diagnostics

MDPI

Predicting Pulmonary Function Testing from Quantified Computed Tomography Using Machine Learning Algorithms in Patients with COPD

Joshua Gawlitza [1,*], Timo Sturm [2], Kai Spohrer [2], Thomas Henzler [1], Ibrahim Akin [3,4], Stefan Schönberg [1], Martin Borggrefe [3,4], Holger Haubenreisser [1] and Frederik Trinkmann [3,5]

[1] Institute of Clinical Radiology and Nuclear Medicine, University Medical Center Mannheim, Medical Faculty Mannheim, Heidelberg University, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany; t.henzler@diagnostik-muenchen.de (T.H.); stefan.schoenberg@umm.de (S.S.); holger.haubenreisser@umm.de (H.H.)

[2] Department of General Management and Information Systems, University of Mannheim, 68131 Mannheim, Germany; timo.sturm@googlemail.com (T.S.); spohrer@uni-mannheim.de (K.S.)

[3] 1st Department of Medicine (Cardiology, Angiology, Pulmonary and Intensive Care), University Medical Center Mannheim, Medical Faculty Mannheim, Heidelberg University, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany; ibrahim.akin@umm.de (I.A.); martin.borggrefe@umm.de (M.B.); frederik.trinkmann@umm.de (F.T.)

[4] DZHK (German Center for Cardiovascular Research), partner site, 68167 Mannheim, Germany

[5] Department of Biomedical Informatics of the Heinrich-Lanz-Center, University Medical Center Mannheim, Medical Faculty Mannheim, Heidelberg University, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany

* Correspondence: Joshua.gawlitza@medma.uni-heidelberg.de; Tel.: +49-621-383-2204

check for updates

Abstract: Introduction: Quantitative computed tomography (qCT) is an emergent technique for diagnostics and research in patients with chronic obstructive pulmonary disease (COPD). qCT parameters demonstrate a correlation with pulmonary function tests and symptoms. However, qCT only provides anatomical, not functional, information. We evaluated five distinct, partial-machine learning-based mathematical models to predict lung function parameters from qCT values in comparison with pulmonary function tests. Methods: 75 patients with diagnosed COPD underwent body plethysmography and a dose-optimized qCT examination on a third-generation, dual-source CT with inspiration and expiration. Delta values (inspiration—expiration) were calculated afterwards. Four parameters were quantified: mean lung density, lung volume low-attenuated volume, and full width at half maximum. Five models were evaluated for best prediction: average prediction, median prediction, *k*-nearest neighbours (kNN), gradient boosting, and multilayer perceptron. Results: The lowest mean relative error (MRE) was calculated for the kNN model with 16%. Similar low MREs were found for polynomial regression as well as gradient boosting-based prediction. Other models led to higher MREs and thereby worse predictive performance. Beyond the sole MRE, distinct differences in prediction performance, dependent on the initial dataset (expiration, inspiration, delta), were found. Conclusion: Different, partially machine learning-based models allow the prediction of lung function values from static qCT parameters within a reasonable margin of error. Therefore, qCT parameters may contain more information than we currently utilize and can potentially augment standard functional lung testing.

Keywords: chronic obstructive pulmonary disease; machine learning; thorax

1. Introduction

Chronic obstructive pulmonary disease (COPD) is a common and largely avoidable disease that is characterized by irreversible airway obstruction, attributable to inhaled noxae and particles. Therapeutic decisions in patients with COPD have largely relied on spirometry. However, given the 2017 GOLD report, spirometry is no longer valid for therapeutic decision-making. Currently, medication is chosen by clinical criteria without widespread availability of objective, quantifiable diagnostic tools [1].

Quantified computed tomography (qCT) is an emerging technique in the complex field of COPD diagnostics. First described in 1988 by Muller et al., they used standardized, visual "density masks" to quantify emphysema in patients with COPD [2]. With advances in scanner technologies and evaluation algorithms, increasing amounts of information can be gathered from non-contrast-enhanced chest CTs in COPD. For example, bronchial airway parameters were shown to correlate not only with lung function parameters but also with exacerbation rates and clinical symptoms [3–5]. Additionally, significant pathological findings in non-contrast-enhanced chest CT in patients with COPD, such as emphysema and air trapping, are associated with an increased mortality [6–8]. On the basis of this evidence, the American Thoracic Society and the European Respiratory Society proposed to evaluate routine chest CT in patients with newly diagnosed COPD [9].

In light of these findings, as well as emergent research in this field, qCT is more likely to contain additionaly, comprehensive COPD-related information than is currently understood and utilized.

Machine-based learning algorithms comprise emergent techniques to evaluate potential connections between imaging data and clinical parameters or outcomes [10,11]. Recently, several commercially available software solutions, using algorithms similar to those in stroke imaging and diagnosis, have been successfully evaluated [12]. However, the potential benefits of these methods have currently not been applied to the field of qCT in patients with COPD.

Therefore, the aim of this study was to evaluate the feasibility of predicting lung function parameters with radiomics features, acquired by non-contrast-enhanced chest qCT, in patients with COPD. Various mathematical models as well as machine learning-based algorithms were evaluated.

2. Methods

2.1. Subjects

The HIPAA-compliant study protocol, which is in accordance to the Declaration of Helsinki, was approved by our local ethics committee (2015-415M-MA).

We prospectively enrolled 75 consecutive patients with previously diagnosed COPD and a clinical indication for non-contrast-enhanced chest CT in a single center approach. Written informed consent was obtained from all patients following a full explanation of the purpose of the study and of potential risks and discomfort associated with their participation.

2.2. Lung Function Testing

All patients underwent whole-body plethysmography (MasterScreen®Body, CareFusion, Höchberg, Germany) which yielded the following parameters: forced expiratory volume in one second (FEV_1), ratio of residual volume to total lung capacity (RV%TLC), and Tiffeneau index (FEV_1/VC). Apart from RV%TLC, all plethysmographic values are represented as the percentage of the expected value, as calculated according to current ATS/ERS recommendations [13,14].

2.3. Computerized Tomography Examinations

A non-contrast chest scan was performed in maximum inspiration and maximum end-expiration using a third-generation dual-source CT system (Somatom FORCE, Siemens Healthineers, Forchheim, Germany) at 100 kVp with an additional tin filter behind the source for dose reduction [15,16]. In addition to a pre-imaging briefing, the patients received voice commands for optimal inspiration

and expiration results. Existing medication was not withheld prior to imaging. The scan parameters were as follows: 100 kVp tube voltage, automated tube current modulation using 96 mAs at 120 kVp as reference (effective mAs = 166.5 ± 105), 0.25 s rotation time, pitch 1.2, 192 mm × 0.6 mm detector collimation. All images were reconstructed with a slice thickness of 1.5 mm, using a suitable reconstruction kernel for quantitative lung analysis (Br32) and a third-generation, iterative reconstruction technique (Adaptive Model-based Iterative reconstruction [ADMIRE], Siemens Healthineers, Germany). The reconstruction algorithm was explained in a recent publication by Gordic et al. [17]. An iterative strength level of four (out of a maximum of five) was chosen for the present study for optimum image noise, as recommended by the CT vendor for quantitative lung analysis. The average CTDI was 0.48 ± 0.19 mGy and the mean DLP 17.2 ± 6.5 mGy·cm.

2.4. Image Analysis

Inspiratory and expiratory datasets were analyzed using dedicated semi-automatic software (SyngoViaVB10, Pulmo3D, Siemens Healthineers, Forchheim, Germany). Lung segmentation was automated and manually revised only if necessary. Four quantitative parameters were acquired: total lung volume (volume), mean lung density (MLD), full-width-half-max (FWHM), and low attenuation volume (LAV). The LAV threshold for emphysema was set to −950 HU. This cut-off has been extensively evaluated in previous studies and strongly correlates with microscopic and gross emphysema [18–20]. FWHM marks the width at the half maximum of the voxel count to a specific HU value curve, representing the density distribution of the lung parenchyma. The difference of the values between inspiratory and expiratory scans were defined as delta values and were used as a distinct dataset in this evaluation. The use of this additional delta data has shown to be beneficial in previous studies [3].

2.5. Model Training and Evaluation

Five models were used for the prediction of lung function values: mean prediction, median prediction, polynomial regression, *k*-nearest neighbor regression (kNN), an additive regression with regression trees in XGBoost, and an artificial neural network multilayer perceptron regression (Tensorflow, Google, Mountain View, CA, USA). All breathing states (inspiration, expiration, delta) and parameters (volume, MLD, LAV and FWHM) were used as input data for every model.

Mean and median prediction was performed, using target lung function values. Therefore, both methods yielded the same results for inspiration, expiration, and delta values. To find best performance, polynomial regression and kNN were performed using different degrees and k values, respectively. The polynomial regression was calculated for every degree from 1 to 10 because the performance peak had already been reached after the third degree. The kNN regression was performed for all k-values between 1 and 74 because 75 patients were included in the study. The approach of finding the ideal k and degree can be seen in Figure 1. Weighted voting based on distance was used to combine the values of the k neighbors.

XGBoost and multilayer perceptron largely remained at their default settings. For XGBoost, this meant a total number of 100 trees, with a maximal depth per tree of three and a boosting learning rate of 0.1. The multilayer perceptron had five neurons building the input layer, two hidden layers with 256 neurons each, and one output neuron.
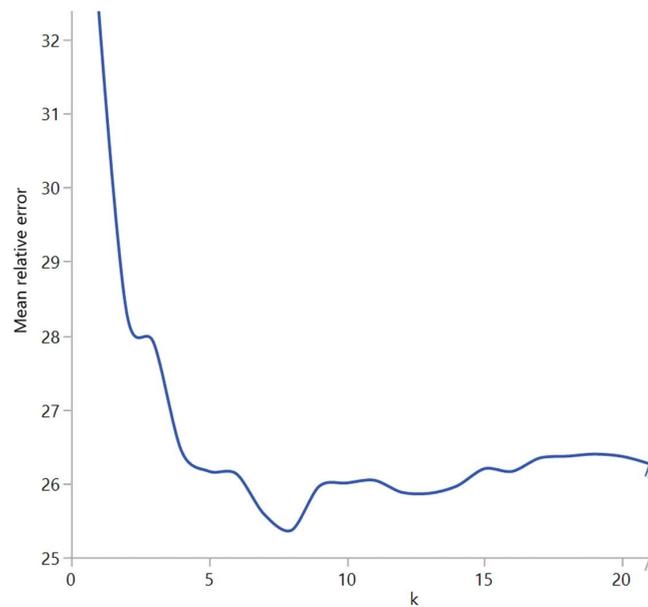
Figure 1. Detail of the brute force approach to find ideal *k* in *k*-nearest neighbour analysis. The mean relative error of the RV%TLC prediction is shown for various *k*s.

2.6. Model Evaluation and Statistical Analysis

"Leave one out" cross validation was used to measure the prediction performance for all models and parameters, which has been shown to be beneficial for small cohorts [21,22]. The mean absolute and the corresponding mean relative error (MRE) were used as markers for prediction performance [23]. Both values were compared for every model, breathing state, and parameter using ANOVA and corrected by Tuckey HSD for multiple testing via dedicated software (JMP 12, SAS, Cary, CA, USA). A *p*-value of less than 0.05 was considered statistically significant.

3. Results

3.1. Data Mining

In the brute force approach to find the ideal *k* for the kNN model and the degree of the polynomial regression, respectively, different values demonstrated the best mean relative error. The degrees are shown in Table 1. These values were used in additional analysis for comparison with the performance parameters of other mathematical models.

Table 1. *k*s and degrees with the lowest mean relative error.

	*k*s and Degrees with Lowest Mean Relative Errors	
	k with Best MRE	Degree with Best MRE
Inspiration %FEV$_1$	23	1
Inspiration %FEV$_1$/VC	14	1
Inspiration RV/TLC	13	2
Expiration %FEV$_1$	6	3
Expiration %FEV$_1$/VC	13	2
Expiration RV/TLC	7	3
Delta %FEV$_1$	7	2
Delta %FEV$_1$/VC	12	1
Delta RV/TLC	8	1

FEV$_1$: forced expiratory volume in one second; VC: vital capacity; RV: residual volume; TLC: total lung capacity; MRE: mean relative error.

3.2. Model Comparison in Prediction Performance

The actual measured values of the lung function testing, the predicted values of the individual algorithm, and the *p*-values of the ANOVA are shown in Table 2.

Significant differences in the ANOVA among both predicted and measured values were found for %FEV$_1$ in all datasets (inspiration ($p = 0.0024$), expiration ($p = 0.0078$), delta (0.002)), and FEV$_1$/VC in the delta dataset ($p = 0.0005$). As shown in subsequent analysis, the predicted lung function values of the median prediction model showed significant differences to all other parameters for %FEV1 in the inspiration and expiration dataset. For FEV$_1$ in the delta dataset, the median prediction values only showed significant differences to the XGBoost values (difference = 8%; $p = 0.0479$). The mean difference between median prediction values and the other models for the %FEV1 was 7.4 % (mean $p = 0.0248$) in the expiration and 7.6% (mean $p = 0.0117$) in the inspiration dataset. In all three cases, the median prediction showed lower values when compared with the actual body plethysmography values as well as the other prediction models.

The ANOVA also showed a difference in variance for the FEV$_1$/VC in the delta dataset. However, the neural network model-based values showed significant lower values when compared to the other models and body plethysmography results (mean difference = 7.7%; mean $p = 0.0001$). For RV/TLC, no significant differences were found in any dataset.

Overall, the predicted values from the kNN model, the polynomial regression, the mean prediction, and the XGBoost model did not show significant differences to the measured lung function values (*p*-range 0.06–1).

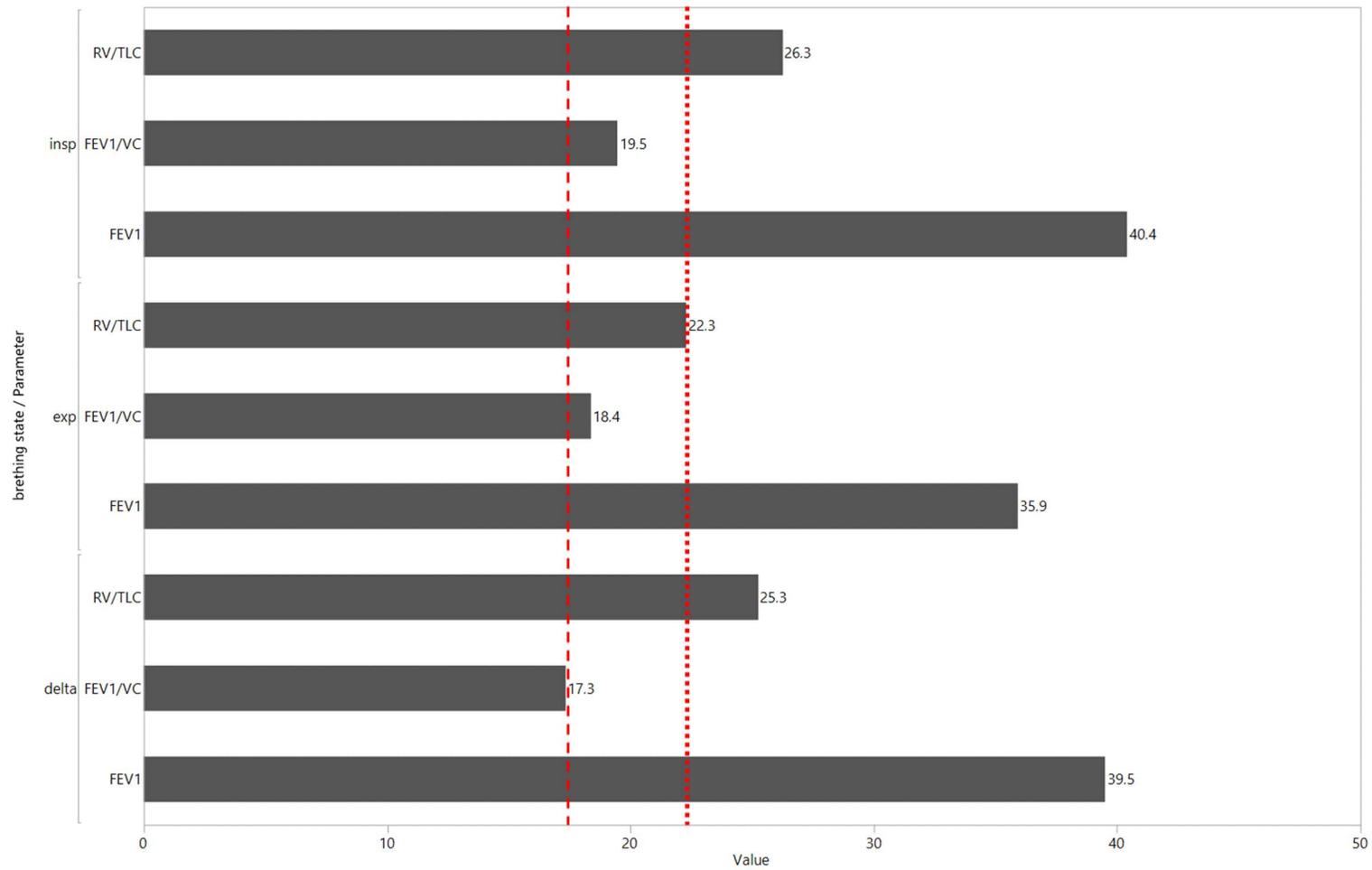### 3.3. Absolute and Relative Errors of Prediction Models

With regard to the mean absolute and relative errors of the predicted lung function values, no significant difference among the used models was found. The MRE ranged between 16 and 47 percent. Overall, kNN and the polynomial regression consistently showed the lowest absolute and relative errors in relation to the measured body plethysmography values.

Despite the lack of significant differences in the statistical analysis, there were distinct differences among the predicted lung function values of the different models. As shown in Figure 2, the MRE of the polynomial regression for RV/TLC (dotted line) was lower in the expiratory dataset (Ø 22.3) in comparison with the inspiratory (Ø 26) and delta datasets (Ø 25). This phenomenon was also noticeable for FEV$_1$/VC. In this case, the lowest MRE was calculated for the delta dataset (Ø 17) in comparison with the expiratory (Ø 18) and inspiratory datasets (Ø 20), with details given in Table 3.

Table 2. Lung function values from body plethysmography and the predicted lung function values from all algorithms used with their standard deviation.

		Measured Values	Mean Prediction	Median Prediction	kNN	Polynomial Regression	Neural Network	XGBoost	*p*-Value
	Lung Function Values from Body Plethysmorgraphy and Predicted Lung Function Values								
inspiration	%FEV$_1$	56 ± 6	56 ± 0.3	48 ± 0.1	56 ± 6	55 ± 13	55 ± 13	56 ± 16	0.0024
	%FEV$_1$/VC	55 ± 7	56 ± 0.1	54 ± 0.1	55 ± 7	56 ± 7	56 ± 7	56 ± 10	0.732
	RV/TLC	59 ± 6	60 ± 0.2	61 ± 0.5	59 ± 6	60 ± 10	60 ± 11	59 ± 11	0.96
expiration	%FEV$_1$	53 ± 12	56 ± 0.3	48 ± 0.1	53 ± 12	56 ± 14	55 ± 14	56 ± 16	0.0078
	%FEV$_1$/VC	55 ± 8	56 ± 0.1	54 ± 0.1	55 ± 8	56 ± 8	56 ± 8	56 ± 11	0.6
	RV/TLC	61 ± 9	60 ± 0.2	61 ± 0.5	61 ± 9	59 ± 11	59 ± 12	59 ± 12	0.7996
delta	%FEV$_1$	56 ± 13	56 ± 0.3	48 ± 0.1	56 ± 13	56 ± 14	49 ± 27	56 ± 16	0.002
	%FEV$_1$/VC	55 ± 8	56 ± 0.1	54 ± 0.1	55 ± 8	56 ± 8	49 ± 25	56 ± 11	0.0005
	RV/TLC	59 ± 8	60 ± 0.2	61 ± 0.5	59 ± 8	59 ± 8	54 ± 26	58 ± 13	0.0522

Measured lung function values from the body plethysmography and predicted lung function values from the mathematical models, both with their standard deviation. For *k*-nearest neigbours and the polynomial regression, only the best performing *k*s and degrees are shown (see Table 1). The p-values of the ANOVA are shown in the right column. FEV1: forced expiratory volume in one second; VC: vital capacity; RV: residual volume; TLC: total lung capacity; kNN: *k*-nearest neighbour.

Figure 2. Mean relative errors of the polynomial regression for all datasets (inspiration, expiration, delta) and predicted parameters. The dotted and dashed line mark the distinct differences in prediction performance with regard to the expiration and delta scan for the FEV1/VC and RV/TLC, respectively.

Table 3. Absolute and relative errors of all algorithms tested.

		Mean Prediction		Median Prediction		kNN		Polynomial Regression		Neural Network		XGBoost	
		MAR	**MRE**	**MAR**	**MRE**	**MAR**	**MRE**	**MAR**	**MRE**	**MAR**	**MRE**	**MAR**	**MRE**
inspiration	%FEV$_1$	21 ± 13	46 ± 43	20 ± 16	39 ± 6	19 ± 13	43 ± 39	19 ± 13	40 ± 37	20 ± 13	41 ± 34	22 ± 18	47 ± 50
	%FEV$_1$/VC	12 ± 8	22 ± 17	11 ± 8	21 ± 12	9 ± 8	17 ± 15	10 ± 7	19 ± 15	12 ± 8	22 ± 17	10 ± 8	19 ± 16
	RV/TLC	14 ± 10	30 ± 40	14 ± 10	31 ± 42	13 ± 9	28 ± 38	12 ± 8	26 ± 36	13 ± 9	27 ± 39	14 ± 12	31 ± 46
expiration	%FEV$_1$	21 ± 13	46 ± 43	20 ± 16	39 ± 6	18 ± 15	36 ± 37	17 ± 13	36 ± 37	19 ± 14	41 ± 36	22 ± 16	46 ± 45
	%FEV$_1$/VC	12 ± 8	22 ± 17	11 ± 8	21 ± 12	9 ± 8	16 ± 14	10 ± 7	18 ± 15	12 ± 8	22 ± 17	10 ± 8	20 ± 17
	RV/TLC	14 ± 10	30 ± 40	14 ± 10	31 ± 42	11 ± 9	25 ± 31	11 ± 8	22 ± 28	11 ± 8	22 ± 25	12 ± 9	24 ± 26
delta	%FEV$_1$	21 ± 13	46 ± 43	20 ± 16	39 ± 6	18 ± 15	39 ± 42	19 ± 15	40 ± 37	21 ± 19	38 ± 35	18 ± 15	40 ± 46
	%FEV$_1$/VC	12 ± 8	22 ± 17	11 ± 8	21 ± 12	9 ± 8	16 ± 14	9 ± 8	17 ± 15	20 ± 16	35 ± 25	9 ± 8	17 ± 16
	RV/TLC	14 ± 10	30 ± 40	14 ± 10	31 ± 42	12 ± 9	25 ± 30	12 ± 10	25 ± 30	26 ± 22	47 ± 49	14 ± 12	29 ± 34

Mean absolute error (MAR) and the mean relative error (MRE) of the predicted lung function values. The MAR is given as the amount of the absolute error needed to compensate the negative signs. FEV1: forced expiratory volume in one second; VC: vital capacity; RV: residual volume; TLC: total lung capacity; kNN: *k*-nearest neighbour; MRE: mean relative error; MAR: mean absolute error.

4. Discussion

This work evaluated the feasibility of predicting lung function values by parameters derived from qCT using different mathematical models. We demonstrated that this prediction can be performed within a reasonable margin of error with regard to differences in target values and prediction models. Overall, kNN and polynomial regression showed the lowest MREs in our study. As shown in previous studies, kNN and polynomial regression are suitable methods for prediction models in small sample sizes [24,25].

Our analysis found only minor significant differences in prediction performance among the evaluated models. Significant differences were shown for the median prediction in the FEV_1. One reason for this might be that the median prediction was only based on the lung function values and was used as a baseline for the other prediction models. This explains the low standard deviation of both as well as mean and median prediction.

The XGBoost-based prediction showed the FEV_1 in the delta values and significantly lower prediction values when compared with the other models. That only significant differences were found in the FEV_1 seems to be consistent when looking at the global results: all prediction models performed better in the prediction of FEV_1/VC and RV/TLC than only FEV_1.

One explanation for this and the aforementioned significant differences might be that the FEV_1 is an explicit dynamical parameter and the qCT, even when performed in two breathing states, contains only static information. Therefore, the TLC, VC, and the RV may be easier to predict with regard to the initial static information. This theory is supported by the different datasets (inspiration, expiration, delta), which showed different MREs with regard to the parameters. This might be attributable to the physiological meaning of the different datasets. In terms of the expiratory dataset, if the patient is in maximal expiration, the volume measured in the expiratory CT is equivalent to the RV. The is similar to the inspiratory CT, which is, at maximal inspiration, equivalent to the TLC. Similarly, if we examine the delta dataset, the subtraction of inspiratory (TLC) and expiratory (RV) data, we receive the maximal mobilizable volume, i.e., the VC. Accordingly, the various MREs seem to be consistent because the lowest MRE for the FEV_1/VC was found in the delta dataset. This was similar to the RV/TLC, which was the lowest in the expiration dataset. Even if the differences were not statistically significant, the calculated results were in line with our physiological understanding of the different qCT datasets. Similar findings have been reported in previous studies [3,26,27]. A larger cohort might confirm our hypothesis to a significant level. In summary, this highlights the importance of an additional scan in maximal expiration, as stated by the Fleischner Society and other studies [28–31].

Most previous studies only correlated lung function values with qCT parameters but did not attempt to predict the lung function itself [3,27,29,31]. Nevertheless, Gu et al. predicted lung function values based on the emphysema acquired by qCT. They calculated MREs in the prediction of the FEV_1% between 17% and 256% [32]. Similar to our study, the prediction of FEV_1/VC performed better when compared with the FEV_1% prediction. In comparison with our approach, they only used one linear prediction model and only different subdivisions of the LAV as initial prediction factors.

In our analysis, the neural-network-based prediction did not perform significantly different when compared with the other mathematical models. With regard to the rise of neural-network-based models, this seems counterintuitive [10,11]. A significant reason for this may be our sample size. As shown in previous work, the size of the initial training set is crucial for the performance of such a neural-network-based prediction model [33,34]. A larger initial dataset will likely significantly improve the performance.

Our study contained several limitations, one of which was our relatively small sample size. A larger cohort will not only improve performance in the neural-network-based model but will also improve generalizability in all models. Further, we evaluated only a limited variety of algorithms with specific settings. A more comprehensive tuning of the profound parameters in XGBoost and the neural network-based model would most likely have improved the prediction performance. Further

studies should evaluate the influence of different factors, for example, tree depths, network layers, or boosting rates.

A third problem was in the initial training dataset itself. We chose a real-life dataset of patients with diagnosed COPD. With regard to the heterogeneity of COPD and its airway obstruction states, a wide standard deviation is present. Therefore, with regard to the small sample size, the models had to adapt to a large range of values. Again, especially for such a heterogenic cohort, a larger sample would be beneficial.

In summary we were able to show that different prediction models can be used to predict lung function values from qCT in patients with COPD. Therefore, the best prediction performance was most likely attributable to the sample size, which was calculated for kNN and polynomial regression. Nevertheless, the prediction performance was generally not significantly different from the other models, suggesting they can also be used for similar objectives. Therefore, computationally incomprehensive models (e.g., KNN or polynomial regression) might offer valid alternatives in lung function prediction when compared with neural networks. Further evaluations with larger sample sizes might benefit the XGBoost and neural-network-based prediction. Nevertheless, the prediction performance of the models used in this work significantly overcomes previous, one-linear regression systems using explorations [32].

Overall, the prediction of static parameters was superior when compared with the ones of dynamic parameters. Additionally, different breathing states showed different prediction performances with regard to individual parameters. In contrast to inspiration-only-based research, this underlines the need for an additional expiratory scan; further studies should take this as well as the calculated delta values into account [35].

Although traditional, inexpensive lung function testing (e.g. spirometry) is not likely to be replaced by quantified computed tomography, there might be a future application for similar prediction models. Sites without a pneumology department or additional information from a noncontrast-enhanced chest CT in outpatient care would benefit from such a prediction system because lung function changes might be recognized earlier. Further, specific pneumological testing could be carried out after these incidental findings. As seen in this proof-of-concept work, quantified computed tomography can be used to predict lung function values in COPD. The distinct differences shown between dynamic and static parameters as well as the importance of separate inspiratory, expiratory, and delta datasets should be useful for research groups for future evaluation.

Author Contributions: Conceptualization, J.G., T.H. and F.T.; methodology, T.S. and K.S.; software, T.S. and J.G.; validation, I.A., S.S. and M.B.; formal analysis: T.S.; resources: S.S. and M.B.; data curation: J.G. and T.S.; original draft preparation: J.G. and F.T.; writing review and editing: H.H. and T.H.; visualization: J.G.; supervision: H.H. and F.T.; project administration: K.S. and H.H.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

COPD	chronic obstructive pulmonary disease
FEV1	forced expiratory volume at one second
kNN	*k*-nearest neighbour
MRE	mean relative error
qCT	quantified computed tomography
RV	residual volume
TLC	total lung capacity
VC	vital capacity

References

1. Vogelmeier, C.F.; Criner, G.J.; Martinez, F.J.; Anzueto, A.; Barnes, P.J.; Bourbeau, J.; Celli, B.R.; Chen, R.; Decramer, M.; Fabbri, L.M.; et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report: Gold executive summary. *Am. J. Respir. Crit. Care Med.* **2017**, *195*, 577–582. [CrossRef] [PubMed]

2. Muller, N.L.; Staples, C.A.; Miller, R.R.; Abboud, R.T. "Density mask". An objective method to quantitate emphysema using computed tomography. *Chest* **1988**, *94*, 782–787.

3. Gawlitza, J.; Trinkmann, F.; Scheffel, H.; Fischer, A.; Nance, J.W.; Henzler, C.; Vogler, N.; Saur, J.; Akin, I.; Borggrefe, M.; et al. Time to exhale: Additional value of expiratory chest CT in chronic obstructive pulmonary disease. *Can. Respir. J.* **2018**, *2018*, 9. [CrossRef] [PubMed]

4. Han, M.K.; Kazerooni, E.A.; Lynch, D.A.; Liu, L.X.; Murray, S.; Curtis, J.L.; Criner, G.J.; Kim, V.; Bowler, R.P.; Hanania, N.A. Chronic obstructive pulmonary disease exacerbations in the COPDGene study: Associated radiologic phenotypes. *Radiology* **2011**, *261*, 274–282. [CrossRef] [PubMed]

5. Grydeland, T.B.; Dirksen, A.; Coxson, H.O.; Eagan, T.M.; Thorsen, E.; Pillai, S.G.; Sharma, S.; Eide, G.E.; Gulsvik, A.; Bakke, P.S. Quantitative computed tomography measures of emphysema and airway wall thickness are related to respiratory symptoms. *Am. J. Respir. Crit. Care Med.* **2010**, *181*, 353–359. [CrossRef] [PubMed]

6. Ostridge, K.; Wilkinson, T.M. Present and future utility of computed tomography scanning in the assessment and management of COPD. *Eur. Respir. J.* **2016**, *48*, 216–228. [CrossRef] [PubMed]

7. Haruna, A.; Muro, S.; Nakano, Y.; Ohara, T.; Hoshino, Y.; Ogawa, E.; Hirai, T.; Niimi, A.; Nishimura, K.; Chin, K.; et al. Ct scan findings of emphysema predict mortality in COPD. *Chest* **2010**, *138*, 635–640. [CrossRef] [PubMed]

8. Johannessen, A.; Skorge, T.D.; Bottai, M.; Grydeland, T.B.; Nilsen, R.M.; Coxson, H.; Dirksen, A.; Omenaas, E.; Gulsvik, A.; Bakke, P. Mortality by level of emphysema and airway wall thickness. *Am. J. Respir. Crit. Care Med.* **2013**, *187*, 602–608. [CrossRef] [PubMed]

9. Celli, B.R.; Decramer, M.; Wedzicha, J.A.; Wilson, K.C.; Agusti, A.A.; Criner, G.J.; MacNee, W.; Make, B.J.; Rennard, S.I.; Stockley, R.A.; et al. An official american thoracic society/european respiratory society statement: Research questions in COPD. *Eur. Respir. Rev.* **2015**, *24*, 159–172. [CrossRef] [PubMed]

10. Parmar, C.; Grossmann, P.; Bussink, J.; Lambin, P.; Aerts, H.J. Machine learning methods for quantitative radiomic biomarkers. *Sci. Rep.* **2015**, *5*, 13087. [CrossRef] [PubMed]

11. Gao, X.; Chu, C.; Li, Y.; Lu, P.; Wang, W.; Liu, W.; Yu, L. The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from 18F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer. *Eur. J. Radiol.* **2015**, *84*, 312–317. [CrossRef] [PubMed]

12. Hoyte, L.C.; Al Sultan, A.S.; Finkelstein, S.; Boyko, M.; Fok, D.; Pordeli, P.; Horn, M.; Neweduk, A.; Yu, A.; Appireddy, R. Reliability of automated software to assign e-ASPECTS to CT scans for acute ischemic changes (s8. 006). *Neurology* **2017**, *88*, S8.006.

13. Quanjer, P.H.; Stanojevic, S.; Cole, T.J.; Baur, X.; Hall, G.L.; Culver, B.H.; Enright, P.L.; Hankinson, J.L.; Ip, M.S.; Zheng, J.; et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: The global lung function 2012 equations. *Eur. Respir. J.* **2012**, *40*, 1324–1343. [CrossRef] [PubMed]

14. Pellegrino, R.; Viegi, G.; Brusasco, V.; Crapo, R.O.; Burgos, F.; Casaburi, R.; Coates, A.; van der Grinten, C.P.; Gustafsson, P.; Hankinson, J.; et al. Interpretative strategies for lung function tests. *Eur. Respir. J.* **2005**, *26*, 948–968. [CrossRef]

15. Weis, M.; Henzler, T.; Nance, J.W., Jr.; Haubenreisser, H.; Meyer, M.; Sudarski, S.; Schoenberg, S.O.; Neff, K.W.; Hagelstein, C. Radiation dose comparison between 70 kvp and 100 kvp with spectral beam shaping for non-contrast-enhanced pediatric chest computed tomography: A prospective randomized controlled study. *Investig. Radiol.* **2017**, *52*, 155–162. [CrossRef] [PubMed]

16. Haubenreisser, H.; Meyer, M.; Sudarski, S.; Allmendinger, T.; Schoenberg, S.O.; Henzler, T. Unenhanced third-generation dual-source chest CT using a tin filter for spectral shaping at 100 kvp. *Eur. J. Radiol.* **2015**, *84*, 1608–1613. [CrossRef] [PubMed]

17. Gordic, S.; Morsbach, F.; Schmidt, B.; Allmendinger, T.; Flohr, T.; Husarik, D.; Baumueller, S.; Raupach, R.; Stolzmann, P.; Leschka, S.; et al. Ultralow-dose chest computed tomography for pulmonary nodule detection: First performance evaluation of single energy scanning with spectral shaping. *Investig. Radiol.* **2014**, *49*, 465–473. [CrossRef] [PubMed]

18. Gevenois, P.A.; de Maertelaer, V.; De Vuyst, P.; Zanen, J.; Yernault, J.C. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. *Am. J. Respir. Crit. Care Med.* **1995**, *152*, 653–657. [CrossRef] [PubMed]

19. Gevenois, P.A.; De Vuyst, P.; de Maertelaer, V.; Zanen, J.; Jacobovitz, D.; Cosio, M.G.; Yernault, J.C. Comparison of computed density and microscopic morphometry in pulmonary emphysema. *Am. J. Respir. Crit. Care Med.* **1996**, *154*, 187–192. [CrossRef] [PubMed]

20. Madani, A.; Zanen, J.; de Maertelaer, V.; Gevenois, P.A. Pulmonary emphysema: Objective quantification at multi-detector row CT—Comparison with macroscopic and microscopic morphometry. *Radiology* **2006**, *238*, 1036–1043. [CrossRef]

21. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the IJCAI, Montreal, QC, Canada, 20–25 August 1995; pp. 1137–1145.

22. Wong, T.-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.* **2015**, *48*, 2839–2846. [CrossRef]

23. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [CrossRef]

24. Shao, L.; Fan, X.; Cheng, N.; Wu, L.; Cheng, Y. Determination of minimum training sample size for microarray-based cancer outcome prediction–an empirical assessment. *PLoS ONE* **2013**, *8*, e68579. [CrossRef]

25. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185.

26. Timmins, S.C.; Diba, C.; Farrow, C.E.; Schoeffel, R.E.; Berend, N.; Salome, C.M.; King, G.G. The relationship between airflow obstruction, emphysema extent, and small airways function in COPD. *Chest* **2012**, *142*, 312–319. [CrossRef]

27. Schroeder, J.D.; McKenzie, A.S.; Zach, J.A.; Wilson, C.G.; Curran-Everett, D.; Stinson, D.S.; Newell, J.D., Jr.; Lynch, D.A. Relationships between airflow obstruction and quantitative CT measurements of emphysema, air trapping, and airways in subjects with and without chronic obstructive pulmonary disease. *Am. J. Roentgenol.* **2013**, *201*, W460–W470. [CrossRef] [PubMed]

28. Lynch, D.A.; Austin, J.H.; Hogg, J.C.; Grenier, P.A.; Kauczor, H.U.; Bankier, A.A.; Barr, R.G.; Colby, T.V.; Galvin, J.R.; Gevenois, P.A.; et al. CT-definable subtypes of chronic obstructive pulmonary disease: A statement of the fleischner society. *Radiology* **2015**, *277*, 192–205. [CrossRef]

29. Gawlitza, J.; Haubenreisser, H.; Henzler, T.; Akin, I.; Schönberg, S.; Borggrefe, M.; Trinkmann, F. Finding the right spot: Where to measure airway parameters in patients with COPD. *Eur. J. Radiol.* **2018**, *104*, 87–93. [CrossRef]

30. Matsuoka, S.; Kurihara, Y.; Yagihashi, K.; Hoshino, M.; Nakajima, Y. Airway dimensions at inspiratory and expiratory multisection CT in chronic obstructive pulmonary disease: Correlation with airflow limitation. *Radiology* **2008**, *248*, 1042–1049. [CrossRef]

31. Camiciottoli, G.; Bartolucci, M.; Maluccio, N.M.; Moroni, C.; Mascalchi, M.; Giuntini, C.; Pistolesi, M. Spirometrically gated high-resolution CT findings in COPD: Lung attenuation vs. lung function and dyspnea severity. *Chest* **2006**, *129*, 558–564. [CrossRef]

32. Gu, S.; Leader, J.; Zheng, B.; Chen, Q.; Sciurba, F.; Kminski, N.; Gur, D.; Pu, J. Direct assessment of lung function in COPD using CT densitometric measures. *Physiol. Meas.* **2014**, *35*, 833–845. [CrossRef] [PubMed]

33. Chan, H.-P.; Sahiner, B.; Hadjiiski, L. *Sample Size and Validation Issues on the Development of CAD Systems*; International Congress Series; Elsevier: Amsterdam, The Netherlands, 2004; pp. 872–877.

34. Fukunaga, K.; Hayes, R.R. Effects of sample size in classifier design. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 873–885. [CrossRef]

35. Baldi, S.; Miniati, M.; Bellina, C.R.; Battolla, L.; Catapano, G.; Begliomini, E.; Giustini, D.; Giuntini, C. Relationship between extent of pulmonary emphysema by high-resolution computed tomography and lung elastic recoil in patients with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **2001**, *164*, 585–589. [CrossRef] [PubMed]