Assessing the Usefulness of Linear Regression Analysis in Finding a Correlation Between Respiration Rate and Oxygen Saturation to Find the Risk of Hypoxemia in Neonates.

April 14, 2023

Word Count: 4,064

CONTEXT

Intensive care units across the United States often admit many patients, with 22,860 neonatal intensive-care unit (NICU) beds used in 2021 (American Hospital Association, 2021). Most often, these patients include infants born prematurely and highly vulnerable to their environment, which increases their risk for severe or life-threatening conditions. Of infants born in the United States, 10% would require time in the NICU (Shellhaas et al., 2019). Additionally, doctors and other healthcare providers responsible for monitoring these neonatal patients often face problems managing large amounts of neonatal patient data, which can negatively impact how they can detect risks for life-threatening conditions. Fortunately, machine learning (ML) can be beneficial for managing several patient data files, which can be utilized to predict dangerous complications that could arise after analyzing patient data patterns. By using ML, healthcare providers caring for NICU patients can have leverage in monitoring the health of their patients so that they can prevent disabilities or even fatalities.

ML is a relatively new form of artificial intelligence that can find patterns in large data sets, including images and records. Because of its novelty, ML's potential in the medical field is ever-increasing. With this in mind, machine learning can be beneficial in finding patterns in the vital signs of ICU patients to find a reasonable prediction for the outcome of their condition. One type of ML often used in data science research is linear regression. This straightforward yet effective tool predicts a numerical value of a particular dependent variable using mathematical analysis based on an input value or feature and also evaluates the impact of each feature on the dependent variable (Gallo, 2015). Three different types of machine learning models can be used as linear regression models to predict a definite output. Support vector regression (SVR) is a machine learning model that can return a continuous output through prediction, which can be visualized in a mathematical equation that approximates the function (Awad & Khanna, 2015).

$$y = f(x) = < w, x > + b = \sum_{j=1}^{M} w_j x_j + b,\ y,\ b \in \mathbb{R},\ x,\ w \in \mathbb{R}^M$$

(Awad & Khanna, 2015)

Decision tree regression models are most commonly used in non-linear regression but are also available as linear regression models, and they predict outcomes by applying different predictor variables (Krzywinski & Altman, 2017). Random forest regression models are similar to decision trees, except that they combine multiple decision trees in an ensemble learning style to predict a continuous output, which is considered a bagging technique (Lyashenko, 2021). Because of the characteristics of these models, they can be used for linear regression tasks that aim to predict a continuous output through a linear graph.

This begs the question: How useful are ML linear regression models in predicting the risk of hypoxemia in neonatal ICU patients in correlation to their respiration rate? This project aims to explain how well linear regression models can predict life-threatening conditions such as hypoxemia. One crucial indicator of hypoxemia is oxygen saturation (SpO2), which the linear regression models should attempt to predict in a set of patient data.

Research investigating the implementation of linear regression analysis of neonatal patient data to identify dangerous conditions is limited, even when the data belongs to highly vulnerable patients who need desperate care. Therefore, this research must be done because doctors can have a clearer image of their patients' risks in critical conditions in the NICU, especially if the patient is very susceptible to unpredictable changes. In addition, the research can identify if linear regression alone can provide accurate predictions for dangerous medical conditions such as hypoxemia, which can then provide an incentive to determine whether they are valuable tools healthcare workers can use. Finally, this research can also provide the NICU patients' families and loved ones an understanding of their predicted condition to identify the best course of action for them, such as supplemental oxygen.

LITERATURE REVIEW

Within the medical field, the use of technology has begun to grow. ML has only been recently implemented, so existing ML models in disease detection and prediction have only started to emerge. There has been very limited research on the effectiveness of linear regression models alone in healthcare, so this research project is among the novel attempts to use linear regression. Currently, ML research includes a variety of linear regression and other non-linear models and has been examined to give an understanding of their reliability in predicting medical conditions.

Hypoxemia in Neonates

The Mayo Clinic defines hypoxemia as an abnormally low amount of SpO2 within the blood, specifically in the arteries (Mayo Clinic, 2018). Hypoxemia can be life-threatening if left unnoticed, especially in newborn infants. Fiore et al. explained that intermittent hypoxemia events are relatively common among neonates, especially if they are born prematurely. However, hypoxemia can result in long-term challenges to their quality of life, as there is a correlation between the time spent having hypoxemia and complications such as cognitive, learning, and motor delays, along with poor control of breathing (Fiore et al., 2021).

Prior to birth, a neonate will start developing control over breathing, which continues after birth. However, certain factors can result in a higher risk factor for hypoxemia, which usually concern the control the newborn has on their breathing. Insufficient levels of oxygen in the bloodstream can be attributed to factors such as maternal breathing disorders, such as sleep apnea, and maternal drug abuse, such as smoking (Mouradian et al., 2021). Preterm infants often struggle to maintain their respiration since their respiratory systems are underdeveloped, which can result in weaker respiration and level of oxygen consumption (Mouradian et al., 2021).

Other major causes of hypoxemia in neonates include respiratory distress syndrome, pneumonia, sepsis, and pulmonary hypertension (Hermansen & Mahajan, 2015). Additionally, congenital heart defects and problems with the neonatal airway are less common causes of hypoxemia, but are still considered factors that could contribute to it (Hermansen & Mahajan, 2015). These types of respiratory distress can result in a greater risk of hypoxemia, as they involve issues with respiration and control of breathing.

A Breakdown of Machine Learning Techniques

Several different ML algorithms are useful in medicine. Specifically, the SVR, DT, and RF have had major promise in predicting life-threatening conditions in patients using clinical data.

An SVR is a supervised machine learning model used for linear and non-linear analysis of various problems. Ideally, it should provide a definite correlation between the independent variable, or feature, and the dependent variable (Zhang & O’Donnell, 2020). Previous research has attempted to use SVR linearly, which was indicated through the performance evaluation using an R^2 score. Zavala-Ortiz et al. aimed to see how well SVR, among two other models–partial least squares regression (PLSR) and artificial neural networks (ANN)–would perform in monitoring CHO cell culture processes. In evaluating the performance of the SVR, the team found that it had an R^2 score of 0.93, indicating that it was successful in producing accurate predictions in a linear manner (Zavala-Ortiz et al., 2022).

The use of DTs have also been applied to medical research. A study by Zhang et al. was conducted to investigate the application of DT and logistic regression to predict the daily activities of patients who had a stroke. The research method extracted the information of rehabilitation therapy data at the researchers' hospital and identified daily activities with a level of independence using the Barthel index (BI) score. To compare the performance of the logistic regression and decision tree models, the researchers used receiver operating characteristic (ROC) curves, and they found that both models were not statistically significant to each other. (Zhang et al., 2022). The results of the study found that both the logistic regression and the DT models were successful in predicting the everyday lives of patients who suffered a stroke, with ROC curves of 0.808 and 0.831 for the logistic regression and DT models respectively. Zhang et al. demonstrated through their research that DT models can be especially useful prediction tools in monitoring the health and stability of patients.

RF models in research have also been useful in the prediction of dangerous diseases, specifically in the prediction of cardiovascular disease. A group of researchers conducted a study in which different models were compared with an RF model for a three year risk assessment for cardiovascular disease (Yang et al., 2020). Similar to the research conducted by Zhang et al., this group used ROC curves and AUC scores to identify the superior model performance of several models: multivariate logistic regression, CART, Naive Bayes, Bagged Trees, ADA Boost, and RF (Yang et al., 2020). The results showed that RF was the best risk assessment model for predicting cardiovascular disease in Eastern China, thus demonstrating that the use of an RF model can provide useful prediction results.

Research using ML to predict various diseases and other aspects of medicine has been increasingly useful, and can contribute to new understandings of how data science can contribute to the management of various diseases. However, there are certain limitations in ML that can inhibit the success of its results. Babic et al. from the Harvard Business Review (HBR) stated that in medical research, there exists many parameters in determining whether the model is truly successful. For example, diagnosing diseases through imaging data relies upon several factors, including the clarity of such images and the quality of the training data (Babic et al., 2021). Additionally, there may be ethical concerns regarding the source of the data, such as data that does not include diverse groups of individuals that are usually underrepresented (Babic et al., 2021). Therefore, ML techniques in research are not free from flaws in data analysis; however, previous works on its applications in medicine can show a promising capability for their usefulness in medical research.

Previous Attempts to use Machine Learning to Predict Hypoxemia

As proven, ML is a relatively novel yet powerful tool in identifying and predicting various conditions to improve healthcare quality. In detecting hypoxemia for vulnerable patients, several attempts were designed to understand whether such ML algorithms can sufficiently predict such a dangerous condition.

Xia et al. have developed a study to produce an ML model that can successfully predict hypoxemia for adults in ICU units after extubation. The research team was able to access data from the MIMIC-IV Database and extract a cohort of data, including patients aged 18 and over who have been treated with mechanical ventilation. After resolving issues with missing data, the team split the dataset into a training set (80%) and a test set (20%), which is an ideal data split. Xia et al. have used several machine learning algorithms not limited to linear regression: K-nearest neighbors (KNN), support vector machine (SVM), logistic regression, RF, XGBoost, and LightGBM. From these models, the team found that the RF and LightGBM models were the best predictors for hypoxemia, allowing them to produce a model that successfully predicts this condition.

Along with predicting hypoxemia in patients from the ICU, previous research has also attempted to predict the risk of hypoxemia in an operative setting. Park et al. have investigated whether machine learning can attempt to predict hypoxemia events for children undergoing surgery. The data they used was collected from a database of intraoperative vital signs (VitalDB), and they used a cohort of 13,130 patients aged 18 and under. Demographic information was collected from the patients' electronic health records. Similar to Xia et al.'s research, the team split their data into an 80% training test and a 20% test set. Three models were considered: gradient boosting machine (GBM), long short-term memory (LSTM), and transformer. Park et al. have discovered that the GBM had the best performance, thus successfully producing a model that can predict hypoxemia events in pediatric patients undergoing an operation.

Perspectives on Using ML on Small Datasets

Researchers who use ML in their studies usually focus on big data problems that analyze the scope of the issue on a larger scale. Therefore, researchers aim to avoid using small datasets so as to not risk poor model performance. Additionally, small data can increase the risk for model overfitting or underfitting, and model selection can be especially difficult since specific models should cater to the quality and quantity of data (Xu et al., 2023). In materials machine learning research, a larger dataset with a combination of diverse sources is ideal for the most versatile model possible. However, there happen to be specific studies in which small data has proven to be suitable for ML research.

For example, in materials science, the performance of the model is reflected by the degree of freedom (DOF) from the dataset, but a research team found that three different studies using small materials datasets were still highly successful in their model performance by implementing crude estimation of property, thus increasing the model prediction and enhanced predictive accuracy (Zhang & Ling, 2018). Even if small data can increase the risk of underfitting or overfitting in the model, they can still be tried and tested in various machine learning methods to identify whether specific ML models follow the trend of the data.

Summary

Based on previous research on predicting both hypoxemia events and other conditions, ML has been particularly successful in accurately predicting such conditions. However, the studies have focused more on using logistic regression analysis rather than linear regression alone and have focused more on larger datasets, thus demonstrating a gap in the ML studies. Therefore, this research will resolve the gap in the research by proposing a method that will use linear regression analysis to predict oxygen saturation, since oxygen saturation levels identify if a hypoxemia event occurred. Additionally, this method will rely on a relatively small cohort of NICU patients.

RESEARCH METHOD

As this is a bioinformatics project, data needed to be collected prior to the execution of the linear regression models. The research aimed to identify two results: find whether there is a correlation between the respiration rate and SpO2, and also determine how well the model performed in predicting the oxygen saturation linearly using an R^2 score analysis.

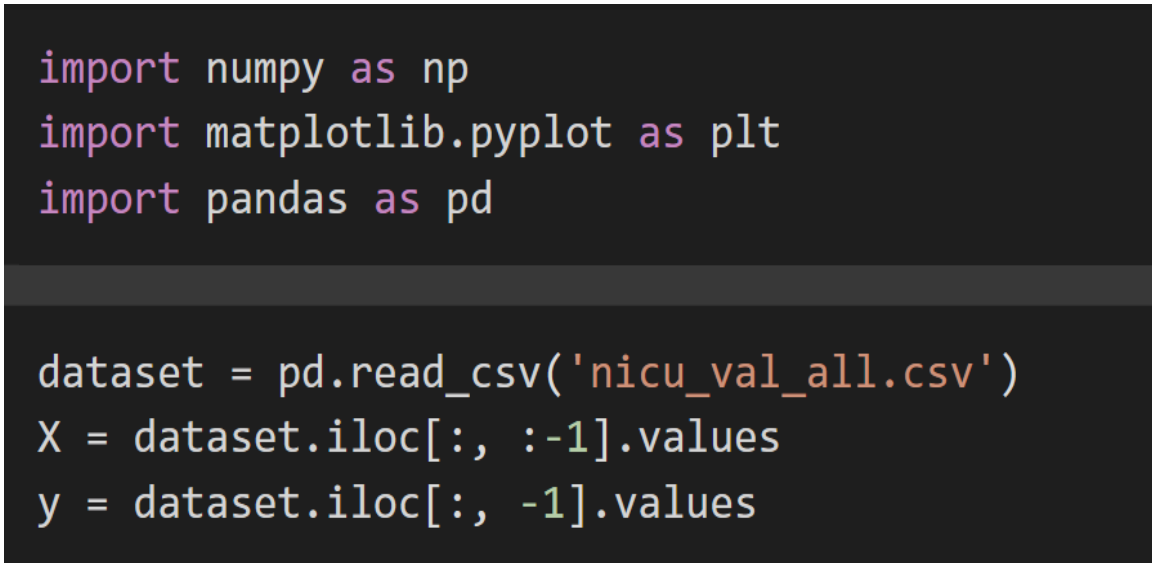
Hypothesis

The respiratory rate and oxygen saturation would have a positive correlation, and the three linear regression models considered can predict hypoxemia.

Data Extraction

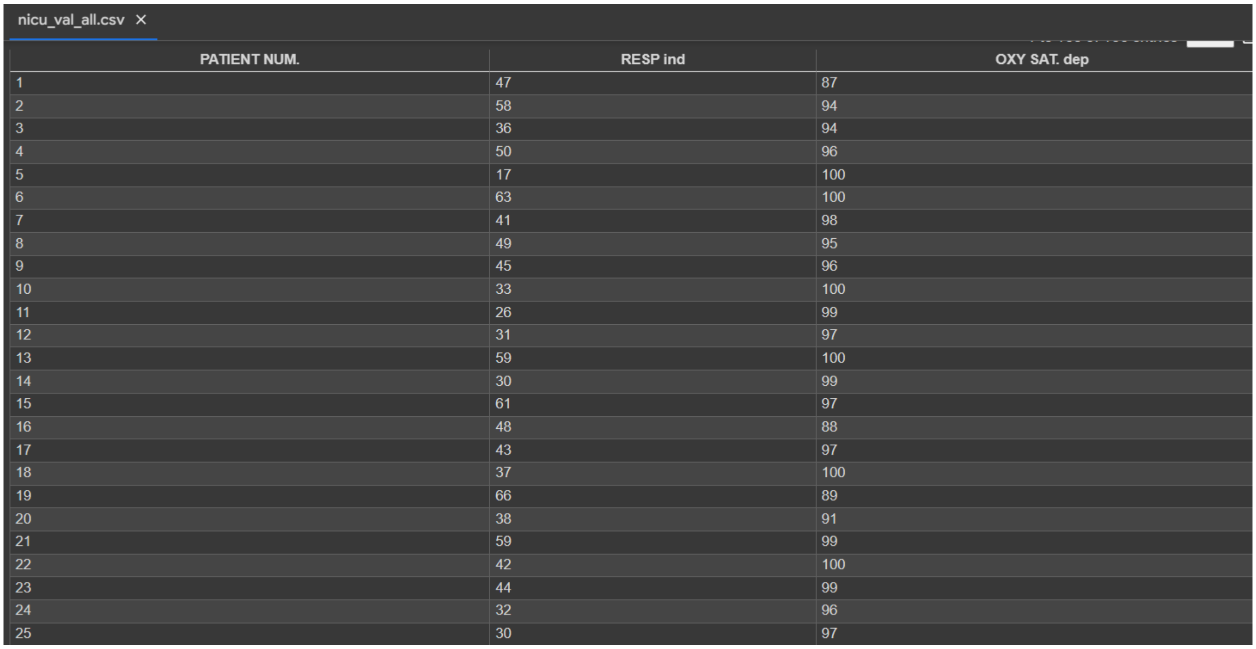
The data was extracted from the numeric records of the MIMIC-III Waveform Database. The MIMIC-III Waveform Database is an open-access repository of numeric and waveform records obtained from various ICU units in the Beth Israel Deaconess Medical Center in Boston, Massachusetts. The data is organized into records containing over 67,000, with waveform data for each record and a matching numerical file. Of the numerical records, one hundred anonymous neonatal records were randomly collected. The defining feature, respiration rate (RESP), was extracted, and the SpO2 was used as the dependent variable. RESP values were considered for the feature because respiratory distress in neonates may exacerbate the risk for hypoxemia, so it may correlate with the level of SpO2. After the data was extracted, the linear regression models were prepared to analyze the data. Three different models were considered for the analysis and prediction of the data: a support vector regression model (SVR), a decision tree regressor (DT), and a random forest regression model (RF). Each model was built on Google Colaboratory, a product derived from Google Research for building machine learning models. The data also had to be implemented with three libraries: NumPY, matplotlib, and pandas. The pandas library manages labeled data within supervised learning models, which can import libraries to assist in data preprocessing. The NumPy library analyzes the data and performs various mathematical and statistical tasks, including arrays. The models were written in Python, so these Python-oriented libraries were especially useful in managing the data used. All three models were preprocessed and were read with the ".csv" data file, and the range of values for the data was organized with both the RESP and SpO2 values. Every model required the data to be split into a training set (80%) and a test set (20%).

(Figure 1.1, Data Preprocessing; Student-produced screenshot)



```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

dataset = pd.read_csv('nicu_val_all.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
```

(Figure 1.2, A fraction of the dataset showing 25 out of 100 patients; Student-produced Screenshot)



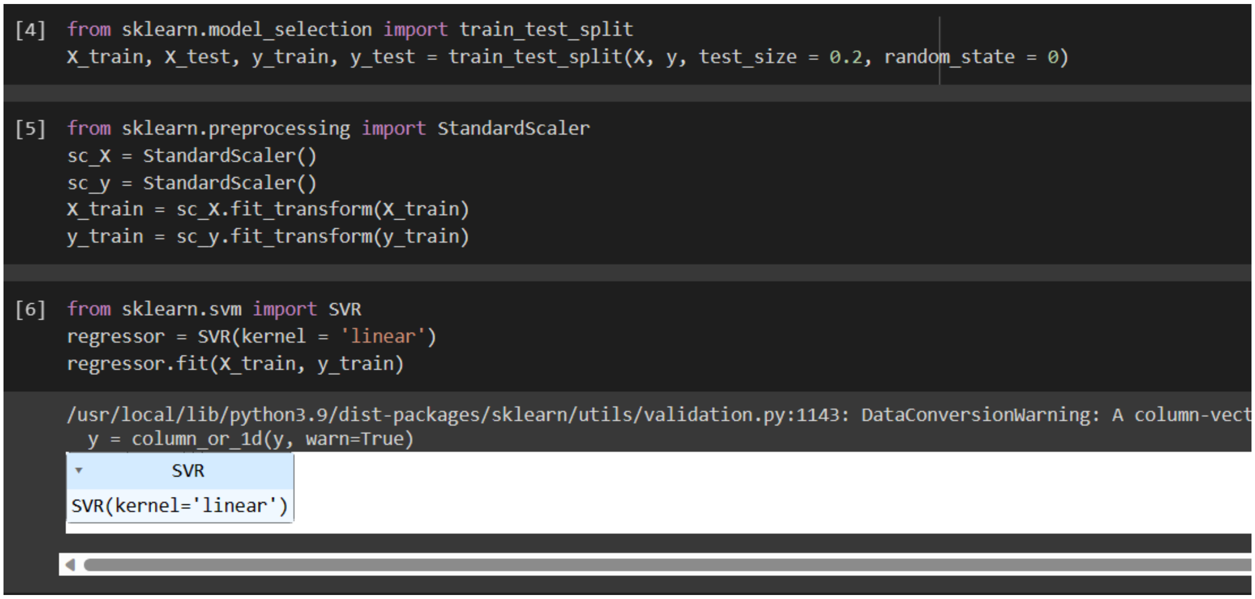
nicu_val_all.csv ×

PATIENT NUM.	RESP ind	OXY SAT. dep
1	47	87
2	58	94
3	36	94
4	50	96
5	17	100
6	63	100
7	41	98
8	49	95
9	45	96
10	33	100
11	26	99
12	31	97
13	59	100
14	30	99
15	61	97
16	48	88
17	43	97
18	37	100
19	66	89
20	38	91
21	59	99
22	42	100
23	44	99
24	32	96
25	30	97

SVR Implementation

As for all the models, the data was read in the notebook with the RESP values as the feature, or independent variable, and the SpO2 value as the dependent variable. The SVR model was trained on the training dataset, which included 80 patients. The data was feature scaled so that the independent variable could be normalized within a specific range of values, which was required for the SVR.

(Figure 1.3, Support Vector Regression background; Student-produced screenshot)



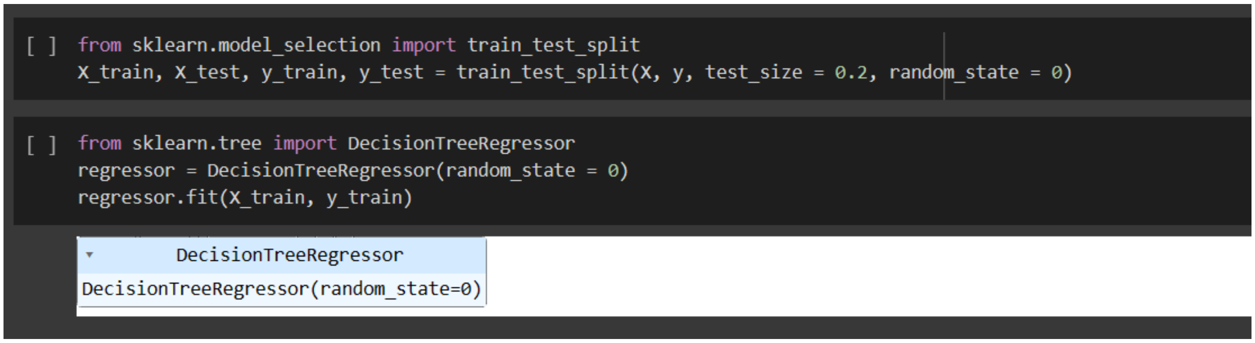
```
[4]  from sklearn.model_selection import train_test_split
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)

[5]  from sklearn.preprocessing import StandardScaler
     sc_X = StandardScaler()
     sc_y = StandardScaler()
     X_train = sc_X.fit_transform(X_train)
     y_train = sc_y.fit_transform(y_train)

[6]  from sklearn.svm import SVR
     regressor = SVR(kernel = 'linear')
     regressor.fit(X_train, y_train)

     /usr/local/lib/python3.9/dist-packages/sklearn/utils/validation.py:1143: DataConversionWarning: A column-vect
       y = column_or_1d(y, warn=True)
     ▾         SVR
     SVR(kernel='linear')
```

After the data was preprocessed and feature scaled, the SVR library was implemented to predict the SpO2 in correlation with RESP. After the SVR was trained on the dataset, the remaining 20 patients left in the dataset were used in the test set, and the SVR aimed to predict the SpO2 values based on the test set values.

DT Implementation

The DT also required data preprocessing to fit within the model. Unlike the SVR, the DT did not require feature scaling to normalize the range of numerical values. The DT was then trained on the training data before finally predicting the SpO2 values in comparison to the test set.

(Figure 1.4, Decision Tree background; Student-produced screenshot)

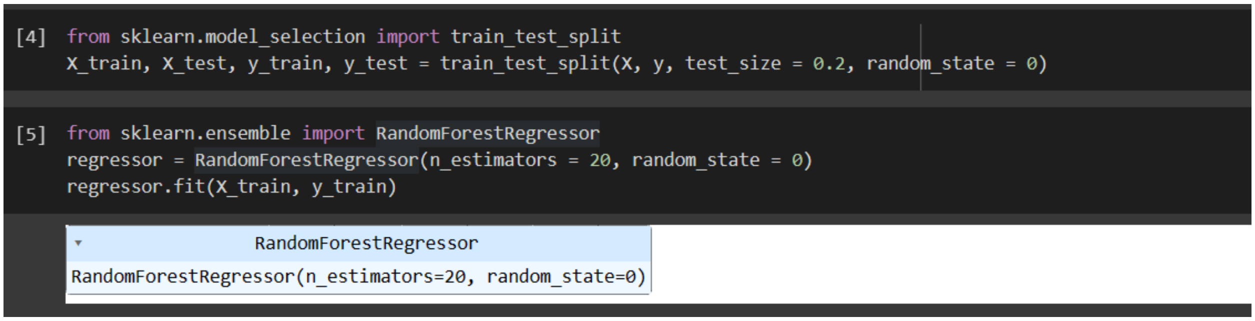


```
[ ]  from sklearn.model_selection import train_test_split
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)

[ ]  from sklearn.tree import DecisionTreeRegressor
     regressor = DecisionTreeRegressor(random_state = 0)
     regressor.fit(X_train, y_train)

     ▾       DecisionTreeRegressor
     DecisionTreeRegressor(random_state=0)
```

RF Implementation

The RF was implemented in a very similar way to the DT, but the only difference is that in addition to the model library being implemented, twenty decision trees dubbed as "n_estimators" were defined. The RF model was run on the training set with 20 estimator trees with a random state of zero. After training, the model performed on the test set and aimed to predict the SpO2 values in accordance with the actual values.

(Figure 1.5, Random Forest Regression background; Student-produced screenshot)



```
[4]  from sklearn.model_selection import train_test_split
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)

[5]  from sklearn.ensemble import RandomForestRegressor
     regressor = RandomForestRegressor(n_estimators = 20, random_state = 0)
     regressor.fit(X_train, y_train)

     ▾              RandomForestRegressor
     RandomForestRegressor(n_estimators=20, random_state=0)
```

After all models were trained and tested on the data, their predictions were visualized in a graph curated by matplotlib, and their accuracies were evaluated using R^2 scores.

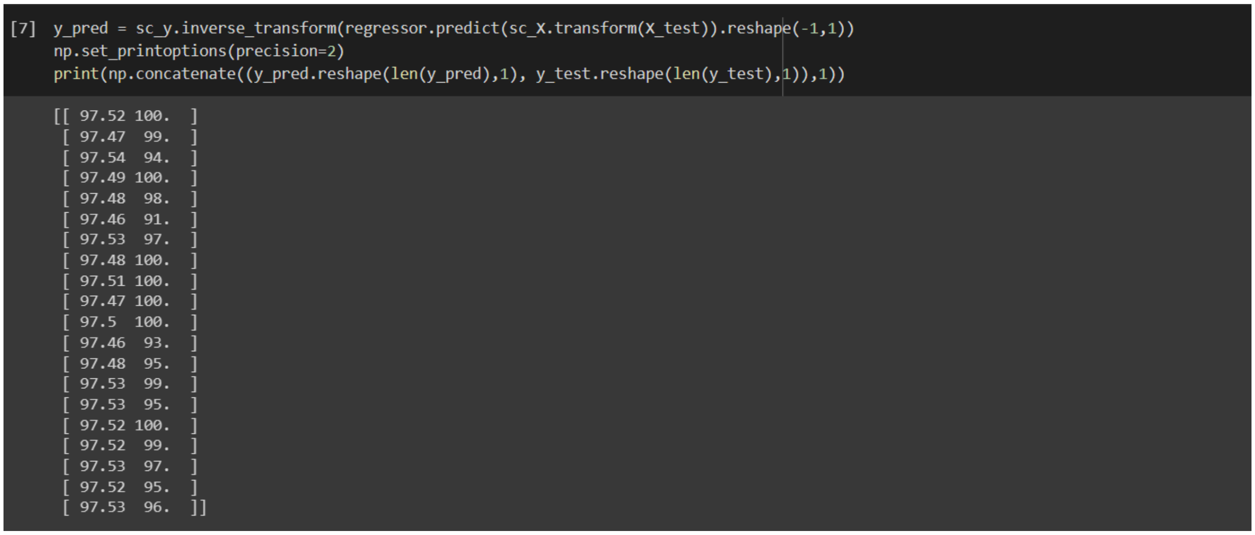
FULFILLMENT OF RESEARCH GAP

Unlike previous studies that identify linear regression models' performance in predicting hypoxemia, I used a simplistic approach to predict the SpO2 on the basis of a single feature by using a relatively smaller cohort of data than other research has utilized and solely linear regression. Additionally, previous works that analyzed the accuracy of model prediction of oxygen saturation did not solely rely on linear regression analysis and often used other methods such as neural networks and various ensemble methods such as XGBoost and LightGBM. Furthermore, other research in model performance has proven that such models were capable of predicting SpO2 acceptably; Xia et al.'s research, for example, used an area-under-the-curve (AUC) analysis to evaluate the performance of their models, which ranged in the acceptable values of 0.7 to 0.78 (Xia et al., 2022). Instead, my research found that the three linear regression models considered alone had a poor correlation between the feature and predicted variable and thus could not predict SpO2 accurately within the dataset, which was a different approach to previous procedures.

RESULTS

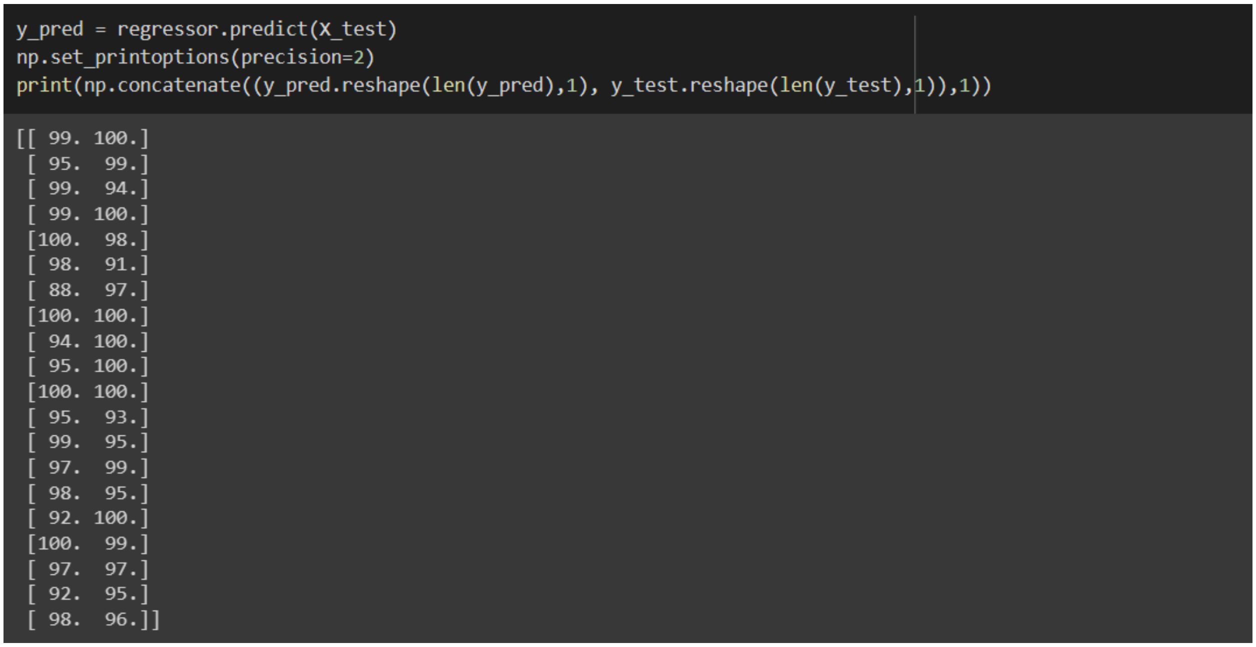
For the SVR, DT, and RF, an array created by NumPy produced the predicted values of the SpO2 alongside the actual values of the dataset. These predicted values were then visualized in three different graphs to provide information about whether the models were able to predict the values in a linear manner.

(Figure 2.1, SVR predicted values; Student-produced screenshot)



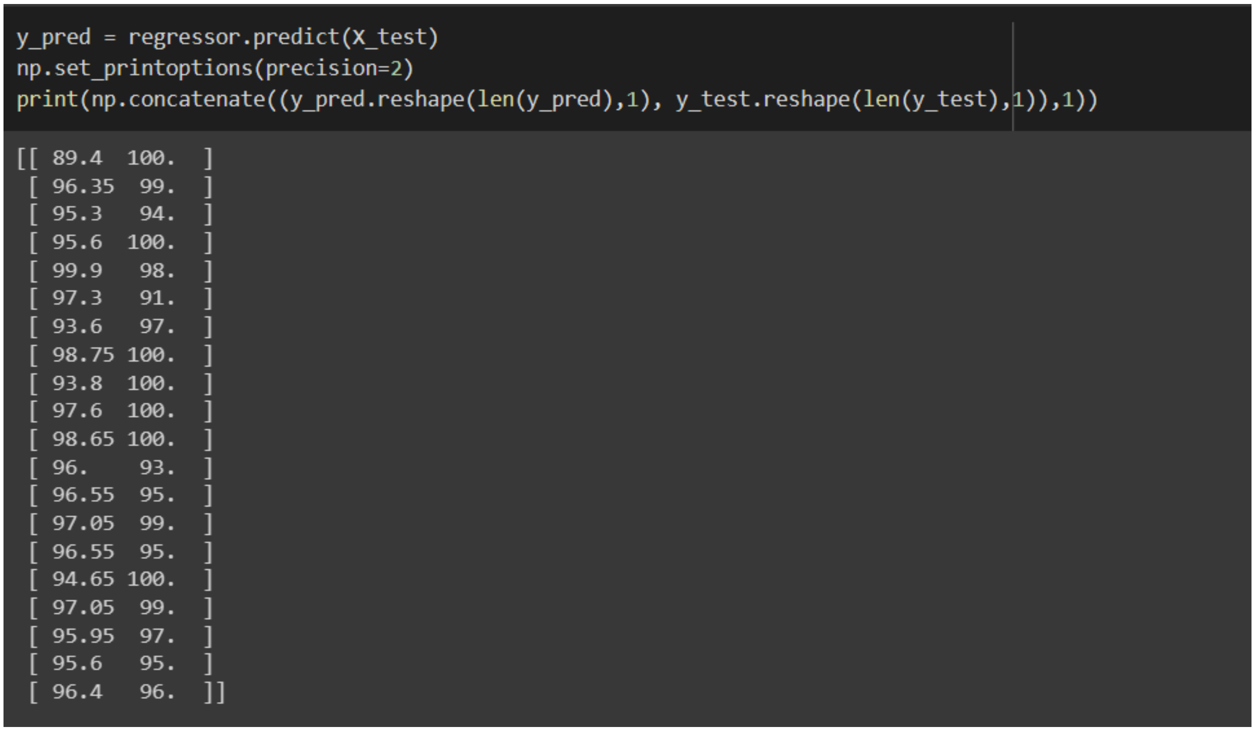
```
[7]  y_pred = sc_y.inverse_transform(regressor.predict(sc_X.transform(X_test)).reshape(-1,1))
     np.set_printoptions(precision=2)
     print(np.concatenate((y_pred.reshape(len(y_pred),1), y_test.reshape(len(y_test),1)),1))

     [[ 97.52 100.  ]
      [ 97.47  99.  ]
      [ 97.54  94.  ]
      [ 97.49 100.  ]
      [ 97.48  98.  ]
      [ 97.46  91.  ]
      [ 97.53  97.  ]
      [ 97.48 100.  ]
      [ 97.51 100.  ]
      [ 97.47 100.  ]
      [ 97.5  100.  ]
      [ 97.46  93.  ]
      [ 97.48  95.  ]
      [ 97.53  99.  ]
      [ 97.53  95.  ]
      [ 97.52 100.  ]
      [ 97.52  99.  ]
      [ 97.53  97.  ]
      [ 97.52  95.  ]
      [ 97.53  96.  ]]
```

(Figure 2.2, DT predicted values; Student-produced screenshot)



```
y_pred = regressor.predict(X_test)
np.set_printoptions(precision=2)
print(np.concatenate((y_pred.reshape(len(y_pred),1), y_test.reshape(len(y_test),1)),1))

[[ 99. 100.]
 [ 95.  99.]
 [ 99.  94.]
 [ 99. 100.]
 [100.  98.]
 [ 98.  91.]
 [ 88.  97.]
 [100. 100.]
 [ 94. 100.]
 [ 95. 100.]
 [100. 100.]
 [ 95.  93.]
 [ 99.  95.]
 [ 97.  99.]
 [ 98.  95.]
 [ 92. 100.]
 [100.  99.]
 [ 97.  97.]
 [ 92.  95.]
 [ 98.  96.]]
```

(Figure 2.3, RF predicted values; Student-produced screenshot)

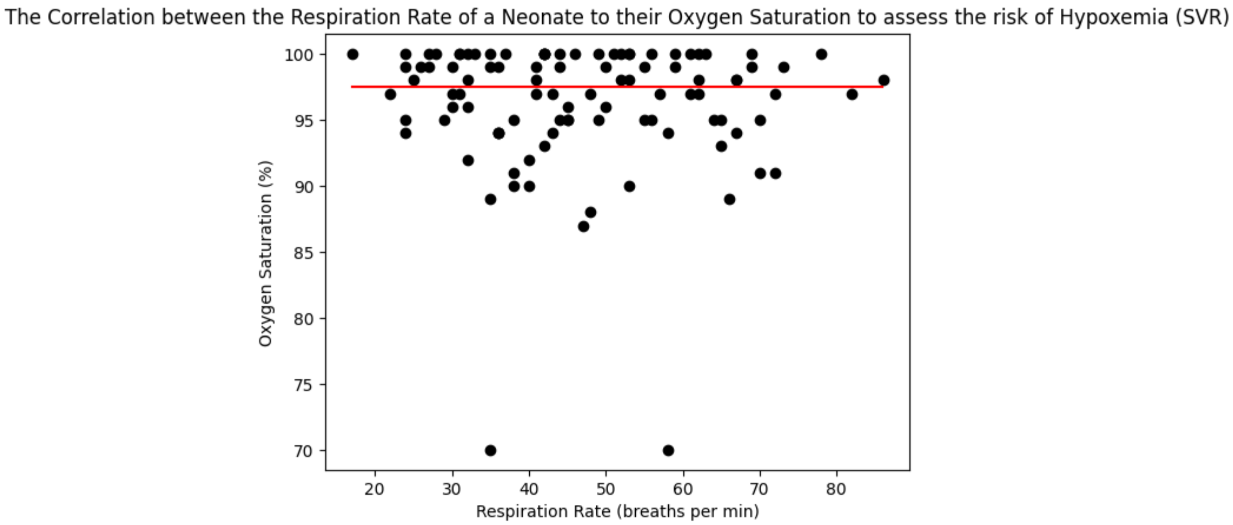


```
y_pred = regressor.predict(X_test)
np.set_printoptions(precision=2)
print(np.concatenate((y_pred.reshape(len(y_pred),1), y_test.reshape(len(y_test),1)),1))
```

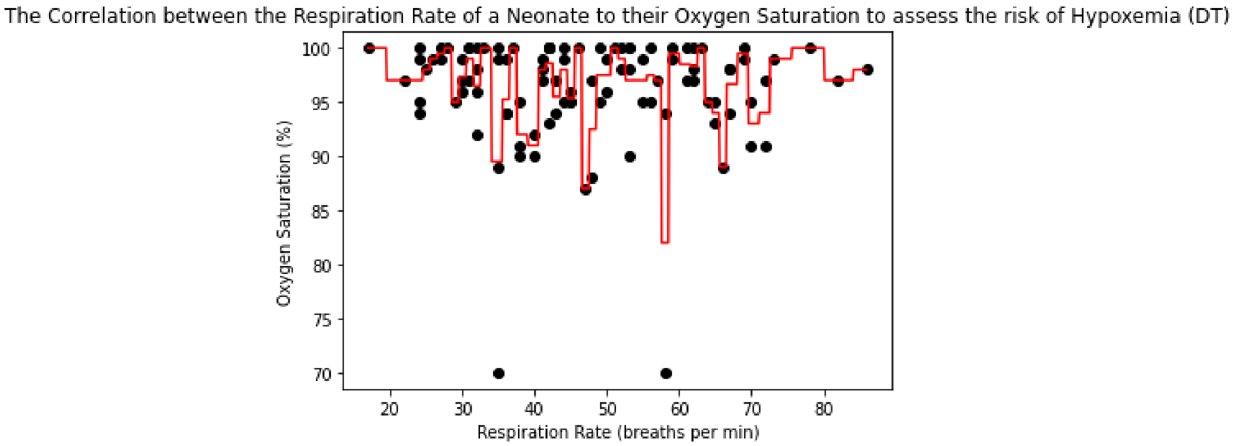
```
[[ 89.4  100.  ]
 [ 96.35  99.  ]
 [ 95.3   94.  ]
 [ 95.6  100.  ]
 [ 99.9   98.  ]
 [ 97.3   91.  ]
 [ 93.6   97.  ]
 [ 98.75 100.  ]
 [ 93.8  100.  ]
 [ 97.6  100.  ]
 [ 98.65 100.  ]
 [ 96.    93.  ]
 [ 96.55  95.  ]
 [ 97.05  99.  ]
 [ 96.55  95.  ]
 [ 94.65 100.  ]
 [ 97.05  99.  ]
 [ 95.95  97.  ]
 [ 95.6   95.  ]
 [ 96.4   96.  ]]
```

As shown, each model had different predicted SpO2 values. Therefore, their prediction curves would be quite different from each other in their visualized results. Each models' prediction is shown as the red line, while the actual dataset values are shown as black dots on each graph.

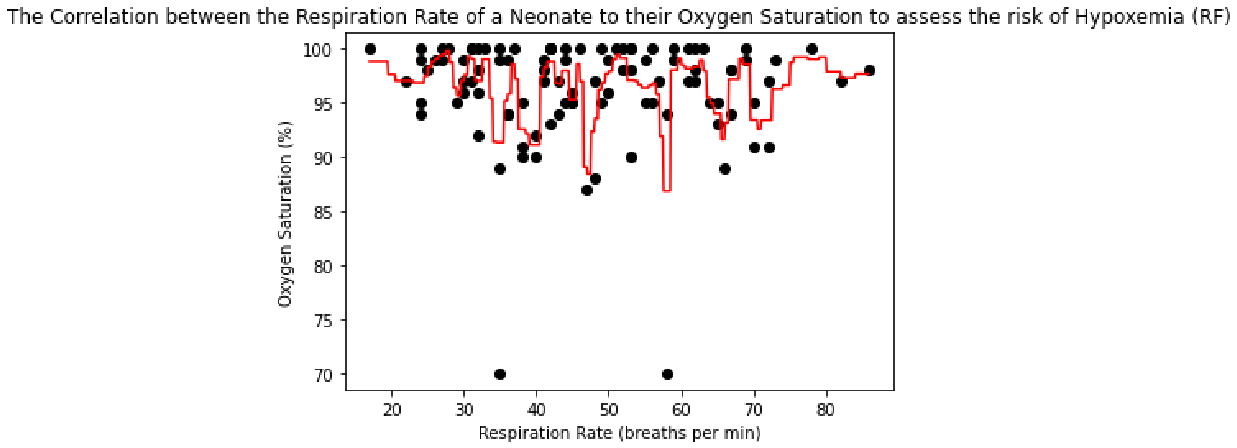
(Figure 2.4, Visualized SVR results; Student-produced screenshot)



(Figure 2.5, Visualized DT results; Student-produced screenshot)



(Figure 2.6, Visualized RF results; Student-produced screenshot)



To simplify the accuracy of each model, the R^2 scores were calculated using scikit methods (sklearn.metrics), importing the “r2_score” library. R^2 scores can evaluate the accuracy of a model’s predictions since it is particularly suited for linear regression.

Typically, the R^2 score is within the range of 0 to 1, with 0 indicating a poorer correlation between the independent and dependent variable and lower model performance and 1 indicating a very high correlation between the variables and the highest model performance. If the R^2 value is negative, it means that the data could not be predicted in a linear manner by the model and the variables considered had a very poor correlation. Surprisingly, for all three models, a negative R^2 value was calculated; the SVR had a value of 0.0006782432160309382, the DT had a value of -1.3773841961852868, and the RF had a value of -1.006352179836513. In terms of best model performance, the SVR was the highest performing model since it had the largest– and only positive– R^2 value of all three models. However, in evaluating general model performance, all three were incapable of predicting the SpO2 values linearly, since two out of three R^2 scores are negative and the SVR score was very low. Additionally, the poor R^2 scores indicate a very poor correlation between RESP and SpO2.

CONCLUSIONS

As shown by each model performance, the SVR, DT, and RF could not predict the SpO2 and were incapable of finding a correlation between RESP and SpO2. Since hypoxemia is identified by SpO2 percentage, these linear regression models could not predict the risk of hypoxemia as well.

Limitations

In this study, there are two certain limitations to the research method that must be addressed. Firstly, there were very few features that were addressed as independent variables that could potentially impact the accuracy of the models' prediction of SpO2; factors such as heart rate and blood pressure could influence the prediction of SpO2. Additionally, with more factors, the performance of the models would be improved. Adding more independent variables to influence the prediction of a dependent variable can add different predictors that could affect the desired output differently, which could ultimately improve the model performance based on several factors.

Additionally, the types of linear regression models used were limited. Although SVR, DT, and RF can be used for linear regression tasks, other models such as simple linear regression (SLR), multiple linear regression (MLR), and polynomial regression (PR) should have been considered. SLR and MLR are excellent tools that can find the correlation between the considered variables, which can also provide a prediction curve by following the trend of the data. PR accomplishes this task as well, but aims to produce a correlation between variables as an nth degree polynomial. If these models were considered, then there would be a higher chance of getting a model that can accurately predict SpO2, which can show which model would be the most effective in predicting the desired outcome linearly.

Further Research

To improve the quality of the research project, I would include more features that could affect the percentage of SpO2, such as pulse, systolic blood pressure, and diastolic blood pressure. Additionally, I would explore other models that could predict the SpO2 without necessarily relying on linear regression analysis. Such models include logistic regression and other ensemble methods, such as LightGBM and XGBoost. Because the results show that predicing SpO2 is not possible with only linear regression, it is important to find different mathematical analyses that can find a positive correlation between the variables and produce an accurate prediction curve.

Additionally, with different features and outputs, this research could be used in predicting other dangerous conditions that could arise from bedside ICU data, such as respiratory distress, hypoxia, and anemia.

The results of this experiment do not mean that linear regression on its own is incapable of being useful predictors in healthcare. Since this study only focused on neonatal data, there is a possibility that linear regression analysis can predict dangerous conditions in adult data depending on their pattern.

Furthermore, the dataset could be expanded using other database information. Such information can come from other ICU datasets, such as the MIMIC-III Clinical Database and the MIMIC-IV Database. By combining each databases' information, the dataset becomes increasingly more varied and could potentially improve model performance.

BIBLIOGRAPHY

American Hospital Association (2021). *Fast Facts on US Hospitals*. American Hospital Association. https://www.aha.org/system/files/media/file/2020/01/2020-aha-hospital-fast-facts-new-Jan-2020.pdf

Awad, M., & Khanna, R. (2015). Support Vector Regression. *Apress eBooks*, 67–80. https://doi.org/10.1007/978-1-4302-5990-9_4

Babic, B. (2020, December 15). *When Machine Learning Goes Off the Rails*. Harvard Business Review. https://hbr.org/2021/01/when-machine-learning-goes-off-the-rails

Di Fiore, J. M., & Raffay, T. M. (2021). The relationship between intermittent hypoxemia events and neural outcomes in neonates. *Experimental neurology*, *342*, 113753. https://doi.org/10.1016/j.expneurol.2021.113753

Gallo, A. (2022, October 12). A Refresher on Regression Analysis. *Harvard Business Review*. https://hbr.org/2015/11/a-refresher-on-regression-analysis

Hermansen, C. L., & Mahajan, A. (2015). Newborn Respiratory Distress. *American family physician*, *92*(11), 994–1002.

Krzywinski, M., & Altman, N. (2017). Classification and regression trees. *Nature Methods*, *14*(8), 757–758. https://doi.org/10.1038/nmeth.4370

Lyashenko, V. (2021). Random Forest Regression - The Definitive Guide. *cnvrg.io*. https://cnvrg.io/random-forest-regression/

Mayo Clinic. (2021, March 24). *Low blood oxygen (hypoxemia)*. https://www.mayoclinic.org/symptoms/hypoxemia/basics/definition/sym-20050930

Moody, B., Moody, G., Villarroel, M., Clifford, G. D., & Silva, I. (2020). MIMIC-III

Waveform Database (version 1.0). *PhysioNet*. https://doi.org/10.13026/c2607m.

Mouradian, G. C., Jr, Lakshminrusimha, S., & Konduri, G. G. (2021). Perinatal Hypoxemia and Oxygen Sensing. *Comprehensive Physiology*, *11*(2), 1653–1677. https://doi.org/10.1002/cphy.c190046

Park, J. B., Lee, H. J., Yang, H. L., Kim, E. H., Lee, H. C., Jung, C. W., & Kim, H. S. (2023). Machine learning-based prediction of intraoperative hypoxemia for pediatric patients. *PloS one*, *18*(3), e0282303. https://doi.org/10.1371/journal.pone.0282303

Shellhaas, R. A., Burns, J. W., Barks, J. D. E., Hassan, F., & Chervin, R. D. (2019). Maternal Voice and Infant Sleep in the Neonatal Intensive Care Unit. *Pediatrics*, *144*(3), e20190288. https://doi.org/10.1542/peds.2019-0288

Xia, M., Jin, C., Cao, S., Pei, B., Wang, J., Xu, T., & Jiang, H. (2022). Development and validation of a machine-learning model for prediction of hypoxemia after extubation in intensive care units. *Annals of translational medicine*, *10*(10), 577. https://doi.org/10.21037/atm-22-2118

Xu, P., Ji, X., Li, M., & Lu, W. (2023). Small data machine learning in materials science. *Npj Computational Materials*, *9*(1). https://doi.org/10.1038/s41524-023-01000-z

Yang, L., Wu, H., Jin, X., Zheng, P., Hu, S., Xu, X., Yu, W., & Yan, J. (2020). Study of cardiovascular disease prediction model based on random forest in eastern China. *Scientific reports*, *10*(1), 5245. https://doi.org/10.1038/s41598-020-62133-5

Zavala-Ortiz, D. A., Denner, A., Aguilar-Uscanga, M. G., Marc, A., Ebel, B., & Guedon, E. (2022). Comparison of partial least square, artificial neural network, and support vector regressions for real-time monitoring of CHO cell culture processes using in situ near-infrared spectroscopy. *Biotechnology and bioengineering*, *119*(2), 535–549. https://doi.org/10.1002/bit.27997

Zeiberg, D., Prahlad, T., Nallamothu, B. K., Iwashyna, T. J., Wiens, J., & Sjoding, M. W. (2019). Machine learning for patient risk stratification for acute respiratory distress syndrome. *PloS one*, *14*(3), e0214465. https://doi.org/10.1371/journal.pone.0214465

Zhang, F., & O'Donnell, L. J. (2020). Chapter 7 - Support vector regression (A. Mechelli & S. Vieira, Eds.). *ScienceDirect; Academic Press*. https://www.sciencedirect.com/science/article/pii/B9780128157398000079#:~:text=Support%20vector%20regression%20%28SVR%29%20is%20a%20supervised%20machine

Zhang, Q., Zhang, Z., Huang, X., Zhou, C., & Xu, J. (2022). Application of Logistic Regression and Decision Tree Models in the Prediction of Activities of Daily Living in Patients with Stroke. *Neural plasticity*, *2022*, 9662630. https://doi.org/10.1155/2022/9662630

Zhang, Y., & Ling, C. (2018). A strategy to apply machine learning to small datasets in materials science. *Npj Computational Materials*, *4*(1). https://doi.org/10.1038/s41524-018-0081-z