1. Introduction:

1.1 Overview:

Operative mortality rates have been a topic of great interest among surgeons, patients, lawyers, and health policy administrators. Post-Operative respiratory complications are the most common fatality following any type of thoracic surgery. The scope of our project is to examine the mortality of patients within a full year after the surgery, that is classification related to the post-operative life expectancy in lung cancer patients. This can significantly help as we are examining the underlying health factors of patients. It could potentially be a powerful predictor for surgically related deaths. Data includes a patients' ICD-10 codes for primary and secondary as well as multiple tumors if any(Diagnosis), Volume that has been exhaled at the end of the first second of forced expiration(FEV), Performance status on Zubrod scale(PERFORMANCE), size of original tumor(TNM), and age at surgery(AGE). In addition we have several classification features such as presence of Pain before surgery, Haemoptysis before surgery,Dyspnoea before surgery, Cough before surgery, Weakness before surgery, whether the patient is a smoker, whether the patient has asthma, and a few others. The classification predicts whether the patient survived the following year-long period. This approach is the crucial step towards understanding the life expectancy post thoracic surgery.

1.2 Purpose:

Lung cancer is the leading cause of cancer-related deaths in the world. Postoperative respiratory complications are the most common fatality following any type of thoracic surgery.

What if you could have a reliable estimator for surgically related deaths given potentially powerful predictor health factors of patients?

The exact incidence is most contingent upon the preoperative health and lung function of the patient, and we would like to explore and understand how those conditions can drive these complications. Despite the serious prognosis of lung cancer, some people with earlier-stage cancers are cured.

2. Literature Survey

2.1 Existing Problem:

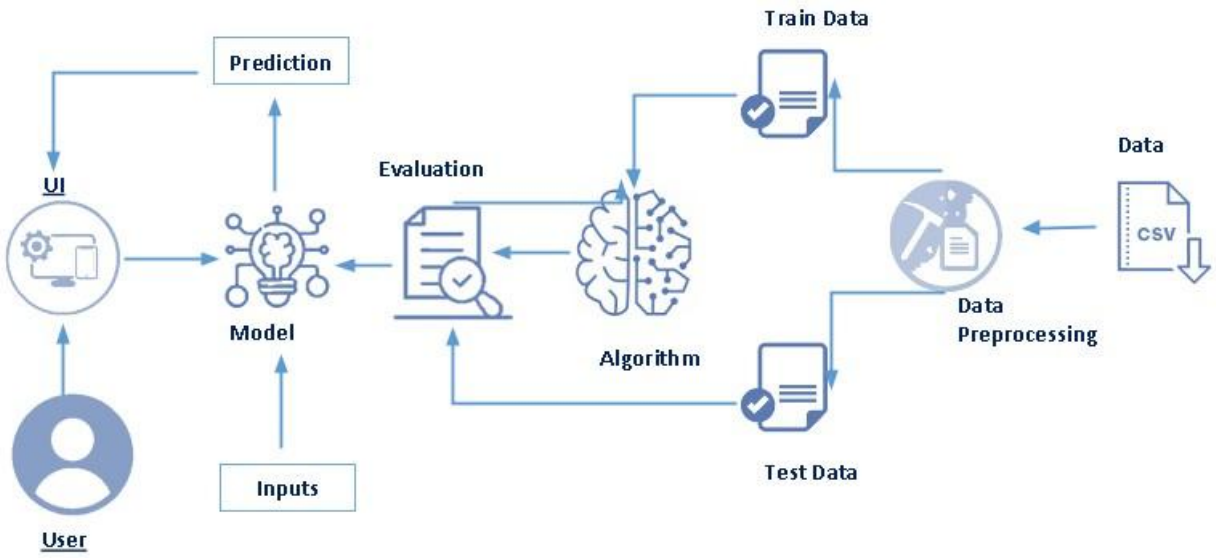
There are just a few systems that attempt to predict survival after thoracic surgery accurately and efficiently.

It is a very time-consuming task to predict the survival after thoracic surgery by manually testing. For one person it could take up to hours of testing and paper work to

do it properly.
Using a machine learning model, we can do the same in seconds.

2.2 Proposed Solution:
The model can be trained to predict underlying health factors of patients that could potentially be a powerful predictor for surgically related deaths. It is using advanced machine learning model which is trained using the verified dataset of life expectancy of post thoracic surgery and its various attributes,
The model can predict the performance with an accuracy of 88% given the fact that the result can be seen in seconds the model is reliable.
Anyone with prior knowledge of using a web browser can operate the application easily. Generally, a model is only as good as the data passed into it, and the data preprocessing we do ensure that the model has as accurate a dataset as possible.

3. Theoretical Analysis
3.1 Block Diagram:



3.2 Hardware and Software Designing:

- Hardware
  Windows 10

- Software
  Anaconda Navigator: Anaconda Navigator is a free and open-source

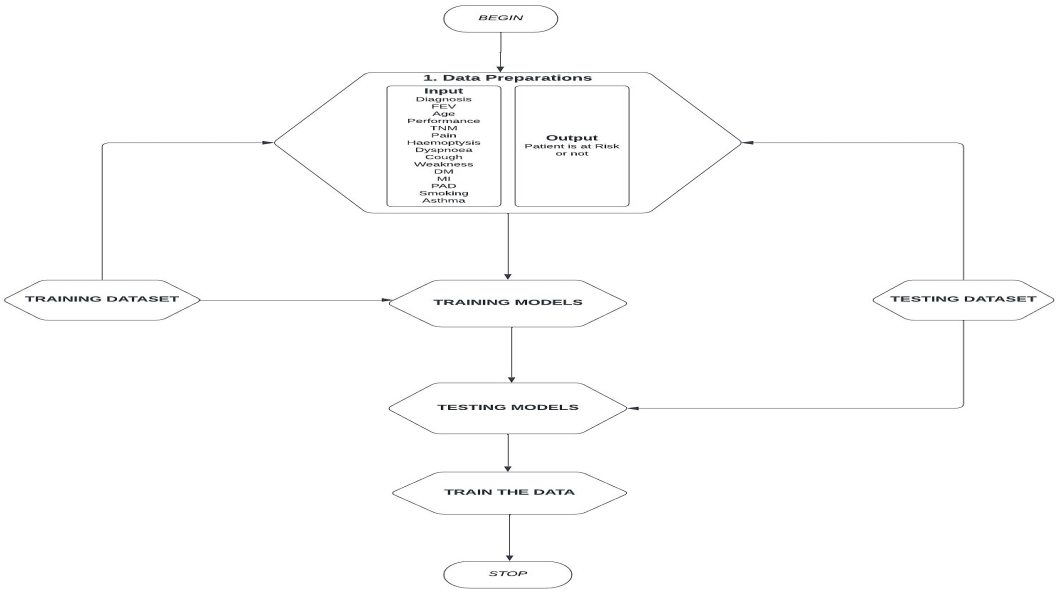
distribution of the Python and R programming languages for data science and machine learning-related applications. It can be installed on Windows, Linux, and macOS.Conda is an open-source, cross-platform, package management system. Anaconda comes with great tools like JupyterLab, Jupyter Notebook, QtConsole, Spyder, Glueviz, Orange, Rstudio, Visual Studio Code. For this project, we will be using Jupyter notebook and Spyder.

To build Machine learning models you must require the following package
• **Sklearn**: Scikit-learn is a library in Python that provides many unsupervised and supervised learning algorithms.
• **NumPy**: NumPy is a Python package that stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n- dimensional array object
• **Pandas**: pandas is a fast, powerful, flexible, and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language.
• **Matplotlib**: It provides an object-oriented API for embedding plots into applications using general-purpose,GUI,toolkits.
• **Flask**: Web framework used for building Web applications.

4. Experimental Investigation

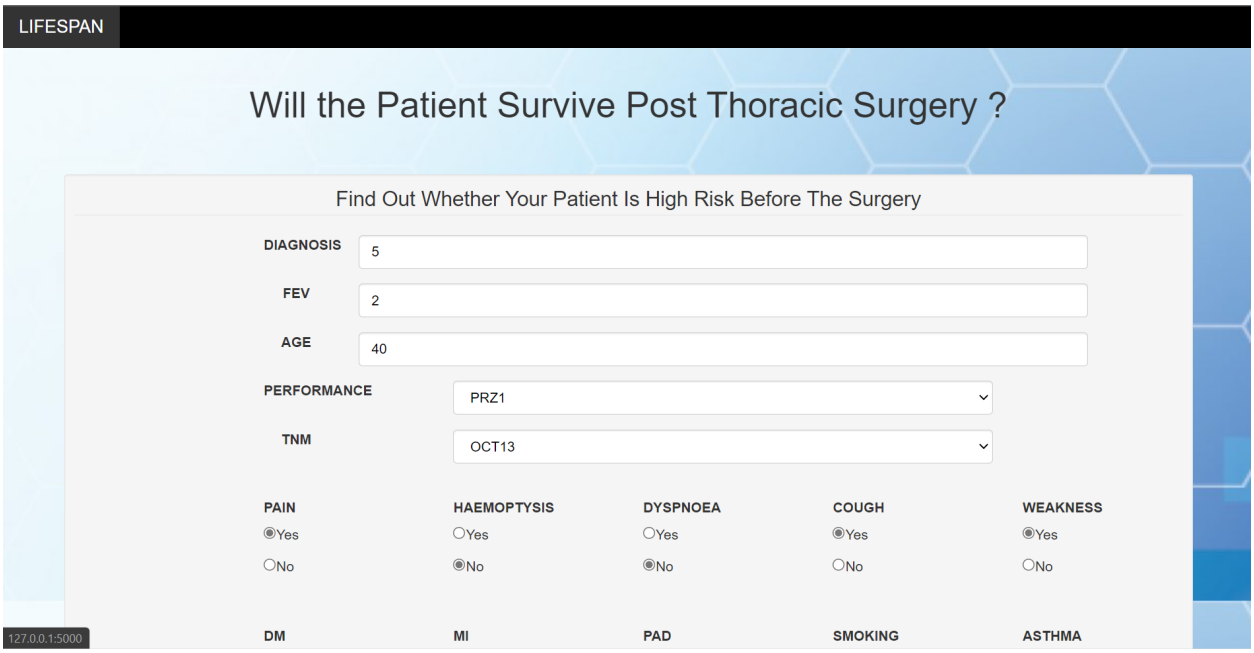
In this section, we will be creating and training our model for predicting life expectancy after thoracic surgery. Since there are multiple algorithms, we can use to build our model, we will compare the accuracy scores after testing and pick the most accurate algorithm.
From this list, we will be using classification algorithms such as Decision tree, Random forest, KNN, and xgboost to perform our predictions. We will train and test the data with these algorithms. We then see which algorithm produces the highest accuracy and select it as our algorithm of choice for future use.
On the results of the following algorithms, we have done the conclusion the Random Forest model is the most accurate out of all the models which we have tested.

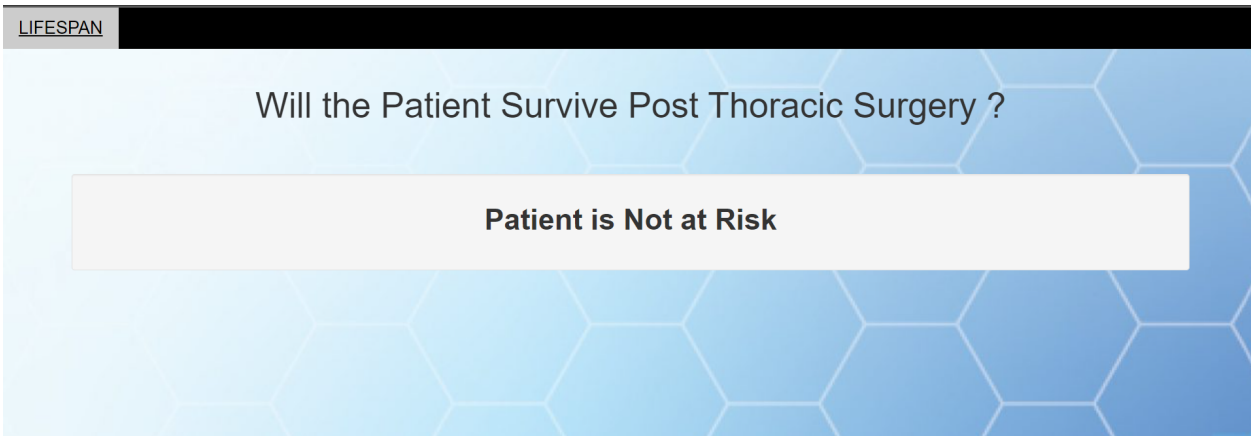
5. Flowchart



6. Result

The result of the project is the prediction of the heart disease. The output would be in this format:





7. Advantages and Disadvantages

Advantages:

The main advantage the proposed model is it that it can predict whether certain attributes like age, FEV, etc. have any effect on the prediction of a life expectancy after thoracic surgery before ever building one.

Therefore, manufactures can have extra opinion on how to create a perfect machine.

It can also help regular people to predict by inputting the values of factors.

Disadvantages:

- Need more datasets, to increase the accuracy of the algorithms.
- The accuracy of the application depends on the dataset used to train the model.
- The proposed application is Web-based, hence cannot be used in Mobile devices.
- The result of the application depends upon the accuracy of the algorithms

8. APPLICATION

By Manufacturers

Using this proposed application manufacturers can predict whether or not to use something. It can work as a simulation of how the person performs if performs the surgery.

Using this data, they can choose whether to include a certain feature or not.

By Professionals

In surgery everything is about performance. By using this model they can predict the performance and they can perform better in the surgery.

There are many factors which can contribute to better performance.

9. CONCLUSION

By the end of all of our iterations and improvements, we were able to achieve fairly good results with the random forest and other methods. These results have large implications in the medical field. An analysis similar to ours could be performed before a patient goes in for surgery to see how high risk they are, which could be crucial information. Our model will be able to perform better with more accurate dataset.

10. FUTURE SCOPE

We can improve our results by averaging a series of identical and independently distributed trees, which we would like to contrast with boosting, in which the trees would be grown in an adaptive manner specific to the bias (not I.I.D.). We would like to recursively train on the residuals of each misclassification. A next possible step would be to implement the following algorithm (bumping):

1. Bootstrap n models (with replacement, forcing even ratios), where number of models = number of features.
2. Train n models, with initially one feature per model.
3. Test all n models on original data set. Pick the model with lowest error on original data set, and define a new residual data set on all misclassified examples.
4. Train your next n models on the residual (i.e. boosting) - but NO averaging at this point.
5. Test on the very original data set and pick the best one. Continue process repeatedly.

Hopefully this will further reduce our variance. In addition, we calculated the optimal feature set as shown above, so it would be interesting to compare different results for all of our implementations if we use only those specific features.

11. BIBLIOGRAPHY

- www.wikipedia.org
- www.google.com
- www.github.org
- Life Expectancy Post Thoracic Surgery, Adam Abdulhamid, Ivaylo Bahtchevanov, Peng Jia, Stanford University