SHAP-prioritised Machine Learning for Diagnostic-grade Prediction of Lung Function

Oliver Pitts[1][0009−0000−4627−8620], Salman Siddiqui[2][0000−0003−3770−7870], Anand Shah[3], Alex Bell[4], Christopher Brightling[5], David Singh[6], Janwillem Kocks[7][0000−0002−2760−0693], Leonardo Fabbri[8], Alberto Papi[8], Klaus Rabe[9][0000−0002−7020−1401], Maarten Van Den Berge[10][0000−0002−9336−7340], Monica Kraft[11][0009−0000−9941−3558], and Rossella Arcucci[1,12][0000−0002−9471−0585]

[1] Data Science Institute, Imperial College London, UK {o.pitts23, r.arcucci}@imperial.ac.uk
[2] National Heart and Lung Institute, Imperial College London, UK s.siddiqui@imperial.ac.uk
[3] School of Public Health & Royal Brompton Hospital, Imperial College London, UK a.shah@imperial.ac.uk
[4] Department of Radiology, University of Michigan, USA {alejbell, cgalban}@med.umich.edu
[5] Institute for Lung Health, National Institute for Health and Care Research Leicester Biomedical Research Centre, University of Leicester, Leicester, UK ceb17@leicester.ac.uk
[6] NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, and University of Manchester, Manchester, UK dsingh@meu.org.uk
[7] General Practitioners Research Institute, Groningen, Netherlands janwillem.kocks@gpri.nl
[8] Department of Translational Medicine, University of Ferrara, Ferrara, Italy leonardo.fabbri20@gmail.com, alberto.papi@unife.it
[9] LungenClinic Grosshansdorf, Grosshansdorf, Germany; Airway Research Center North, German Center for Lung Research, Grosshansdorf, Germany; Christian-Albrechts University of Kiel, Kiel, Germany k.f.rabe@lungenclinic.de
[10] Department of Pulmonary Diseases, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands m.van.den.berge@umcg.nl
[11] Department of Medicine, Icahn School of Medicine at Mount Sinai, Mount Sinai Health System, New York, USA monica.kraft@mssm.edu
[12] Department of Earth Science and Engineering, Imperial College London, UK

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

Abstract. Automated machine learning (ML) can streamline the characterisation and management of chronic airway conditions. With the advent of quantitative CT (qCT) imaging allowing precise extraction of structural features from scans, assessment of airway obstruction levels could be automated to compliment traditional testing. This "feature known" approach has the added potential benefit of identifying key structure-function relationships through explainability measures. We therefore aimed to develop inverse models to estimate spirometry parameters from high-dimensional quantitative data using these structural metrics as constraints. With the ATLANTIS (NCT02123667) dataset, this paper experiments with a selection of ML methods, specifically k-nearest neighbours (kNN), random forest (RF) and support vector machine (SVM), to predict spirometry values (Forced Expiratory Volume (FEV1), Forced Vital Capacity (FVC) and FEV1/FVC). The dynamic ratio FEV1/FVC was predicted better by all models than FEV1 or FVC. Results show effective counteraction to high-dimensionality through iterative feature refinement guided by SHapley Additive exPlanations (SHAP), and to limited training data through dynamic Gaussian noise (DGN). Diagnostic-grade prediction accuracy was achieved with DGN SHAP sequential feature selection (SFS)-kNN at 1.64% MRE with 37/76 features. A selection of typical variables including expiratory tissue density and lung volume, vasculature and airway geometries were seen to be important for prediction. This approach therefore can not only predict pulmonary function, but also extract useful structural information in a dynamic airway system through linking back to personalised abnormalities.

Keywords: Machine Learning · SHAP · Quantitative CT · Airway Diseases

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

1 Background & Motivation

Asthma is a common chronic airway disorder with more than 300 million people estimated to have been diagnosed worldwide [1]. In the UK, it affects approximately 12% of the population, resulting in an annual public healthcare burden of £1.9 billion, in part through the ∼6 million related primary care consultations and ∼100,000 hospital admissions [2]. According to the 'Global Initiative for Asthma' (GINA), it can be defined as a heterogeneous condition characterised by chronic inflammation and variable airway obstruction [1]. Patients are burdened with symptoms such as wheezing, shortness of breath, chest tightness and a cough, with these arising in episodes over time known as exacerbations [1, 3]. This results in impaired lung function and accelerated deterioration thereof.

Several challenges persist around asthma and the personalised management of it, including accurate characterisation of underlying structural and functional abnormalities. Currently this is done through a series of clinical tests, questionnaires and standard computed tomography (CT) scans [4, 5], which can holistically take several hours, and although efficacious for broad diagnostics, often lacks specificity. As a result, there is a need for more insightful assessment of lung function, particularly for determining disease severity and tracking pathogenesis.

Currently, the clinically predominant lung function test is spirometry, which measures the volume of air forcibly expelled in one second - 'forced expiratory volume' (FEV1), and the maximum air a patient can expel from the lungs following a maximal inhale - 'forced vital capacity' (FVC) [6]. The dynamic ratio of these values (FEV1/FVC) is a key diagnostic indicator of airway obstruction [7]. A low ratio of <0.7 tends to indicate some level of ventilation hindrance through narrowing of the trachea and bronchi, or small airway dysfunction [6]. This test comes with several limitations including variable spirometer accuracy and human error, resulting in an acceptable margin of $\pm 2.5\%$ [8]. It is also an effort-dependent test that requires patient cooperation, making it less reliable in young children and certain older populations. Moreover, although spirometry indicates the presence of airflow limitation, it does not localise to, or provide information on the specific malignancies [6].

Quantitative computed tomography (qCT) has emerged as a powerful tool for evaluating lung structure, offering a more detailed evaluation of airway morphology, lung parenchyma and vascular features. Through qCT we can capture three-dimensional airway changes, detect signs of small airway disease and identify precise heterogeneities contributing to airflow limitation [9]. CT scans from both cycles provide complementary information, with inspiratory scans highlighting lung inflation and tissue density, while expiratory scans show key residual features like air trapping [10]. Despite the routine imaging of asthma patients, qCT remains underutilised in asthma care due to the inherent complexity of its manual analysis, leaving a need for advanced computational methods to derive

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

meaningful insights.

Machine learning (ML) offers a promising avenue for extracting such clinically useful information from qCT data to streamline asthma classification and management. However, feature panels tend to be extensive, which added to the small patient cohorts often typical of respiratory datasets, makes training a generalisable model to effectively capture complex relationships challenging. A way to circumvent this is dimensionality reduction to an optimal feature subset, whilst synthetic data generation can expand and diversify the training data [11, 12]. Existing artificial intelligence (AI) applications in respiratory imaging, such as convolutional neural networks (CNNs) to detect COVID-19 or cancer nodules, have tended to draw on inspiratory scans alone, which does not fully capture patient-specific breathing dynamics [13, 14]. Further, current best attempts to predict pulmonary function in specific airway diseases have also drawn on CNNs, with none yet reaching diagnostic-grade accuracy [15, 16]. These deep learning techniques are of course black-box, whereas a 'feature known' approach allows inverse modelling of structure-function relationships, offering inherent interpretability advantages, including over traditional spirometry.

This study therefore looked to bridge gaps in asthma healthcare and AI, by applying ML techniques to qCT-derived features to provide interpretable, diagnostic-grade predictions of spirometry outcomes. We aimed to overcome the challenges presented by a small, high-dimensional dataset through guided iterative feature elimination and dynamic introduction of synthetic noise.

The rest of the paper is structured as follows: Section 2 describes the data we used as well as the pre-analysis and normalisation implemented. Following initial data handling and exploration, Section 3 introduces the ML models tested as well as the experimented adaptations thereof. Results are presented in Section 4 and discussed in Section 5, with concluding remarks and perspectives on future works presented in Section 6.

2 Data and Preprocessing

This work was carried out using the (Assessment of Small Airways Involvement in Asthma) ATLANTIS ((NCT02123667) dataset, a multinational prospective cohort study which was collated to assess the link between small airways disease and severity of asthma [17]. Of the 363 patients for which there was complete and quality controlled imaging data; 310 were from asthma patients (collected at least 6 months from their diagnosis) and 53 from healthy controls. This dataset also held extensive electronic health records including a full complement of demographic and clinical data.

Using Mi-TAP, a deep learning-based qCT algorithm developed by the Galban lab (`https://websites.umich.edu/~cgalbanlab/index.html`), a panel of 67

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

variables was extracted including values and ratios relating to physiological presentations like lung volume, tissue density, vasculature and emphysema, from inspiratory and expiratory CT scans. To avoid missing values or their imputation in the resulting dataframe distorting the distribution of the target variable and introducing significant bias, rows with incomplete dependent data were dropped (9 subjects). Of the sparse missing value instances in the predictor variables, mean imputation was used. The feature set was then normalised using the skLearn standard scaling module to ensure comparability across variables. Finally, prior to running models the dataset was split into training (70%), validation (15%), and test (15%) sets to facilitate development and evaluation.

3 Methods

Following initial data exploration to examine the relationship between each variable and the pulmonary function metrics of interest: FEV1, FVC and FEV1/FVC; a collection of ML models were tested on the whole qCT dataset:

1. **Mean Prediction** - Used as a simple benchmark for model performance as per
$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{1}$$
where N is the total number of observations and y_i is the actual target value.

2. **Linear Regression** - A more sophisticated benchmark for model performance as per
$$Y_i = f(X_i, \beta) + \varepsilon_i \tag{2}$$
where Y_i is the dependent variable, X_i is the independent variable vector, β is the vector of weight coefficients and ϵ_i is an error term.

3. **kNN** - Predictions were derived using a weighted voting mechanism, where the contribution of each of the k nearest neighbours was inversely proportional to their Euclidean distance from the query point. Bayesian optimisation was applied to tune k and the distance metric, ensuring optimal local interpolation.

4. **RF** - A random forest ensemble was constructed using multiple decision trees trained on different subsets of the data. Each tree contributed to the final prediction via averaging, reducing variance and improving generalisation. Key hyperparameters such as the number of trees, maximum depth, and feature selection strategy were optimised using Bayesian methods.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

5. **SVM** - A support vector approach was employed looking at linear, radial basis function (RBF), and polynomial kernels to map input features into a higher-dimensional space. The model sought to optimise the epsilon-insensitive loss function, ensuring robust predictions by balancing margin width and prediction error. Bayesian optimisation was applied to tune the kernel type (linear, RBF, polynomial), the regularisation parameter C, and kernel-specific parameters such as the RBF kernel coefficient γ and polynomial degree d, all evaluated using mean squared error (MSE) as the objective function. For example, the RBF kernel was defined as:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{3}$$

and the polynomial kernel as:

$$K(x_i, x_j) = (x_i^\top x_j + c)^d \tag{4}$$

where γ and d were included in the optimisation process alongside C.

Mean relative error (MRE) was used as the main accuracy metric and models incorporated learning curves to look at performance on the training vs validation sets. Loss function convergence was examined during optimisation. After evaluation of initial model performances, the dynamic ratio FEV1/FVC was carried forward as the primary output variable. Feature selection was explored through the iterative selection techniques - recursive feature elimination (RFE) and sequential feature selection (SFS), applied post-optimisation. These were used to reduce dimensionality while retaining the most informative features.

Interpretability was enhanced through use of SHapley Additive exPlanations (SHAP), enabling detailed, model agnostic insights into the contributions of individual qCT variables. These were subsequently ranked in descending order and used to adapt the standard SFS and RFE approaches. Instead of relying solely on internal model coefficients or traditional selection criteria, features were explicitly prioritised for addition or removal based on their SHAP-derived importance scores. This hierarchical sorting post-optimisation ensured retention of globally influential features while systematically eliminating variables with minimal predictive contributions. To explore meaningful associations with biologically relevant demographic and clinical parameters, variables such as height, weight, age, BMI, blood pressure, and gender were then incorporated. Challenges posed by the risk of overfitting to limited training data were mitigated through dynamic Gaussian noise augmentation for regularisation. This approach systematically produced multiple noise-enhanced copies of the original data. Specifically, the training dataset was iteratively augmented by adding Gaussian noise with zero mean and incrementally increasing variance:

$$X_{\text{new},step} = X + \epsilon_{step}, \quad \epsilon_{step} \sim \mathcal{N}(0, \sigma^2_{step}) \tag{5}$$

where the standard deviation for each augmentation step was defined as:

$$\sigma_{step} = \text{initial\_noise} + \frac{\text{step} \times (\text{max\_noise} - \text{initial\_noise})}{\text{steps} - 1} \tag{6}$$

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

4 Results

4.1 Model performance for FEV1/FVC vs FEV1 or FVC

All models demonstrated superior performance when predicting FEV1/FVC compared to FEV1 or FVC, with the best from the latter metric being kNN at 15.46% MRE and 15.73% for FEV1. Prediction of pre-bronchodilator FEV1/FVC was likely better as this dynamic ratio incorporates information from both inspiratory and expiratory phases of breathing, thus models could better capture relationships between features from both scan types and the output variable. For FEV1/FVC, initial application of the models to the full qCT dataset showed some efficacy with all models surpassing mean prediction - 10.91% MRE but only kNN beating the linear regression benchmark - 9.71% (Table 1).

Table 1. Overview table showing the MRE % results for FEV1/FVC prediction by each model, with the data types used (qCT), Demo=Demographic and Clin=Clinical) and method setup.

	Model (MRE - %)		
Data & Setup	**kNN**	**Random Forest**	**SVM**
qCT	9.30	10.73	10.71
qCT with RFE/SFS	8.87	9.87	9.08
qCT with SHAP RFE/SFS	8.74	8.88	7.89
qCT/Demo/Clin with SHAP RFE/SFS	6.95	9.91	10.00
qCT/Demo/Clin with DGN SHAP RFE/SFS	1.64	4.66	5.16

4.2 Efficacy of iterative feature refinement

The iterative feature selection methods, RFE and SFS, consistently enhanced predictive accuracy for all models by systematically refining the feature space (Table 1). Similarly, integrating SHAP-based feature prioritisation further enhanced predictive accuracy compared to RFE/SFS alone (Table 1). Among the evaluated methods, SFS-SVM and RFE-RF achieved the greatest accuracy improvement when combined with SHAP-prioritisation, with the latter showing error reduction from 9.87% to 8.88% MRE, while SFS-SVM improved from 9.08% to 7.89% (Table 1). These results strongly suggest that ranking globally important features enhances traditional RFE/SFS methods by potentially better capturing non-linear relationships.

4.3 Integration of demographic and clinical data

The inclusion of demographic and clinical variables yielded mixed results. Both the SFS-SVM and SFS-RF models exhibited slightly reduced predictive accuracy. This is possibly due to overfitting caused by the increased feature space,

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

with models potentially capturing the additional noise rather than meaningful patterns. Conversely, the datapoint-greedy SHAP SFS-kNN model yielded a lower MRE of 6.95% with 37 of 76 selected features (Table 1).

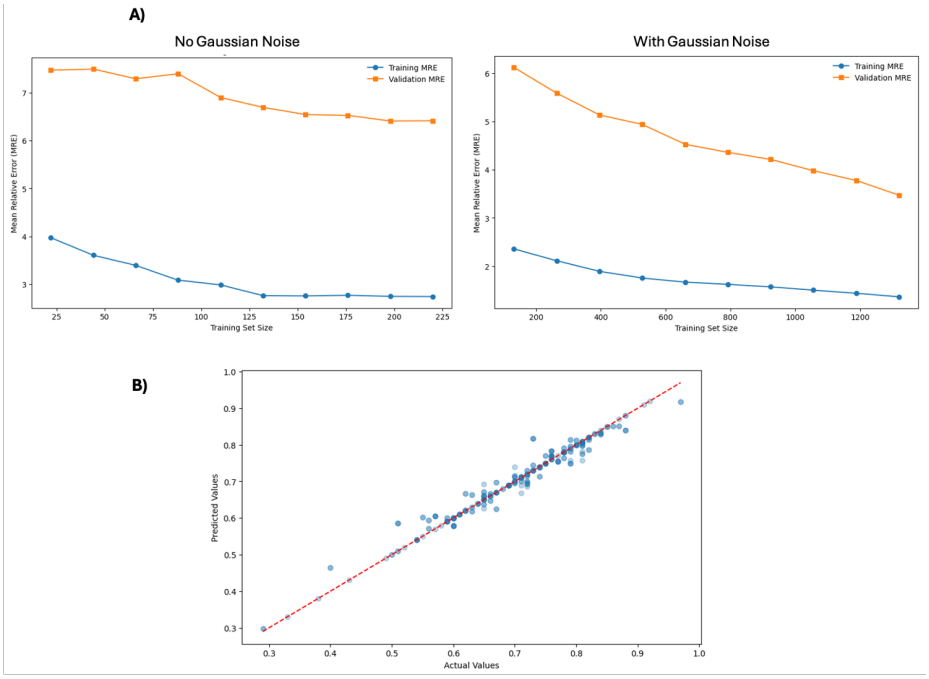

Fig. 1. A) Learning curves for the SHAP RFE-RF model as applied to the same qCT and demographic/clinical dataset (76 features) with and without Gaussian noise, showing reduced overfitting. **B)** Predicted vs actual plot for the diagnostic grade SHAP SFS-kNN showing tight grouping of points to the line of best fit.

4.4 Dynamic Gaussian noise optimisation

High-dimensional feature sets posed a risk of overfitting in some models. To address this, dynamic Gaussian noise augmentation was introduced during training, significantly benefiting SHAP-prioritised models. As shown in Figure 1A, the inclusion of Gaussian noise improved model generalisability by stabilising training and validation errors, reducing sensitivity to small perturbations, and mitigating overfitting. The most pronounced improvement was observed in SHAP SFS-kNN, where MRE decreased to 1.64% with 37/76 selected features (Table 1). This effect is evident in the predicted vs actual plot (Figure 1B), where values,

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

including those near extremes, tightly fit the ground truth. Similarly, SFS-SVM with SHAP prioritisation improved to 5.16% MRE using 44/67 features (qCT alone), while SHAP SFS-RFE demonstrated a substantial error reduction from 9.91% to 4.66% with 54/76 features. All models significantly outperformed both the mean and linear regression baselines ($p<0.01$) by paired t-test, reinforcing the utility of iterative noise regularisation in preventing overfitting.

4.5 Linking back to structure with SHAP

SHAP values provided key insights into the structural determinants of lung function predictions, enhancing the clinical interpretability of the models. The most influential features identified were vascular parameters (vessel volume, number of components, vessel volume 5 down), parenchymal remodelling metrics (PRM2 - normal tissue %), airway structural markers (pi10 - normalised airway wall thickness metric), and expiratory tissue density/lung volume - all of which are potentially closely associated with functional decline (Figure 2). This demonstrates the potential for SHAP-derived feature importance to guide mechanistic investigations from routine disease characterisation and inform personalised treatment strategies.

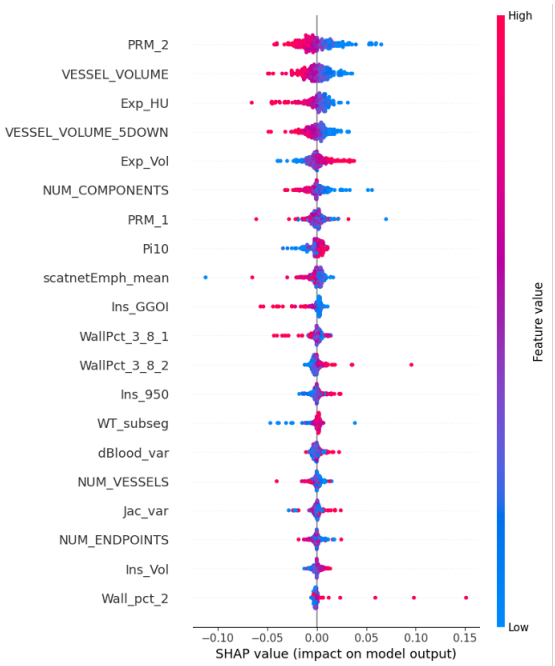

Fig. 2. Plot of SHAP values for each feature from the SHAP SFS-kNN model (with dynamic Gaussian noise) applied to the held-out test set.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

5 Discussion

In this work, we explored the effectiveness of combining explainable feature selection and dynamic noise augmentation (synthetic data generation) to elucidate complex relationships between numerous qCT-derived features and lung function outputs within a limited dataset. Comparative analysis of model performance across different spirometry output labels showed ubiquitous superior prediction of the dynamic ratio FEV1/FVC. This is perhaps unsurprising given that it intrinsically integrates both inspiratory and expiratory breathing phases, highlighting the benefit of incorporating the static data from both scan types. The only current attempt to use machine learning in a similar manner was by Gawlitza et al., (2019), which applied several models to four manually selected features in chronic obstructive pulmonary disease (COPD) patients, achieving an MRE of ∼14% with a kNN model. Comparatively, our optimised DGN SHAP SFS-kNN method markedly improved prediction accuracy, at just 1.64% MRE [18]. This was also significantly better than current deep learning benchmarks by Park et al., (2023) and the literature best by Yoshida et al., (2024) with CNN based models at 9.7% and 5.21% MRE respectively [15, 16]. Indeed, all three model types (kNN, RF and SVM) achieved greater mean accuracy (Table 1).

Justification for our model choices stems from a need to leverage different model architectures whilst maintaining comparability across setups and improving on deep learning attempts through explicit features. The inclusion of RF, SVM, and kNN enabled a comprehensive evaluation across diverse algorithmic paradigms, each with distinct advantages in handling high-dimensional qCT data. Random forest, being an ensemble of decision trees, effectively integrates categorical and continuous data while inherently ranking feature importance, making it highly adaptable. SVM, a margin-based classifier, benefits significantly from informative feature selection since it constructs decision boundaries in high-dimensional spaces, which explains its marked improvement when coupled with SHAP-prioritisation. Conversely, a distance-based method like kNN treats all features equally and is less sensitive to feature selection, which likely accounts for its marginal improvement in performance when SFS/SHAP SFS were applied. By comparing all three models, we ensured robust and comprehensive experimentation across different methodologies for handling qCT-derived features and clinical variables.

Moreover, the complementary use of RFE and SFS was motivated by their capacity to systematically refine feature sets, mitigate overfitting and improve generalisation. RFE is a backward elimination technique in which models are iteratively retrained, progressively removing the least important features based on model coefficients or feature importance scores until an optimal subset is identified, eliminating non-contributory features. It was thus only applied to the the RF

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

model which inherently ranks features. In contrast, SFS is a model-agnostic approach that incrementally adds features based on their individual contributions to model performance until the optimal subset is achieved ensuring adaptability to varying feature importances. The fact that the iterative feature selection methods significantly improved model performance meant they effectively mitigated the curse of dimensionality that is inherent to qCT data, particularly given the logistical and financial challenges of collating results from large numbers of patients and processing such cohorts.

The integration of SHAP-prioritisation further refined feature selection by capturing global feature importance and complex non-linear interactions beyond the means of traditional RFE/SFS. Notably, the SVM model exhibited the greatest improvement (MRE reduction from 9.08% to 7.89%), likely due to its kernel-based structure, which benefited from well-curated feature sets that improved the separability of decision boundaries. Ground was also gained for the RF model from 9.87% MRE (with just RFE) to 8.88% with SHAP-prioritisation, building on its inherent feature ranking to further mitigate dimensionality. Little improvement was seen for kNN since it lacks an internal weighting mechanism to properly leverage feature importance. Taken holistically however, these results represent a notable benefit of being able to use the whole qCT dataset, a marked advantage on the previous machine learning attempt by Gawlitza et al., (2019) who only used a select four features. Since abnormalities in airways and parenchymal structural variables are likely to be interlinked, this affords better capture of underlying structure-function relationships.

Indeed, from SHAP seen in Figure 2 and other models, key structural features for determining lung function included vasculature and parenchymal features as well as parameters reflective of airway wall thickness and lumen geometry. No individual features were strongly important in FEV1/FVC predictions, perhaps as a result of the large panel of variables, further highlighting the requirement for holistic input to effectively capture relationships with the output variable. Moreover, these were neither precisely localised nor inherently defective, however they do open an avenue of investigation into precise defects associated with impaired lung function. By linking abnormal lung function results (e.g., FEV1/FVC $<$ 0.5) to quantifiable anatomical changes, this approach facilitates the integration of omics data for a deeper investigation into airway disease mechanisms. The identification of key qCT-derived structural markers, opens avenues for personalised disease phenotyping. Such insights could support targeted therapeutic interventions, enabling clinicians to tailor treatments based on patient-specific structural abnormalities, potentially improving early detection and monitoring of airway conditions.

As demonstrated in figure 1A, the introduction of dynamic Gaussian noise improved the convergence of learning curves for the models, leading to a more stable alignment between training and validation errors. Minimal convergence

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

was seen between curves and the test set error was typically far higher, emphasising the necessity of noise augmentation. Typically, the introduction of noise can be used to test generalisation of a model [18]. The novelty here lies in incremental Gaussian noise in training, playing a critical role in stabilising model performance by reducing sensitivity to minor perturbations through controlled variability, thereby mitigating overfitting. In the tree ensemble RF model this likely served to decorrelate trees and help capture more diverse patterns. Notably, despite not being directly depicted in figure 1A, the RF error on the held-out test set was also far closer to the validation error, further underscoring the enhanced generalisability benefits. SFS-kNN also significantly benefited from this synthetically expanded training data, as it relies on Euclidean distance between points. Furthermore, SHAP SFS-SVM demonstrated notable improvement, achieving approximately $\sim 5\%$ MRE, likely identifying more generalisable support vectors, contributing to improved margins and stabilising better decision boundaries. To the best of our knowledge, this was the first study to apply Gaussian noise to mitigate the challenges posed by high-dimensional qCT data. It is also a rare instance of synthetic noise successfully enhancing model training in a biomedical ML study.

The performance here of synthetic noise augmentation underscores a key limitation of this work: its relatively small training set. While Gaussian noise improved model robustness by mitigating overfitting, it cannot fully substitute for a larger, more diverse training cohort. Expanding the dataset would likely further reduce MRE, potentially approaching or surpassing the $\pm 2.5\%$ threshold across all model types. However, challenges in generalisability remain, particularly across diverse populations with varying demographic and clinical characteristics. Moreover, the current models were trained on a single cohort and results have not yet had external validation. Variability in qCT image quality due to factors such as scanner differences or acquisition protocols may also have impacted feature consistency and model reliability. Additional model refinements, such as the integration of custom loss functions, could help adapt to these complexities. The predictive improvement of the kNN model upon inclusion of demographic/clinical data, as well as the RF upon Gaussian noise augmentation for the same extended data, further points to leveraging an ensembling approach to better utilise the disparate information held in continuous and categorical feature types.

6 Future work and Conclusion

Ultimately the goal of this work is to integrate automated lung function prediction into clinical diagnostic practise given that CT imaging is routinely used and qCT software readily available. To this end, although our SHAP-prioritised SFS-kNN model demonstrated diagnostic-grade accuracy as defined as within $\pm 2.5\%$ by Graham et al., (2019), external validation with a cohort of scans processed through the same qCT platform (and thus the same panel of features)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

is needed [8]. Future work will also explore incorporating a time aware element to longitudinally map lung function trajectories from initial CT imaging. Accurate tracking of pulmonary decline, particularly in abnormal cases, would be to a major clinical benefit in streamlining the management of asthma and airway diseases holistically. Successful application of this setup to dynamic ratios of other pulmonary tests more relevant to other conditions, such as residual volume/total lung capacity (RV/TLC) from body plethysmography, would allow extension to multiple labels. Finally, in a held-out abnormal test set, the inverse modelling to precise structural heterogeneities with SHAP could be an interesting approach to investigating overlapping features in airflow obstruction.

To conclude, this study demonstrates the viability of automated machine learning techniques to characterise the functionality of patient breathing dynamics from static qCT data, with the best performing SHAP SFS-kNN model showing diagnostic grade accuracy (1.64% MRE). Dynamic Gaussian noise and iterative feature selection were shown to be effective in capturing underlying complexities in a small, high-dimensional dataset. Inverse modelling, guided by SHAP explainability measures effectively refines feature selection and offers a unique approach to understanding structure-function relationships in chronic airway conditions. These findings are preliminary proof-of-concept for a novel method in future precision respiratory diagnostics and disease management.

Acknowledgments. This research has received financial support from the UKRI AI for Health grant EP/S023283/1.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aegerter, H., Lambrecht, B.N.: The pathology of asthma: what is obstructing our view? *Annu. Rev. Pathol.* **18**(1), 387–409 (2023)
2. Jacobs, A., Wu, R., Tomini, F., De Simoni, A., Mihaylova, B.: Strong and graded associations between level of asthma severity and all-cause hospital care use and costs in the UK. *BMJ Open Respir. Res.* **10** (2023)
3. Habib, N., Pasha, M.A., Tang, D.D.: Current understanding of asthma pathogenesis and biomarkers. *Cells* **11**, 2764 (2022)
4. Lommatzsch, M., Criee, C.P., de Jong, C.C.M., et al.: Diagnosis and treatment of asthma: a guideline for respiratory specialists 2023, published by the German Respiratory Society (DGP) e.V. *Pneumologie* (2023)
5. Louis, R., Satia, I., Ojanguren, I., et al.: European Respiratory Society guidelines for the diagnosis of asthma in adults. *Eur. Respir. J.* **60**(3) (2022)
6. Maldonado-Franco, A., Giraldo-Cadavid, L.F., Tuta-Quintero, E., Bastidas Goyes, A.R., Botero-Rosas, D.A.: The challenges of spirometric diagnosis of COPD. *Can. Respir. J.* **2023**(1), 6991493 (2023)
7. Toren, K., Schioler, L., Lindberg, A., et al.: The ratio FEV1/FVC and its association to respiratory symptoms, a Swedish general population study. *Clin. Physiol. Funct. Imaging* **41**(2), 181–191 (2021)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7

8. Graham, B.L., Steenbruggen, I., Miller, M.R., et al.: Standardization of spirometry 2019 update. An official American Thoracic Society and European Respiratory Society technical statement. *Am. J. Respir. Crit. Care Med.* **200**(8), e70–e88 (2019)
9. Foy, B.H., Soares, M., Bordas, R., et al.: Lung computational models and the role of the small airways in asthma. *Am. J. Respir. Crit. Care Med.* **200**(8), 982–991 (2019)
10. Ram, S., Hoff, B.A., Bell, A.J., et al.: Improved detection of air trapping on expiratory computed tomography using deep learning. *PLoS One* **16**(3), e0248902 (2021)
11. Jia, W., Sun, M., Lian, J., Hou, S.: Feature dimensionality reduction: a review. *Complex Intell. Syst.* **8**(3), 2663–2693 (2022)
12. Lu, Y., Shen, M., Wang, H., et al.: Machine learning for synthetic data generation: a review. *arXiv preprint* arXiv:2302.04062 (2023)
13. Aslan, M.F., Sabanci, K., Durdu, A., Unlersen, M.F.: COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian optimization. *Comput. Biol. Med.* **142**, 105244 (2022)
14. Zhao, D., Liu, Y., Yin, H., Wang, Z.: An attentive and adaptive 3D CNN for automatic pulmonary nodule detection in CT image. *Expert Syst. Appl.* **211**, 118672 (2023)
15. Park, H., Yun, J., Lee, S.M., et al.: Deep learningâbased approach to predict pulmonary function at chest CT. *Radiology* **307**(2), e221488 (2023)
16. Yoshida, A., Kai, C., Futamura, H., et al.: Spirometry test values can be estimated from a single chest radiograph. *Front. Med.* **11**, 1335958 (2024)
17. Postma, D.S., Brightling, C., Baldi, S., et al.: Exploring the relevance and extent of small airways dysfunction in asthma (ATLANTIS): baseline data from a prospective cohort study. *Lancet Respir. Med.* **7**(5), 402–416 (2019)
18. Gawlitza, J., Sturm, T., Spohrer, K., Henzler, T., Akin, I., SchÃ¶nberg, S., Borggrefe, M., Haubenreisser, H., Trinkmann, F.: Predicting pulmonary function testing from quantified computed tomography using machine learning algorithms in patients with COPD. *Diagnostics (Basel)* **9**(1), 33 (2019). `https://doi.org/10.3390/diagnostics9010033`
19. Chen, X., Wang, Q.: A novel dimensionality reduction approach by integrating dynamics theory and machine learning. *Math. Comput. Simul.* (in press)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_7