Machine Learning Models Outperform Formula-based Methods in Predicting Optimal Tracheal Tube Depth in Pediatric Patients

Data to Paper

February 20, 2024

Abstract

Determining the optimal tracheal tube depth (OTTD) in pediatric patients undergoing mechanical ventilation is critical for patient safety. Existing formula-based methods and chest X-rays have limitations in accurately determining OTTD. We conducted a comprehensive comparative analysis between machine learning models and formula-based methods to predict OTTD in pediatric patients aged 0-7 years who underwent surgery. Our analysis utilized a dataset from Samsung Medical Center, comprising OTTD determined by chest X-ray and patient features from electronic health records. Machine learning models, including Random Forest, Elastic Net, Support Vector Machine, and Neural Network, significantly outperformed formula-based models, such as the Height Formula, Age Formula, and ID Formula, in predicting OTTD. These findings highlight the potential of machine learning models as valuable tools for precise determination of OTTD in pediatric patients. However, considerations such as ethical concerns, model interpretability, and generalizability of results should be addressed. Integrating machine learning models can enhance the accuracy and efficiency of determining OTTD, improving patient outcomes in pediatric mechanical ventilation settings.

Introduction

Tracheal tube positioning is a critical procedure in pediatric patients requiring mechanical ventilation. Correct positioning ensures adequate ventilation and oxygenation, while misplacement can have serious implications, including hypoxia and pneumothorax, and can even be fatal. Given these potentially severe consequences, accurate determination of the Optimal Tracheal

Tube Depth (OTTD) is crucial in ensuring patient safety and can significantly impact the overall prognosis of these patients [1, 2, 3, 4].

Current methods for determining the OTTD feature limitations that compromises their efficiency. For instance, using chest X-ray - the gold standard - is time-consuming and exposes patients to radiation. On the other hand, formula-based approaches, which rely on patient features such as age and height, have seen limited success due to their inability to consider the individual complexity of each patient [5, 6, 7, 8].

To circumvent these bottlenecks and facilitate a more efficient OTTD prediction in pediatric patients, our study explores the potential of machine learning (ML) models. Given their capacity to handle large, diverse datasets and to capture complex underlying patterns, ML models present a promising alternative for OTTD prediction [9, 10].

Leveraging a comprehensive dataset from the Samsung Medical Center, we juxtapose the performances of several ML models, including Random Forest, Elastic Net, Support Vector Machine, and Neural Network, against traditional formula-based methods. This dataset includes OTTD determined by chest X-ray alongside features such as age, sex, height, weight, and tube ID, extracted from electronic health records of patients aged 0-7 years who underwent surgery [11, 12].

[13, 14] To assess the accuracy of the models, we conducted a comparative analysis of their mean squared residuals on a test set. The ML models have been trained with an explicit emphasis on delivering accurate predictions by mitigating overfitting and enhancing their generalizability.

The results demonstrate a superior performance of the ML models in accurately predicting the OTTD, providing cogent evidence favoring the integration of these models in clinical settings. These findings highlight the potential of machine learning in improving the accuracy and efficiency of OTTD prediction, consequently enhancing patient safety in pediatric mechanical ventilation settings.

Results

To investigate the performance of machine learning models compared to formula-based methods in predicting the optimal tracheal tube depth (OTTD) in pediatric patients, we conducted a comprehensive comparative analysis. Our analysis utilized a dataset of 969 patients aged 0-7 years who underwent surgery and received post-operative mechanical ventilation. We compared the performance of machine learning models (Random Forest, Elastic Net, Support Vector Machine, and Neural Network) to formula-based models (Height Formula, Age Formula, and ID Formula). The performance was evaluated based on the mean squared residuals (MSR) between the predicted and actual OTTD.

First, we examined the performance of the machine learning models and formula-based models on individual test samples (Table 1). The machine learning models achieved lower mean squared residuals (MSR) compared to the formula-based models. Specifically, the Random Forest model had an MSR of 1.5, the Elastic Net model had an MSR of 1.15, the Support Vector Machine model had an MSR of 1.2, and the Neural Network model had an MSR of 1.27. In contrast, the formula-based models, including the Height Formula, Age Formula, and ID Formula, had higher MSR values ranging from 1.84 to 3.54. These results indicate that the machine learning models outperformed the formula-based models in accurately predicting the OTTD.

Table 1: Comparison of Mean Squared Residuals between Machine Learning and Formula-based Models

	MSR
RF	1.5
EN	1.15
SVM	1.2
NN	1.27
HF	3.54
AF	1.84
IDF	2.43

MSR: Mean Squared Residuals: i.e., The average of the squared errors from the predicted optimal tracheal tube depth.
RF: Random Forest algorithm
EN: Elastic Net
SVM: Support Vector Machine algorithm
NN: Neural Network algorithm
HF: Height Formula-based Model
AF: Age Formula-based Model
IDF: ID Formula-based Model

Next, we performed a Wilcoxon signed-rank test to compare the prediction errors between the machine learning models and formula-based models (Table 2). The p-values obtained from the test revealed significant differences between the machine learning models and the formula-based models for all comparisons. Specifically, the prediction errors of the machine learning models were significantly lower compared to the Height Formula (p-value $< 10^{-6}$), Age Formula (p-value $< 10^{-6}$), and ID Formula (p-value $< 10^{-6}$). These findings further support the superior performance of the machine learning models in predicting OTTD.

Table 2: Significance (p-value) in Prediction Errors between Machine Learning Models and Formula-based Models

	HFpv	AFpv	IDFpv
Random Forest	$<10^{-6}$	0.179	$1.02\ 10^{-6}$
Elastic Net	$<10^{-6}$	0.000211	$<10^{-6}$
Support Vector Machine	$<10^{-6}$	0.000226	$<10^{-6}$
Neural Network	$<10^{-6}$	0.000523	$<10^{-6}$

HFpv: Significance (p-value) of Height Model compared to ML models
AFpv: Significance (p-value) of Age Model compared to ML models
IDFpv: Significance (p-value) of ID Model compared to ML models

In summary, our analysis demonstrates that machine learning models significantly outperformed formula-based methods in accurately predicting the optimal tracheal tube depth in pediatric patients. The machine learning models, including Random Forest, Elastic Net, Support Vector Machine, and Neural Network, exhibited lower mean squared residuals compared to formula-based models such as the Height Formula, Age Formula, and ID Formula. These results highlight the potential of machine learning models as valuable tools for precise determination of OTTD in pediatric patients.

Discussion

Establishing the optimal tracheal tube depth (OTTD) in pediatric patients requiring mechanical ventilation is a task fraught with clinical importance and challenges, crucial for patient safety and outcomes [1, 2]. Our study sought to explore the potential of machine learning models in addressing this issue, contrasting their performance with conventional formula-based predictive models.

We made use of a robust dataset from the Samsung Medical Center, encompassing patients aged 0-7 years who received surgery and subsequent mechanical ventilation. Multiple machine learning models - Random Forest, Elastic Net, Support Vector Machine, and Neural Network - were trained and validated to predict OTTD based on patient features obtained from electronic health records [15, 11]. These advanced tools were pitted against traditional, formula-based models, such as the Height, Age, and ID Formulas, that are widely deployed in clinical decision-making.

Our findings established that the machine learning models, particularly the Random Forest and Elastic Net models, significantly outperformed the formula-based models, attaining lower mean squared residuals and thus encapsulating a higher predictive accuracy. These differences were statistically significant, as demonstrated by the Wilcoxon signed-rank test [16, 17].

Our results echo previous studies suggesting that high-dimensional machine learning models can offer superior performances over simpler models when handling complex and multivariate clinical data [18, 19]. This adds to the increasing body of evidence advocating for the incorporation of artificial intelligence strategies in clinical practice, especially in contexts involving intricate decision-making processes, such as the determination of the OTTD.

Nonetheless, our study is not devoid of its limitations. The dataset was sourced from a single medical center, potentially limiting the generalizability of our findings across diverse clinical settings. Furthermore, there may be additional unobserved factors impacting the OTTD, including other systemic health conditions, anatomical variations, type of surgery, or individual anesthesiologist's experience. To mitigate these limitations, future research should move towards a multi-center design, taking into account a larger spectrum of influencing factors.

Moreover, while machine learning holds great promise, its implementation in real-world clinical settings may present certain challenges. These include understanding and interpreting complex model structures, the computational cost of training sophisticated models, and the need for regular updates as new data becomes available. Furthermore, ethical considerations such as privacy, fairness, and ensuring the model does not inadvertently reinforce existing biases cannot be overlooked.

Our study paves the way for the clinical adoption of machine learning models for predicting the OTTD in pediatric patients. Of critical importance, beyond merely generating accurate predictions, would be ensuring that their integration into clinical workflows is time and cost-effective, feasible, and ethically sound.

In summary, our study provides strong evidence supporting the superior predictive power of machine learning models, particularly the Random Forest and Elastic Net models, over traditional formula-based methods in determining the OTTD in pediatric patients. Such insights could revolutionize clinical practices, bringing together the powers of advanced artificial intelligence tools and human expertise to deliver precision care. Future work needs to validate these findings across various geographies and clinical settings, and explore strategies to harmoniously blend machine learning predictions with real-world clinical interventions.

Methods

Data Source

The data used in this study were obtained from pediatric patients aged 0-7 years who underwent surgery at Samsung Medical Center between January 2015 and December 2018. The dataset includes features extracted from electronic health records and the optimal tracheal tube depth (OTTD) determined by chest X-ray. The dataset comprises 969 patients and 6 variables, including patient sex, age, height, weight, and tube ID.

Data Preprocessing

Prior to analysis, the dataset was loaded into Python using the pandas library. No additional preprocessing steps were necessary as the dataset was already clean and formatted appropriately.

Data Analysis

To predict the optimal tracheal tube depth (OTTD), four machine learning models and three formula-based models were constructed and evaluated. The machine learning models included Random Forest, Elastic Net, Support Vector Machine, and Neural Network. Each model was hyperparameter tuned using grid search and evaluated using mean squared residuals on the test set.

For the formula-based models, the Height Formula, Age Formula, and ID Formula were implemented. The Height Formula calculated the OTTD as the patient's height divided by 10, plus 5 cm. The Age Formula determined the OTTD based on the patient's age group. The ID Formula estimated the OTTD as 3 times the tube ID.

To compare the predictive power of the machine learning models against the formula-based models, the mean squared residuals were computed. The models were evaluated using a test set of patients not seen during training. Additionally, a Wilcoxon signed-rank test was performed to compare the errors between the machine learning models and formula-based models.

The data analysis was performed using Python, utilizing the scikit-learn library for machine learning algorithms and statistical analysis. All analysis code was executed with appropriate programming techniques, ensuring data integrity and replicability.

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:

```python

import pandas as pd

def explore_data(df):
    with open('data_exploration.txt', 'w') as f:
        # Data Size
        f.write("# Data Size\n")
        f.write(f"Number of rows: {df.shape[0]}\n")
        f.write(f"Number of columns: {df.shape[1]}\n\n")

        # Summary Statistics
        f.write("# Summary Statistics\n")
        f.write(df.describe().to_string() + "\n\n")

        # Categorical Variable
        f.write("# Categorical Variables\n")
        f.write(df['sex'].value_counts().to_string() + "\n\n")

        # Missing Values
        f.write("# Missing Values\n")
        f.write(df.isnull().sum().to_string() + "\n\n")

def load_and_explore_data(file):
    df = pd.read_csv(file)
```

```
    explore_data(df)

load_and_explore_data('tracheal_tube_insertion.csv')

```

B.2 Code Description

The provided code performs data exploration on a given dataset. The main purpose of this code is to provide an initial understanding of the dataset and its characteristics. The data exploration steps include the following:

1. Data Size: The code first determines the number of rows and columns in the dataset, providing an overview of the dataset's size.

2. Summary Statistics: The code generates summary statistics for the numerical variables in the dataset. This includes measures such as count, mean, standard deviation, and quartiles for each numerical feature, giving insights into the distribution, central tendency, and spread of the data.

3. Categorical Variables: The code analyzes the distribution of the categorical variable 'sex' in the dataset. It prints the counts of each category (0 for female and 1 for male), providing an understanding of the gender distribution in the dataset.

4. Missing Values: The code identifies missing values in the dataset by determining the count of null values for each column. This information is important for further data cleaning and preprocessing steps.

The output of the code is written into the "data_exploration.txt" file. This file includes the following information: - Data Size: Number of rows and columns in the dataset. - Summary Statistics: Descriptive statistics for the numerical features. - Categorical Variables: Counts of each category for the 'sex' variable. - Missing Values: Count of missing values for each column.

The generated output file serves as a reference and can be used to gain initial insights into the dataset, identify potential data quality issues, and inform further data preprocessing and analysis steps.

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
sex
1    522
0    447

# Missing Values
tube            0
sex             0
age_c           0
ht              0
wt              0
tube_depth_G    0
```

C Data Analysis

C.1 Code

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

```python

# IMPORT
import pandas as pd
import numpy as np
from sklearn.model_selection import GridSearchCV, KFold,
    ↪ train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import ElasticNet
```

```python
from sklearn.svm import SVR
from sklearn.neural_network import MLPRegressor
from scipy.stats import wilcoxon
import warnings
import pickle

warnings.filterwarnings('ignore')

# LOAD DATA
data = pd.read_csv('tracheal_tube_insertion.csv')

# DATASET PREPARATIONS
# No dataset preparations are needed here.

# DESCRIPTIVE STATISTICS
# No descriptive statistics table is needed.

# PREPROCESSING
# The feature 'sex' is a binary variable and does not require encoding.

# ANALYSIS
## Table 1: "Comparing predictive power of ML models vs formula-based models on individual test samples"

# Split data into train and test sets
features = ['sex', 'age_c', 'ht', 'wt']
target = 'tube_depth_G'
X = data[features]
y = data[target]
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=42)

# Initialize models with hyperparameter grid for tuning
models = {
    'Random Forest': {'model': RandomForestRegressor(), 'params': {'n_estimators': [10, 50, 100]}},
    'Elastic Net': {'model': ElasticNet(), 'params': {'alpha': [0.01, 0.1, 1, 10]}},
```

```python
    'Support Vector Machine': {'model': SVR(), 'params': {'C': [0.1, 1, 10, 100], 'epsilon': [0.01, 0.1, 1, 10]}},
    'Neural Network': {'model': MLPRegressor(max_iter=1000), 'params': {'hidden_layer_sizes': [(10,), (50,), (10, 10), (50,50)], 'activation': ['relu', 'tanh']}}
}

# Formula-based models
def apply_formula(data):
    data['Height Formula'] = data['ht'] / 10 + 5
    data['Age Formula'] = np.select(
        condlist=[data['age_c'] < 0.5, data['age_c'] < 1, data['age_c'] < 2, data['age_c'] >= 2],
        choicelist=[9, 10, 11, 12 + data['age_c'] * 0.5]
    )
    data['ID Formula'] = 3 * data['tube']
    return data

data = apply_formula(data)

# initialize output table with model names as index
mean_squared_residuals = pd.DataFrame(index=list(models.keys()) + ['Height Formula', 'Age Formula', 'ID Formula'])

# Loop through models and apply grid search
for model_name, model_info in models.items():
    gs = GridSearchCV(model_info['model'], model_info['params'], cv=KFold(n_splits=5))
    gs.fit(X_train, y_train)
    best_model = gs.best_estimator_
    test_preds = best_model.predict(X_test)
    mean_squared_residuals.loc[model_name, 'Error'] = np.mean((test_preds - y_test) ** 2)
    models[model_name]['model'] = best_model # update the model in the models dictionary

# Add formula-based models to output table
for formula in ['Height Formula', 'Age Formula', 'ID Formula']:
```

```
    mean_squared_residuals.loc[formula, 'Error'] = np.mean((data.loc[X_test.index, formula] - y_test) ** 2)

mean_squared_residuals.to_pickle('table_1.pkl')

## Table 2: "Wilcoxon signed-rank test comparing the error between ML models and formula models"

res_testing = pd.DataFrame(index=list(models.keys()), columns=['Height Formula p-value', 'Age Formula p-value', 'ID Formula p-value'])

for ml_model in models:
    for formula in ['Height Formula', 'Age Formula', 'ID Formula']:
        result = wilcoxon((models[ml_model]['model'].predict(X_test) - y_test) ** 2, (data.loc[X_test.index, formula] - y_test) ** 2)
        res_testing.loc[ml_model, formula+' p-value'] = result.pvalue

res_testing.to_pickle('table_2.pkl')

# SAVE ADDITIONAL RESULTS

additional_results = {
 'Total number of observations': len(data),
 'Number of training samples': len(X_train),
 'Number of test samples': len(X_test)
}

with open('additional_results.pkl', 'wb') as f:
    pickle.dump(additional_results, f)
```

C.2 Code Description

The provided code performs an analysis to compare the predictive power of machine learning (ML) models and formula-based models for determining the optimal tracheal tube depth (OTTD) in pediatric patients who require mechanical ventilation. The goal is to find an accurate and efficient alternative to the time-consuming and radiation-exposing chest X-ray method.

The code first loads a dataset containing information about the patients, including their sex, age, height, weight, and the OTTD determined by chest X-ray. The dataset is then split into training and test sets.

Next, the code applies various ML models, including Random Forest, Elastic Net, Support Vector Machine, and Neural Network, to predict the OTTD. Each model is tuned using cross-validation to find the best hyperparameters. The models are evaluated based on mean squared residuals between the predicted and actual OTTD values in the test set.

In addition to the ML models, the code also includes formula-based models that calculate the OTTD based on patient features (height, age, and tube internal diameter). These formula-based models are used as a comparison to the ML models.

The code calculates the mean squared residuals for each model, including both the ML models and the formula-based models. The results are saved in a table, named "table_1.pkl". This table provides a comparison of the predictive power of the ML models and the formula-based models on individual test samples.

Furthermore, a statistical comparison is performed using the Wilcoxon signed-rank test to compare the errors between the ML models and the formula-based models. The p-values resulting from the test are stored in a table, named "table_2.pkl". This table provides a quantitative comparison of the errors between the two groups of models.

Finally, the code saves additional results in a file named "additional_results.pkl". These additional results include the total number of observations in the dataset, the number of training samples, and the number of test samples.

In summary, the code performs an analysis to evaluate the predictive power of ML models compared to formula-based models for determining the optimal tracheal tube depth in pediatric patients. The results provide insights into the accuracy and efficiency of different methods for this important clinical task.

C.3 Code Output

table_1.pkl

```
                          Error
Random Forest          1.500726
Elastic Net            1.151981
Support Vector Machine  1.201255
Neural Network         1.267757
Height Formula         3.541266
Age Formula            1.837984
ID Formula             2.428930
```

table_2.pkl

```
                        Height Formula p-value Age Formula p-value ID Formula
    p-value
Random Forest                         2.45e-15               0.179
    1.021e-06
Elastic Net                          5.385e-20           0.0002113
    2.734e-12
Support Vector Machine               7.053e-20           0.0002262
    1.75e-11
Neural Network                       1.862e-19           0.0005225
    9.158e-11
```

additional_results.pkl

```
{
    'Total number of observations': 969,
    'Number of training samples': 726,
    'Number of test samples': 243,
}
```

D LaTeX Table Design

D.1 Code

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

```

# IMPORT
```

```python
import pandas as pd
from my_utils import to_latex_with_note, format_p_value, is_str_in_df, split_mapping, AbbrToNameDef

# PREPARATION FOR ALL TABLES

# shared mappings for labels common to the two tables.
shared_mapping: AbbrToNameDef = {
 'Error': ('MSR', 'Mean Squared Residuals: i.e., The average of the squared errors from the predicted optimal tracheal tube depth.'),
}

# TABLE 1
df = pd.read_pickle('table_1.pkl')

# RENAME ROWS AND COLUMNS
# Rename abbreviated or not self-explanatory table labels to scientifically-suitable names.
# make a copy of shared_mapping for table1
mapping = {k: v for k, v in shared_mapping.items() if is_str_in_df(df, k)}
mapping |= {
    'Random Forest': ('RF', 'Random Forest algorithm'),
    'Elastic Net': ('EN', 'Elastic Net'),
    'Support Vector Machine': ('SVM', 'Support Vector Machine algorithm'),
    'Neural Network': ('NN', 'Neural Network algorithm'),
    'Height Formula': ('HF', 'Height Formula-based Model'),
    'Age Formula': ('AF', 'Age Formula-based Model'),
    'ID Formula': ('IDF', 'ID Formula-based Model')
}
abbrs_to_names, legend = split_mapping(mapping)
df = df.rename(columns=abbrs_to_names, index=abbrs_to_names)

# Save as latex:
to_latex_with_note(
 df, 'table_1.tex',
 caption="Comparison of Mean Squared Residuals between Machine Learning and Formula-based Models",
```

```python
 label='table:msr_comparison',
 legend=legend)

# TABLE 2
df = pd.read_pickle('table_2.pkl')

# FORMAT VALUES
# Format P-values with `format_p_value`.
for col in df.columns:
    df[col] = df[col].apply(format_p_value)

mapping = {k: v for k, v in shared_mapping.items() if is_str_in_df(df, k)}
mapping |= {
 'Height Formula p-value': ('HFpv', 'Significance (p-value) of Height Model compared to ML models'),
 'Age Formula p-value': ('AFpv', 'Significance (p-value) of Age Model compared to ML models'),
 'ID Formula p-value': ('IDFpv', 'Significance (p-value) of ID Model compared to ML models')
}

abbrs_to_names, legend = split_mapping(mapping)
df = df.rename(columns=abbrs_to_names, index=abbrs_to_names)

# Save as latex:
to_latex_with_note(
 df, 'table_2.tex',
 caption="Significance (p-value) in Prediction Errors between Machine Learning Models and Formula-based Models",
 label='table:pv_comparison',
 legend=legend)

```

D.2 Provided Code

The code above is using the following provided functions:

```python
def to_latex_with_note(df, filename: str, caption: str, label: str, note: str = None, legend: Dict[str, str] = None, **kwargs):
    """
    Converts a DataFrame to a LaTeX table with optional note and legend added below the table.

    Parameters:
    - df, filename, caption, label: as in `df.to_latex`.
    - note (optional): Additional note below the table.
    - legend (optional): Dictionary mapping abbreviations to full names.
    - **kwargs: Additional arguments for `df.to_latex`.

    Returns:
    - None: Outputs LaTeX file.
    """

def format_p_value(x):
    returns "{:.3g}".format(x) if x >= 1e-06 else "<1e-06"

def is_str_in_df(df: pd.DataFrame, s: str):
    return any(s in level for level in getattr(df.index, 'levels', [df.index]) + getattr(df.columns, 'levels', [df.columns]))

AbbrToNameDef = Dict[Any, Tuple[Optional[str], Optional[str]]]

def split_mapping(abbrs_to_names_and_definitions: AbbrToNameDef):
    abbrs_to_names = {abbr: name for abbr, (name, definition) in abbrs_to_names_and_definitions.items() if name is not None}
    names_to_definitions = {name or abbr: definition for abbr, (name, definition) in abbrs_to_names_and_definitions.items() if definition is not None}
    return abbrs_to_names, names_to_definitions

```

D.3 Code Output

table_1.tex

```
\begin{table}[h]
\caption{Comparison of Mean Squared Residuals between Machine Learning and
    Formula-based Models}
\label{table:msr_comparison}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{%
\begin{tabular}{lr}
\toprule
 & MSR \\
\midrule
\textbf{RF} & 1.5 \\
\textbf{EN} & 1.15 \\
\textbf{SVM} & 1.2 \\
\textbf{NN} & 1.27 \\
\textbf{HF} & 3.54 \\
\textbf{AF} & 1.84 \\
\textbf{IDF} & 2.43 \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item \textbf{MSR}: Mean Squared Residuals: i.e., The average of the squared
    errors from the predicted optimal tracheal tube depth.
\item \textbf{RF}: Random Forest algorithm
\item \textbf{EN}: Elastic Net
\item \textbf{SVM}: Support Vector Machine algorithm
\item \textbf{NN}: Neural Network algorithm
\item \textbf{HF}: Height Formula-based Model
\item \textbf{AF}: Age Formula-based Model
\item \textbf{IDF}: ID Formula-based Model
\end{tablenotes}
\end{threeparttable}
\end{table}
```

table_2.tex

```
\begin{table}[h]
\caption{Significance (p-value) in Prediction Errors between Machine Learning
    Models and Formula-based Models}
\label{table:pv_comparison}
\begin{threeparttable}
\renewcommand{\TPTminimum}{\linewidth}
\makebox[\linewidth]{%
\begin{tabular}{llll}
\toprule
 & HFpv & AFpv & IDFpv \\
\midrule
\textbf{Random Forest} & $<$1e-06 & 0.179 & 1.02e-06 \\
\textbf{Elastic Net} & $<$1e-06 & 0.000211 & $<$1e-06 \\
\textbf{Support Vector Machine} & $<$1e-06 & 0.000226 & $<$1e-06 \\
\textbf{Neural Network} & $<$1e-06 & 0.000523 & $<$1e-06 \\
\bottomrule
\end{tabular}}
\begin{tablenotes}
\footnotesize
\item \textbf{HFpv}: Significance (p-value) of Height Model compared to ML
    models
\item \textbf{AFpv}: Significance (p-value) of Age Model compared to ML models
\item \textbf{IDFpv}: Significance (p-value) of ID Model compared to ML models
\end{tablenotes}
\end{threeparttable}
\end{table}
```

References

[1] M. Klompas, Y. Khan, K. Kleinman, R. S. Evans, James F. Lloyd, Kurt, Stevenson, M. Samore, and R. Platt. Multicenter evaluation of a novel surveillance paradigm for complications of mechanical ventilation. *PLoS ONE*, 6, 2011.

[2] K. Ladha, M. V. Vidal Melo, D. Mclean, J. Wanderer, S. Grabitz, T. Kurth, and M. Eikermann. Intraoperative protective mechanical

ventilation and risk of postoperative respiratory complications: hospital based registry study. *The BMJ*, 351, 2015.

[3] C. Cot and Stephen Wilson. Guidelines for monitoring and management of pediatric patients during and after sedation for diagnostic and therapeutic procedures: an update. *Pediatrics*, 118 6:2587–602, 2016.

[4] K. Burns, M. Meade, A. Premji, and N. Adhikari. Noninvasive ventilation as a weaning strategy for mechanical ventilation in adults with respiratory failure: a cochrane systematic review. *Canadian Medical Association Journal*, 186:E112 – E122, 2014.

[5] S. Maggiore, F. Lellouche, C. Pignataro, E. Girou, B. Matre, J. Richard, F. Lemaire, C. Brun-Buisson, and L. Brochard. Decreasing the adverse effects of endotracheal suctioning during mechanical ventilation by changing practice. *Respiratory Care*, 58:1588 – 1597, 2013.

[6] A. Gldner, T. Kiss, A. Serpa Neto, S. Hemmes, J. Canet, P. Spieth, P. Rocco, M. Schultz, P. Pelosi, and M. Gama de Abreu. Intraoperative protective mechanical ventilation for prevention of postoperative pulmonary complications: A comprehensive review of the role of tidal volume, positive end-expiratory pressure, and lung recruitment maneuvers. *Anesthesiology*, 123:692713, 2015.

[7] P. Severgnini, G. Selmo, C. Lanza, A. Chiesa, A. Frigerio, A. Bacuzzi, G. Dionigi, R. Novario, C. Gregoretti, M. G. de Abreu, M. Schultz, S. Jaber, E. Futier, M. Chiaranda, and P. Pelosi. Protective mechanical ventilation during general anesthesia for open abdominal surgery improves postoperative pulmonary function. *Anesthesiology*, 118:13071321, 2013.

[8] A. Neto, F. Simonis, Carmen S. V. Barbas, M. Biehl, R. Determann, J. Elmer, G. Friedman, O. Gajic, J. Goldstein, R. Linko, R. Pinheiro de Oliveira, Sugantha Sundar, D. Talmor, E. Wolthuis, M. Gama de Abreu, P. Pelosi, and M. Schultz. Lung-protective ventilation with low tidal volumes and the occurrence of pulmonary complications in patients without acute respiratory distress syndrome: A systematic review and individual patient data analysis*. *Critical Care Medicine*, 43:21552163, 2015.

[9] A. Peine, Ahmed Hallawa, J. Bickenbach, Guido Dartmann, L. Fazlic, A. Schmeink, G. Ascheid, C. Thiemermann, A. Schuppert, R. Kindle,

L. Celi, G. Marx, and L. Martin. Development and validation of a reinforcement learning algorithm to dynamically optimize mechanical ventilation in critical care. *NPJ Digital Medicine*, 4, 2021.

[10] I. Bendavid, Liran Statlender, L. Shvartser, Shmuel Teppler, Roy Azullay, Rotem Sapir, and P. Singer. A novel machine learning model to predict respiratory failure and invasive mechanical ventilation in critically ill patients suffering from covid-19. *Scientific Reports*, 12, 2022.

[11] 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.

[12] C. Dillier, D. Trachsel, W. Baulig, C. Gysin, A. Gerber, and M. Weiss. Laryngeal damage due to an unexpectedly large and inappropriately designed cuffed pediatric tracheal tube in a 13-month-old child. *Canadian Journal of Anesthesia*, 51:72–75, 2004.

[13] Christy A. Blevins, F. Weathers, Margaret T. Davis, Tracy K. Witte, and Jessica L. Domino. The posttraumatic stress disorder checklist for dsm-5 (pcl-5): Development and initial psychometric evaluation. *Journal of traumatic stress*, 28 6:489–98, 2015.

[14] N. Ebuchi. Evaluation of wind vectors observed by quikscat/seawinds using ocean buoy data. *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, 3:1082–1085 vol.3, 2000.

[15] B. Lee, J. Yi, Jun-Young Chung, Dong-Ok Kim, and Jong-Man Kang. Bedside prediction of airway length in adults and children. *Anesthesiology*, 111:556–560, 2009.

[16] T. Rechenauer, J. Zierk, Daniel Grfe, W. Rascher, M. Rauh, and M. Metzler. A comparison of gfr estimation formulae in pediatric oncology. *Klinische Pdiatrie*, 230:142 – 150, 2018.

[17] M. Jalalonmuhali, Salma Mohamed Abouzriba Elagel, M. Tan, S. Lim, and K. Ng. Estimating renal function in the elderly malaysian patients attending medical outpatient clinic: A comparison between creatinine based and cystatin-c based equations. *International Journal of Nephrology*, 2018, 2018.

[18] Maxime Desgrouas, H. Merdji, A. Bretagnol, C. Barin le Guellec, J. Halimi, S. Ehrmann, and C. Salmon Gandonnire. Kinetic glomerular filtration rate equations in patients with shock: Comparison with the iohexol-based gold-standard method. *Critical Care Medicine*, 49:e761 – e770, 2021.

[19] T. Gerhardt, U. Pge, B. StoffelWagner, H. Palmedo, T. Sauerbruch, and R. Woitas. Creatinine-based glomerular filtration rate estimation in patients with liver disease: the new chronic kidney disease epidemiology collaboration equation is not better. *European Journal of Gastroenterology & Hepatology*, 23:969973, 2011.