. Symptom burden was assessed with the COPD Assessment Test (CAT) [14] and dyspnoea with the modified Medical Research Council (mMRC) scale [15]; both were administered with investigator assistance.

Spirometry

Spirometry was performed according to American Thoracic Society/European Respiratory Society recommendations [16] with a turbine spirometer (Spirolab; MIR – Medical International Research, Rome, Italy). At least three acceptable manoeuvres were obtained, and the best values meeting repeatability criteria were analysed. Measurements were obtained before bronchodilator administration, and inhaled therapies were not withheld. Spirometry was performed in the morning by an operator blinded to clinical and video data, and the 6MWT was performed in the evening of the same day. Percent-predicted values were calculated with the European Community for Steel and Coal/Knudson reference equations embedded in the

device [17, 18]. FEV1 (L), FEV1 (%), FVC (L), FVC (%), FEV1/FVC (%) and FEF25–75 (L/s) were the target variables.

Six-minute walk test and video recording

The 6MWT was performed once per participant in a flat 30-m corridor with standardized encouragement, according to American Thoracic Society and European Respiratory Society/American Thoracic Society standards [19, 20]. Walking was recorded with the built-in camera of a standard smartphone positioned to capture the corridor and the participant. Recordings were stored in a secure environment accessible only to the research team.

Video-based gait feature extraction

Videos were processed with MediaPipe Pose, which provides the coordinates of 33 anatomical landmarks in each frame [7, 8]. Hip, knee, ankle and heel landmarks were evaluated together with trunk movements, and steps were identified from positional changes across consecutive frames and periodic movement patterns. Step counts (N1, N2 and N3) were determined in 10-s sampling windows at the beginning, middle and end of the test (Fig. 2). The 6MWD was measured manually. Four functional features were entered into the models:

- 6MWD: total distance walked during the test, measured manually.
- Estimated total step count: the mean of N1–N3 multiplied by 36 (the ratio of the 360-s test duration to the 10-s window).
- Fatigue ratio: N3/N1, representing the relative change in walking cadence over the test.
- Step-count variability: the standard deviation of N1–N3.

Derived variables duplicating the same information were not entered: walking speed and cadence are constant transformations of 6MWD and estimated total step count, respectively, and between-window differences are redundant with the fatigue ratio and step-count variability. 6MWD and estimated total step count were both retained because together they provide indirect information on step length.

Input variables and modelling

Inputs were selected on the basis of clinical parameters used in COPD assessment, the components of the BODE index and their availability in routine practice [6]. The 18 inputs were demographic (age, sex, BMI); smoking and disease history (pack-years, current smoking, time since cessation, COPD duration); clinical (comorbidity, allergy, pulse rate, respiratory rate, oxygen saturation, CAT, mMRC); and functional/video-based (6MWD, estimated total step count, fatigue ratio, step-count variability). Group membership (COPD/control) was not used as an input. Each spirometric parameter was modelled as a separate regression problem, and other spirometric values were never used as inputs, to prevent target leakage.

Thirteen regression models representing different learning principles were compared: linear and regularized models (linear regression, ridge, lasso, elastic net); instance- and kernel-based models (k-nearest neighbours [KNN], kernel ridge regression, support vector regression [SVR]); tree-based models (decision tree, random forest, extra trees); and boosting models (gradient boosting, histogram-based gradient boosting, XGBoost). Scaling was model-specific (StandardScaler for linear models and kernel ridge, RobustScaler for KNN and SVR, none for tree-based models) and was performed within a single pipeline together with the model. Sex (male = 1), current smoking, comorbidity and allergy (present = 1) were coded as binary variables. COPD duration (months) was set to 0 in participants without a prior diagnosis, and time since cessation (weeks) to 0 in current and never smokers. Hyperparameters were fixed across targets (Additional file 1: Table S1). Analyses were performed in Python (version 3.9.19) using NumPy (version 1.26.4), pandas (version 2.2.3), scikit-learn (version 1.6.1) [21], XGBoost (version 2.1.4) [22] and SHAP (version 0.49.1) [23].

Model evaluation

For each target, the data were split into 90% training and 10% test sets using quantile-based stratification on the target, repeated with five fixed random seeds (approximately 90 training and 10 test participants per repeat). All models were evaluated on identical splits for a given target. Performance was measured with R^2, mean absolute error (MAE), root mean squared

error (RMSE) and median absolute error (MedAE) and reported as mean ± standard deviation across the five splits; mean R^2 was the primary comparison criterion.

Explainability analysis

Variable contributions in the model with the highest mean R^2 for each target were examined with SHAP [23]. Global importance was defined by mean absolute SHAP values, and the direction of effect was assessed with SHAP summary plots. Explainable AI is recommended to facilitate clinician interpretation and acceptance [9].

Statistical analysis

Normality was assessed with the Shapiro–Wilk test. Data are presented as mean ± standard deviation, median (interquartile range) or n (%). Groups were compared with the independent-samples t-test or the Mann–Whitney U test, and categorical variables with Pearson's χ^2 test or Fisher's exact test. $P < 0.05$ was considered statistically significant. Analyses were performed in Python (SciPy).

Use of large language models

During the preparation of this work the authors used Claude (Anthropic) to translate the manuscript from Turkish into English and to edit its language and format it according to journal requirements. The tool was not used for study design, data analysis or interpretation of the results. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Results

Participant characteristics

The median age was 66.0 (57.5–70.0) years, and 82% of participants were men. Compared with controls, the COPD group had a higher proportion of men (93.9% vs 76.1%; P = 0.029), more pack-years [50.0 (40.0–60.0) vs 35.0 (0.0–50.0); P = 0.004] and a lower BMI [25.7 (22.7–28.4) vs 28.4 (25.7–32.2) kg/m²; P = 0.016]; age was similar (P = 0.052). Forty-eight controls

(71.6%) had a smoking history (Table 1). Characteristics by sex are given in Additional file 1: Table S3.

Of the 33 patients with COPD, 9 (27.3%) had a prior diagnosis and 24 (72.7%) were diagnosed for the first time by study spirometry. Based on FEV1 (%), 9 (27.3%) were classified as Global Initiative for Chronic Obstructive Lung Disease (GOLD) grade 1, 20 (60.6%) as GOLD 2 and 4 (12.1%) as GOLD 3; none were GOLD 4. FEV1 (L) ($P = 0.006$), FEV1 (%), FEV1/FVC (%) and FEF25–75 (all $P < 0.001$) were lower in the COPD group, whereas FVC (L) ($P = 0.315$) and FVC (%) ($P = 0.774$) did not differ. 6MWD ($P = 0.190$), CAT ($P = 0.083$) and mMRC ($P = 0.278$) were also similar (Table 1).

Model performance

The best model for each target is shown in Table 2 and mean R^2 values for all model–target combinations in Fig. 3; full results are given in Additional file 1: Table S2, and split-wise R^2 distributions in Additional file 1: Figs. S1–S6. The highest mean R^2 was obtained for FVC (L) with elastic net (0.782 ± 0.049), followed by FEF25–75 with histogram-based gradient boosting (0.730 ± 0.078), FEV1/FVC with gradient boosting (0.701 ± 0.135) and FEV1 (L) with lasso (0.701 ± 0.047). Performance was lower for percent-predicted values: 0.595 ± 0.103 for FVC (%) with XGBoost and 0.555 ± 0.083 for FEV1 (%) with linear regression. MAE was 0.237 L for FEV1 and 0.242 L for FVC (Table 2).

Linear or regularized models performed best for FEV1 (L), FEV1 (%) and FVC (L), and boosting models for FVC (%), FEV1/FVC and FEF25–75. Averaged across the six targets, the highest mean R^2 values were observed for elastic net (0.656), linear regression (0.655), lasso (0.653) and ridge (0.641), and the lowest for SVR (0.286), decision tree (0.353) and KNN (0.410) (Fig. 3). Between-split standard deviations were generally larger for boosting models (e.g. gradient boosting for FEV1 [L], 0.610 ± 0.173).

Explainability analysis

The five most important variables for each target are listed in Table 3, and SHAP summary plots are shown in Fig. 4. Sex and age dominated the predictions of FEV1 (L) and FVC (L).

mMRC, age, pack-years, respiratory rate and 6MWD were most important for FEV1 (%); BMI, time since cessation, age, estimated total step count and pack-years for FVC (%); pack-years, BMI, COPD duration, pulse rate and time since cessation for FEV1/FVC; and age, pack-years, mMRC, BMI and 6MWD for FEF25–75. Overall, clinical and demographic variables were more influential, although 6MWD ranked among the top five variables for FEV1 (%) and FEF25–75, and estimated total step count for FVC (%). Male sex was generally associated with higher FEV1 (L) and FVC (L) predictions, whereas older age (absolute volumes and FEF25–75) and greater smoking exposure (several models) were associated with lower predictions. SHAP values describe contributions to model predictions and do not indicate causality or independent incremental performance.

Discussion

Combining demographic and clinical variables with functional and video-based 6MWT features predicted six spirometric parameters with moderate-to-good accuracy. Mean R^2 was 0.70–0.78 for FVC (L), FEF25–75, FEV1/FVC and FEV1 (L) and 0.55–0.60 for percent-predicted values, and SHAP indicated that predictions relied mainly on physiologically and clinically meaningful variables.

The difference between absolute and percent-predicted targets is clinically informative. Absolute FEV1 and FVC depend largely on sex, age and body size irrespective of disease, the same determinants on which reference equations are built [17, 18]; accordingly, sex and age were the leading predictors for these targets. Percent-predicted values normalize these determinants, so the remaining variance mainly reflects disease-related loss of function. Lower R^2 values were therefore expected, but the prominence of mMRC, smoking exposure and 6MWD for these targets suggests that the models captured information on disease severity. High R^2 values for absolute volumes may thus reflect the capture of physiological determinants more than the detection of disease.

Linear and regularized models performed best for three of the six targets and were the most consistent family overall. This is in line with a systematic review showing no performance

benefit of machine learning over logistic regression for clinical prediction models in comparisons at low risk of bias [24], and with a meta-analysis of machine learning prognostic models in COPD noting that models requiring many inputs have limited clinical applicability compared with simple scores such as BODE [10]. Coefficient shrinkage by lasso and elastic net limits overfitting [25, 26], which may have been advantageous with approximately 90 training samples and 18 inputs. Conversely, boosting models performed best for FEV1/FVC and FEF25–75, the targets that most directly reflect obstruction, suggesting more pronounced non-linear relationships. The higher between-split variability of boosting models and test sets of approximately 10 participants mean that single-split estimates carry wide uncertainty [27]; model selection was therefore based on mean rather than maximum performance.

Many AI-based approaches use spirometry itself or imaging [4, 5], which limits their use for pre-assessment where spirometry is unavailable. Other non-spirometric approaches have been reported: an impulse oscillometry-derived equation detected spirometry-defined COPD with 88% sensitivity and 77% specificity in external validation and was proposed for community and primary-care screening [28], and hybrid AI models using photoplethysmography signals reported high accuracy in a small sample [29]. Together with these studies, our findings suggest that low-cost functional data requiring no additional equipment could, after further validation, contribute to pre-assessment approaches that complement spirometry.

The SHAP findings were broadly consistent with clinical knowledge. Pack-years ranked among the top five variables for all six targets and first for FEV1/FVC, consistent with the central role of smoking in airflow limitation [1]. mMRC was the most important variable for FEV1 (%); although dyspnoea correlates only weakly with FEV1 and adds information beyond it [6, 15], it is an important marker of disease burden and functional limitation [15]. For FVC (%), higher BMI lowered predictions, consistent with the restrictive effect of obesity on lung volumes, whereas older age raised predictions; this may relate to how reference equations estimate predicted values in older adults and should be interpreted cautiously. The prominence of 6MWD for FEV1 (%) and FEF25–75 and of estimated total step count for FVC (%) suggests that walking performance carries target-specific information, in keeping with the

relationship between exercise capacity and lung function [30] and with the altered gait patterns reported in COPD [31]. The recurrence of mMRC, 6MWD and BMI, which correspond to the dyspnoea, exercise capacity and body mass components of BODE [6], suggests that the models exploit variables used in the multidimensional assessment of COPD; it does not imply that BODE was reconstructed or its prognostic validity confirmed. Because SHAP explains contributions to individual predictions [23], these findings should not be interpreted as causal. The finding that 72.7% of patients with COPD were newly diagnosed is consistent with reports of substantial underdiagnosis [3] and highlights the need for pre-spirometry prioritization tools, although our sample was a referred clinical population without post-bronchodilator confirmation. AI has predicted an obstructive pattern from pre-bronchodilator spirometry with high accuracy in previously undiagnosed smokers in primary care [32], and electronic health record algorithms can identify individuals who should be referred for spirometry [33]. Combining clinical data with features obtained from a single video of a standard 6MWT, without additional devices, could offer a feasible decision-support approach in primary care and in settings with limited access to spirometry. However, an MAE of approximately 0.24 L for FEV1 indicates that the models are not currently accurate enough to replace spirometry for diagnosis, individual monitoring or preoperative risk stratification. They are better viewed as research tools that may support pre-spirometry assessment and prioritization, and their clinical utility requires independent validation and analyses around clinical decision thresholds. The approach is intended for use by staff who already perform the standard 6MWT, with a smartphone or camera and without additional devices or specialized expertise. When video quality is inadequate or an input variable is missing, spirometry should be preferred.

Limitations

This study has limitations. First, it was a single-centre study of 100 participants; with 18 inputs, this increases the risk of overfitting and limits the stability of performance estimates, and larger development and validation samples are needed [13]. Second, test sets of approximately 10 participants yield wide uncertainty, confidence intervals were not calculated, no external

validation was performed, and selecting the best model on the same splits may have biased the reported performance optimistically. Third, modelling patients with COPD and controls together widened the range of spirometric values and may have inflated R^2; performance in patients with COPD alone should be assessed. Fourth, most participants were men, COPD was predominantly mild-to-moderate with no GOLD 4 patients, and model performance could not be assessed separately by sex. Fifth, hyperparameters were fixed; the incremental contribution of gait features was not quantified by ablation against clinical-only models; the models were not compared with a simple baseline using age, sex and body size or reference-equation predicted values; and height was not a separate input. Sixth, spirometry was pre-bronchodilator only, whereas GOLD requires a post-bronchodilator FEV1/FVC < 0.70 [1]; individuals with reversible obstruction such as asthma could not be fully excluded and some may have been misclassified, although pre-bronchodilator measurements are used in some primary-care screening studies [32]. Inhaled therapies were also not withheld, which may have led to underestimation of obstruction. Seventh, the 6MWT was performed once rather than twice to reduce learning effects [20], possibly underestimating 6MWD. Eighth, video features were derived from three 10-s windows (30 s in total), and agreement between video-based and manual step counts was not reported, which may have increased measurement noise, particularly for the fatigue ratio and step-count variability. Ninth, correlated inputs (smoking variables; 6MWD and step count) may affect the stability of SHAP rankings, and COPD duration reflects a prior spirometry-based diagnosis and may therefore carry information indirectly related to the target, especially for FEV1/FVC. Finally, videos were recorded under controlled clinical conditions, and performance under other recording conditions and at home requires investigation.

Conclusions

Integrating demographic and clinical variables with functional and video-based 6MWT features predicted FEV1, FVC, FEV1/FVC and FEF25–75 with moderate-to-good accuracy, and the best-performing algorithm varied by target. Linear and regularized models were the most

consistent, whereas boosting models performed best for some targets. Age, sex, smoking exposure and dyspnoea were the main determinants of predictions, and gait features ranked highly for selected targets, although their independent contribution remains to be quantified. These models should be regarded not as an alternative to diagnostic spirometry but as research tools that may support pre-spirometry assessment and prioritization. Multicentre external validation, ablation analyses of video-derived features, evaluation around clinically relevant thresholds and quantification of model uncertainty are priorities for future work.

List of abbreviations

6MWD	six-minute walk distance
6MWT	six-minute walk test
AI	artificial intelligence
BMI	body mass index
CAT	COPD Assessment Test
COPD	chronic obstructive pulmonary disease
FEF25–75	forced expiratory flow between 25% and 75% of forced vital capacity
FEV1	forced expiratory volume in 1 s
FVC	forced vital capacity
GOLD	Global Initiative for Chronic Obstructive Lung Disease
KNN	k-nearest neighbours
MAE	mean absolute error
MedAE	median absolute error
mMRC	modified Medical Research Council
R^2	coefficient of determination
RMSE	root mean squared error
SHAP	SHapley Additive exPlanations
SpO_2	peripheral oxygen saturation
SVR	support vector regression

Declarations

Ethics approval and consent to participate

The study was approved by the Ondokuz Mayıs University Clinical Research Ethics Committee (decision no. OMÜ KAEK 2025/164, 26 March 2025) and was conducted in

accordance with the Declaration of Helsinki. Written informed consent covering video recording and the scientific use and publication of the data was obtained from all participants.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are not publicly available to protect the privacy of the participants but are available from the corresponding author on reasonable request. The analysis code and fitted models are also available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Ondokuz Mayıs University Scientific Research Projects Coordination Unit (grant number BAP01-2025-6067). The funder had no role in the study design, data collection, analysis, interpretation or writing of the manuscript.

Authors' contributions

Burçin Çelik: Conceptualization; Funding acquisition; Methodology; Project administration; Supervision; Writing – review & editing. Sema Gül: Methodology; Software; Validation; Writing – review & editing. Muammer Türkoğlu: Formal analysis; Methodology; Software; Visualization; Writing – review & editing. Öykü Çelik: Investigation; Validation; Writing – review & editing. Melda İşevi: Data curation; Investigation; Writing – review & editing. Caner İşevi: Conceptualization; Data curation; Investigation; Project administration; Writing – original draft; Writing – review & editing. All authors read and approved the final manuscript.

Acknowledgements

Preliminary results of this study were presented as an abstract at the 14th National Congress of Thoracic Surgery (Turkish Society of Thoracic Surgery), Antalya, Türkiye, 15–18 October 2026.

References

1. Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: 2026 report. https://goldcopd.org. Accessed 2 Nov 2026.
2. Brunelli A, Charloux A, Bolliger CT, et al. ERS/ESTS clinical guidelines on fitness for radical therapy in lung cancer patients (surgery and chemo-radiotherapy). Eur Respir J. 2009;34:17–41.
3. Diab N, Gershon AS, Sin DD, et al. Underdiagnosis and overdiagnosis of chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2018;198:1130–9.
4. Sunjaya AP, Edwards GD, Harvey J, et al. Validation of artificial intelligence spirometry diagnostic support software in primary care: a blinded diagnostic accuracy study. ERJ Open Res. 2025;11:00116–2025.
5. Li R, Guo H, Wu Q, Han J, Kang S. Deep learning based CT images for lung function prediction in patients with chronic obstructive pulmonary disease. BMC Pulm Med. 2025;25:484.
6. Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. N Engl J Med. 2004;350:1005–12.
7. Lugaresi C, Tang J, Nash H, et al. MediaPipe: a framework for building perception pipelines. arXiv. 2019. https://doi.org/10.48550/arXiv.1906.08172.
8. Bazarevsky V, Grishchenko I, Raveendran K, Zhu T, Zhang F, Grundmann M. BlazePose: on-device real-time body pose tracking. arXiv. 2020. https://doi.org/10.48550/arXiv.2006.10204.

9. Pinheira A, Casal-Guisande M, Represas-Represas C, et al. Artificial intelligence applications in chronic obstructive pulmonary disease: a global scoping review of diagnostic, symptom-based, and outcome prediction approaches. Biomedicines. 2025;13:3053.
10. Smith LA, Oakden-Rayner L, Bird A, et al. Machine learning and deep learning predictive models for long-term prognosis in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. Lancet Digit Health. 2023;5:e872–81.
11. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378.
12. Haider NS, Singh BK, Periyasamy R, Behera AK. Respiratory sound based classification of chronic obstructive pulmonary disease: a risk stratification approach in machine learning paradigm. J Med Syst. 2019;43:255.
13. Riley RD, Snell KIE, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: Part I – Continuous outcomes. Stat Med. 2019;38:1262–75.
14. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD Assessment Test. Eur Respir J. 2009;34:648–54.
15. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. Thorax. 1999;54:581–6.
16. Graham BL, Steenbruggen I, Miller MR, et al. Standardization of spirometry 2019 update. An official American Thoracic Society and European Respiratory Society technical statement. Am J Respir Crit Care Med. 2019;200:e70–88.
17. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. Eur Respir J Suppl. 1993;16:5–40.

458 18. Knudson RJ, Lebowitz MD, Holberg CJ, Burrows B. Changes in the normal maximal
459 expiratory flow-volume curve with growth and aging. Am Rev Respir Dis.
460 1983;127:725–34.
461 19. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories.
462 ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med.
463 2002;166:111–7.
464 20. Holland AE, Spruit MA, Troosters T, et al. An official European Respiratory
465 Society/American Thoracic Society technical standard: field walking tests in chronic
466 respiratory disease. Eur Respir J. 2014;44:1428–46.
467 21. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J
468 Mach Learn Res. 2011;12:2825–30.
469 22. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the
470 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data
471 Mining. New York: ACM; 2016. p. 785–94.
472 23. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural
473 Inf Process Syst. 2017;30:4765–74.
474 24. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A
475 systematic review shows no performance benefit of machine learning over logistic
476 regression for clinical prediction models. J Clin Epidemiol. 2019;110:12–22.
477 25. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B
478 Stat Methodol. 1996;58:267–88.
479 26. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc
480 Series B Stat Methodol. 2005;67:301–20.
481 27. Varoquaux G. Cross-validation failure: small sample sizes lead to large error bars.
482 Neuroimage. 2018;180:68–77.
483 28. Liu M, Yao X, Shi Y, et al. Impulse oscillometry-derived equation for prediction of
484 abnormal FEV1/FVC ratio for COPD screening in Chinese population: a multicenter
485 cross-sectional study. Lancet Reg Health West Pac. 2025;56:101501.

29. Melekoğlu E, Kocabıçak Ü, Uçar MK, Bilgin C, Bozkurt MR, Çunkaş M. A new diagnostic method for chronic obstructive pulmonary disease using the photoplethysmography signal and hybrid artificial intelligence. PeerJ Comput Sci. 2022;8:e1188.
30. Spruit MA, Watkins ML, Edwards LD, et al. Determinants of poor 6-min walking distance in patients with COPD: the ECLIPSE cohort. Respir Med. 2010;104:849–57.
31. Yentes JM, Rennard SI, Schmid KK, Blanke D, Stergiou N. Patients with chronic obstructive pulmonary disease walk with altered step time and step width variability as compared with healthy control subjects. Ann Am Thorac Soc. 2017;14:858–66.
32. Moreno Méndez R, Marín A, Ferrando JR, et al. Artificial intelligence applied to forced spirometry in primary care. Open Respir Arch. 2024;6(Suppl 2):100313.
33. Wells BJ, Saha AK, Ohar JA. Development and validation of an electronic health record algorithm to predict the presence of chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis. 2026;21:546651.

Figure titles and legends

Fig. 1 Study workflow. Eighteen input variables obtained from clinical assessment and the video-recorded six-minute walk test were modelled separately for each spirometric target with 13 regression models; the model with the highest mean coefficient of determination (R^2) was interpreted with SHapley Additive exPlanations (SHAP). BMI: body mass index; CAT: COPD Assessment Test; COPD: chronic obstructive pulmonary disease; FEF25–75: forced expiratory flow between 25% and 75% of forced vital capacity; FEV1: forced expiratory volume in 1 s; FVC: forced vital capacity; MAE: mean absolute error; MedAE: median absolute error; mMRC: modified Medical Research Council; RMSE: root mean squared error; SD: standard deviation; SpO2: peripheral oxygen saturation; 6MWD: six-minute walk distance.

Fig. 2 MediaPipe Pose-based gait analysis and feature extraction. Step counts in 10-s windows at the beginning (N1), middle (N2) and end (N3) of the test were used to calculate the estimated total step count (mean of N1–N3 × 36, the ratio of the 360-s test to the 10-s

window), the fatigue ratio (N3/N1) and step-count variability (standard deviation of N1–N3). The six-minute walk distance was measured manually.

Fig. 3 Heat map of mean coefficient of determination (R^2) values of the regression models by spirometric target. Cells show the mean and standard deviation over five data splits; red frames indicate the model with the highest mean R^2 for each target. The right-hand column shows the mean R^2 across the six targets, by which models are ordered. Negative values are coloured as 0. FEF25–75: forced expiratory flow between 25% and 75% of forced vital capacity; FEV1: forced expiratory volume in 1 s; FVC: forced vital capacity; KNN: k-nearest neighbours; SVR: support vector regression.

Fig. 4 SHapley Additive exPlanations (SHAP) summary plots for the best model for each spirometric target: (a) forced expiratory volume in 1 s (FEV1, L), lasso; (b) FEV1 (% predicted), linear regression; (c) forced vital capacity (FVC, L), elastic net; (d) FVC (% predicted), XGBoost; (e) FEV1/FVC (%), gradient boosting; (f) forced expiratory flow between 25% and 75% of FVC (L/s), histogram-based gradient boosting. Each dot represents one participant; positive SHAP values increase and negative values decrease the prediction. Colour indicates the value of the variable (red, high; blue, low).

Table 1: Demographic, clinical, functional and spirometric characteristics of the participants

Variable	Total (n = 100)	COPD (n = 33)	Control (n = 67)	P-value
Age (years)	66.0 (57.5–70.0)	67.0 (62.0–71.0)	65.0 (55.0–70.0)	0.052
Male sex, n (%)	82 (82.0)	31 (93.9)	51 (76.1)	0.029
BMI (kg/m²)	27.5 (24.2–31.3)	25.7 (22.7–28.4)	28.4 (25.7–32.2)	0.016
Smoking (pack-years)	40.0 (9.5–50.0)	50.0 (40.0–60.0)	35.0 (0.0–50.0)	0.004
Never smoker, n (%)	21 (21.0)	2 (6.1)	19 (28.4)	0.010
Current smoker, n (%)	24 (24.0)	9 (27.3)	15 (22.4)	0.591
Comorbidity, n (%)	79 (79.0)	24 (72.7)	55 (82.1)	0.280
Allergy, n (%)	7 (7.0)	2 (6.1)	5 (7.5)	1.000
Pulse rate (/min)	86.5 (81.0–90.0)	87.0 (81.0–90.0)	86.0 (81.5–90.0)	0.777
Respiratory rate (/min)	18.0 (16.0–20.0)	19.0 (16.0–22.0)	18.0 (16.0–20.0)	0.274
SpO_2 (%)	97.0 (96.0–97.0)	96.0 (96.0–97.0)	97.0 (96.0–97.0)	0.759

Variable	Total (n = 100)	COPD (n = 33)	Control (n = 67)	P-value
CAT score	5.0 (3.0–7.0)	5.0 (4.0–8.0)	5.0 (3.0–6.0)	0.083
mMRC score	1.0 (0.0–1.0)	1.0 (0.0–1.0)	1.0 (0.0–1.0)	0.278
6MWD (m)	500 (420–600)	450 (400–580)	500 (428–600)	0.190
FEV1 (L)	2.38 (1.98–2.76)	2.17 (1.69–2.58)	2.47 (2.15–2.88)	0.006
FEV1 (% predicted)	83.1 ± 18.6	71.9 ± 16.1	88.7 ± 17.4	<0.001
FVC (L)	3.17 (2.64–3.73)	3.34 (2.65–3.96)	3.15 (2.65–3.56)	0.315
FVC (% predicted)	86.3 ± 16.3	85.6 ± 17.9	86.6 ± 15.5	0.774
FEV1/FVC (%)	77.0 (69.6–81.0)	66.1 (63.2–69.6)	79.6 (77.0–83.3)	<0.001
FEF25–75 (L/s)	1.96 (1.40–2.44)	1.31 (0.91–1.68)	2.21 (1.94–2.94)	<0.001

Values are median (interquartile range), mean ± standard deviation or n (%). Continuous variables were compared with the independent-samples t-test or Mann–Whitney U test according to distribution, and categorical variables with Pearson's χ^2 or Fisher's exact test. 6MWD was measured manually. 6MWD: six-minute walk distance; BMI: body mass index; CAT: COPD Assessment Test; COPD: chronic obstructive pulmonary disease; FEF25–75: forced expiratory flow between 25% and 75% of FVC; FEV1: forced expiratory volume in 1 s; FVC: forced vital capacity; mMRC: modified Medical Research Council; SpO2: peripheral oxygen saturation.

Table 2: Best-performing model for each spirometric target

Target	Model	R^2	MAE	RMSE	MedAE
FEV1 (L)	Lasso	0.701 ± 0.047	0.237 ± 0.027	0.277 ± 0.029	0.222 ± 0.051
FEV1 (%)	Linear regression	0.555 ± 0.083	9.04 ± 1.61	11.26 ± 2.09	8.77 ± 2.73
FVC (L)	Elastic net	0.782 ± 0.049	0.242 ± 0.044	0.295 ± 0.052	0.209 ± 0.077
FVC (%)	XGBoost	0.595 ± 0.103	6.40 ± 1.23	8.02 ± 1.57	4.66 ± 1.72
FEV1/FVC (%)	Gradient boosting	0.701 ± 0.135	3.63 ± 0.96	4.30 ± 1.12	3.12 ± 1.31
FEF25–75 (L/s)	HistGradientBoosting	0.730 ± 0.078	0.444 ± 0.067	0.551 ± 0.078	0.369 ± 0.157

Values are mean ± standard deviation over five data splits. Error metrics are in the unit of the target (L, L/s or percentage points). FEF25–75: forced expiratory flow between 25% and 75% of FVC; FEV1: forced expiratory volume in 1 s; FVC: forced vital capacity; HistGradientBoosting: histogram-based gradient boosting; MAE: mean absolute error; MedAE: median absolute error; R^2: coefficient of determination; RMSE: root mean squared error.

Table 3: Five most important variables in the SHapley Additive exPlanations analysis for each spirometric target

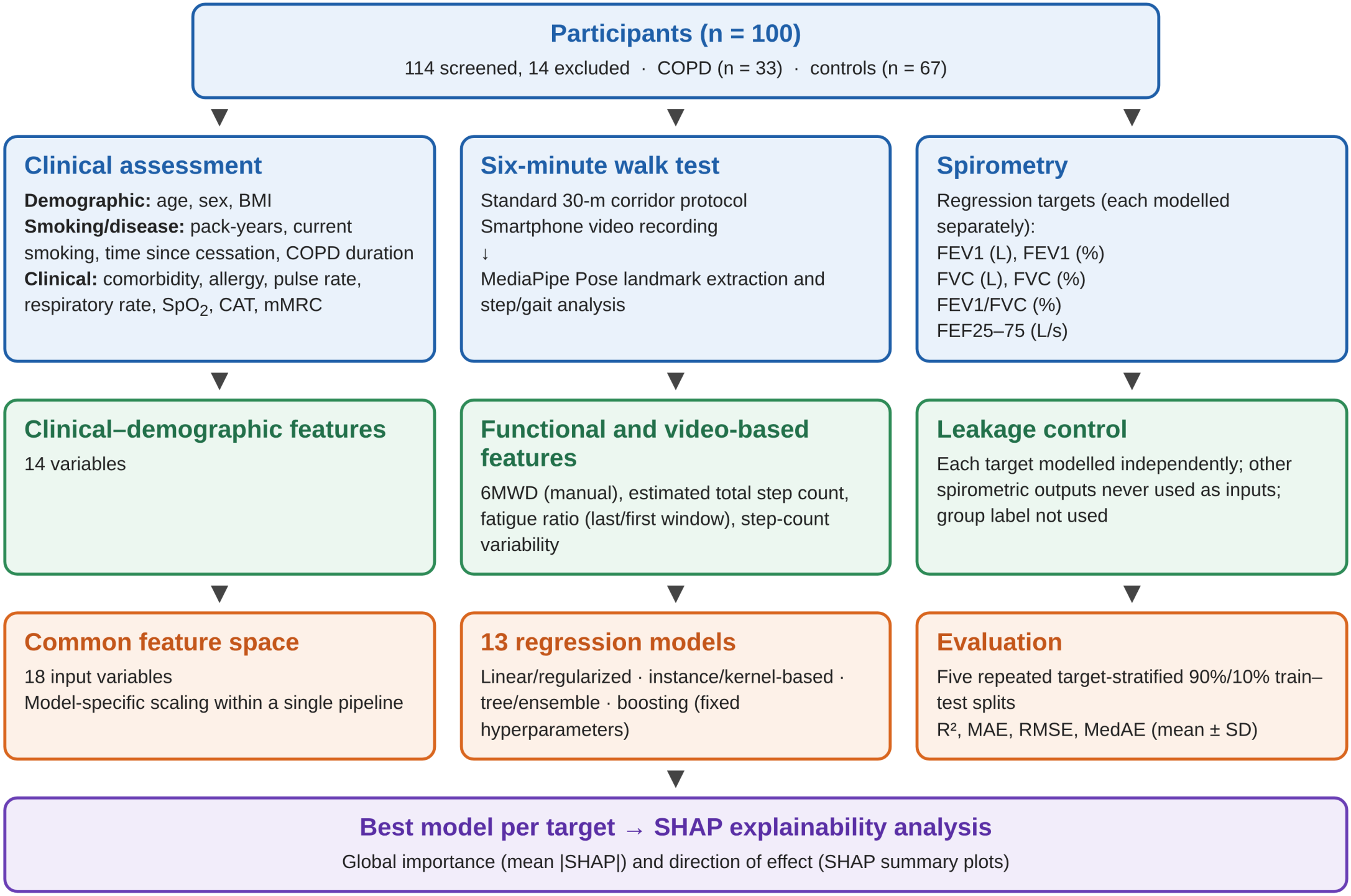
Target	Model	1st	2nd	3rd	4th	5th
FEV1 (L)	Lasso	Sex	Age	mMRC	Pack-years	Respiratory rate
FEV1 (%)	Linear regression	mMRC	Age	Pack-years	Respiratory rate	6MWD
FVC (L)	Elastic net	Sex	Age	mMRC	Comorbidity	Pack-years
FVC (%)	XGBoost	BMI	Time since cessation	Age	Estimated total step count	Pack-years
FEV1/FVC (%)	Gradient boosting	Pack-years	BMI	COPD duration	Pulse rate	Time since cessation
FEF25–75 (L/s)	HistGradientBoosting	Age	Pack-years	mMRC	BMI	6MWD

Rankings reflect the relative contribution of variables to model predictions (mean absolute SHAP value) and do not indicate statistical significance or causal effect. 6MWD: six-minute walk distance; BMI: body mass index; COPD: chronic obstructive pulmonary disease; FEF25–75: forced expiratory flow between 25% and 75% of FVC; FEV1: forced expiratory volume in 1 s; FVC: forced vital capacity; HistGradientBoosting: histogram-based gradient boosting; mMRC: modified Medical Research Council; SHAP: SHapley Additive exPlanations.

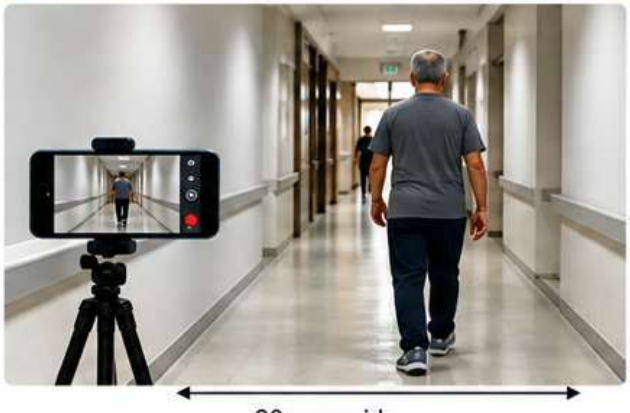
Additional files

Additional file 1 (.docx): Supplementary tables and figures. Table S1. Regression models and main hyperparameters. Table S2. Performance of all regression models by spirometric target. Table S3. Baseline and spirometric characteristics by sex. Figs. S1–S6. Coefficient of

556 determination (R^2) of the regression models for each spirometric target (maximum R^2 across
557 data splits and distribution over the five splits).
558 **Additional file 2 (.docx):** TRIPOD+AI checklist.

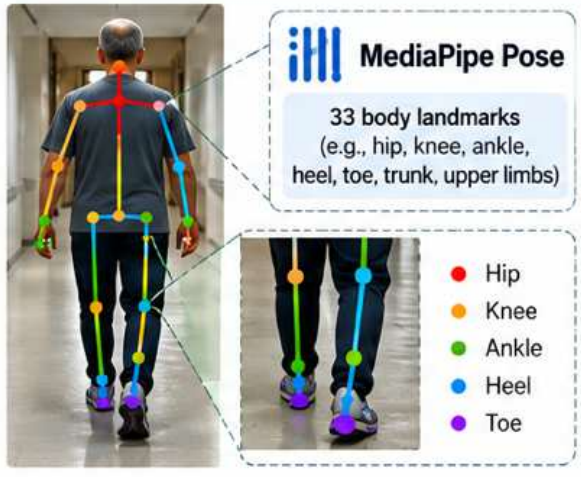

Participants (n = 100)
114 screened, 14 excluded · COPD (n = 33) · controls (n = 67)
Clinical assessment
Demographic: age, sex, BMI
Smoking/disease: pack-years, current smoking, time since cessation, COPD duration
Clinical: comorbidity, allergy, pulse rate, respiratory rate, SpO2, CAT, mMRC
Six-minute walk test
Standard 30-m corridor protocol
Smartphone video recording
↓
MediaPipe Pose landmark extraction and step/gait analysis
Spirometry
Regression targets (each modelled separately):
FEV1 (L), FEV1 (%)
FVC (L), FVC (%)
FEV1/FVC (%)
FEF25–75 (L/s)
Clinical–demographic features
14 variables
Functional and video-based features
6MWD (manual), estimated total step count, fatigue ratio (last/first window), step-count variability
Leakage control
Each target modelled independently; other spirometric outputs never used as inputs; group label not used
Common feature space
18 input variables
Model-specific scaling within a single pipeline
13 regression models
Linear/regularized · instance/kernel-based · tree/ensemble · boosting (fixed hyperparameters)
Evaluation
Five repeated target-stratified 90%/10% train–test splits
R², MAE, RMSE, MedAE (mean ± SD)
Best model per target → SHAP explainability analysis
Global importance (mean |SHAP|) and direction of effect (SHAP summary plots)

1 6-Minute Walk Test and Video Recording

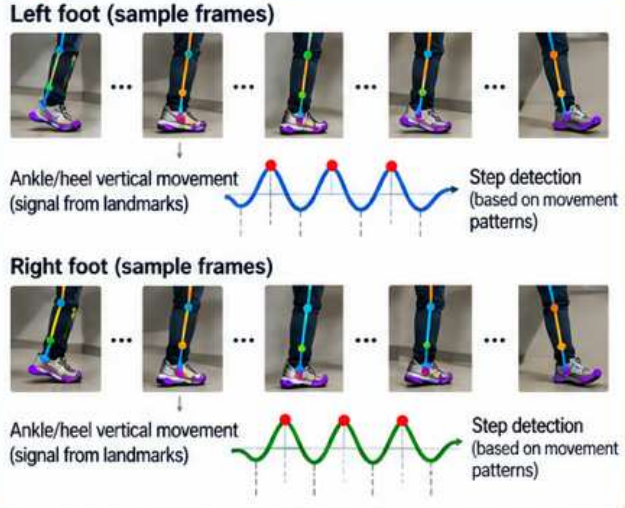


6-minute walk test was performed on a 30 m corridor and recorded with a smartphone camera.

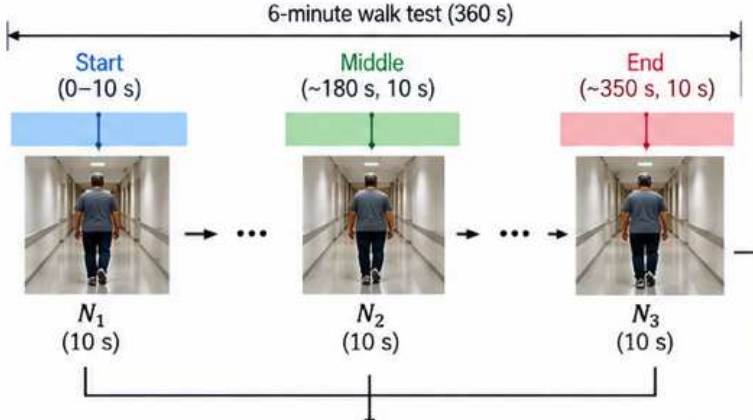
2 Pose Estimation with MediaPipe Pose



3 Step Detection from Landmark Movements



4 Sampling and Feature Calculation



6-minute walk distance (6DWD)
Total distance manually measured on a 30 m corridor (meters)

Estimated total step count

$$\hat{N}_{total} = \frac{N_1 + N_2 + N_3}{3} \times 36$$

(36 = 360 s / 10 s)

Fatigue ratio

$$R_{fatigue} = \frac{N_3}{N_1}$$

Step count variability

$$SD_{step} = SD(N_1, N_2, N_3)$$

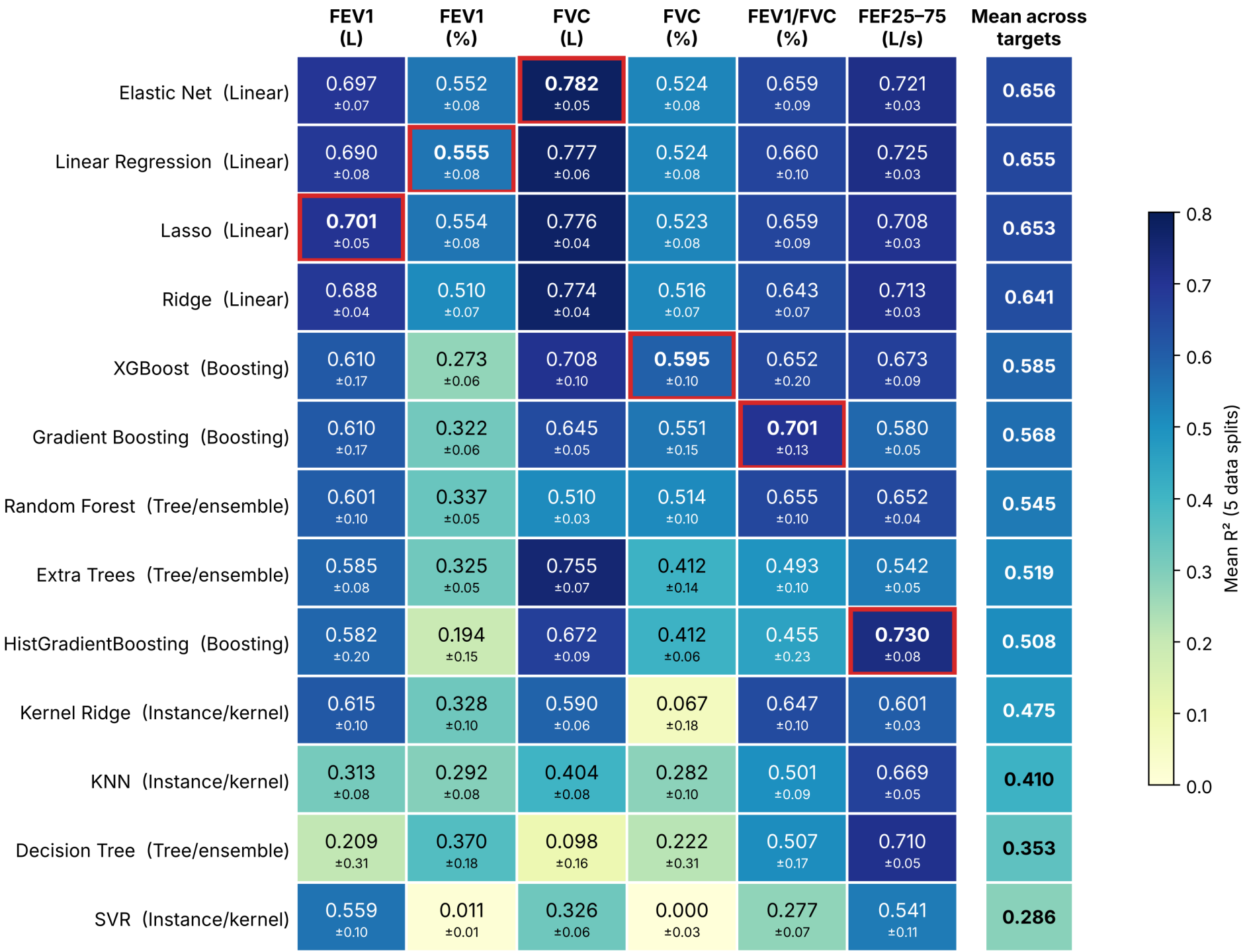
5 Gait Features for Modeling

6-minute walk distance (6DWD)
Total functional capacity (meters)

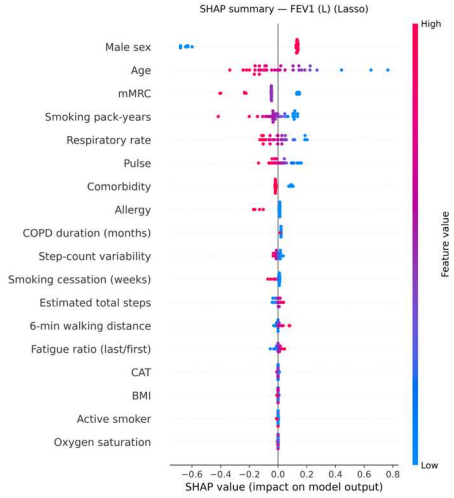
Estimated total step count
Total amount of movement during the test

Fatigue ratio
Ratio of step count in the last 10 s to the first 10 s

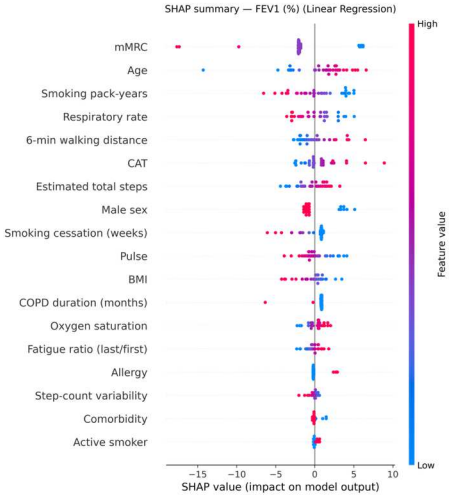
Step count variability
Standard deviation of step counts across the three 10-second intervals


FEV1 (L)
FEV1 (%)
FVC (L)
FVC (%)
FEV1/FVC (%)
FEF25–75 (L/s)
Mean across targets
Elastic Net (Linear)
0.697 ±0.07
0.552 ±0.08
0.782 ±0.05
0.524 ±0.08
0.659 ±0.09
0.721 ±0.03
0.656
Linear Regression (Linear)
0.690 ±0.08
0.555 ±0.08
0.777 ±0.06
0.524 ±0.08
0.660 ±0.10
0.725 ±0.03
0.655
Lasso (Linear)
0.701 ±0.05
0.554 ±0.08
0.776 ±0.04
0.523 ±0.08
0.659 ±0.09
0.708 ±0.03
0.653
Ridge (Linear)
0.688 ±0.04
0.510 ±0.07
0.774 ±0.04
0.516 ±0.07
0.643 ±0.07
0.713 ±0.03
0.641
XGBoost (Boosting)
0.610 ±0.17
0.273 ±0.06
0.708 ±0.10
0.595 ±0.10
0.652 ±0.20
0.673 ±0.09
0.585
Gradient Boosting (Boosting)
0.610 ±0.17
0.322 ±0.06
0.645 ±0.05
0.551 ±0.15
0.701 ±0.13
0.580 ±0.05
0.568
Random Forest (Tree/ensemble)
0.601 ±0.10
0.337 ±0.05
0.510 ±0.03
0.514 ±0.10
0.655 ±0.10
0.652 ±0.04
0.545
Extra Trees (Tree/ensemble)
0.585 ±0.08
0.325 ±0.05
0.755 ±0.07
0.412 ±0.14
0.493 ±0.10
0.542 ±0.05
0.519
HistGradientBoosting (Boosting)
0.582 ±0.20
0.194 ±0.15
0.672 ±0.09
0.412 ±0.06
0.455 ±0.23
0.730 ±0.08
0.508
Kernel Ridge (Instance/kernel)
0.615 ±0.10
0.328 ±0.10
0.590 ±0.06
0.067 ±0.18
0.647 ±0.10
0.601 ±0.03
0.475
KNN (Instance/kernel)
0.313 ±0.08
0.292 ±0.08
0.404 ±0.08
0.282 ±0.10
0.501 ±0.09
0.669 ±0.05
0.410
Decision Tree (Tree/ensemble)
0.209 ±0.31
0.370 ±0.18
0.098 ±0.16
0.222 ±0.31
0.507 ±0.17
0.710 ±0.05
0.353
SVR (Instance/kernel)
0.559 ±0.10
0.011 ±0.01
0.326 ±0.06
0.000 ±0.03
0.277 ±0.07
0.541 ±0.11
0.286
Mean R² (5 data splits)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

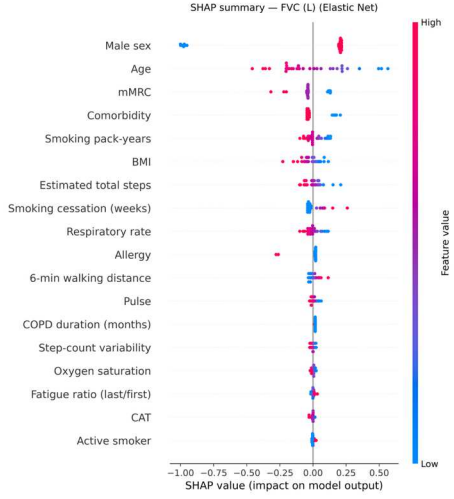
(a) FEV1 (L) – lasso



(b) FEV1 (%) – linear regression



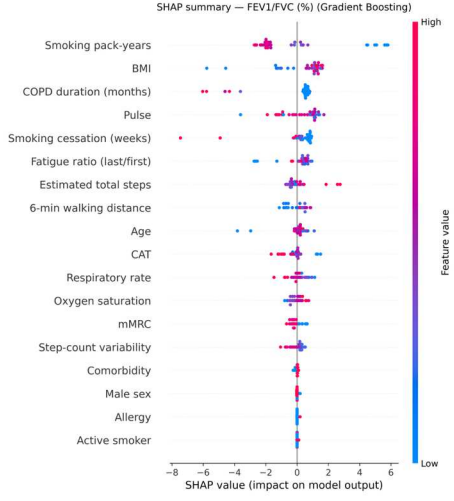
(c) FVC (L) – elastic net



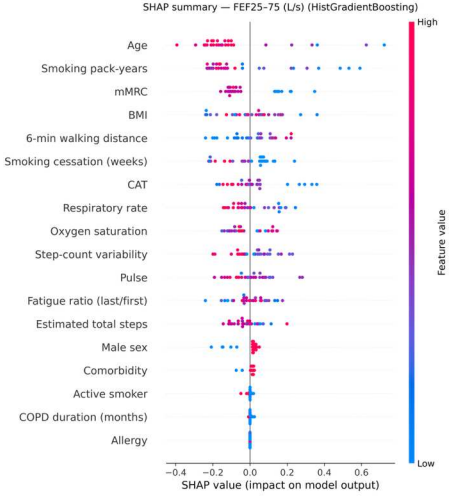
(d) FVC (%) – XGBoost



(e) FEV1/FVC (%) – gradient boosting



(f) FEF25–75 (L/s) – HistGradientBoosting



Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx
- Additionalfile2TRIPODAI.docx