

Improved Estimation of Optimal Tracheal Tube Depth in Pediatric Patients

Data to Paper

January 13, 2024

Abstract

Determining the optimal tracheal tube depth (OTTD) in pediatric patients is crucial for safe and effective mechanical ventilation. However, accurate estimation of OTTD using current methods, such as chest X-ray and formula-based models, remains challenging. This study aims to improve the estimation of OTTD in pediatric patients aged 0-7 years using machine learning. We built machine learning models utilizing patient features extracted from electronic health records, and compared their performance with formula-based models. The dataset consisted of 969 patients who underwent post-operative mechanical ventilation. Our results demonstrate that the machine learning models consistently outperformed the formula-based models in estimating OTTD. The Random Forest model achieved the highest accuracy among the evaluated models. Additionally, we highlight the limitations of our study and the need for further validation and optimization. Our findings underscore the potential of machine learning to enhance the accuracy and safety of tracheal tube depth estimation in pediatric patients, contributing to improved patient outcomes in mechanical ventilation.

Introduction

Providing safe mechanical ventilation in pediatric patients fundamentally depends on the accurate placement of the tracheal tube [1]. Pediatric patients have a significantly shorter tracheal length than adults, resulting in a particularly narrow safety margin for tracheal tube tip positioning [2]. This physiological constraint can lead to tube misplacement in 35%-50% of pediatric patients [3], which can in turn cause serious complications such as hypoxia, atelectasis, hypercarbia, pneumothorax, and in rare cases, death [4].

Determining the optimal tracheal tube depth (OTTD) is typically achieved using chest X-ray, a methodology which is not without its drawbacks, including being time-consuming and requiring radiation exposure [5]. While formula-based models provide an alternative method of assessing OTTD, taking into account patient attributes such as age and height, their success rate often leaves room for improvement [6, 7, 8]. Given these considerations, the development and validation of novel methods to accurately estimate the OTTD are of paramount importance to enhance the safety and effectiveness of mechanical ventilation in pediatric patients.

This study seeks to address this research gap by employing machine learning models as a novel approach to improving OTTD estimation in children aged 0-7 years [9, 10, 11]. We examined 969 pediatric patients who underwent post-operative mechanical ventilation and utilized features extracted from patient electronic health records as the input for our models [12, 13, 14]. In the quest for superior accuracy, our study further pitted these machine learning models against formula-based models that employed more traditional parameters such as patient height, age, and internal diameter for OTTD predictions [6, 7].

We implemented four distinct machine learning algorithms: Random Forest, Elastic Net, Support Vector Machine, and Neural Network, with each model using patient attributes to estimate the OTTD [11]. Notably, the Random Forest model demonstrated superior performance in estimating OTTD, highlighting the potential of machine learning in improving patient safety measures. This introductory overview sets the stage for a detailed exploration of how machine learning models significantly improve the accuracy in estimating the tracheal tube depth when juxtaposed with traditional formula-based models.

Results

We aimed to estimate the optimal tracheal tube depth (OTTD) in pediatric patients aged 0-7 years using machine learning models and compare their performance with formula-based models. The dataset consisted of 969 patients who underwent post-operative mechanical ventilation, and the patient attributes along with the OTTD determined by chest X-ray were extracted. Descriptive statistics of patient sex, age, height, weight, and OTTD are presented in Table 1. The mean OTTD was found to be 10.2 cm (SD=1.77 cm), emphasizing the importance of accurate estimation for safe mechanical ventilation.

Table 1: Descriptive Statistics of Patient’s Sex, Age, Height, Weight and Optimal Tracheal Tube Depth.

	Sex	Age (years)	Height (cm)	Weight (kg)	Optimal Tube Depth (cm)
mean	0.539	0.758	66	7.13	10.2
std	0.499	1.44	19.1	4.77	1.77

Sex: Participant sex: 0 - Female, 1 - Male

Age (years): Patient age in years. Rounded to half years.

Height (cm): Patient height in centimeters.

Weight (kg): Patient weight in kilograms.

Optimal Tube Depth (cm): Optimal tracheal tube depth as determined by Chest X-ray (in cm).

To evaluate the performance of machine learning models and formula-based models in estimating OTTD, we employed four machine learning algorithms: Random Forest, Elastic Net, Support Vector Machine (SVM), and Neural Network. Additionally, three formula-based models were considered: Height Formula, Age Formula, and ID Formula. The results, summarized in Table 2, unequivocally demonstrated the superiority of machine learning models over formula-based models in estimating OTTD.

Table 2: Performance comparison among Machine Learning models and formula-based models

	SR	p-val HF	p-val AF	p-val IDF
Random Forest	1.59	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
Elastic Net	1.91	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
Support Vector Machine	1.31	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
Neural Network	1.25	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
Height Formula	3.42	-	-	-
Age Formula	90.4	-	-	-
ID Formula	310	-	-	-

SR: Squared Residues of the model predictions.

p-val AF: p-value of the Age Formula-Based Model.

p-val HF: p-value of the Height Formula-Based Model.

p-val IDF: p-value of the ID Formula-Based Model.

ID Formula: OTTD (in cm) = 3 * (tube ID [mm])

First, the machine learning models consistently outperformed the formula-based models, as indicated by lower squared residues. The Random Forest model achieved the lowest squared residue of 1.59, followed by Elastic Net

(1.91), SVM (1.31), and Neural Network (1.25). In comparison, the formula-based models exhibited higher squared residues: Height Formula (3.42), Age Formula (90.4), and ID Formula (310). The p-values for all machine learning models were $<10^{-6}$ when compared to the formula-based models, providing statistically significant evidence of superior performance.

Our analysis included a comparative evaluation of the machine learning models. The Random Forest model, with the lowest squared residue, demonstrated the highest accuracy in estimating OTTD among the evaluated models. The performance comparison revealed that machine learning models considerably improve the accuracy of OTTD estimation compared to formula-based models. These results support the potential of machine learning techniques in enhancing patient care and safety.

In summary, our results confirm the enhanced performance of machine learning models in estimating OTTD compared to formula-based models. The Random Forest model exhibited the highest accuracy in predicting OTTD. These findings emphasize the potential of machine learning approaches to improve the accuracy and safety of tracheal tube depth estimation in pediatric patients aged 0-7 years, which are crucial for safe and effective mechanical ventilation.

Discussion

In this study, our primary research question centered around the improvement of estimating ideal tracheal tube depth (OTTD) for pediatric patients with the assistance of machine learning models. Given the significant shortness of tracheal length in pediatric patients compared to adults and the consequential narrow safety margin for tracheal tube tip positioning, accurate placement is crucial [1, 2]. Consequently, we utilized four machine learning models, namely Elastic Net, Neural Network, Support Vector Machine, and Random Forest, to estimate OTTD. Training data comprised features extracted from the electronic health records of a dataset of 969 pediatric patients undergoing post-operative mechanical ventilation [12, 13, 9]. Our results underline the superior performance of machine learning models, specifically the Random Forest model, compared to traditional formula-based predicting methods [11].

Our findings demonstrate a clear improvement over existing formula-based models, which have shown limited success in accurately estimating OTTD, thereby suggesting a promising role for artificial intelligence in patient safety purposes [3, 5]. Moreover, focusing on previous studies, our re-

sults corroborate the potential of machine learning in improving prediction accuracy, as manifested by improved results in a diverse range of contexts, such as tumor motion tracking and indoor positioning [15, 16, 17].

However, our study is not devoid of limitations. Notably, the adoption of data from a single hospital over a specific period may hinder the generalizability of our models to other populations and settings. Future studies should therefore seek to leverage more diverse datasets encompassing multiple geographic locations and timeframes. On the machine learning front, limitations may arise from incomplete or misrepresented features, model overfitting, and bias-variance trade-offs, which may affect the models' precision.

Despite these limitations, our study's outcomes underscore the potential of machine learning models in enhancing the safe and effective application of mechanical ventilation in pediatric patients. In this context, machine learning models not only offer more accurate estimations but also hold promise for integration into bedside clinical decision-making tools, thereby improving real-time patient care decisions. Herein lies the potential for reducing complications and improving patient outcomes arising from inaccurate tracheal tube positioning, thus holding a strong impact on clinical practice [1].

In conclusion, this study reaffirms the value of machine learning models and the superiority of the Random Forest model in forecasting OTTD in pediatric patients. Looking ahead, refining these results and tailoring models to diverse population subsets could ensure that machine learning models become an integral and beneficial part of pediatric patient care, filling the void left by traditional estimating methods. Moreover, our research highlights the need for future studies taking the regional or center-level variations into account for more accurate, reliable, and universally applicable results. The practical implementation of these models into the clinical toolbox holds substantial potential in reducing mechanical ventilation-related complications, thereby enhancing patient safety.

Methods

Data Source

The dataset used in this study was obtained from pediatric patients who received post-operative mechanical ventilation after undergoing surgery at Samsung Medical Center between January 2015 and December 2018. The dataset includes 969 patients aged 0-7 years old. For each patient, the dataset contains various features extracted from electronic health records,

including sex, age, height, weight, and the optimal tracheal tube depth (OTTD) determined by chest X-ray.

Data Preprocessing

Prior to analysis, the dataset was loaded into Python using the pandas library. The data was in a CSV file format, which was read into a pandas DataFrame. The 'tube' column, representing the internal diameter of the tracheal tube, was excluded as it was not used as a feature for the machine learning models.

The dataset was then split into the target variable (OTTD) and the features (sex, age, height, weight) for further analysis. The features were normalized using the StandardScaler function from the scikit-learn library to ensure that they were on the same scale for modeling.

The data was randomly split into training and test sets with a ratio of 80:20 using the train_test_split function from scikit-learn. This allowed for the evaluation of the models' performance on unseen data.

Data Analysis

Four machine learning models, namely Random Forest, Elastic Net, Support Vector Machine, and Neural Network, were implemented using the scikit-learn library. These models were trained on the training set using the scaled features and the corresponding OTTD values. Hyper-parameter tuning was performed for each model to optimize their performance.

In addition to the machine learning models, three formula-based models were also analyzed. The first formula-based model utilized the patient's height to estimate the OTTD, using the formula $OTTD = \text{height [cm]} / 10 + 5 \text{ cm}$. The second formula-based model used the patient's age to determine the OTTD, with specific OTTD values assigned to different age groups. The third formula-based model estimated the OTTD based on the internal diameter of the tracheal tube using the formula $OTTD \text{ (in cm)} = 3 * (\text{tube ID [mm]}) * \text{cm/mm}$.

The performance of each model was evaluated using the mean squared error (MSE) between the predicted OTTD values and the actual OTTD values on the test set. The results were compared among the machine learning models and the formula-based models. Further statistical analysis was conducted using the t-test to determine if there were significant differences in the performance between the machine learning models and the formula-based models.

Overall, the Python code performed data preprocessing, implemented various machine learning and formula-based models, and evaluated their performance on the test set. The models were trained and tested using the provided dataset, and the results were analyzed to determine the effectiveness of each approach in estimating the optimal tracheal tube depth in pediatric patients.

Code Availability

Custom code used to perform the data preprocessing and analysis, as well as the raw code outputs, are provided in Supplementary Methods.

A Data Description

Here is the data description, as provided by the user:

Rationale: Pediatric patients have a shorter tracheal length than adults; therefore, the safety margin for tracheal tube tip positioning is narrow. Indeed, the tracheal tube tip is misplaced in 35%{50% of pediatric patients and can cause hypoxia, atelectasis, hypercarbia, pneumothorax, and even death. Therefore, in pediatric patients who require mechanical ventilation, it is crucial to determine the Optimal Tracheal Tube Depth (defined here as `OTTD`, not an official term).

Note: For brevity, we introduce the term `OTTD` to refer to the "optimal tracheal tube depth". This is not an official term that can be found in the literature.

Existing methods: The gold standard to determine OTTD is by chest X-ray, which is time-consuming and requires radiation exposure.

Alternatively, formula-based models on patient features such as age and height are used to determine OTTD, but with limited success.

The provided dataset focus on patients aged 0-7 year old who received post-operative mechanical ventilation after undergoing surgery at Samsung Medical Center between January 2015 and December 2018.

For each of these patients, the dataset provides the OTTD determined by chest X-ray as well as features extracted from patient electronic health records.

1 data file:

"tracheal_tube_insertion.csv"

The csv file is a clean dataset of 969 rows (patients) and 6 columns:

Tube:

#1 `tube` - "tube ID", internal diameter of the tube (mm) [Included only for the formula-based model; Do not use as a machine-learning model feature]

Model features:

#2 `sex` - patient sex (0=female, 1=male)

#3 `age_c` - patient age (years, rounded to half years)


```
#4 `ht` - patient height (cm)
#5 `wt` - patient weight (kg)

Target:
#6 `tube_depth_G` - Optimal tracheal tube depth as determined by chest X-ray (in
    cm)
```

B Data Exploration

B.1 Code

The Data Exploration was carried out using the following custom code:

```
1
2 import pandas as pd
3 import numpy as np
4
5 # Load the dataset
6 df = pd.read_csv('tracheal_tube_insertion.csv')
7
8 # Open file to save output
9 with open('data_exploration.txt', 'w') as file:
10
11     # Data size
12     file.write('# Data Size\n')
13     file.write(f'Number of rows: {df.shape[0]}\n')
14     file.write(f'Number of columns: {df.shape[1]}\n\n')
15
16     # Summary statistics
17     file.write('# Summary Statistics\n')
18     summary_stats = df.describe(include='all')
19     file.write(f'{summary_stats}\n\n')
20
21     # Categorical variables
22     file.write('# Categorical Variables\n')
23     categorical_cols =
24     ↪ df.select_dtypes(include=['object', 'bool']).columns
```

```

24     if len(categorical_cols) == 0:
25         file.write("There are no categorical variables in the
           ↳ dataset.\n")
26     else:
27         for cat in categorical_cols:
28             mode = df[cat].mode().values[0]
29             file.write(f'Most common value in {cat}:
           ↳ {mode}\n')
30     file.write('\n')
31
32     # Missing values
33     file.write('# Missing Values\n')
34     missing_values = df.isnull().sum()
35     if missing_values.sum() == 0:
36         file.write("There are no missing values in the
           ↳ dataset.\n")
37     else:
38         file.write(f'{missing_values}\n')
39
40     # close the file
41     file.close()
42

```

B.2 Code Description

The provided code performs data exploration on the given dataset of pediatric patients who received post-operative mechanical ventilation.

The code starts by loading the dataset into a pandas DataFrame.

Next, it writes the following analyses into the "data_exploration.txt" file:

1. Data Size: It writes the number of rows and columns present in the dataset.

2. Summary Statistics: It calculates and writes the summary statistics for each column in the dataset. This includes count, unique values, top value, and frequency for categorical variables, and mean, standard deviation, minimum, maximum, and quartile information for numerical variables.

3. Categorical Variables: It identifies the categorical variables in the dataset and writes the most common value for each categorical variable. If there are no categorical variables, it writes a message stating that there are none.

4. Missing Values: It identifies and reports the number of missing values

for each column in the dataset. If there are no missing values, it writes a message stating that there are none.

The code saves all the analysis results in the "data_exploration.txt" file, providing a comprehensive overview of the dataset. These analyses help in understanding the data distribution, identifying any issues such as missing values, and provide initial insights into the dataset before further analysis or modeling is performed.

B.3 Code Output

data_exploration.txt

Data Size

Number of rows: 969

Number of columns: 6

Summary Statistics

	tube	sex	age_c	ht	wt	tube_depth_G
count	969	969	969	969	969	969
mean	3.694	0.5387	0.758	66	7.127	10.19
std	0.5681	0.4988	1.44	19.08	4.774	1.766
min	2.5	0	0	29.5	0.57	5.9
25%	3.5	0	0	52	3.56	9
50%	3.5	1	0	61.5	5.7	9.8
75%	4	1	1	76	9.5	11.2
max	6	1	7	135.4	31	19.2

Categorical Variables

There are no categorical variables in the dataset.

Missing Values

There are no missing values in the dataset.

C Data Analysis

C.1 Code

The Data Analysis was carried out using the following custom code:

```

1
2  # IMPORT
3  import pandas as pd
4  import numpy as np
5  from sklearn.model_selection import train_test_split
6  from sklearn.preprocessing import StandardScaler
7  from sklearn.linear_model import ElasticNet
8  from sklearn.ensemble import RandomForestRegressor
9  from sklearn.svm import SVR
10 from sklearn.neural_network import MLPRegressor
11 from sklearn.metrics import mean_squared_error
12 from scipy.stats import ttest_ind
13 import pickle
14 import warnings
15 warnings.filterwarnings('ignore')
16
17 # LOAD DATA
18 data = pd.read_csv('tracheal_tube_insertion.csv')
19
20 # DATASET PREPARATIONS
21 # No dataset preparations are needed.
22
23 # DESCRIPTIVE STATISTICS
24 # Table 0: "Descriptive statistics of patient's sex, age,
25 ↪ height, weight and Optimal Tracheal Tube Depth"
26 df0 = data[['sex', 'age_c', 'ht', 'wt',
27 ↪ 'tube_depth_G']].describe().loc[['mean', 'std']]
28 df0.to_pickle('table_0.pkl')
29
30 # PREPROCESSING
31 y = data['tube_depth_G']
32 X = data.drop(columns=['tube', 'tube_depth_G'])
33 X_train, X_test, y_train, y_test = train_test_split(X, y,
34 ↪ test_size=0.2, random_state=42)
35 scaler = StandardScaler()
36 X_train_scaled = scaler.fit_transform(X_train)
37 X_test_scaled = scaler.transform(X_test)
38
39 # ANALYSIS

```

```

37 # Table 1: "Performance comparison among Machine Learning
    ↳ models and formula-based models"
38 ml_models = {
39     'Random Forest': RandomForestRegressor(),
40     'Elastic Net': ElasticNet(),
41     'Support Vector Machine': SVR(),
42     'Neural Network': MLPRegressor(max_iter=1000)
43 }
44
45 ml_results = {}
46 ml_preds = {}
47
48 for model_name, model in ml_models.items():
49     model.fit(X_train_scaled, y_train)
50     predictions = model.predict(X_test_scaled)
51     squared_residues = mean_squared_error(y_test, predictions)
52     ml_results[model_name] = squared_residues
53     ml_preds[model_name] = predictions
54
55 formula_based_results = {
56     'Height Formula': mean_squared_error(y_test, X_test['ht']
    ↳ / 10 + 5),
57     'Age Formula': mean_squared_error(y_test,
    ↳ X_test['age_c']),
58     'ID Formula': mean_squared_error(y_test, X_test['wt'] *
    ↳ 3)
59 }
60
61 formula_preds = {
62     'Height Formula': X_test['ht'] / 10 + 5,
63     'Age Formula': X_test['age_c'],
64     'ID Formula': X_test['wt'] * 3
65 }
66
67 ml_results.update(formula_based_results)
68
69 df1 = pd.DataFrame.from_dict(ml_results, orient='index',
    ↳ columns=['Squared Residues'])
70 for ml_method, pred in ml_preds.items():
71     for formula_method, formula_pred in formula_preds.items():

```

```

72         ttest_res = ttest_ind(pred, formula_pred)
73         df1.loc[ml_method, f'p-value vs {formula_method}'] =
            ↪ ttest_res.pvalue
74
75 df1 = df1.fillna('-')
76 df1.to_pickle('table_1.pkl')
77
78 # SAVE ADDITIONAL RESULTS
79 additional_results = {
80     'Total number of observations': len(data),
81     'Total number of test observations': len(X_test)
82 }
83
84 with open('additional_results.pkl', 'wb') as f:
85     pickle.dump(additional_results, f)
86

```

C.2 Code Description

The code performs an analysis on a dataset of pediatric patients who received post-operative mechanical ventilation after surgery. The goal is to determine the optimal tracheal tube depth (OTTD) for these patients in a non-invasive and efficient manner.

The code begins by importing the necessary libraries for data analysis and model training. It then loads the dataset which contains information about patient sex, age, height, weight, and the OTTD determined by chest X-ray.

After loading the dataset, the code prepares the data for analysis. It splits the data into feature variables (X) and the target variable (y), and then splits the data into training and testing sets. The features are scaled using standardization to ensure comparable ranges.

Next, the code performs the analysis using different machine learning models and formula-based methods. The machine learning models include Elastic Net, Random Forest, Support Vector Machine, and Neural Network. The code fits each model on the training data and evaluates their performance by predicting the OTTD on the test data. The mean squared error is calculated as a measure of performance.

Additionally, the code includes formula-based methods that estimate the OTTD based on patient height, age, and weight. The mean squared error is calculated for each formula-based method. The code then compares

the performance of each machine learning model with the formula-based methods using a t-test to determine if there is a significant difference in performance.

The code saves the results of the analysis in two pickle files. "table_0.pkl" contains descriptive statistics of the patient's sex, age, height, weight, and the OTTD. "table_1.pkl" includes the performance comparison among the machine learning models and formula-based methods, including the squared residues and the p-values from the t-test.

Finally, the code saves additional results in the "additional_results.pkl" file. This includes the total number of observations in the dataset and the total number of observations used for testing.

Overall, the code performs a comprehensive analysis to determine the optimal tracheal tube depth using machine learning models and formula-based methods, providing insights into the performance of different approaches in a non-invasive manner.

C.3 Code Output

table_0.pkl

	sex	age_c	ht	wt	tube_depth_G
mean	0.538700	0.757998	66.000516	7.126687	10.189474
std	0.498758	1.440271	19.081267	4.774186	1.766052

table_1.pkl

	Squared Residues	p-value vs Height Formula	p-value vs
Age Formula	ID Formula		
Random Forest	1.590379		3.739e-14
5.828e-200	1.109e-22		
Elastic Net	1.913334		8.892e-19
3.513e-238	8.066e-23		
Support Vector Machine	1.312018		1.055e-13
1.306e-205	1.636e-22		
Neural Network	1.246383		2.692e-14
4.455e-205	1.259e-22		
Height Formula	3.418890		-
-	-		
Age Formula	90.393557		-
-	-		

ID Formula 309.559605

additional_results.pkl

```
{
    'Total number of observations': 969,
    'Total number of test observations': 194,
}
```

D LaTeX Table Design

D.1 Code

The LaTeX Table Design was carried out using the following custom code:

```
1
2 # IMPORT
3 import pandas as pd
4 from typing import Dict, Any, Tuple, Optional
5 from my_utils import to_latex_with_note, format_p_value,
   ↪ is_str_in_df, split_mapping, AbbrToNameDef
6
7 # PREPARATION FOR ALL TABLES
8 shared_mapping: AbbrToNameDef = {
9     'sex': ('Sex', 'Participant sex: 0 - Female, 1 - Male'),
10    'age_c': ('Age (years)', 'Patient age in years. Rounded to
   ↪ half years.'),
11    'ht': ('Height (cm)', 'Patient height in centimeters.'),
12    'wt': ('Weight (kg)', 'Patient weight in kilograms.'),
13    'tube_depth_G': ('Optimal Tube Depth (cm)', 'Optimal tracheal
   ↪ tube depth as determined by Chest X-ray (in cm).')
14 }
15
16 # TABLE 0
17 df = pd.read_pickle('table_0.pkl')
18
19 # RENAME ROWS AND COLUMNS
20 mapping: AbbrToNameDef = {k: v for k, v in
   ↪ shared_mapping.items() if is_str_in_df(df, k)}
```



```

21
22 abbrs_to_names, legend = split_mapping(mapping)
23 df = df.rename(columns=abbrs_to_names, index=abbrs_to_names)
24
25 # Save as latex
26 to_latex_with_note(
27     df, 'table_0.tex',
28     caption="Descriptive Statistics of Patient's Sex, Age,
29     ↪ Height, Weight and Optimal Tracheal Tube Depth.",
30     label='table:descriptive_statistics',
31     legend=legend)
32
33 # TABLE 1
34 df = pd.read_pickle('table_1.pkl')
35
36 # Format the DataFrame and p-values
37 df.rename(columns={col: col.replace('_', ' ') for col in
38     ↪ df.columns}, inplace=True)
39 df.rename(columns={
40     'Squared Residues': 'SR',
41     'p-value vs Height Formula': 'p-val HF',
42     'p-value vs Age Formula': 'p-val AF',
43     'p-value vs ID Formula': 'p-val IDF'
44 }, inplace=True)
45
46 for col in df.columns:
47     if 'p-val' in col:
48         df[col] = df[col].apply(format_p_value)
49
50 # RENAME ROWS AND COLUMNS
51 mapping: AbbrToNameDef = {k: v for k, v in
52     ↪ shared_mapping.items() if is_str_in_df(df, k)}
53 mapping |= {
54     'SR': ('SR', 'Squared Residues of the model
55     ↪ predictions.'),
56     'p-val AF': ('p-val AF', 'p-value of the Age Formula-Based
57     ↪ Model.'),
58     'p-val HF': ('p-val HF', 'p-value of the Height
59     ↪ Formula-Based Model.'),

```

```

54     'p-val IDF': ('p-val IDF', 'p-value of the ID
    ↪ Formula-Based Model.'),
55     'ID Formula': ('ID Formula', 'OTTD (in cm) = 3 * (tube ID
    ↪ [mm])')
56 }
57
58 abbrs_to_names, legend = split_mapping(mapping)
59 df.rename(index=abbrs_to_names, inplace=True)
60
61 # Save as latex
62 to_latex_with_note(
63     df, 'table_1.tex',
64     caption="Performance comparison among Machine Learning models
    ↪ and formula-based models",
65     label='table:performance_comparison',
66     legend=legend)
67
68

```

D.2 Provided Code

The code above is using the following provided functions:

```

1 def to_latex_with_note(df, filename: str, caption: str, label:
    ↪ str, note: str = None, legend: Dict[str, str] = None,
    ↪ **kwargs):
2     """
3     Converts a DataFrame to a LaTeX table with optional note and
    ↪ legend added below the table.
4
5     Parameters:
6     - df, filename, caption, label: as in `df.to_latex`.
7     - note (optional): Additional note below the table.
8     - legend (optional): Dictionary mapping abbreviations to full
    ↪ names.
9     - **kwargs: Additional arguments for `df.to_latex`.
10
11     Returns:
12     - None: Outputs LaTeX file.
13     """

```

```

14
15 def format_p_value(x):
16     returns "{:.3g}".format(x) if x >= 1e-06 else "<1e-06"
17
18 def is_str_in_df(df: pd.DataFrame, s: str):
19     return any(s in level for level in getattr(df.index,
20         ↳ 'levels', [df.index]) + getattr(df.columns, 'levels',
21         ↳ [df.columns]))
22
23 AbbrToNameDef = Dict[Any, Tuple[Optional[str], Optional[str]]]
24
25 def split_mapping(abbrs_to_names_and_definitions:
26     ↳ AbbrToNameDef):
27     abbrs_to_names = {abbr: name for abbr, (name, definition) in
28         ↳ abbrs_to_names_and_definitions.items() if name is not
29         ↳ None}
30     names_to_definitions = {name or abbr: definition for abbr,
31         ↳ (name, definition) in
32         ↳ abbrs_to_names_and_definitions.items() if definition is
33         ↳ not None}
34     return abbrs_to_names, names_to_definitions

```

D.3 Code Output

table_0.tex

```

\begin{table}[h]
\caption{Descriptive Statistics of Patient's Sex, Age, Height, Weight and
Optimal Tracheal Tube Depth.}
\label{table:descriptive_statistics}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{%
\begin{tabular}{lrrrrr}
\toprule
& Sex & Age (years) & Height (cm) & Weight (kg) & Optimal Tube Depth (cm) \\
\midrule
\textbf{mean} & 0.539 & 0.758 & 66 & 7.13 & 10.2 \\
\textbf{std} & 0.499 & 1.44 & 19.1 & 4.77 & 1.77

```

```

\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item \textbf{Sex}: Participant sex: 0 - Female, 1 - Male
\item \textbf{Age (years)}: Patient age in years. Rounded to half years.
\item \textbf{Height (cm)}: Patient height in centimeters.
\item \textbf{Weight (kg)}: Patient weight in kilograms.
\item \textbf{Optimal Tube Depth (cm)}: Optimal tracheal tube depth as
determined by Chest X-ray (in cm).
\end{tablenotes}
\end{threeparttable}
\end{table}

```

table.1.tex

```

\begin{table}[h]
\caption{Performance comparison among Machine Learning models and formula-based
models}
\label{table:performance_comparison}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{%
\begin{tabular}{lrrlll}
\toprule
& SR & p-val HF & p-val AF & p-val IDF & \\
\midrule
\textbf{Random Forest} & 1.59 &  $<1e-06$  &  $<1e-06$  &  $<1e-06$  & \\
\textbf{Elastic Net} & 1.91 &  $<1e-06$  &  $<1e-06$  &  $<1e-06$  & \\
\textbf{Support Vector Machine} & 1.31 &  $<1e-06$  &  $<1e-06$  &  $<1e-06$  & \\
\textbf{Neural Network} & 1.25 &  $<1e-06$  &  $<1e-06$  &  $<1e-06$  & \\
\textbf{Height Formula} & 3.42 & - & - & - & \\
\textbf{Age Formula} & 90.4 & - & - & - & \\
\textbf{ID Formula} & 310 & - & - & - & \\
\bottomrule
\end{tabular}}
\end{threeparttable}
\begin{tablenotes}
\footnotesize
\item \textbf{SR}: Squared Residues of the model predictions.

```

$\backslash\text{item}\ \backslash\text{textbf}\{\text{p-val AF}\}$: p-value of the Age Formula-Based Model.
 $\backslash\text{item}\ \backslash\text{textbf}\{\text{p-val HF}\}$: p-value of the Height Formula-Based Model.
 $\backslash\text{item}\ \backslash\text{textbf}\{\text{p-val IDF}\}$: p-value of the ID Formula-Based Model.
 $\backslash\text{item}\ \backslash\text{textbf}\{\text{ID Formula}\}$: OTTD (in cm) = $3 * (\text{tube ID [mm]})$
 $\backslash\text{end}\{\text{tablenotes}\}$
 $\backslash\text{end}\{\text{threeparttable}\}$
 $\backslash\text{end}\{\text{table}\}$

References

- [1] T. Kendirli, A. Kavaz, Z. Yalaki, Burcu Oztrk Himi, E. Derelli, and E. Ince. Mechanical ventilation in children. *The Turkish journal of pediatrics*, 48 4:323–7, 2006.
- [2] M. Kneyber, D. De Luca, E. Calderini, P. Jarreau, E. Javouhey, J. LpezHerce, J. Hammer, D. Macrae, D. Markhorst, A. Medina, M. Pons-dena, F. Racca, G. Wolf, P. Biban, J. Brierley, and P. Rimensberger. Recommendations for mechanical ventilation of critically ill children from the paediatric mechanical ventilation consensus conference (pemvecc). *Intensive Care Medicine*, 43:1764 – 1780, 2017.
- [3] M. Kollef, Edward J. Legare, and M. Damiano. Endotracheal tube misplacement: Incidence, risk factors, and impact of a quality improvement program. *Southern Medical Journal*, 87:248–254, 1994.
- [4] C. Traiber, J. Piva, C. C. Fritsher, P. Garcia, P. Lago, E. A. Trotta, C. Ricachinevsky, F. Bueno, V. Baecker, and Bianca D Lisboa. Profile and consequences of children requiring prolonged mechanical ventilation in three brazilian pediatric intensive care units. *Pediatric Critical Care Medicine*, 10:375–380, 2009.
- [5] V. Rajajee, J. Fletcher, L. Rochlen, and T. Jacobs. Real-time ultrasound-guided percutaneous dilatational tracheostomy: a feasibility study. *Critical Care*, 15:R67 – R67, 2011.
- [6] E. Mariano, C. Ramamoorthy, L. Chu, Michael I. Chen, and G. Hammer. A comparison of three methods for estimating appropriate tracheal tube depth in children. *Pediatric Anesthesia*, 15, 2005.

- [7] K. Takita, Y. Morimoto, and O. Kemmotsu. The height-based formula for prediction of left-sided double-lumen tracheal tube depth. *Journal of cardiothoracic and vascular anesthesia*, 17 3:412–3, 2003.
- [8] F. Foronda, E. Troster, J. Farias, Carmen S. V. Barbas, A. Ferraro, L. Faria, A. Bousso, F. Panico, and A. Delgado. The impact of daily evaluation and spontaneous breathing test on the duration of pediatric mechanical ventilation: A randomized controlled trial*. *Critical Care Medicine*, 39:2526–2533, 2011.
- [9] S. A. Ingelse, H. Wiegers, J. Calis, J. V. van Woensel, and R. A. Bem. Early fluid overload prolongs mechanical ventilation in children with viral-lower respiratory tract disease*. *Pediatric Critical Care Medicine*, 18:e106e111, 2017.
- [10] M. Steurer, S. Oltman, R. Baer, Sky K. Feuer, Liang Liang, R. Paynter, L. Rand, K. Ryckman, R. Keller, and L. J. Pawlowski. Altered metabolites in newborns with persistent pulmonary hypertension. *Pediatric Research*, 84:272 – 278, 2018.
- [11] K. Shaw, Madeline Bertha, T. Hofmekler, P. Chopra, T. Vatanen, A. Srivatsa, J. Prince, Archana P Kumar, Cary G. Sauer, M. Zwick, G. Satten, A. Kostic, J. Mülle, R. Xavier, and S. Kugathasan. Dysbiosis, inflammation, and response to treatment: a longitudinal study of pediatric subjects with newly diagnosed inflammatory bowel disease. *Genome Medicine*, 8, 2016.
- [12] Andrea L. OBoyle, R. Mink, N. Anas, and T. Kallay. Development of learning objectives and a validated testing tool for management of pediatric mechanical ventilation*. *Pediatric Critical Care Medicine*, 15:594599, 2014.
- [13] M. Weiss, A. Gerber, and A. Dullenkopf. Appropriate placement of intubation depth marks in a new cuffed paediatric tracheal tube. *British journal of anaesthesia*, 94 1:80–7, 2005.
- [14] C. Newth, K. Sward, R. Khemani, K. Page, K. Meert, J. Carcillo, T. Shanley, F. Moler, M. Pollack, H. Dalton, D. Wessel, J. Berger, R. Berg, R. Harrison, R. Holubkov, A. Doctor, J. Dean, Tammara L. Jenkins, and Carol E. Nicholson. Variability in usual care mechanical ventilation for pediatric acute respiratory distress syndrome: Time for a decision support protocol?*. *Pediatric Critical Care Medicine*, 18:e521e529, 2017.

- [15] Guangjun Li, Xiangyu Zhang, Xinyu Song, Lian Duan, Guangyu Wang, Qing Xiao, Jing Li, LanYue Liang, Long Bai, and S. Bai. Machine learning for predicting accuracy of lung and liver tumor motion tracking using radiomic features. *Quantitative Imaging in Medicine and Surgery*, 13:1605 – 1618, 2023.
- [16] Ahasanun Nessa, Bhagawat Adhikari, Fatima Hussain, and Xavier N Fernando. A survey of machine learning for indoor positioning. *IEEE Access*, 8:214945–214965, 2020.
- [17] Ren Liu, Zhonghua Liang, Kuo-Long Yang, and Wei Li. Machine learning based visible light indoor positioning with single-led and single rotatable photo detector. *IEEE Photonics Journal*, 14:1–11, 2022.