Article

A Novel Machine Learning Prediction Model for Aerosol Transport in Upper 17-Generations of the Human Respiratory Tract

Mohammad S. Islam [1,*], Shahid Husain [2], Jawed Mustafa [3] and Yuantong Gu [4]

[1] School of Mechanical and Mechatronic Engineering, Faculty of Engineering and Information Technology, University of Technology Sydney (UTS), Ultimo, NSW 2007, Australia

[2] Department of Mechanical Engineering, Zakir Husain College of Engineering & Technology, Aligarh Muslim University, Aligarh 202002, India

[3] Mechanical Engineering Department, College of Engineering, Najran University, Najran P.O Box 1988, Saudi Arabia

[4] Faculty of Engineering, School of Mechanical, Medical and Process Engineering, Queensland University of Technology, Brisbane, QLD 4000, Australia

[*] Correspondence: mohammadsaidul.islam@uts.edu.au

Abstract: The main challenge of the health risk assessment of the aerosol transport and deposition to the lower airways is the high computational cost. A standard large-scale airway model needs a week to a month of computational time in a high-performance computing system. Therefore, developing an innovative tool that accurately predicts transport behaviour and reduces computational time is essential. This study aims to develop a novel and innovative machine learning (ML) model to predict particle deposition to the lower airways. The first-ever study uses ML techniques to explore the pulmonary aerosol TD in a digital 17-generation airway model. The ML model uses the computational data for a 17-generation airway model and four standard ML regression models are used to save the computational cost. Random forest (RF), *k*-nearest neighbour (*k*-NN), multi-layer perceptron (MLP) and Gaussian process regression (GPR) techniques are used to develop the ML models. The MLP regression model displays more accurate estimates than other ML models. Finally, a prediction model is developed, and the results are significantly closer to the measured values. The prediction model predicts the deposition efficiency (DE) for different particle sizes and flow rates. A comprehensive lobe-specific DE is also predicted for various flow rates. This first-ever aerosol transport prediction model can accurately predict the DE in different regions of the airways in a couple of minutes. This innovative approach and accurate prediction will improve the literature and knowledge of the field.

Keywords: machine learning regression; tracheobronchial airways; deposition prediction; drug delivery; inhalation toxicology; aerosol therapy

1. Introduction

Particulate matter emission into the atmospheric air is a global health challenge, and emissions are increasing every day [1]. The increasing rate of respiratory health patients is evidence of the health hazards of atmospheric pollution [1]. Precise knowledge of atmospheric aerosol transport and deposition to the airways of the human lung is essential for health assessment and drug delivery purposes [2]. To date, researchers have employed wide-ranging approaches (in silico, in vivo and in vitro) to understand aerosol transport to the airways. The computational fluid dynamics (CFD) approach is the most popular method for airflow and particle transport in airways [3,4] and most of the studies to date used CFD for aerosol transport in airways. Almost all of the published literature considers the upper airways and analyses aerosol transport in airways [5–7]. Cheng et al. [8] performed an in vivo study for ultrafine aerosol transport and deposition in human oral and nasal airways. The study employed 10 adult healthy males cast for the analysis. The in vivo study reports the smaller cross-section area of the extrathoracic airways and

that the complex airway shape influences the overall ultrafine particle deposition pattern. Later, Zhang et al. [6] studied ultrafine and microparticles transport in an idealised mouth–throat and upper airways. The symmetric three-generation (G0–G3) model investigates the aerosol deposition pattern for different flow rates. The study reports a higher deposition concentration of the aerosol at the carinal angle region of the airways than in the bifurcating airways. In 2008, Zhang et al. [9] considered a 16-generation airway model and analysed the nanoparticle's transport and deposition. The study did not consider the asymmetric whole lung model for the 16 generations, and several 3-generation models were used to analyse the particle deposition.

To date, the best available and large-scale asymmetric airway model with maximum bifurcating branches has been developed by Schmidt et al. [10]. The authors developed a large-scale and highly complex airway model for the first 17 generations of human airways. Gemci et al. [5] used the model of Schmidt et al. [10] and analysed the airflow for the large-scale model for the first time. The digital reference model had 1453 bronchi up to the 17th Horsfield order. This study analysed the airflow for the first time for a large-scale model and improved the knowledge of the field. Later, in 2017, Islam et al. [7] used the same 17-generation airway model and comprehensively analysed the microparticle transport and deposition in a 17-generation model. This first-ever study considered different physical activity conditions (resting, light activity and heavy physical activity) and analysed the deposition pattern for the first time. However, the CFD simulation of a 3-dimensional anatomically realistic 17-generation lung model based on high-resolution computer tomography (HRCT) data requires huge time and computational resources. The 17-generation model used by Islam et al. [7] took approximately 55 days for a single simulation in a high-performance computing system. This study did not consider all branches of the 17-generation model and would significantly increase the computational cost. Islam et al. [11] also studied ultrafine particle transport and deposition to the lower airways of the 17-generation model, and the first-ever study comprehensively analysed the lobe-specific deposition pattern. The study calculated the deposition hot spot for the ultrafine particles, and the computational cost for ultrafine particle transport for the 17-generation model was significantly high. The computational cost for the CFD model is extremely high due to the huge number of computational cells, and it is nearly impossible to simulate a whole lung model (23 generations) with all bifurcating branches. Therefore, coalescing machine learning algorithms (MLAs) with CFD simulation has attracted wide attention because it can accelerate the prediction and save time and computational cost [12].

Recently, many researchers have used MLAs in combination with CFD to save time and cost required in the CFD simulation. Ref. [13] employed the random forest (RF) technique to predict the viscosity of nanofluids. Kwon et al. [14] discovered that the RF model agreed with the CFD results in estimating convective heat transfer in a channel. Jamei et al. [15] used different MLAs to estimate the specific heat capacity of nanofluids. Although researchers have used MLAs for different engineering applications using CFD data, no ML model has been developed for aerosol transport and deposition to the lower airways.

Therefore, the objective of this study is to develop an innovative ML prediction model for pharmaceutical aerosol transport to the lower airways of a 17-generation model. The main goal of this study is to develop ML regression models based on the CFD dataset generated in our previous work. Next, the different ML models are compared using statistical parameters, and the best ML model is selected. For this purpose, four different MLAs are adopted to significantly decrease the computational cost and find the impact of each parameter on estimated results. Then using the best MLA model, particle transport and deposition (TD) in the transitional and respiratory zones (deeper airways) of the human lung for different particle diameter and flow rates are predicted and analysed.

2. Problem Definition and Numerical Methods

The present study extends our previous work in which we considered a highly-asymmetric 17-generation airway model (Figure 1) for aerosol transport prediction. The

airway model is constructed from the airway dimensions of Schmidt et al. [10] and the final model consists of 1453 bronchioles. The large-scale 17-generation model is one of the best available airway models in the literature. The airflow and particle transport in a whole lung model is highly computationally expensive as it needs to solve millions of computational cells. The 17-generation model consists of 7 million cells, and the multiphase flow for a single case takes approximately 55 days in high-performance computing systems. To get a precise understanding of the various shape and size-specific parameters of aerosol transport to the lower airways of the large-scale model, it needs to run a large number of simulations, and it needs a significant amount of time. Our previous study analysed particle transport numerically and developed a database for aerosol transport for various flow rates and diameter particles [11]. In the present work, the database is used to develop an ML model which can accurately predict aerosol transport to the lower airways for various aerosol sizes and flow rates.
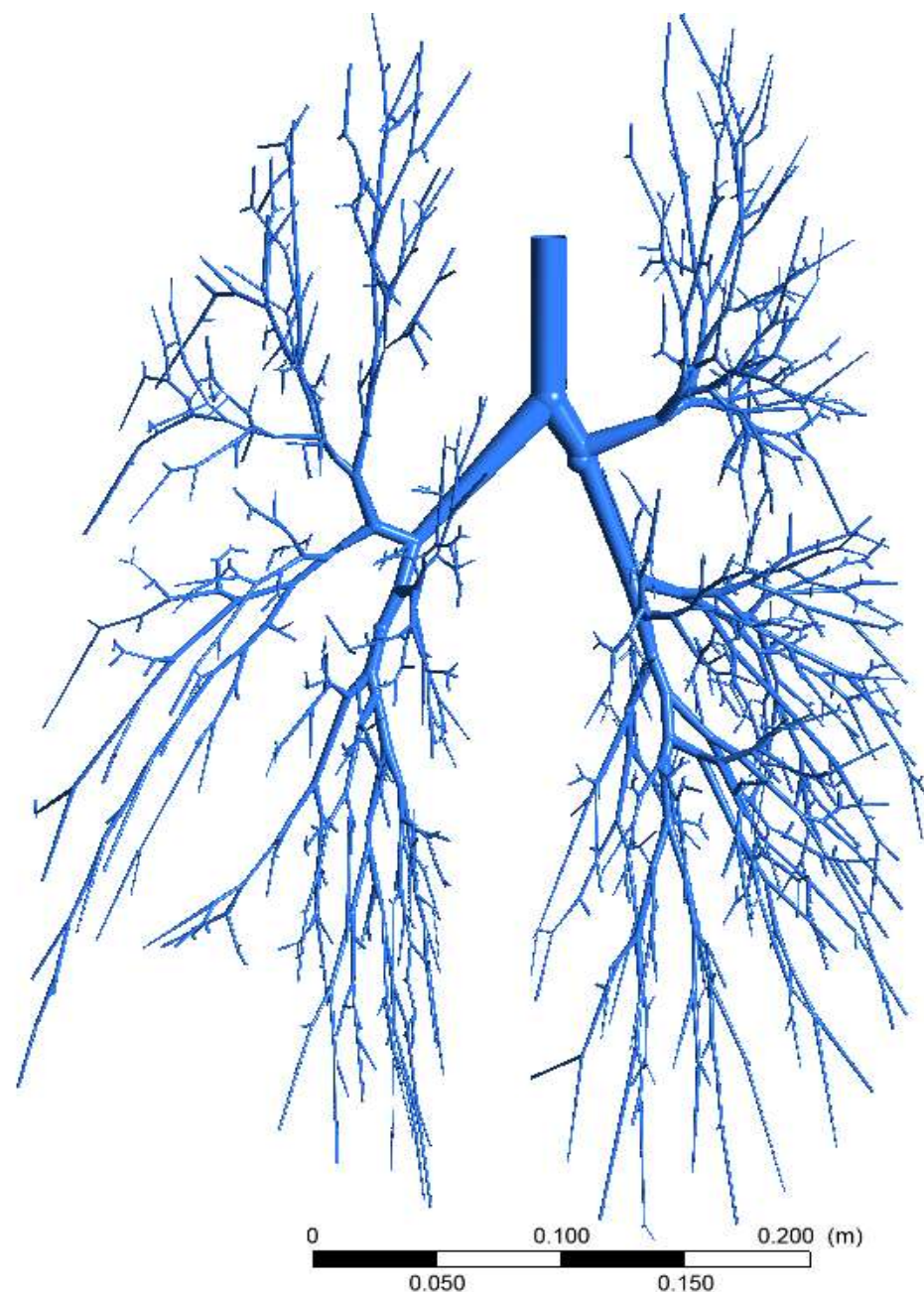

Figure 1. Three-dimensional 17-generation bifurcating airway model of a healthy adult.

The numerical study solved mass and momentum equations for the primary and dispersed phase [7]. A proper grid independence and validation with the benchmark experimental, numerical and theoretical data was conducted by a numerical model. The detailed numerical methods, grid refinement and validation of the numerical data were available in the author's previous studies [7,11]. That study developed the database for the deposition efficiency at the right lung and the left lung of the airway model. Table 1 shows the deposition data at the right lung and the left lung for different flow rates. Table 2 shows the deposition data at the right (Upper, Middle and Lower) and left (Upper and Left) lungs for different flow rates.

Table 1. Numerical data of DE for various size particles at different flow rates.

	10-μm		5-μm		1-μm	
	Left Lung	**Right Lung**	**Left Lung**	**Right Lung**	**Left Lung**	**Right Lung**
15 lpm	25.81081	45.27027	16.48649	27.63514	12.90541	24.18919
30 lpm	32.97297	59.45946	20.60811	35.06757	15.27027	27.97297
60 lpm	32.02703	67.36486	26.89189	47.09459	16.68919	31.75676

Table 2. Numerical DE data for different lung sections for various size particles at different flow rates.

	Right Upper	**Right Middle**	**Right Lower**	**Left Upper**	**Left Lower**
	10-micron				
15 lpm	13.12	11.31	39.16	7.51	28.9
30 lpm	11.11	9.8	43.35	3.8	31.94
60 lpm	6.66	4.07	56.97	1.34	30.93
	5-micron				
15 lpm	16.54	13.78	32.16	10.87	26.65
30 lpm	13.32	14.72	33.12	9.5	29.3
60 lpm	10.96	13.61	39	7.49	28.95
	1-micron				
15 lpm	17.49	14.39	33.15	11.11	23.86
30 lpm	17.19	15	33.34	12.5	22.97
60 lpm	16.83	12.87	35.22	12.78	27.3

3. Methodology

The procedure adopted in the present work is presented in the flowchart, as shown in Figure 2. Firstly, we collected the CFD data from our previous work for the ML step. The goal of using MLAs with CFD simulations was to decrease the computational time and, thus, the computational cost. The four MLAs used in the present analysis are k-nearest neighbours (k-NN) [16], Gaussian process regression (GPR) [17], random forest (RF) [18] and multi-layer perceptron (MLP). The best MLA among all was then used for optimisation. MLAs application requires hyper-parameters tuning to obtain the optimum result. The GridsearchCV tool was used for obtaining the optimised values of the important parameters. It is available in the Scikit-learn library [19].
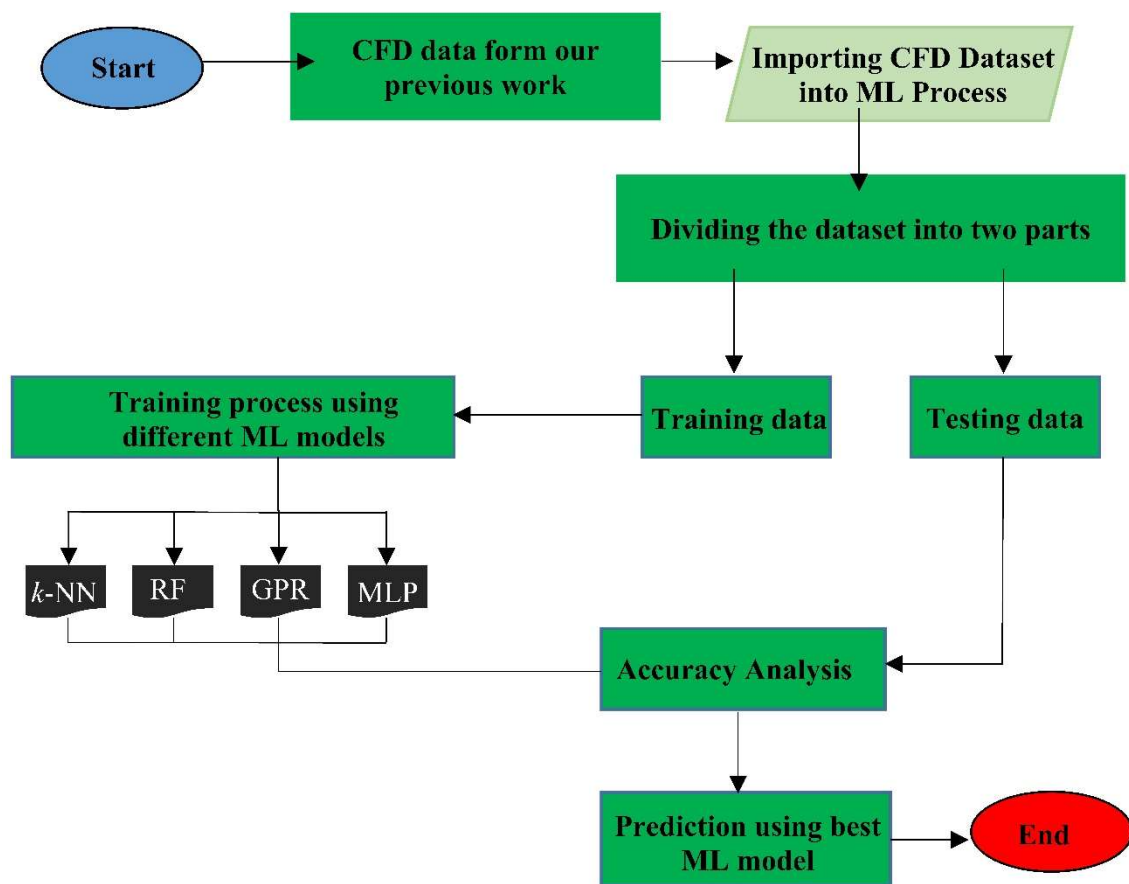
Figure 2. Flowchart showing methodology used in the present work.

Machine Learning Regression Models

The CFD problems involving multi-variables can be easily solved by using MLAs. Further, the number of variables can be easily increased in MLAs (Jalalifar et al., 2020) as they are proficient in handling many input and output quantities. The present study has three independent variables (particle size, flow rate and lung position) and one dependent variable (deposition efficiency). The functional form of the model is presented below:

$$DE = f(particle\ size,\ fow\ rate,\ lung\ position) \tag{1}$$

Two categories of models are developed in the present study. In category-I models, we selected left and right lungs only, while in Category-II, we selected left lower, left upper, right lower, right middle and right upper lungs to achieve a more detailed analysis.

The four ML techniques selected in this work have been extensively used for regression problems [20,21].

k-nearest neighbour (*k*-NN) [22]: *k*-NN MLA is robust, computationally inexpensive and easy to implement. The outcome of a test point is computed based on the interpolated value of the test point's adjacent neighbours. The optimum number of neighbours, **n**, is indicated by the user and it is defined by the **n** training points with minimum Euclidean distance in the input feature space. *k*-NN MLA can estimate an outcome by performing data points weighted average with analogous input features.

Random forest (RF) [23]: RF is an ensemble model that increases estimating ability by combining models. The method it works on is called bagging. In this method several trees are created from input data and then data are reinstated and extracted from all data. The final result is the average of the results of all trees. The RF method is widely used because of its prediction capability.

Gaussian process regression (GPR) [24]: Gaussian processes can represent random variables in ML. It may be an unidentified error in a linear regression model or anything else. The data can be represented similar to a normal distribution curve in a GPR distribution. Gaussian distribution, similar to a normal distribution, has two key parameters, i.e., mean (μ) and variance (σ^2) ("Gaussian Distribution for Machine Learning and Data Science (Normal Distribution) | by Hemanth Nhs | Medium"). The following equation gives the Gaussian distribution for a dataset:

$$f\left(x\middle|\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2}$$

where μ is the average value and σ is the skewness of any dataset.

Multi-layer perceptron (MLP) [25]: MLP can offer improved output by learning the non-linear data pattern through network modelling. MLP has a minimum three layers: an input layer, a hidden layer and an output layer. Each node contains a neuron along with non-linear activation function, excluding the input nodes. MLP uses a Levenberg–Marquardt algorithm (back propagation) for training. MLP comprises of hyperbolic tangent sigmoid transfer function for hidden layers and a linear function at outer layer.

4. Results and Discussion

4.1. Regression Model Results

- Performance evaluation

In the current study, 80% of the data was selected for training, while 20% of the data was used to validate the ML models as recommended by Nguyen et al. [26]. The training and testing data were randomly selected to confirm that the datasets were typical samples of the original dataset.

The MLAs performance was evaluated with the help of some of the most commonly used statistical indicators. The particulars of the statistical parameters are revealed in Table 3 [27].

Table 3. Equations of statistical parameters.

S. No.	Statistical Parameter	Equation
1.	Correlation coefficient (R)	$R = \frac{\sum_{i=1}^{n}\left(M_i - M_{avg}\right)\left(E_i - E_{avg}\right)}{\sqrt{\sum_{i=1}^{n}\left(M_i - M_{avg}\right)^2 \sum_{i=1}^{n}\left(E_i - E_{avg}\right)^2}}$
2.	Mean Bias Error (MBE)	$MBE = \frac{1}{n}\sum_{i=1}^{n}(E_i - M_i)$
3.	Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(E_i - M_i)^2}$
4.	Mean Percentage Error (MPE)	$MPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(E_i - M_i)}{M_i}\right) \times 100$
5.	Mean Absolute Percentage error (MAPE)	$MAPE = \frac{100}{m}\sum_{i=1}^{m}\left|\frac{(E_i - M_i)}{M_i}\right|$
6.	Maximum Absolute Relative Error (erMAX)	$erMAX = max\left(\left|\frac{(E_i - M_i)}{M_i}\right|\right)$
7.	Mean Absolute Error (MAE)	$MAE = \frac{1}{n}\sum_{i=1}^{n}|(E_i - M_i)|$
8.	Uncertainty at 95% (U_{95})	$U_{95} = 1.96\left(SD^2 + RMSE^2\right)^{0.5}$
9.	T-Statistics (t-stats)	$t = \left[\frac{(n-1)MBE^2}{RMSE^2 - MBE^2}\right]^{0.5}$

Table 4 depicts the configuration parameters for all the considered MLAs. The Grid-searchCV tool in Scikit-learn was used to obtain the significant hyperparameters for each MLA. The hyper-parameters of ML models were calculated based on the fit accuracy.

Table 4. Hyper-parameters of the MLAs used in the present work.

No.	MLAs	Hyper-Parameters	Category 1	Category 2
1	*k*-NN	No. of Neighbors:	4	12
2	RF	No. of estimators:	4	4
3	GPR	Kernel type:	DotProduct (sigma = 1) + RBF(length scale = 1),	DotProduct (sigma = 1) + RBF(length scale = 1),
		No. of restarts optimiser	2	2
4	MLP	Sizes of hidden layer:	(100,)	(100,)
		No. of hidden layers:	1	1
		Max iterations:	10000	10000
		No. of hidden layers: learning rate:	0.001	0.001

Figure 3 shows the performance of developed MLAs. There is a good agreement between the estimated and the measured (from CFD simulations) values for most MLAs. The deposition efficiency was estimated extremely well by all the MLAs. The MLP and GPR models performed exceptionally well, while the RF and *k*-NN models performed reasonably.
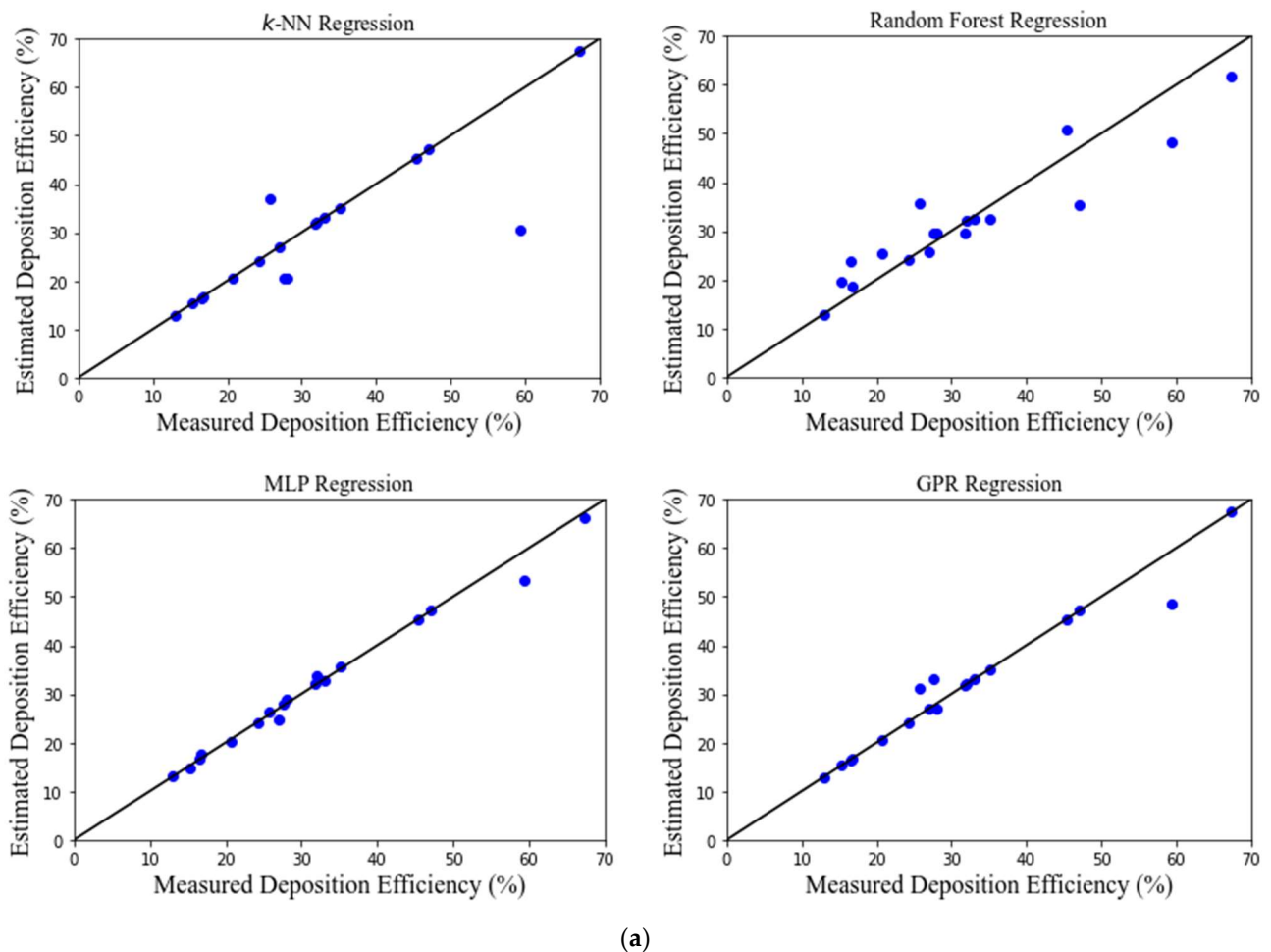

(**a**)

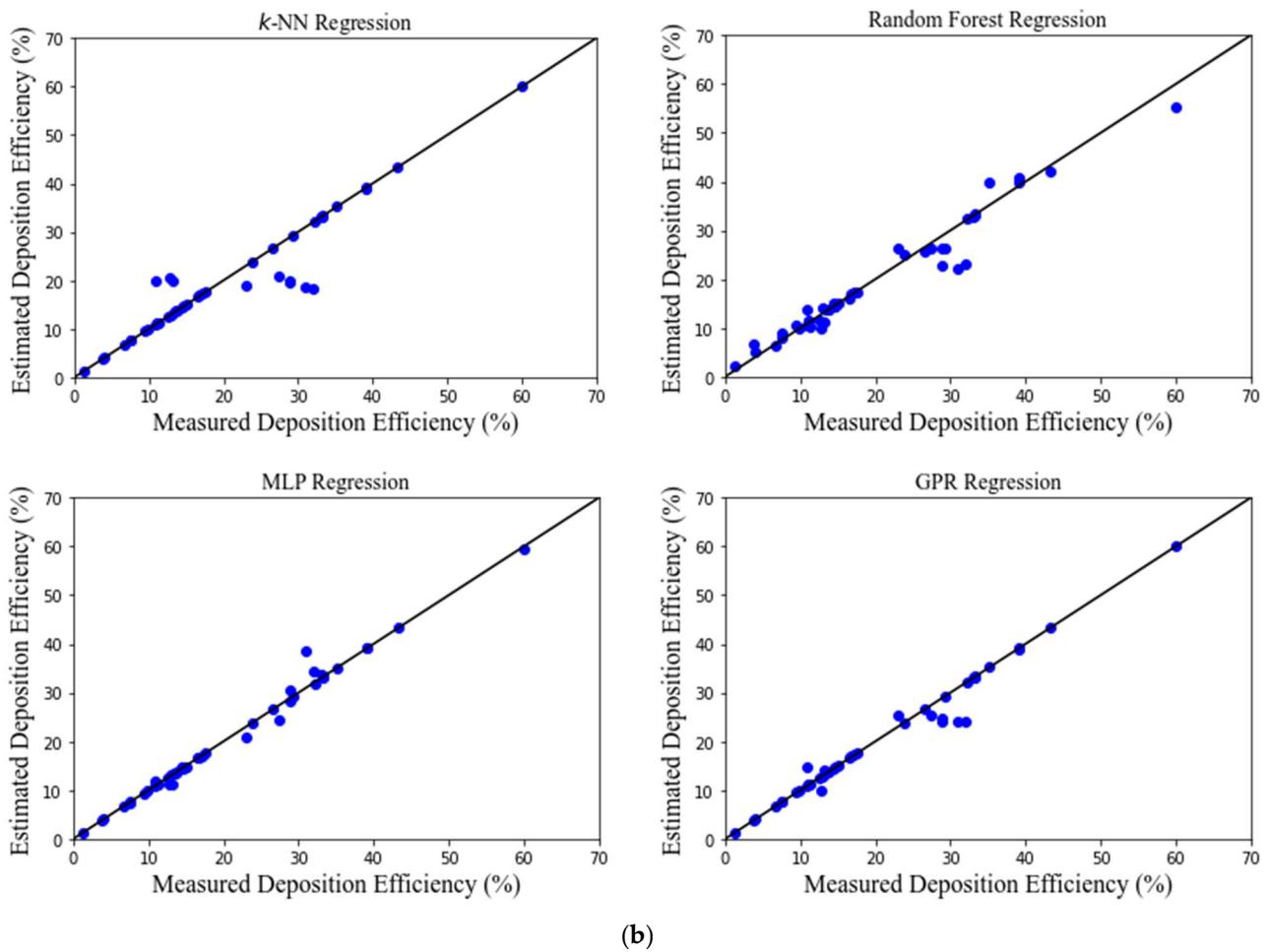**Figure 3.** *Cont.*

Figure 3. (**a**) Scatter diagrams of estimated and predicted deposition efficiency for only left and right lung. (**b**) Scatter diagrams of estimated and predicted deposition efficiency for different sections of left and right lung.

Table 5 shows the values of different statistical pointers, with the most significant value shown in bold. MBE values are in the range from −1.810 to 0.090 for category 1 and 0.066 to −0.711 for category 2. Most models have negative MBE values that suggest underprediction, whereas the MLP model has a positive value that suggests over-prediction. The GPR model possesses the most significant value of MBE.

The observed RMSE values are from 1.675 to 7.754 for category 1 and from 1.419 to 4.095 for category 2, with the MLP model showing the smallest value for both categories. MAPE values are detected in the range from 0.027% to 0.122% for category 1 and from 0.026% to 0.096% for category 2, with a minimum value seen for the MLP model in both categories. The coefficient of determination (R) values lies within 0.716–0.986 in category 1 and 0.889–0.986 in category 2, demonstrating an excellent data fit for all models. The MLP model shows the maximum value of R as 0.986 in both categories. The statistical pointer U_{95} values are from 25.308–30.208 in category 1 and from 23.431–24.791 in category 2, with the RF model showing the smallest values in both categories.

In MAE, the values range from 0.9–4.025 in category 1 and from 0.577–1.743 in category 2, with the MLP model again having a minimum value in both categories. For t-stat, the values range from 0.067–0.989 in category 1 and from 0.312–1.602 in category 2, with the minimum for the RF model in category 1 and MLP model in category 2. The erMAX value ranges from 6.221 to 29.115 in category 1 and from 7.778–13.283 in category 2, with the minimum value for the MLP model in both categories.

Table 5. Statistical Indicator Values for different ML algorithms.

	MBE	RMSE	MAPE	R	U95	MAE	t-Stats	erMAX
Category 1								
RF	0.090	5.477	0.122	0.858	25.308	4.025	0.067	11.891
KNN	−1.810	7.754	0.110	0.716	30.208	3.059	0.989	29.115
GPR	−0.079	3.164	0.033	0.954	27.225	1.272	0.104	11.016
MLP	−0.230	1.675	0.027	0.986	27.356	0.915	0.573	6.221
Category 2								
RF	−0.469	2.636	0.096	0.954	23.429	1.588	1.200	8.855
KNN	−0.711	4.095	0.090	0.889	24.284	1.743	1.170	13.283
GPR	−0.475	2.026	0.039	0.972	23.786	0.796	1.602	7.919
MLP	0.066	1.419	0.026	0.986	24.791	0.577	0.312	7.778

- Global Performance Indicator (GPI)

GPI was employed to improve our outcomes and eliminate anomalies in the statistical analysis and model ranking. Despotovic et al. [28] were the first to introduce GPI as a novel aspect; it is a fantastic way to integrate the effects of many statistical pointers. The particulars of the GPI technique can be found in Ref [29].

Scaled values of statistical parameters (0 to 1) and GPI and ML model ranking are depicted in Table 6. Figure 4 represents the overall GPI estimation. The MLP model ranks 1st with GPI = 1.603 then the GPR model with GPI = 1.308, followed by the RF model with GPI = −0.946. The MLP model shows the best performance among all the models in category 1. In category 2, the MLP model again shows top performance, followed by the GPR and RF models. Thus, we selected the MLP model for further analysis.

Table 6. Scaled statistical values and GPI for different ML models.

| MODEL	MBE	RMS	MAPE	R	U95	MAE	t-Stat	erMAX	GPI	Rank
Category 1										
RF	1.000	0.625	1.000	0.526	0.000	1.000	0.000	0.248	−0.946	3
KNN	0.000	1.000	0.874	0.000	1.000	0.689	1.000	1.000	−3.162	4
GPR	0.911	0.245	0.063	0.881	0.391	0.115	0.040	0.209	1.308	2
MLP	0.832	0.000	0.000	1.000	0.418	0.000	0.549	0.000	1.603	1
Category 2										
RF	0.311	0.455	1.000	0.670	0.000	0.867	0.688	0.196	−0.652	3
KNN	0.000	1.000	0.914	0.000	0.628	1.000	0.665	1.000	−3.012	4
GPR	0.304	0.227	0.186	0.856	0.262	0.188	1.000	0.026	0.859	2
MLP	1.000	0.000	0.000	1.000	1.000	0.000	0.000	0.000	1.195	1
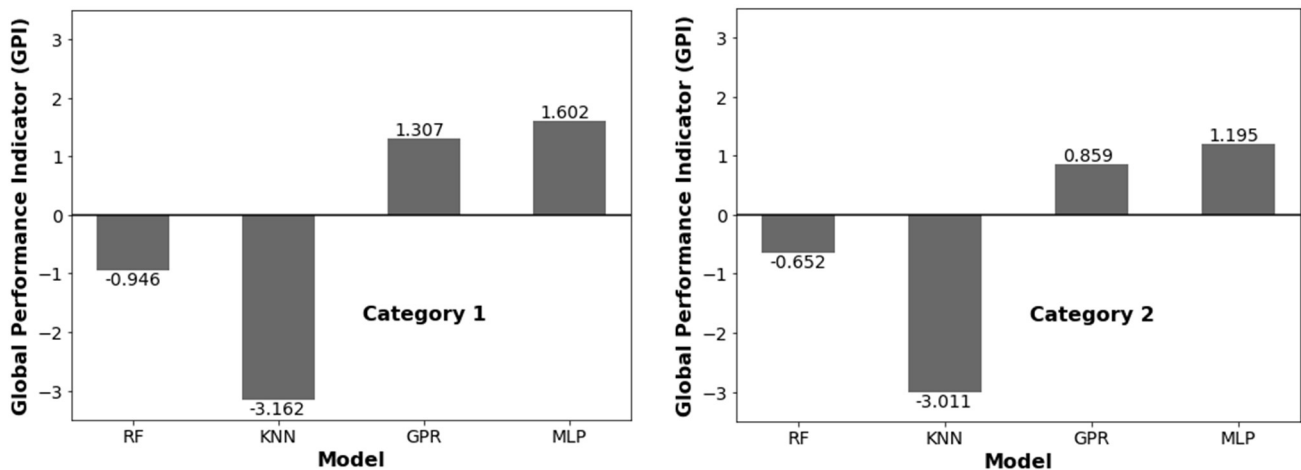
Figure 4. GPI values for different ML regression models.

4.2. Prediction Using ML Regression

The ML prediction model calculated the DE of the right lung and left lung for the 17-generation model. The DE was calculated for a wide range of flow rates and particle sizes, which was unavailable from the CFD measurement. The overall DE was calculated for the left and right lung and the five different lobes for the large-scale model. The schematic in Figure 5 shows the areas for the different parts of the 17-generation model.

(**a**)

(**b**)
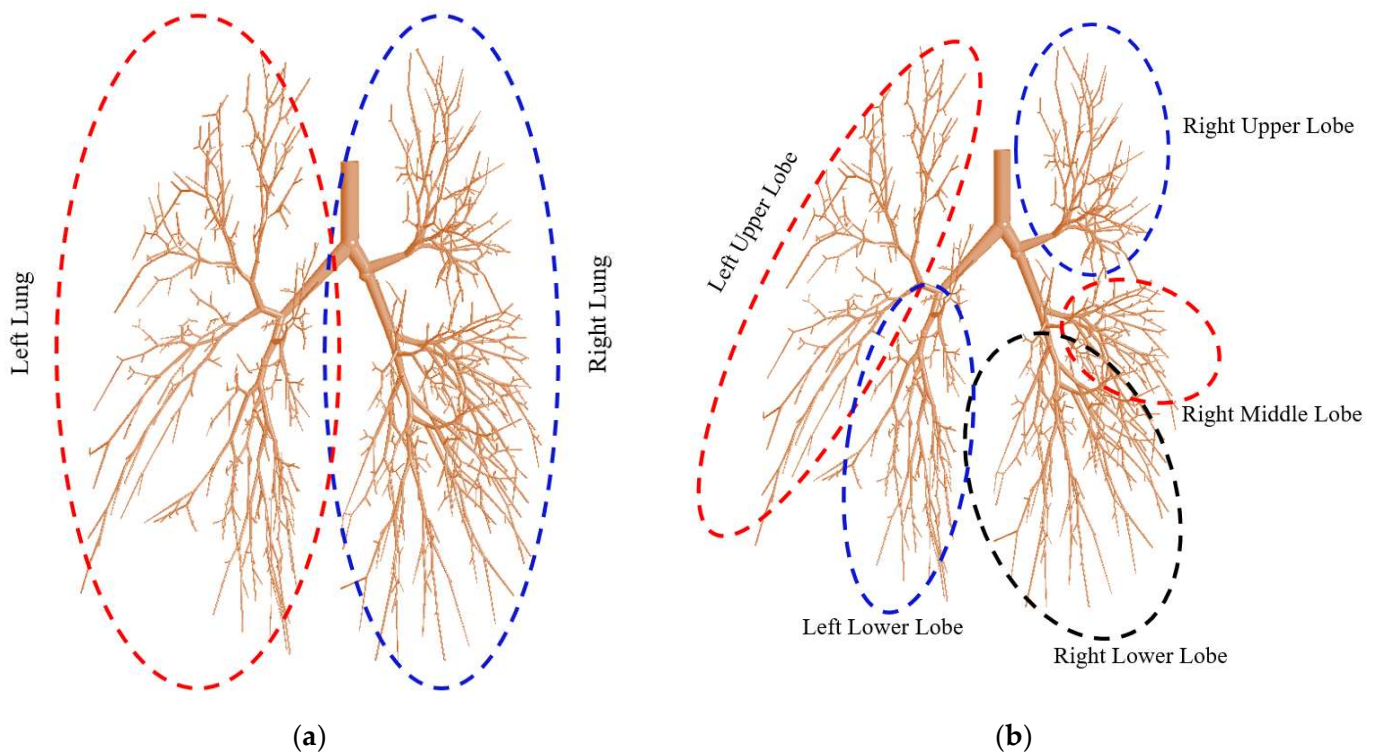
Figure 5. Definition of the local regions of the 17-generation airway model: (**a**) left and right lung, and (**b**) five different lobes.

Figure 6a shows the DE for 1 μm diameter particles at different flow rates. The overall DE of 1 μm particle in the right lung is higher than in the left, irrespective of the flow rates. For micron-sized particles, inertia plays an important role in overall transport and deposition. At high flow rates, larger particles cannot follow the air path line and deviates from the air streamline. At high flow conditions, once the particles are displaced from the

original pathline, they touch the airway wall and become trapped on the highly viscous surface of the airway wall. However, aerosols can follow the air streamline at low flow inlet conditions, minimising the overall DE at the upper airways. The right lung anatomical diameter is higher than the left lung, and the right lung consists of three airway lobes while the left contains two lobes. The complex anatomical shape and branching structure, higher flow distribution to the right lung, flow rates and particle inertia influence the overall DE at the right lung. The DE for the right and the left lung increased with the flow rates, which closely aligns with the CFD data, and it also satisfies the general aerosol deposition hypothesis in airways. Figure 6b–f show a similar trend of DE for different diameter particles. Figure 6d–f show the DE for 7 μm, 8.5 μm and 10 μm diameter particles, which is significantly higher than the smaller particles. Larger diameter particles mean the particle inertia will be higher, and the impaction mechanism is dominant for larger particles. At high flow condition, the inertial impaction became more dominant for the larger diameter particles and influenced the overall DE.
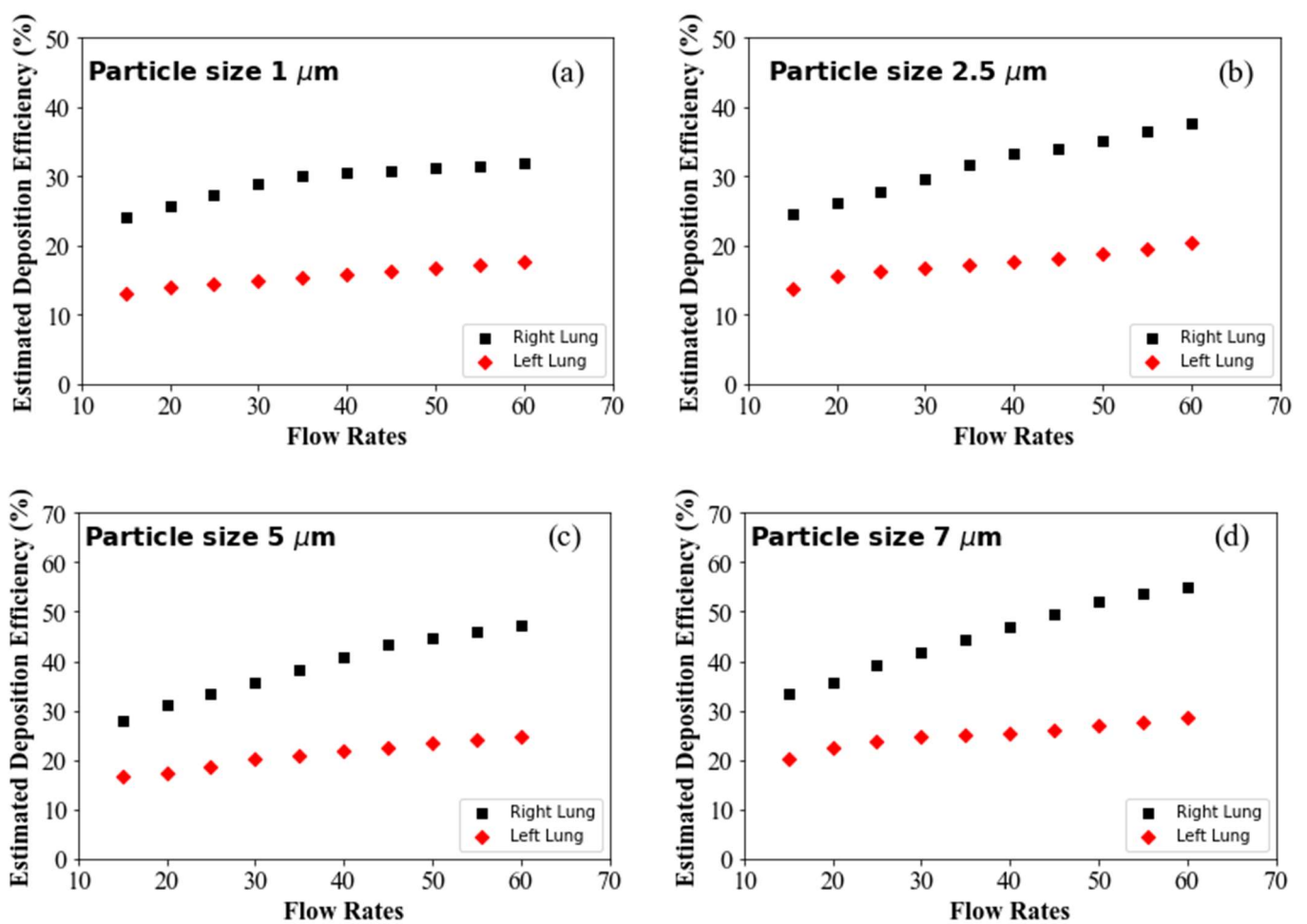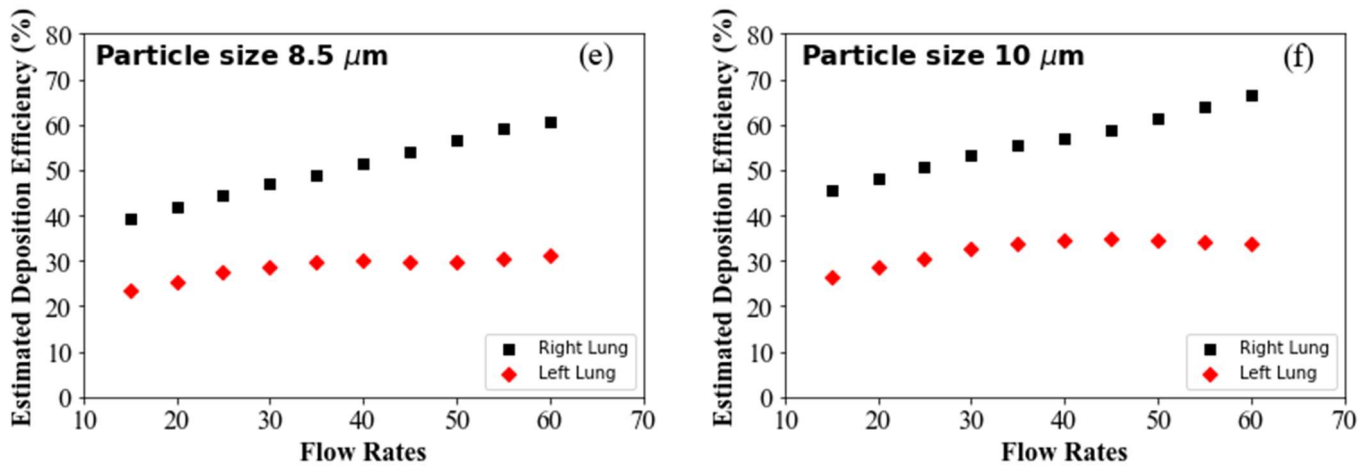

Figure 6. *Cont.*

Figure 6. Predicted DE for different flow rates: (**a**) 1 μm particle, (**b**) 2.5 μm particle, (**c**) 5 μm particle, (**d**) 7 μm particle, (**e**) 8.5 μm particle and (**f**) 10 μm particle.

The comparison of DE in different lobes with respect to flow rate and particle diameter for particle sizes of 2.5 μm and 7.5 μm is shown in Figure 7. For a 2.5 μm diameter particle, the DE is higher at the right lower lobe and lower at the left upper lobe irrespective of the deposition parameter. It is also noticed that the DE of 2.5 μm particles at the right lower lobe increased with the flow rate. The overall DE at the right and left upper lobes is lower than the right and left lower lobes. The upper lobes' branches are in the upward direction and opposite to the gravitational force, while the lower lobes are in the downward direction. Micron-sized particles mostly travel downward due to the high inertia, which transports fewer aerosols to the upper lobes of the airways, and less deposition occurs at the upper lobes. Figure 7b shows the 7.5 μm aerosol DE at different lobes, and a similar trend is observed at various lobes. At the flow rate (15 L/min), DE at the left upper lobe is different from other lobes. The overall DE at the left upper lobe is higher for low flow rate and lower for high flow rate. At a high flow rate, larger diameter particles deviate from the streamline and transport to the lower airways, while at a low flow rate, more particles transport to the upper airways. The overall DE at the right middle lobe shows a similar trend for 2.5 μm and 7.5 μm particles.
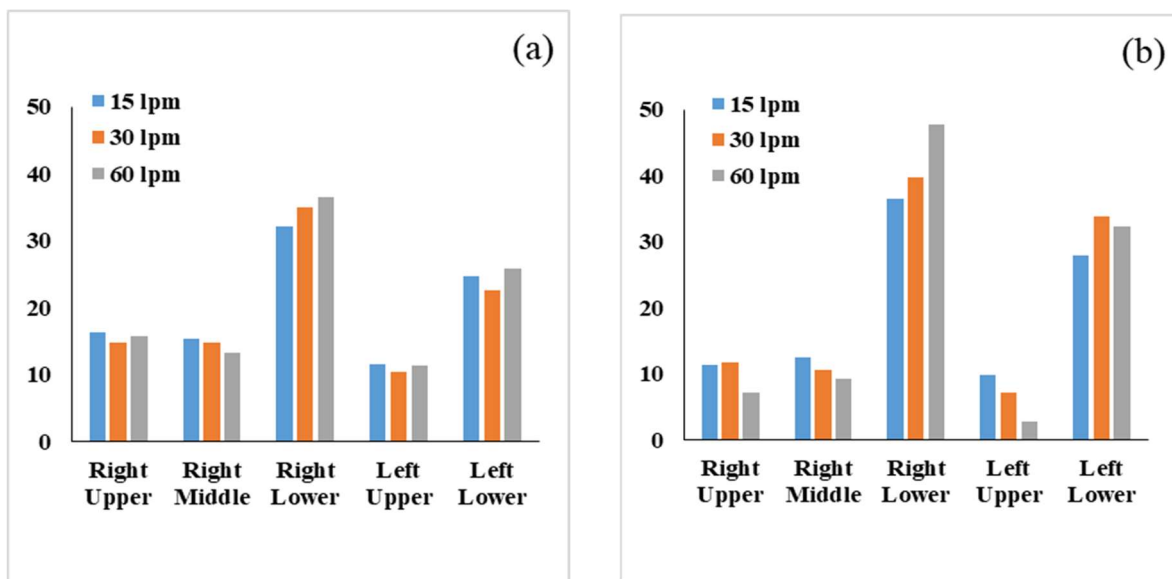

Figure 7. Comparison particle DE in the different lobes of the left and right lungs for flow rates of (**a**) 2.5 μm and (**b**) 7.5 μm.

A comprehensive lobe-specific DE was calculated for various diameter particles. Figure 8 shows the DE at different lobes of the 17-generation model for 15 L/min, 30 L/min and 60 L/min flow rates. The general trend shows higher DE at the lower lobes of the left and the right lung for all cases. The DE at the left lower lobe is lower than the remaining lobes in all cases. At the right upper lobe, the overall DE shows a fluctuating trend for different diameter particles. These specific findings are crucial for the health risk assessment of lung diseases and targeted drug delivery. The prediction model clearly indicates the deposition hot spot of the different diameter particles at various flow rates, which could potentially improve the efficiency of the targeted drug delivery. In conventional drug delivery tools, drug particles are mostly deposited in the upper airways. A minimum amount of drugs can reach the targeted areas of the lower part of the airways. The ML prediction model will improve the knowledge of the various diameter aerosol deposition hot spots and DE in the lower airways. An innovative drug delivery device can be developed for different lobes, and after diagnosis, a patient-specific drug delivery tool can be used, which will improve overall drug delivery efficiency.
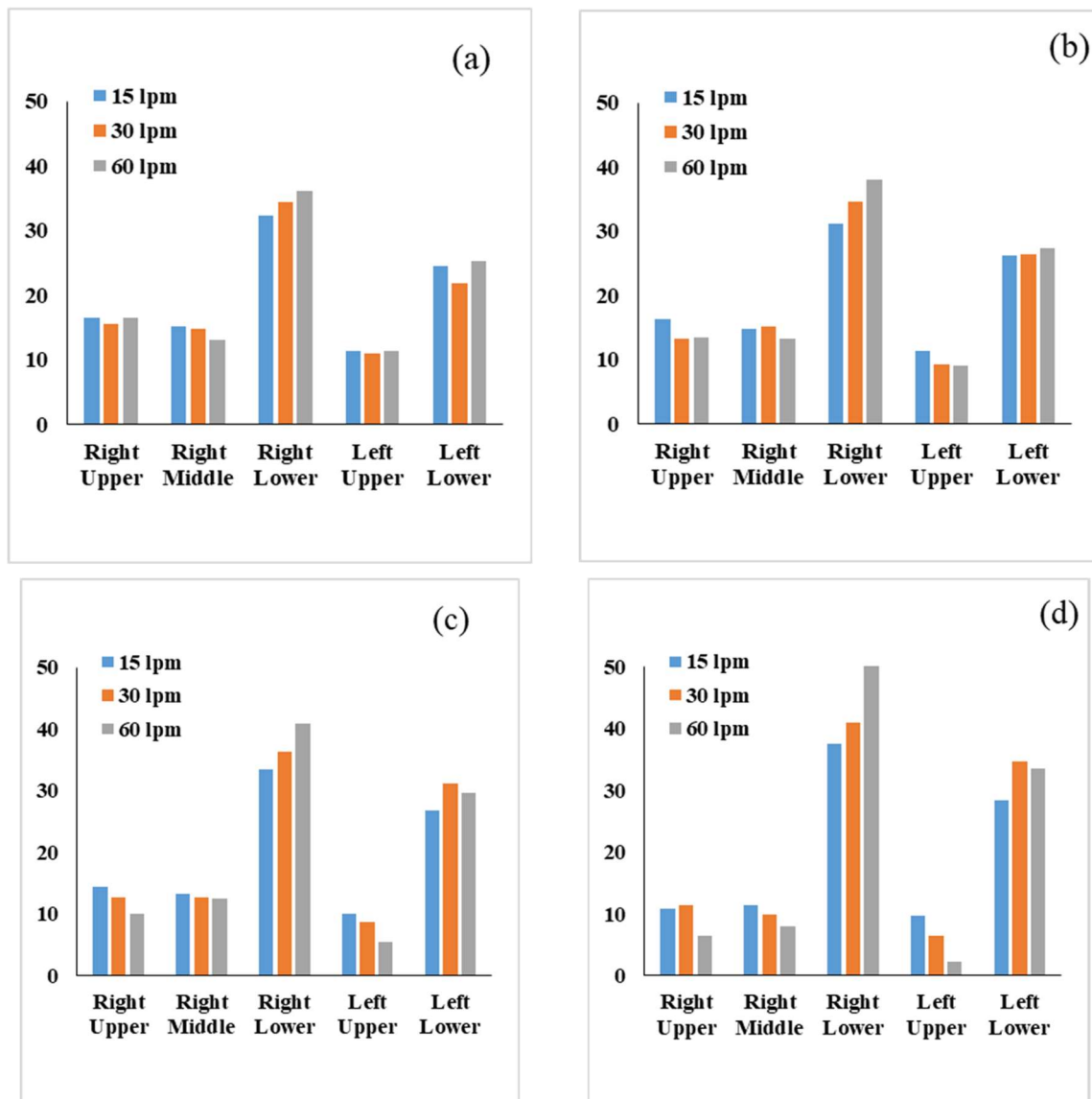

Figure 8. Particle DE comparison in the different lobes of the right and left lungs for different flow rates for (**a**) 2 μm, (**b**) 4 μm, (**c**) 6 μm and (**d**) 8 μm.

Figure 9 presents the comparison of DE for the 17-generation model. The ML prediction data is compared with the CFD measurements at a 25 L/min flow rate. The overall DE from the ML model shows a good agreement with the CFD data. A second-order polynomial equation is developed from the trendline of the CFD and ML DE curves. The R^2 value for the CFD and ML trendline is 0.9637 and 0.9751, respectively. The curve fitting indicates the accuracy of the ML model.
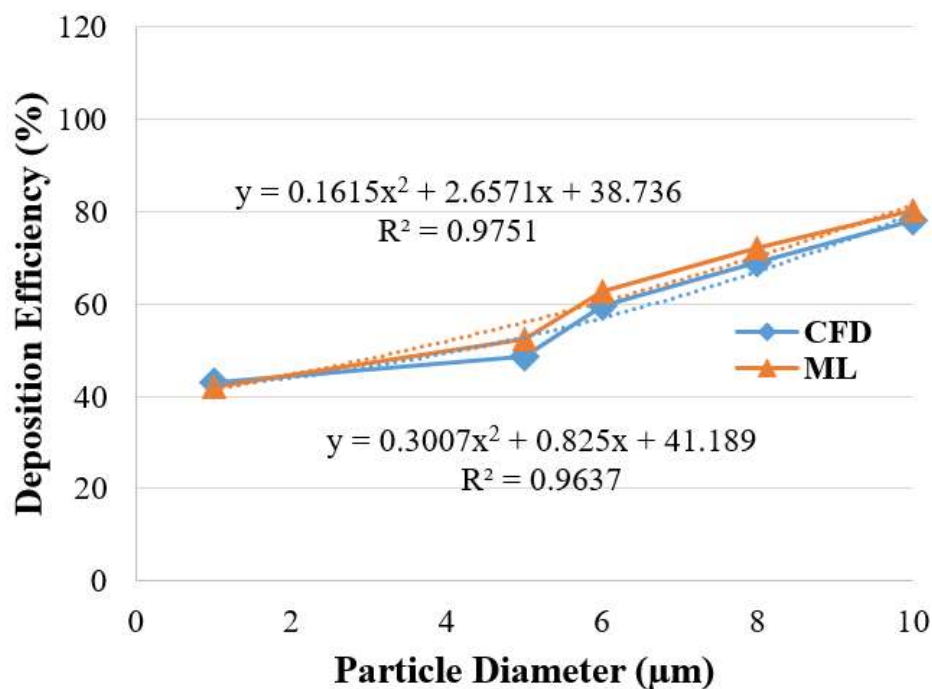
$$y = 0.1615x^2 + 2.6571x + 38.736$$
$$R^2 = 0.9751$$

$$y = 0.3007x^2 + 0.825x + 41.189$$
$$R^2 = 0.9637$$

Figure 9. DE comparison for CFD and ML measurement.

5. Conclusions

The present work advances the deposition analysis in the upper 17 generations of the human respiratory tract. Previously, researchers used CFD analysis, while a novel ML-based model is developed in the present work for the first-time. The present study developed ML prediction models from the CFD data of a 17-generation large-scale model. Different ML regression models are developed, and statistical analysis is performed to determine the best ML model for predicting pulmonary aerosol transport and deposition analysis in the upper 17 generations of the human respiratory tract. The ML models are trained and tested with the CFD data. It is witnessed that the MLP model performed well, with an overall GPI of 1.603 compared to other regression models. Furthermore, out of eight statistical indicator values, the MLP model has significant values for six. Therefore, the MLP regression model is used to predict the DE for a wide range of flow rates and particle sizes.

The MLP model predicts the DE in both lungs. The prediction shows higher DE in the right lung than in the left lung. The MLP model also reports higher DE at the left and the right lung irrespective of the inlet conditions and particle diameter. It is also observed that the MLP model gives excellent predictions with a similar trend of deposition efficiency as observed in our CFD work.

A comprehensive lobe-specific DE is calculated by using the ML prediction model. The overall DE at the right lower lobe is higher than the remaining lobes. The DE of various diameter particles is different at the lobes for different flow rates. The MLP prediction model analysed the deposition hot spot at various lobes for the first time, which would improve the knowledge of the aerosol deposition in the lower airways. These specific findings would improve the knowledge of the field and could potentially improve the efficiency of the targeted drug delivery to the lower airways. Thus, the MLP model can

be used to predict deposition efficiency for flow rates and particle sizes not considered in CFD analysis, resulting in considerable savings in time and cost. The developed prediction model will save a significant amount of computational time. The present study, along with a physics-informed ML modelling for the airflow and particle transport in airways, would improve the knowledge of the field.

References

1. Bass, K.; Farkas, D.; Hassan, A.; Bonasera, S.; Hindle, M.; Longest, W. High-efficiency dry powder aerosol delivery to children: Review and application of new technologies. *J. Aerosol Sci.* **2020**, *153*, 105692. [CrossRef] [PubMed]
2. Koullapis, P.; Nicolaou, L.; Kassinos, S.C. In silico assessment of mouth-throat effects on regional deposition in the uppertracheo-bronchial airways. *J. Aerosol Sci.* **2018**, *117*, 164–188. [CrossRef]
3. Cheng, Y.-S.; Smith, S.M.; Yeh, H.-C.; Kim, D.-B.; Cheng, K.-H.; Swift, D.L. Deposition of ultrafine aerosols and thoron progeny in replicas of nasal airways of young children. *Aerosol Sci. Technol.* **1995**, *23*, 541–552. [CrossRef]
4. Hofmann, W.; Steinhausler, F.; Pohl, E. Dose Calculations For the Respiratory Tract From Inhaled Natural Radioactive NuclidesAs a Function of Age-I: Compartmental Deposition, Retention and Resulting Dose. *Health Phys.* **1979**, *37*, 517–532. [CrossRef] [PubMed]
5. Gemci, T.; Ponyavin, V.; Chen, Y.; Chen, H.; Collins, R. Computational model of airflow in upper 17 generations of human respiratory tract. *J. Biomech.* **2008**, *41*, 2047–2054. [CrossRef]
6. Zhang, Z.; Kleinstreuer, C.; Donohue, J.F.; Kim, C.S. Comparison of micro- and nano-size particle depositions in a human upper airway model. *J. Aerosol Sci.* **2005**, *36*, 211–233. [CrossRef]
7. Islam, M.S.; Saha, S.C.; Sauret, E.; Gemci, T.; Gu, Y.T. Pulmonary aerosol transport and deposition analysis in upper 17 generations of the human respiratory tract. *J. Aerosol Sci.* **2017**, *108*, 29–43. [CrossRef]
8. Cheng, K.-H.; Cheng, Y.-S.; Yeh, H.-C.; Guilmette, R.A.; Simpson, S.Q.; Yang, Y.-H.; Swift, D.L. In vivo measurements of nasal airway dimensions and ultrafine aerosol deposition in the human nasal and oral airways. *J. Aerosol Sci.* **1996**, *27*, 785–801. [CrossRef]
9. Zhang, Z.; Kleinstreuer, C.; Kim, C.S. Airflow and nanoparticle deposition in a 16-generation tracheobronchial airway model. *Ann. Biomed. Eng.* **2008**, *36*, 2095–2110. [CrossRef]
10. Schmidt, A.; Zidowitz, S.; Kriete, A.; Denhard, T.; Krass, S.; Peitgen, H.O. A digital reference model of the human bronchial tree. *Comput. Med. Imaging Graph.* **2004**, *28*, 203–211. [CrossRef]
11. Islam, M.S.; Saha, S.C.; Sauret, E.; Gemci, T.; Yang, I.A.; Gu, Y. Ultrafine particle transport and deposition in a large scale 17-generation lung model. *J. Biomech.* **2017**, *64*, 16–25. [CrossRef] [PubMed]
12. Jalalifar, S.; Masoudi, M.; Abbassi, R.; Garaniya, V.; Ghiji, M.; Salehi, F. A hybrid SVR-PSO model to predict a CFD-based optimised bubbling fluidised bed pyrolysis reactor. *Energy* **2020**, *191*, 116414. [CrossRef]
13. Gholizadeh, M.; Jamei, M.; Ahmadianfar, I.; Pourrajab, R. Prediction of nanofluids viscosity using random forest (RF) approach. *Chemom. Intell. Lab. Syst.* **2020**, *201*, 104010. [CrossRef]
14. Kwon, B.; Ejaz, F.; Hwang, L.K. Machine learning for heat transfer correlations. *Int. Commun. Heat Mass Transf.* **2020**, *116*, 104694. [CrossRef]
15. Jamei, M.; Ahmadianfar, I.; Olumegbon, I.A.; Karbasi, M.; Asadi, A. On the assessment of specific heat capacity of nanofluids for solar energy applications: Application of Gaussian process regression (GPR) approach. *J. Energy Storage* **2021**, *33*, 102067. [CrossRef]
16. Modaresi, F.; Araghinejad, S.; Ebrahimi, K. A Comparative Assessment of Artificial Neural Network, Generalized Regression Neural Network, Least-Square Support Vector Regression, and K-Nearest Neighbor Regression for Monthly Streamflow Forecasting in Linear and Nonlinear Conditions. *Water Resour. Manag.* **2018**, *32*, 243–258. [CrossRef]
17. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2018. [CrossRef]
18. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

19. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2012**, *12*, 2825–2830. [CrossRef]
20. Husain, S.; Khan, U.A. Machine learning models to predict diffuse solar radiation based on diffuse fraction and diffusion coefficient models for humid-subtropical climatic zone of India. *Clean. Eng. Technol.* **2021**, *5*, 100262. [CrossRef]
21. Mohammadpour, J.; Salehi, F.; Sheikholeslami, M.; Masoudi, M.; Lee, A. Optimisation of nanofluid heat transfer in a microchannel heat sink with multiple synthetic jets based on CFD-DPM and MLA. *Int. J. Therm. Sci.* **2021**, *167*, 107008. [CrossRef]
22. Hu, H. Enhanced Gabor Feature Based Classification Using a Regularized Locally Tensor Discriminant Model for Multiview Gait Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 1274–1286. [CrossRef]
23. Vinci, A.; Zoli, L.; Sciti, D.; Melandri, C.; Guicciardi, S. Understanding the Mechanical Properties of Novel UHTCMCs through Random Forest and Regression Tree Analysis. *Mater. Des.* **2018**, *145*, 97–107. [CrossRef]
24. Williams, C.K.I. *Gaussian Processes Formachine Learning*; Taylor & Francis Group: Abingdon, UK, 2006.
25. Mukherjee, S.; Mishra, P.; Ali, N.; Aljuwayhel, N.; Ebrahim, S.; Chaudhuri, P. Thermo-Physical Properties and Heat Transfer Potential of Novel Silica-Ethylene Glycol Mono Nanofluid: Experiments and Multi-Layer Perceptron (MLP) Modelling. In *Colloids and Surfaces A: Physicochemical and Engineering Aspects*; Elsevier: Amsterdam, The Netherlands, 2022; p. 129412. [CrossRef]
26. Nguyen, Q.H.; Ly, H.-B.; Ho, L.S.; Al-Ansari, N.; Le, H.; Van Tran, V.Q.; Prakash, I.; Pham, B.T. Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Math. Probl. Eng.* **2021**, *2021*, 1–15. [CrossRef]
27. Jamil, B.; Akhtar, N. Empirical models for estimation of diffuse solar radiation based on measured data for humid-subtropical climatic region of India. *J. Renew. Sustain. Energy* **2017**, *9*, 033702. [CrossRef]
28. Despotovic, M.; Nedic, V.; Despotovic, D.; Cvetanovic, S. Review and statistical analysis of different global solar radiation sunshine models. *Renew. Sustain. Energy Rev.* **2015**, *52*, 1869–1880. [CrossRef]
29. Husain, S.; Khan, U.A. Development of machine learning models based on air temperature for estimation of global solar radiation in India. *Environ. Prog. Sustain. Energy* **2021**, *41*, e13782. [CrossRef]