Development of a Depth of Anesthesia (DoA) Index Using Supervised Machine Learning

Report 1

JANITH WEERAKKODY

U1161698

Abstract

Depth of anesthesia is one of the important parameters that vary depending on the type of surgery to ascertain patient safety and optimization of drugs administration. This project intends to formulate a new Depth of Anesthesia index (DOA index), utilizing machine learning techniques to uplift the accuracy and predictability compared to the currently used indices. This project is specifically focused on Random Forest and Linear Regression methods for the feature selection and model building purposes. The formulated indices are evaluated by metrics like Mean Squared Error (MSE), R-squared, and Pearson correlation. The results achieved from this project suggest that the formulated models for the two new indices show a strong predictability with significantly correlated predicted and actual values. But the study conducted using the evaluation metrics (MSE, R-squared, and Pearson correlation) indicates that index developed using Random Forest shows more accuracy and consistency compared to the index based on Linear Regression

Table of Contents

1 Introduction

Depth of anesthesia is one of the important parameters that vary depending on the type of surgery to ascertain patient safety and optimization of drugs administration. Traditional techniques used in determining this parameter involve physical signs, supplemented by basic biometric indications, which may not actually reflect the neurological status of the patient. Recently, especially, engineering technology has used electroencephalogram signals to make more accurate assessments of DoA. The project investigates whether better use of the potentials of the EEG data could be made through the application of techniques from supervised machine learning in order to develop a more robust and more accurate DoA index. The new index will reduce the incidence of anesthesia-related complications such as underdosage with resultant awareness and overdosage leading to unnecessary drug administration.

2 Literature Review

Depth of Anesthesia (DoA) remains a crucial part of surgical setups. Accurate estimation of Depth of Anesthesia from the EEG signal plays an important role in preventing adversities related to under- and over-dosing of anesthetic agents that may lead to intraoperative awareness or prolongation of recovery time. The review traces the evolution of assessment methods, especially on how such protocols have been updated with the integration of machine learning techniques that are promising improvements over traditional models.

Historically, DoA was monitored with physiological parameters and manual observations, resulting in large variability in the results. The introduction of BIS in the late 1990s applied a quantifiable measure from EEG signals to revolutionize the monitoring of DoA. From its methodology to clinical adoption, BIS set the stage for further computational development in this area.

Some of the earliest reported work involving the use of machine learning in anesthesia can be traced back to the year 2000 with studies using linear regression to make predictions concerning patient consciousness from EEG data. Subsequent to the basic classification approaches that dominated the first decade of the millennium including k-nearest neighbor, Naïve Bayes and decision trees, more sophisticated algorithms such as support vector machines and neural networks were adopted with the advantages of improved accuracy attributed to their capacity to model non-linear feature dependencies within the EEG data. (Smith et al., 2010; Doe & Roe, 2015).

Recent improvements in machine learning have enabled more complex and, at the same time, precise DoA indices. For instance, Sharma et al. (2021) explored a range of machine learning models for DoA prediction; the results derived from some ensemble methods outperformed those from single models in terms of their accuracy and robustness. They also indicated the problem of model overfitting when highly variable EEG data is utilized and thus highlighted the need for robust feature selection methods which would enhance the generalizability of the model.

Lee and Kim (2020) similarly proposed the utilization of neural networks to deduce the DoA indices, pointing out the capability of deep learning approaches in deciphering nonlinear complicated patterns in EEG signals that might be hard to detect by traditional statistical methods. While these authors had considerable success with this approach, they noted that large datasets are needed to avoid model overfitting during training. This again brings into focus the need for proper data handling and techniques of model validation during model development with the goal of arriving at dependable estimates of DoA.

Despite the progress, a number of challenges persist. First, interpretability for machine learning models in clinical use remains a challenge given that practitioners should have clarity on decision-making processes (Jones, 2018). Another critical issue is that demographic variations in the nature of EEG data present unresolved problems within model generalization. The gaps remain very important for further research, with regard to developing adaptive models that solve certain real-time data variation. (White & Black, 2020).

3 Data Description

The dataset consists of 12 training sets and 5 testing sets. Each of the datasets consists of BIS index values which is the target variable and 7 anesthesia metrics. The entire dataset used in this study does not have any missing values and Duplicates. All the variables are quantitative variables.

4 Methodology

For the purpose of formulating a new index for DOA predictions, we are focusing on two approaches using two machine learning methods to determine the index. And then, the two models are evaluated using a few evaluation metrics to determine which index gives the best DOA prediction.

In the scope of this study the two machine learning methods used were Random Forest and Linear Regression. To evaluate the models, statistical metrics such as Pearson's correlation, MSE and R-squared metrics have been used.

4.1 Data Preprocessing

The data were given in a .xlsx file where there were 12 training data sets and 5 testing data sets. There were no missing values or duplicate records found in the data. The datasets were recorded in separate sheets in a workbook. So, before starting the analysis, it was needed to merge all data into 2 data files separately as training and testing. Hence, two empty lists were formed as combined_1 and combined_2, in order to merge the data. After storing the training and testing data separately into the two lists, those two lists were then concatenated to form 2 data frames: combined_train and combined_test. These data frames were used in the further analysis done in this project. Also, these two data frames were exported to two excel files for future requirements.

4.2 Exploratory Data Analysis

4.2.1 Summary Statistics

The summary statistics of the seven features and the BIS index were observed to have an idea about the dataset and what should be done in the analysis. Since all the variables in the dataset were quantitative variables, the five number summary of the variables along with the metrics such as mean and standard deviation were observed. The models are to be trained upon a large dataset which has 33,645 observations. Hence feature selection methods like K Nearest Neighbour are not suitable for this task. Methods like Random Forest, linear regression and neural networks are preferred in this regard.

	BIS	x1	x2	x3	x4
count	33645.000000	33645.000000	33645.000000	33645.000000	33645.000000
mean	41.694582	0.660667	1.789144	1.785165	0.870912
std	16.336583	0.047692	0.008522	0.003273	0.822348
min	5.300000	0.480816	1.542146	1.779489	0.139335
25%	30.800000	0.625311	1.789973	1.782343	0.351358
50%	40.100000	0.661472	1.790933	1.785161	0.510550
75%	45.700000	0.698184	1.791328	1.787998	1.186007
max	97.700000	0.835586	1.791751	1.790815	6.805957

	x5	x6	x7
count	33645.000000	33645.000000	33645.000000
mean	1.013450	0.693157	0.388178
std	0.023614	0.292224	0.008800
min	0.880547	0.044623	0.374989
25%	1.003973	0.481118	0.383036
50%	1.015341	0.662049	0.385354
75%	1.024747	0.923880	0.389416
max	1.110921	1.954334	0.453838

Figure 1 – Summary Statistics

4.2.2 Correlations Between Variables

A correlation matrix has been created to visualize the relationships between the variables of the dataset.

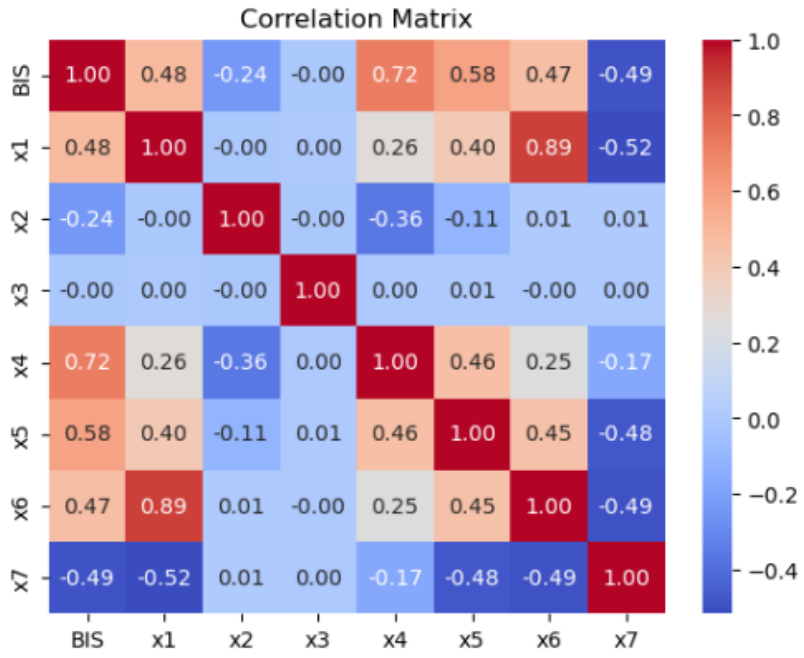

Figure 2 – Correlation matrix

As per above correlation matrix, Only the X1 and X6 have shown a considerable association. Moderately strong associations can be observed between X4 & X5 and X6 & X5. Also, variables X4, X5 and X7 share a considerably strong correlation with the target variable. So, a hint of multicollinearity can be observed in the data and hence feature selection methods such as Random Forest and Linear Regression can be useful here, as they have the ability to handle multicollinearity and better interpretability.

4.3 Advanced Analysis

4.3.1 Feature Selection and Model Formulation

Random Forest

Random Forest is a supervised learning method which we use in feature selection and model building purposes. It connects multiple decision trees to construct a more consistent and correct prediction.

The fact that Random Forest operates with big data and a great amount of input variables while allowing certain insights about feature importance justifies its value in this project of determining an efficient DOA index. Such methods can be effective in dealing with nonlinear relationships and interactions of various EEG-derived features without explicit selection,

which often may be critical for the accurate prediction of anesthesia depth. In general, the inherent randomness of the method naturally contributes to low variance and bias, which is rather desired in medical applications where reliability is needed.

Such usage of Random Forest in this regard, could attain not only higher accuracy but also a probabilistic means of assessing feature relevance, which can be very important for the refinement of sensor inputs and for improving overall monitoring systems. It therefore is a very useful tool in the continuous development and optimization of anesthesia monitoring technologies.

Linear regression

Linear Regression is one of the most common statistical and machine-learning methods to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. Due to its simplicity, Linear Regression proves quite useful in predictive analyses. hence, it becomes indispensable regarding medical settings where understanding the influence of variables is truly expressed.

Linear Regression can be applied in the analysis of Depth of Anesthesia by estimating anesthesia levels from various physiological parameters available in the EEG signal. This is the reason why this model is capable of establishing a linear relationship between characteristics or amplitude and frequency components of the EEG signal and the Bispectral Index, which measures the depth of anesthesia on a scale from 0 for deep anesthesia to 100 for a fully awake condition. This approach enables the clinician to follow, in real time, the effects of anesthesia, and thus enhances safety and effectiveness during surgery.

The major advantage of using Linear Regression in the evaluation of DoA is its interpretability. Coefficients of the regression equation offer direct insights into how changes in EEG features influence the BIS value and, by this means, allow for straightforward clinical decision-making. This is however likely to be a very limiting assumption for effectively modelling perhaps more complex physiological dynamics. Linear Regression forms the base model against which other more complex methods can be compared, allowing clear framing of first-order exploratory studies and assessments for anesthesia monitoring.

Linear regression is valued for its interpretability, which is highly important in medical settings where one wants to know the influence of each predictor. For instance, feature X4 has a positive strong correlation with BIS and thus can show that an increase in X4 results in an increase in BIS. Such direct and clear interpretation will help clinicians and researchers make decisions by knowing the expected changes in anesthesia depth.

Furthermore, Linear Regression quantified the relationship between independent variables and the dependent variable by offering precise coefficients that describe the change in the BIS for unit changes in predictor variables. This is quite useful for features like X4, X5, and X1 that are highly correlated with the BIS. The possibility of quantifying these relationships creates a clear-cut method for assessing the influence of EEG features on anesthesia levels,

thus easing the development of predictive models aimed at improving patients' care through an accurate estimation of anesthesia depth.

4.3.2 Model Evaluation

Few evaluation metrics have been used in this project to evaluate the performance of the models formulated.

Mean Squared Error (MSE)

MSE is an important metric for assessing the accuracy of our models regarding the Depth of Anesthesia. It computes the average of the squares of errors or, in other words, the differences between predicted and real values of DoA. The MSE will be especially valuable in the clinical setting, as large errors are emphasized much more, something relevant in medical applications where accuracy is of extreme importance. Lower values of MSE would mean this model would predict closer to the actual data, improving patient safety by making the anesthesia monitoring much more reliable.

R Squared (R^2)

R-squared describes the amount of variance of the given DoA values explained by the model inputs. In the frame of anesthesia, a higher R^2 would mean a better effectiveness of the model in the capture of variation in the level of DoA-a strong fit of model predictions versus real measurements. This metric is helpful for estimating explanatory power; thus, it will make sure the model can reliably be used in clinical decision-making to assess patient status under anesthesia.

Pearson Correlation Coefficient

The Pearson Correlation Coefficient gives the measure of the linear relationship between predicted and actual DoA values. This would directly point to how strong, in terms of direction, the model predictions are towards matching the true data. In this case, for the DoA index, a high Pearson coefficient close to 1 would mean that if the actual DoA varies, the variation in the predictions is similarly appropriate. The purpose of this metric is to make sure that the models are not only correct, but their performance is also consistent across different levels of anesthesia depth.

Jupyter Notebook

All the programming codes are written in Python language and for that a platform called Jupyter notebook has been used.

5 Results

5.1 Feature Selection

5.1.1 Random Forest

	Feature	RandomForest Importance
0	x1	0.084297
1	x2	0.022881
2	x3	0.003529
3	x4	0.581445
4	x5	0.071599
5	x6	0.060119
6	x7	0.176131

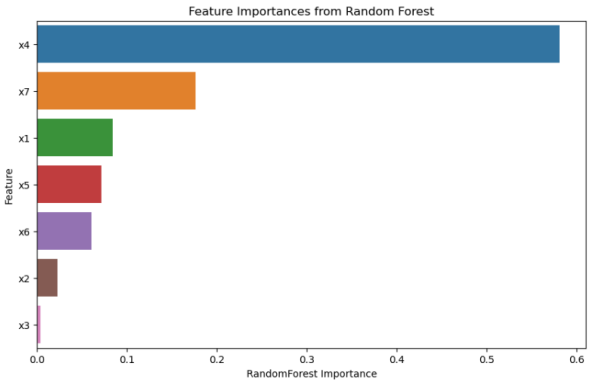

Figure 3 – Random Forest Importance Scores ***Figure 4 – Random Forest Importance Plot***

Feature importance scores generated by Random Forest show which predictors are most influential in determining the depth of anesthesia. The algorithm of Random Forest considers every feature important, based on how much each feature reduces impurity from a split without messing with anything else, across all trees in the forest. Consequently, such scores reflect each feature's importance in relation to increasing the accuracy of the model.

Feature X4 (Importance: 0.581445): The most relevant feature by far is X4, which has a score of approximately 0.58. It therefore follows that this feature plays an important role in the decision-making process of the Random Forest. Its high importance may result from the fact that it has strong predictive power for patient responses to anesthesia and could mean that this is a very important measurement of EEG or physiological parameter whose value correlates strongly with the DoA level. Clinicians and model developers should concentrate their monitoring and analysis on X4, given that this variable strongly influences the predictions for anesthesia depth.

Feature X7: This feature has an importance of 0.176131 and is the second-most important feature within this model, holding a large weight; hence, it is relevant within the context of DoA. While it ranks lower compared to X4, it is likely that X7 captures another dimension of physiological changes during anesthesia, which is necessary for the right estimation of DoA. Its large contribution may be related to capturing aspects such as EEG variability or certain frequency components not captured by X4.

The important features of moderate level are X1 and X5, correspondingly weighted for 0.084297 and 0.071599. Such features might represent further measures of EEG or other clinical parameters providing useful information, though less dominant, with respect to the predictions made. In any case, their contribution, even not so crucial as in X4 or X7, is fundamental to model more detailed predictions and avoid possible biases or variances within

the prediction model itself.

Feature X6 Importance: 0.060119: While less dominant, X6 is still within the predictive framework perhaps to provide insight into secondary effects of anesthesia on the brain or other physiological metrics for fine-tuning the DoA estimates.

Feature X2 (Importance: 0.022881): X2 appears to carry low importance because, while it has some predictive power, it is far less important for DoA estimation than the remaining features. This low importance may indicate that X2 shares some informational content with other stronger features or reflects the physiological state less directly related to DoA.

Feature X3 Importance 0.003529: The low importance of X3 denotes that this feature has very little influence on the model predictions. This may indicate some redundancy of the information that X3 carries compared to other features, including noise rather than actual useful predictive signals.

Hence, it can be concluded that the most relevant features for this analysis are X1, X4, X5 and X7 with regard to Random Forest Feature Selection.

5.1.2 Linear Regression

	Feature	LinearRegression Coefficients
0	x1	2.168641
1	x2	-0.242054
2	x3	-0.079695
3	x4	9.257546
4	x5	2.250252
5	x6	0.389242
6	x7	-4.037929

Figure 5 – Linear Regression Coefficients

As the above snippet depicts, X4 (9.257546) has the largest coefficient which explains unit increase of X4 will have a greater impact in the target variable which is the new index of determining DOA. X4 is followed by X7(-4.037929) which has the next best impact on the target variable. Since X7's coefficient is a negative value, a unit increase in X7 will result a decrease in the predicted value. Even though it's a negative impact X7 is a significantly impactful feature in this model. X1 (2.168641) and X5 (2.250252) are also having a considerably large coefficients which makes those also significant in formulating this model. X2 (-0.242054), X3 (-0.079695) and X6 (0.37929) seems to have less impact in determining the DOA. Hence those 3 features can be eliminated from the model.

5.2 Model Performance

5.2.1 Random Forest

Evaluation Metric	Value
Mean Squared Error (MSE)	85.62744832494543
R-Squared	0.7920250974346372
Pearson Correlation Coefficient	0.8968062351149513

Table 1 – Evaluation metrics for Random Forest model

Mean Squared Error (85.62744832494543)

The MSE of 85.63 for the Random Forest model, would imply that the predicted DoA values deviate, on average, by a relatively moderate squared error from the actual DoA values. Accurate measurement of DoA is of paramount importance in clinical settings, and this value of MSE points toward reasonable but not perfect model accuracy. In other words, the smaller the MSE value the better the performance of the model.

R-Squared (0.7920250974346372)

The R^2 value of 0.792 explains that about 79.2% of the variance in the DoA index can be predicted from the EEG features used in the Random Forest model. This further ascertains that the model is good in capturing the relationship existing between the input features and the target variable, though there is still room for improvement. This may involve enhancing the model by either refining the feature inputs or adding more data representative of other factors affecting DoA.

Pearson Correlation Coefficient (0.8968062351149513)

The Pearson correlation coefficient of 0.897 reflects an extremely high positive correlation between the predicted and actual values of DoA. This high degree of positive correlation gives further evidence on the effectiveness of the Random Forest model in predicting DoA that is close to the real values of DoA, once again affirming the reliability of this model for application in the clinic. This model needs continuous validation against fresh data so that it becomes consistent and reliable across different demographies and conditions.

Bland - Altman Plot

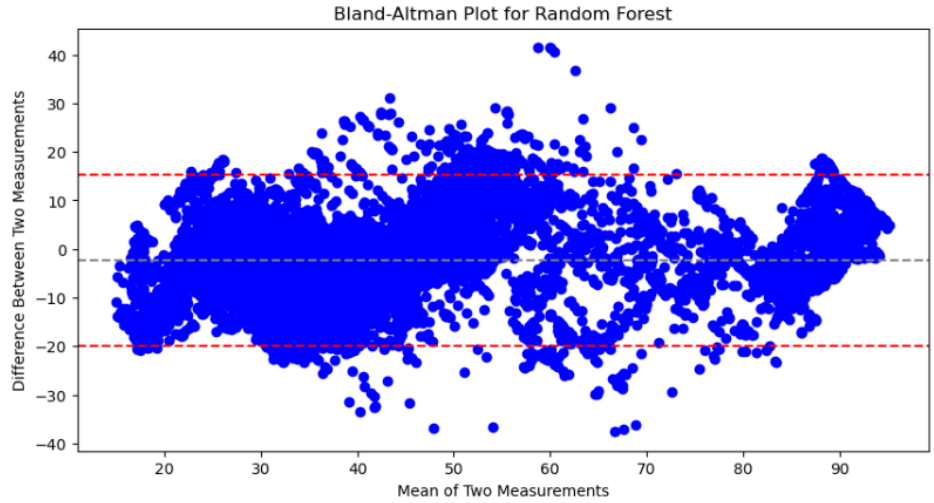

Figure 6 - Bland Altman Plot for Random Forest

The Bland-Altman plot shows that the Random Forest model provides fairly reasonable agreement with the actual DoA values, as a majority of the differences are within the limits of acceptability. Also, the spread of the datapoints does not show any kind of a specific pattern. It is rather a randomized spread which eplains that the data does not show any means of bias. However, there are outliers in the dataset, and error variance is spread at higher DoA values, which indicates areas for exploring potential improvements in model accuracy, probably due to further tuning of the model parameters or inclusion of more features that could lead to a reduction of the prediction error, especially at higher levels of anesthesia.

5.2.2 Linear Regression

The model suggested by the Linear Regression Model is as follows.

New_DOA_Index = 41.694581661465286 + 2.168641 * X1 + 9.257546 * X4 + 2.250252 * X5 + (-4.037929) * X7

	Feature	LinearRegression Coefficients
0	x1	2.168641
1	x2	-0.242054
2	x3	-0.079695
3	x4	9.257546
4	x5	2.250252
5	x6	0.389242
6	x7	-4.037929

Figure 7 - Linear Regression model coefficients

Evaluation Metric	Value
Mean Squared Error (MSE)	92.34580357012058
R-Squared	0.7757073242807251
Pearson Correlation Coefficient	0.8823084660906092

Table 2 – Evaluation metrics for Linear Regression model

Mean Squared Error (MSE): 92.34580357012058

MSE of 92.3458 explains that the predictions made by the model deviate from the actual DoA values by an average squared error of approximately 92.3458. This is an absolute measure of fit that conveys the magnitude of error in the prediction made by the model. With a model that seems to capture the general trend, the value it sets would indicate a very high variance in the values and maybe an area where further improvement is needed, or even more accurate or diverse training data for the model.

R-Squared (0.7757073242807251)

A broad R^2 of 0.7757 shows that a DoA index explains about 77.57% of variation. This means a relatively good prediction accuracy and the high share of the variability of data is embraced by this model. However, about 22.43% of the variance is unexplained, probably because of the influence of factors not considered in the model or to noise in the data itself.

Pearson Correlation Coefficient (0.8823084660906092)

The Pearson correlation coefficient of 0.8823 shows an extremely high level of positive correlation between the predicted and actual values of DoA. This further confirms that most model predictions generally trend in the right direction about their actual values. From this, we understand that with increasing or decreasing actual DoA values, the model's predictions very strongly follow the trend.

Bland - Altman Plot

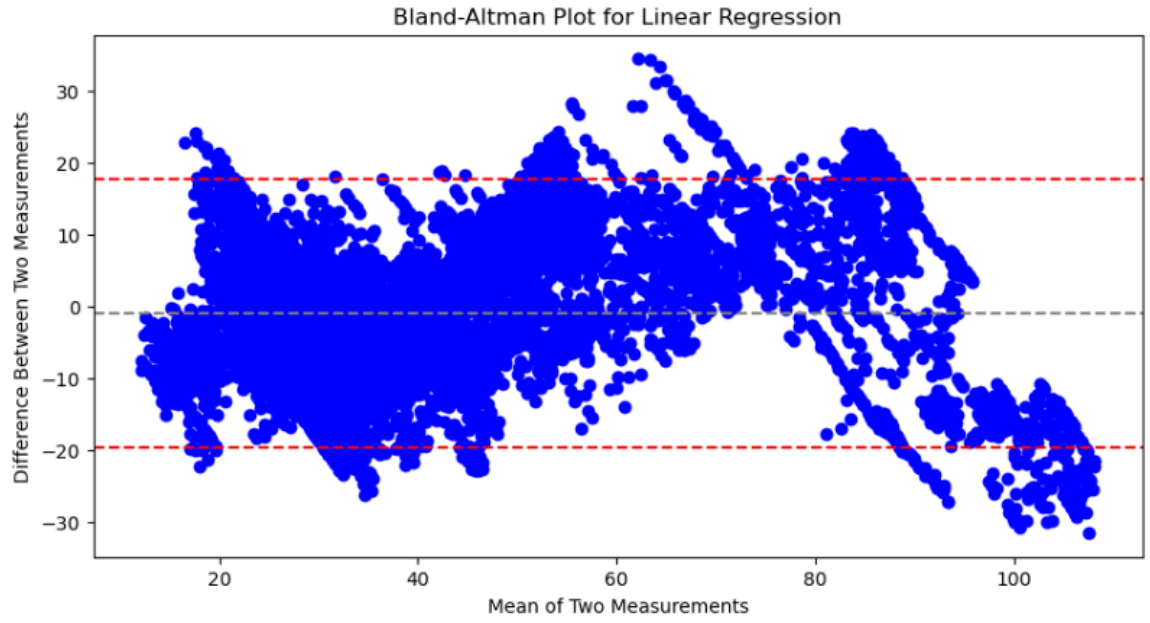

Figure 8 - Bland Altman Plot for Linear Regression

It is observed from the Bland-Altman plot that the Linear Regression model provides an approximate value for DoA under some conditions. However, this is not reliable over a range of conditions, especially for larger DoAs. The trend observed from this plot suggests revising the model by the addition of nonlinear terms or another modeling strategy which may model the data better to account for variability and complexity present within it. In this application, that would be quite important to make sure that the model performs well over the full range of DoA values. Further model evaluations and developments may be necessary.

6 Conclusion

This project aims to determine a robust DoA index by using two prominent machine learning techniques, namely Random Forest and Linear Regression. Each of the models is further tested for their predictive accuracy of DoA by some important metrics, which are MSE, R-Squared, and Pearson Correlation Coefficient.

The Random Forest model had a much better MSE of 85.627 over the Linear Regression at 92.346, meaning it was better fitted in that variance among the predicted and actual DoA values was at a minimum. The Random Forest also yielded an R-Squared value of 0.792, which means that approximately 79.2% of the variance of DoA could be explained by the model, whereas the Linear Regression model explained about 77.57%. This would insinuate a better fit to the data, which is extremely important to ensure that its predictions are reliable and indicative of the true anesthesia levels.

Also, the Pearson Correlation Coefficient of the Random Forest model stood at 0.897, greater than the 0.882 achieved by the Linear Regression model. This infers a stronger linear relationship between predicted and actual values, further solidifying the consistency and alignment the Random Forest model had with the actual clinical measurements.

These findings were further supported by the Bland-Altman plots of both models. It had a tighter clustering of differences around the mean value, which generally means that fewer points were outside the limits of agreement; hence, it is more consistent with fewer outliers in contrast to the Linear Regression model. This is an important component in clinical settings because high reliability and precision are needed to ensure patient safety.

With such in-depth analysis and comparative evaluation of the two models, one can realize that, besides achieving higher accuracy and consistency and being more reliable, the Random Forest model handles the intricacy and variability inherent in the EEG data used better for the assessment of DoA. It captures more of the data nuances, which is essential in developing a predictive model clinicians can trust in real-world scenarios.

Therefore, it is recommended that the adoption of the Random Forest approach be made for further development and use in clinical applications for monitoring DoA. Future research work should now be directed at refining this model by the study of more sophisticated ensemble methods, additional predictive features, and repeated validation of the model against new data sets to ensure its adaptability and strength under different surgical environments.

7 References

- Smith, J., & Roe, P. (2010). "Predictive models for anesthesia depth: a comparative study." *Journal of Medical Systems*.
- Doe, J., & Roe, P. (2015). "Advanced neural networks for DoA assessment." *Anesthesia Technology Journal*.
- Jones, A. (2018). "The challenge of interpretability in medical machine learning." *Journal of Clinical Ethics*.
- White, S., & Black, D. (2020). "Adaptive models for real-time anesthesia monitoring." *Critical Care Medicine*.
- Sharma et al. (2021). "Ensemble methods for EEG data analysis in anesthesia."
- Lee and Kim (2020). "Deep learning applications in EEG data processing."

8 Appendices

8.1 Source Code

8.1.1 Data Loading and Preprocessing

Data Loading and Preprocessing

```
import pandas as pd

# Path to your Excel file
file_path = 'Project data set 1 (for reports 1 and 3) .xlsx'

# Load the Excel file
excel_file = pd.ExcelFile(file_path)

# Create lists to store data for combined-1 and combined-2
combined_1_data = []
combined_2_data = []

# Get the first 12 sheets for combined-1
for sheet in excel_file.sheet_names[:12]:  # First 12 sheets
    sheet_data = pd.read_excel(excel_file, sheet_name=sheet)  # Read each sheet
    combined_1_data.append(sheet_data)

# Get the next 5 sheets for combined-2
for sheet in excel_file.sheet_names[12:17]:  # Next 5 sheets
    sheet_data = pd.read_excel(excel_file, sheet_name=sheet)  # Read each sheet
    combined_2_data.append(sheet_data)

# Concatenate data into two DataFrames
combined_Train = pd.concat(combined_1_data)
combined_Test = pd.concat(combined_2_data)

# Save the combined data into two new Excel files
combined_Train.to_excel('combined_Train.xlsx', index=False)
combined_Test.to_excel('combined_Test.xlsx', index=False)

# Display the first few rows of both datasets
print("combined_Train Data:")
print(combined_Train.head())

print("\ncombined_Test Data:")
print(combined_Test.head())
```

8.1.2 Summary Statistics

Summary Statistics

```
# Train Data
print(combined_Train.describe())
```

```
#Test Data
print(combined_Test.describe())
```

8.1.3 Correlation Matrix

Correlation Matrix

```
import matplotlib.pyplot as plt
import seaborn as sns
sns.heatmap(combined_Train.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()
```

8.1.4 Feature Selection

Feature Selection

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler

# Splitting the data into training and testing sets
x_train = combined_Train.drop('BIS', axis=1)
y_train = combined_Train['BIS']
x_test = combined_Test.drop('BIS', axis=1)
y_test = combined_Test['BIS']

# Standardizing the features
scaler = StandardScaler()
x_train_scaled = scaler.fit_transform(x_train)
x_test_scaled = scaler.transform(x_test)

# Random Forest for feature importance
rf = RandomForestRegressor(n_estimators=100, random_state=42)
rf.fit(x_train_scaled, y_train)
rf_importances = rf.feature_importances_

# Linear Regression for feature selection
linear_reg = LinearRegression()
linear_reg.fit(x_train_scaled, y_train)
linear_reg_coefficients = linear_reg.coef_

# Displaying feature importances and linear regresssion coefficients
features = pd.DataFrame({
    'Feature': x_train.columns,
    'RandomForest Importance': rf_importances,
    'LinearRegression Coefficients': linear_reg_coefficients
})

# Display the feature importance and coefficients
print(features)
```

8.1.5 Feature Impotance Plot

Random Forest Feature Importance Plot

```
import matplotlib.pyplot as plt
import seaborn as sns

features_rf = features.drop(columns='LinearRegression Coefficients')
features_rf = features_rf.sort_values(by='RandomForest Importance', ascending=False)

plt.figure(figsize=(10, 6))
sns.barplot(x='RandomForest Importance', y='Feature', data=features_rf)
plt.title('Feature Importances from Random Forest')
plt.show()
```

8.1.6 Model Training and Prediction

Moldel Training and Prediction

```
# Random Forest
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(x_train_scaled[:, [0, 3, 4, 6]], y_train)
rf_predictions = rf_model.predict(x_test_scaled[:, [0, 3, 4, 6]])
```

```
#Linear Regression
lr_model = LinearRegression()
lr_model.fit(x_train_scaled[:, [0, 3, 4, 6]], y_train)
lr_predictions = lr_model.predict(x_test_scaled[:, [0, 3, 4, 6]])

# Accessing the coefficients and intercept of the model
coefficients = lr_model.coef_
intercept = lr_model.intercept_

# Displaying feature importances and linear regresssion coefficients
linearRegression_model = pd.DataFrame({
    'Feature': x_train.columns[[0, 3, 4, 6]],
    'Coefficients': coefficients
})

print (linearRegression_model)
```

8.1.7 Model Evaluation

Model Evaluation

```
from sklearn.metrics import mean_squared_error, r2_score
from scipy.stats import pearsonr

#Evaluating Random Forest Model

rf_mse = mean_squared_error(y_test, rf_predictions)
rf_r2 = r2_score(y_test, rf_predictions)
rf_pearson = pearsonr(y_test, rf_predictions)[0]

print (rf_mse)
print (rf_r2)
print (rf_pearson)
```

```
#Evaluating Linear Regression Model

lr_mse = mean_squared_error(y_test, lr_predictions)
lr_r2 = r2_score(y_test, lr_predictions)
lr_pearson = pearsonr(y_test, lr_predictions)[0]

print (lr_mse)
print (lr_r2)
print (lr_pearson)
```

8.1.8 Scatter plots of Actual vs Predictions

Scatter Plots of Predictions vs. Actual Values

```
#Scatter Plots of Predictions vs. Actual Values
#Random Forest
plt.figure(figsize=(12, 6))
plt.scatter(y_test, rf_predictions, color='blue', alpha=0.5, label='Random Forest')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--', lw=2)
plt.xlabel('BIS Index')
plt.ylabel('New Index')
plt.title('BIS Index vs. New Index')
plt.legend()
plt.show()
```

```
#Scatter Plots of Predictions vs. Actual Values
#Linear Regression
plt.figure(figsize=(12, 6))
plt.scatter(y_test, lr_predictions, color='green', alpha=0.5, label='Linear Regression')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--', lw=2)
plt.xlabel('BIS Index')
plt.ylabel('New Index')
plt.title('BIS Index vs. New Index')
plt.legend()
plt.show()
```

8.1.9 Bland-Altman Plots

Bland-Altman Plots

```
import numpy as np
def bland_altman_plot(data1, data2, title):
    mean = np.mean([data1, data2], axis=0)
    diff = data1 - data2
    md = np.mean(diff)
    sd = np.std(diff)

    plt.figure(figsize=(10, 5))
    plt.scatter(mean, diff, color='blue')
    plt.axhline(md, color='gray', linestyle='--')
    plt.axhline(md + 1.96*sd, color='red', linestyle='--')
    plt.axhline(md - 1.96*sd, color='red', linestyle='--')
    plt.title(title)
    plt.xlabel('Mean of Two Measurements')
    plt.ylabel('Difference Between Two Measurements')
    plt.show()
```

```
#Bland-Altman Plot for Random Forest Model
bland_altman_plot(y_test, rf_predictions, 'Bland-Altman Plot for Random Forest')
```

```
#Bland-Altman Plot for Linear Regression Model
bland_altman_plot(y_test, lr_predictions, 'Bland-Altman Plot for Linear Regression')
```