Iowa Initiative for Artificial Intelligence
Final Report

Project title:	Using machine learning to develop symptom prediction models in patients with cancer and multimorbidity
Principal Investigator:	Stephanie Gilbertson-White PhD, ARNP-BC (College of Nursing) Sanvesh Srivastava PhD (Department of Statistics and Actuarial Sciences) Nick Street PhD (Tippie College of Business) Patrick Fan PhD (Tippie College of Business) Heath Davis (ICTS, Biomedical Informatics)
Prepared by (IIAI):	Avinash Mudireddy
Other investigators:	
Date:	09/01/2023

Were specific aims fulfilled:	Partially
Readiness for extramural proposal?	No
If yes … Planned submission date	N/A
Funding agency	N/A
Grant mechanism	N/A
If no … Why not? What went wrong?	

Brief summary of accomplished results:

Research report:

Aims (provided by PI):

Increasingly, cancer is recognized as a chronic disease requiring management of symptoms associated with the disease and the complex sequelae of intensive therapies. Cancer related symptoms are associated with decreased quality of life, decreased functional status, increased health care utilization, and shorter life expectancy. A persistent problem for clinicians and researchers alike is the challenge of identifying which patients will develop which symptoms or clusters of symptoms. In addition, there is a wide range in the symptom severity, distress, and chronicity reported by patients with chronic cancer when symptoms occur. This variability makes it extremely challenging to deliver symptom management interventions to the "right patients" at the "right time". If/how multiple chronic conditions (MCCs) contribute to variability in symptoms experiences in the context of chronic cancer has not be explored.

Multimorbidity is the complex health condition in which one is living with MCCs for one year or more and has been associated with increased symptom burden, decreased functioning, and quality of life (QOL) in non-cancer populations. The chance of being diagnosed with MCCs, including cancer increases as we age. Emerging research has shown that people with cancer and MCCs have decreased functional status, poorer survival and often do not receive standard cancer care and/or supportive care therapies.

The majority of research on multimorbidity and chronic cancer has evaluated multimorbidity as a count of the number of concurrent diagnoses. Epidemiologically, we do not know the natural history of the symptoms in cancer patients with different patterns of multimorbidity. Improved understanding of which patients will develop severe/distressing symptoms and the patterns of those symptoms will help us

anticipate who is at risk for the most severe/distressing symptoms and improve our ability to precisely deliver symptom management interventions to the right patients at the right time. Data from the electronic health record (EHR) can be leveraged to understand how diagnoses cluster. However, patient reported symptoms are not routinely documented in the EHR in a manner that is amendable to data science methodologies. Therefore to understand the symptoms of patients with chronic cancer and MCCs, prospective data collection that carefully caputres the patient experience in the form of questionnaires and patient interviews is needed.

Purpose of this study is to describe how MCCs cluster with chronic cancer diagnoses and to predict the symptom experience of patients with chronic cancer and MCCs.

Specific aim 1: Describe the most common MCCs co-occurring in patients with chronic cancer and identify subgroups of patients that aggregate based different multimorbidity patterns. Differences in demographic (including sex) and clinical characteristics as well as health care utilization and life expectancy will be evaluated.

Approach for aim 1: Diagnoses, demographics, health care utilization, and life expectancy from the time of diagnosis of a large cohort (n>250,000) of patients with chronic cancer and MCCs will be extracted from the EHR from 36 health care organizations. Subgroups of patients will identify using several clustering patterns among MCCs and cancer diagnoses.

Specific aim 2: Predict the occurrence patterns and trajectories of symptoms in subgroups of patients with chronic cancer and MCCs from the time of diagnosis to up to five years post diagnosis.

Approach for aim 2: Text mining of clinical notes will be used to identify symptoms experienced within the context of cancer and MCC. (n>1M) notes from the dataset will be mined using natural language processing. Prediction models with be evaluated to…(need help here)

Aims defined for IIAI:

While the above goals are pursued by Professor Sanvesh's team,
The team wanted to explore the publicly available MIMIC IV dataset to look at two specific problems.
Aim 1:
Given the admissions, diagonis_icds, patient demographics, hospitalization data predict the length of stay per hospital visit of a patient.
Aim2:
Given the same data, predict the mortality of a given patient.

Summary of findings:(still working on this)

1) **Regression approaches (Length of stay Prediction)**
    a. The published results of a recent paper on the MIMIC IV dataset (https://www.mdpi.com/2076-3417/13/12/6930), which focused on a subset of the dataset, considering only Length of Stay (LoS) prediction for 1-21 days, specifically for ICU stays/hospitalization, achieved an R2 score of 0.24.
    b. However, our objective is to explore various subsets of the MIMIC IV dataset, not limited to ICU stays.
        i. Full MIMIC IV dataset
        ii. chronic-condition-dataset (explained below)
        iii. Dataset with LOS only ranging from 1-31 days.

c. Our best-performing model is Linear Regression with Principal Component Analysis (PCA), which achieved an R2 score of 0.295 on the dataset with LoS ranging from 1-31 days. Similarly, the best scores for the full MIMIC IV dataset and chronic condition dataset are 0.181 and 0.135, respectively.
d. See below for more details

2) **Classification approaches**
   a. Most recent publications focus exclusively on ICU stays within the MIMIC IV dataset. The main reason is the unavailability of outpatient mortality information in the MIMIC datasets, which may lead to incorrect mortality rates for any given hospitalization.
      i. https://pubmed.ncbi.nlm.nih.gov/35626224/
      ii. https://bmcanesthesiol.biomedcentral.com/articles/10.1186/s12871-023-02138-5
      iii. https://arxiv.org/pdf/2110.08949.pdf
   b. However, we worked on Full MIMIC IV dataset
   c. Our best-performing model achieved an Area Under the Receiver Operating Characteristic (AUROC) score of 0.892 using Multilayer Perceptron-based neural networks. (Results might be comparable to the other papers.)
   d. However, F1 score is poor.
   e. See detailed results below

3) Additionally, the IIAI team aimed to explore the prediction of Length of Stay as a classification problem, treating LoS ≤ 3 days as short stay and >3 days as long stay, similar to this paper (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8135024/)..
   a. Best AUROC is about 0.87 (Comparable with paper)
   b. See detailed results below

Data for Aims:

Admissions: This component refers to the records of patient admissions to the hospital. It includes information such as admission and discharge dates, admission type (e.g., emergency, elective), admission location, and other administrative details related to the patient's stay in the hospital.

Diagnosis_ICDs: This component consists of International Classification of Diseases (ICD) codes assigned to patients to describe their medical diagnoses. ICD codes are standardized codes used globally to classify and categorize various health conditions, diseases, and injuries. In the MIMIC-IV dataset, the Diagnosis_ICDs component provides information on the specific diagnoses assigned to patients during their hospitalization.

Patient Demographics: This component contains demographic information about the patients in the MIMIC-IV dataset. It includes details such as patient age, gender, ethnicity, marital status, and other relevant personal information. Patient demographics are crucial for analyzing patient populations, studying health disparities, and understanding the impact of demographic factors on healthcare outcomes.

The above three datasets are combined to form a combined dataset for the project. It has 521111 hospital admissions × 19943 feature(ICD, demographics and admissions).

AI/ML Approach:

In our project, we employ a holistic approach to address the problem at hand, which involves benchmarking the performance of various machine learning models against traditional linear and logistic regression models. This approach is aimed at assessing the effectiveness of alternative algorithms in

tackling our specific research objectives. In Aim 1, we evaluate the predictive power of models in a regression context, where we compare linear regression (LnR) to more advanced techniques such as a Multi-layer Perceptron-based Artificial Neural Network for Regression (NNR) and a novel approach referred to as Multi-task Autoencoder Regression (MAER). Aim 2, on the other hand, focuses on classification tasks, comparing logistic regression (LogR) with Naïve Bayes (NB), Multi-layer Perceptron-based Artificial Neural Network for Classification (NNC), and another method we've developed, Multi-task Autoencoder Classification (MAEC).

Data Pre-processing:

In the MIMIC-IV dataset, the ICD codes are very diverse. But the ultimate goal of the project is to analyze cancer specific data. Hence the following data preprocessing steps are applied.

- First converting the ICD9 codes in the data to Corresponding ICD10 codes. (We have this mapping)
- Remove all those features whose proportion of zeros in each column (as computed in step 3) is greater than 0.99999 (almost all zeros)
- Added back all the cancer list ICD codes which belongs to the family of "C00", "C01", "C02", "C03", "C04", "C05", "C06", "C07", "C08", "C09",

"C10", "C11", "C12", "C13", "C14", "C15", "C16", "C17", "C18", "C19",
"C20", "C21", "C22", "C23", "C24", "C25", "C26", "C30", "C31", "C32",
"C33", "C34", "C37", "C38", "C39", "C40", "C41", "C43", "C44", "C45",
"C46", "C47", "C48", "C49", "C50", "C51", "C52", "C53", "C54", "C55",
"C56", "C57", "C58", "C59", "C60", "C61", "C62", "C63", "C64", "C65",
"C66", "C67", "C68", "C69", "C70", "C71", "C72", "C73", "C74", "C75",
"C76", "C77", "C78", "C79", "C80", "C81", "C82", "C83", "C84", "C85",
"C86", "C7A", "C7B",
"D00", "D01", "D02", "D03", "D04", "D05", "D06", "D07", "D09",
"D10", "D11", "D12", "D13", "D14", "D15", "D16", "D17", "D18", "D19",
"D20", "D21", "D22", "D23", "D24", "D25", "D26", "D27", "D28", "D29",
"D30", "D31", "D32", "D33", "D34", "D35", "D36", "D37", "D38", "D39","D3A"

We call this dataset a full-dataset which has a total of 4315 features.
However, we repeated all our experiments with a smaller dataset which contains ICD codes of only cancer list in addition to admissions and demographics data. We call this chronic-condition-dataset with 903 features.
Other filters applied to both the dataset include:

- Remove all those hospitalizations with length of stay less than or equal to 0.
- Remove all those hospitalizations whose mortality of a given patient is not available

Note that the length of stay is in days and mortality is a binary target.
Finally the dataset for the datasets are:
full-dataset = 520943 x 4315
chronic-condition-dataset = 520943 x 903

Model design:

While the classical regression and classification models are self-explanatory, it's worth noting the unique attributes of MAER and MAEC. Autoencoders play a role in our approach, serving to transform high-dimensional data into a lower-dimensional representation while retaining essential information. Initially,

we envisioned encoding features into lower dimensions and then feeding these encoded features into separate Multi-layer Perceptron (MLP) networks for training. However, we transitioned to a multi-task approach, wherein the last layer of the encoder serves as input not only to the decoder but also to a regression block (MLP) in MAER and a classification block (MLP) in MAEC. This design results in each algorithm having two distinct loss functions: one for the decoder and another for either the regression or classification block.

In summary, our research approach involves a rigorous comparison of traditional linear and logistic regression models with advanced techniques, such as NNR, NNC, MAER, and MAEC. This comprehensive assessment framework aims to offer a deeper understanding of the strengths and weaknesses of these models, ultimately enhancing the quality and relevance of our project's findings.

Likewise, the NNC and NNR models are defined below:
At a high level, the model consists of two primary functions: residual_block and simple_ann_model. The residual_block function is a building block used within the neural network, while the simple_ann_model function defines the overall architecture of the model.

The residual_block function defines a single residual block. A residual block is a structural unit that consists of several layers. Within this function, a dense layer is created with a specified number of units and activation function. This dense layer is followed by batch normalization and dropout layers. Importantly, a skip connection is established, which is a core feature of residual networks. This skip connection helps the model propagate gradients effectively during training, even in very deep networks. If the input and output shapes do not match, a projection shortcut is applied to ensure compatibility.

The simple_ann_model function serves as the main architecture of the neural network. It starts with an input layer that is shaped to match the dimensions of your input data. This input layer is followed by an initial dense layer with 512 units, a ReLU activation function, and regularization techniques (L1 and L2) to prevent overfitting. Dropout is also applied to this layer to further regularize the model.

Following the initial layer, several residual blocks are stacked together. Each residual block is connected to the previous one, and together they form a deep network. These residual blocks capture hierarchical features and patterns in the data. The specific number of residual blocks and their hyperparameters can be adjusted based on the complexity of your problem.

Finally, the output layer of the model is defined with a linear/sigmoid activation function for NNR and NNC respectively. Depending on the task (regression or classification), the number of output units can be adjusted by setting OUTPUT_CHANNELS. This output layer produces the final predictions of the model.

In summary, the model combines the power of residual networks with dropout and regularization techniques to create a deep neural network architecture capable of learning complex relationships in your data. It is a versatile architecture that can be customized and fine-tuned to suit the specific requirements of your research.

Experimental methods:

In our research project, we have designed a comprehensive set of experiments to evaluate the performance of various machine learning models in addressing our research objectives. These experiments encompass both regression and classification tasks and involve different datasets, including PCA (Principal Component Analysis) versions of the data. Additionally, we have considered the impact of outliers, defined as hospitalizations with a length of stay exceeding six months, by conducting experiments with and without them. Here's a detailed list of the variations of model experiments we designed:

Regression Experiments (with and without outliers):

1. **Linear Regression (LnR):**
- Full Dataset (with outliers)
- Chronic Condition Dataset (with outliers)
- PCA Version of Full Dataset (with outliers)
- PCA Version of Chronic Condition Dataset (with outliers)
- Full Dataset (without outliers)
- Chronic Condition Dataset (without outliers)
- PCA Version of Full Dataset (without outliers)
- PCA Version of Chronic Condition Dataset (without outliers)

2. **Multi-layer Perceptron-based Artificial Neural Network for Regression (NNR):**
- Full Dataset (with outliers)
- Chronic Condition Dataset (with outliers)
- PCA Version of Full Dataset (with outliers)
- PCA Version of Chronic Condition Dataset (with outliers)
- Full Dataset (without outliers)
- Chronic Condition Dataset (without outliers)
- PCA Version of Full Dataset (without outliers)
- PCA Version of Chronic Condition Dataset (without outliers)

3. **Multi-task Autoencoder Regression (MAER):**
- Full Dataset (without outliers)
- Chronic Condition Dataset (without outliers)

Classification Experiments:

4. **Logistic Regression (LogR):**
- Full Dataset
- Chronic Condition Dataset
- PCA Version of Full Dataset
- PCA Version of Chronic Condition Dataset

5. **Naïve Bayes (NB):**
- Full Dataset
- Chronic Condition Dataset
- PCA Version of Full Dataset
- PCA Version of Chronic Condition Dataset

6. **Multi-layer Perceptron-based Artificial Neural Network for Classification (NNC):**
- Full Dataset
- Chronic Condition Dataset
- PCA Version of Full Dataset
- PCA Version of Chronic Condition Dataset

7. **Multi-task Autoencoder Classification (MAEC):**
- Full Dataset (Non-PCA)
- Chronic Condition Dataset (Non-PCA)

In summary, we have conducted a thorough set of experiments by varying the machine learning models (LnR, NNR, MAER, LogR, NB, NNC, MAEC) and the datasets (Full Dataset, Chronic Condition

Dataset) as well as considering the application of PCA for dimensionality reduction in some cases. This comprehensive approach allows us to assess the performance of these models under various conditions, providing valuable insights into their effectiveness for both regression and classification tasks in our research.

The train: test split is 70:30.

The code for all the models can be found here:
https://research-git.uiowa.edu/iiai-consultants/Nursing/Ai_in_nursing

Results:

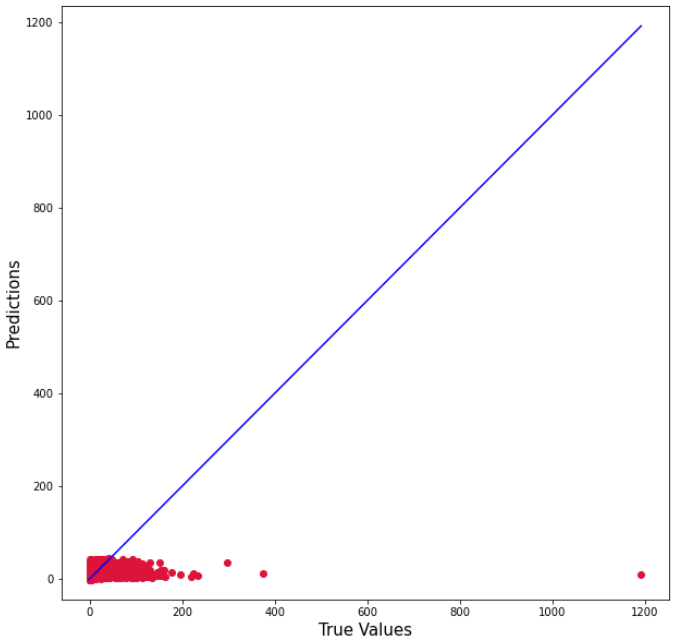
Regression Test Results- R^2 values:

Model Name	Full dataset with Outliers	Chronic Condition Dataset with outliers	Full dataset without Outliers	Chronic Condition Dataset with without outliers
PCA+ LnR	0.166	0.128	0.196	0.130
Non PCA+ LnR	-0.007	0.129	-0.008	0.132
PCA+ NNR	0.181	0.135	0.19386	0.135
Non PCA+ NNR	-.0000069	0.139	-.000006	0.135
MAER	0.137	0.126	0.160	0.115

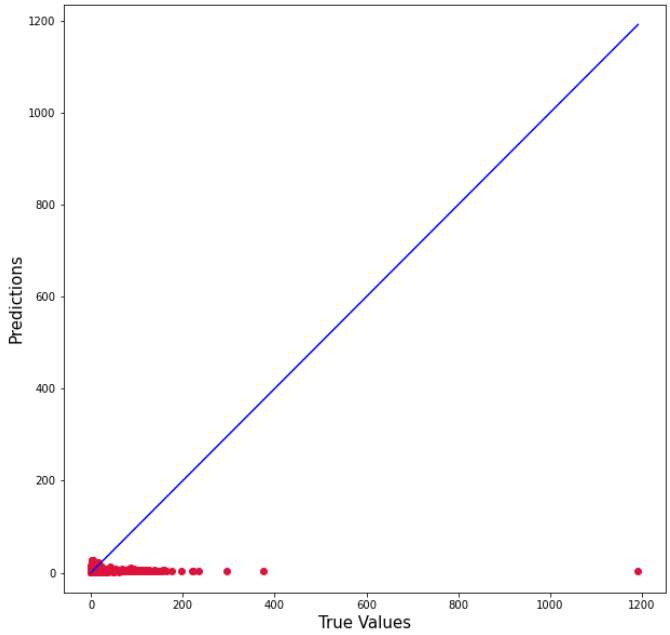
The best of all experiments above is PCA+LNR for full dataset with outliers. Which is 0.196

It is difficult to compare these results with existing papers because almost all of the papers worked only with ICU stay data and not with full hospitalization data.
If we compare the above results with a similar paper
(Prediction of Intensive Care Unit Length of Stay in the MIMIC-IV Dataset) that focused only on LOS for 1-21 days, the results are comparable, as their R2 values range between 0.16 and 0.23.
It is also noteworthy that most of the papers predict LOS as a classification variable rather than as a regression variable

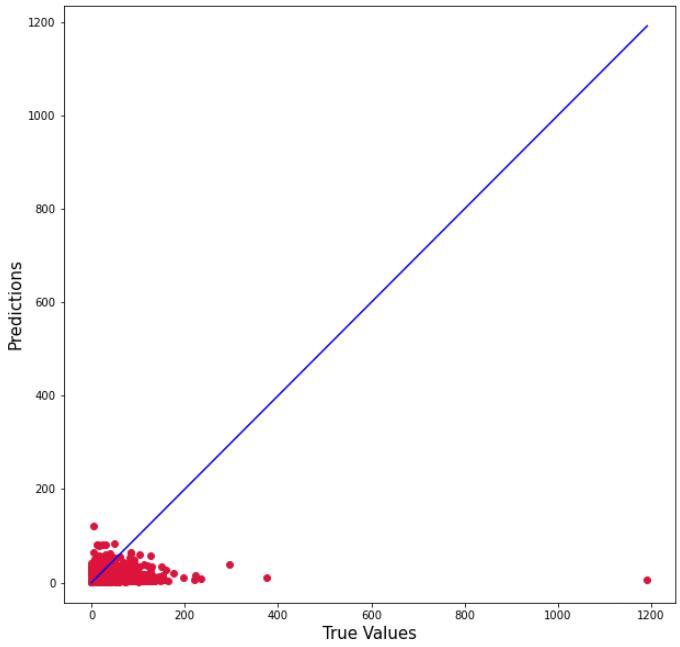
PCA+ LnR - Full dataset with Outliers:



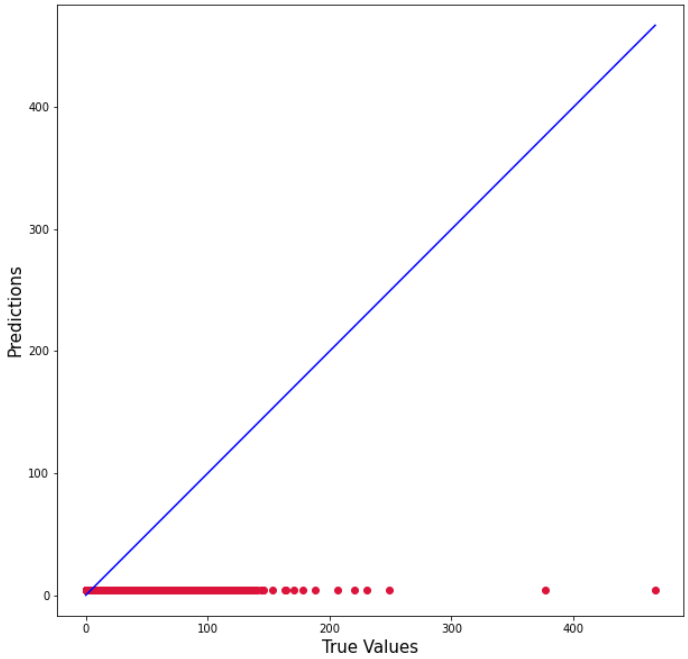
Non-PCA- Full dataset with Outliers:



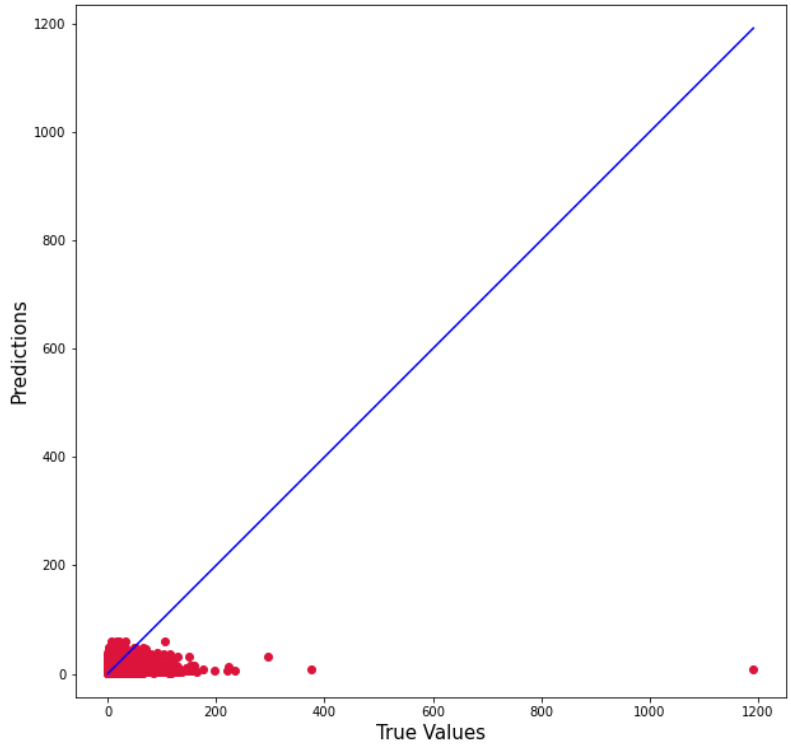
PCA+ NN- Full dataset with Outliers :



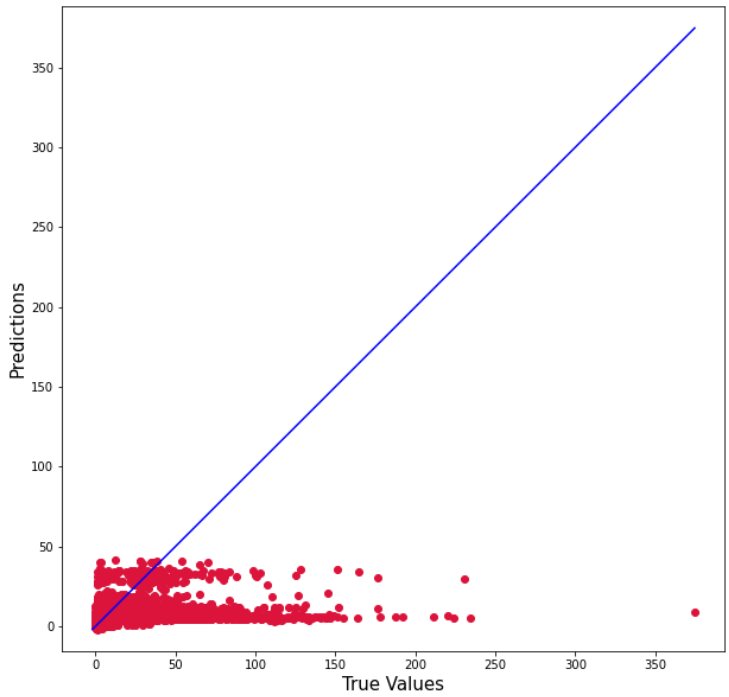
Non PCA+ NN- Full dataset with Outliers:



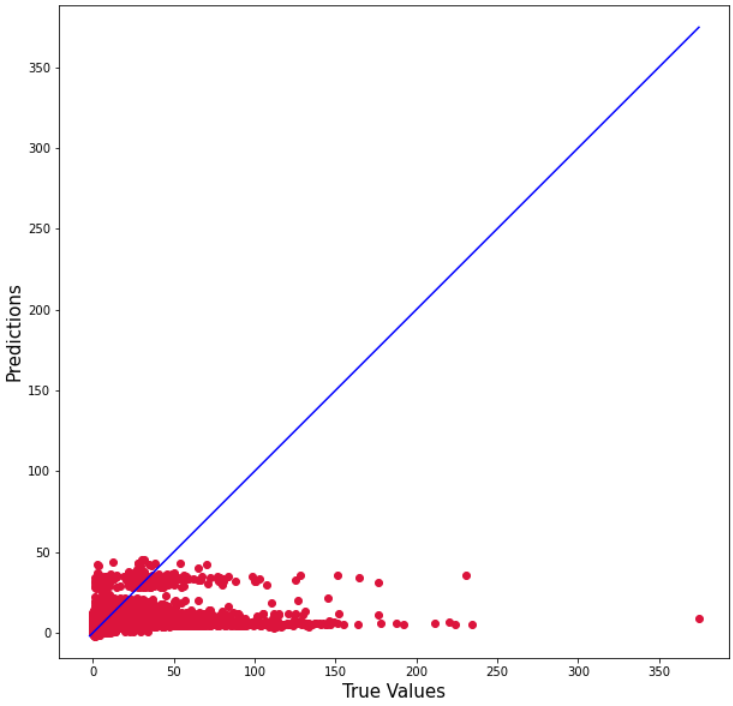
MAER - Full dataset with Outliers :



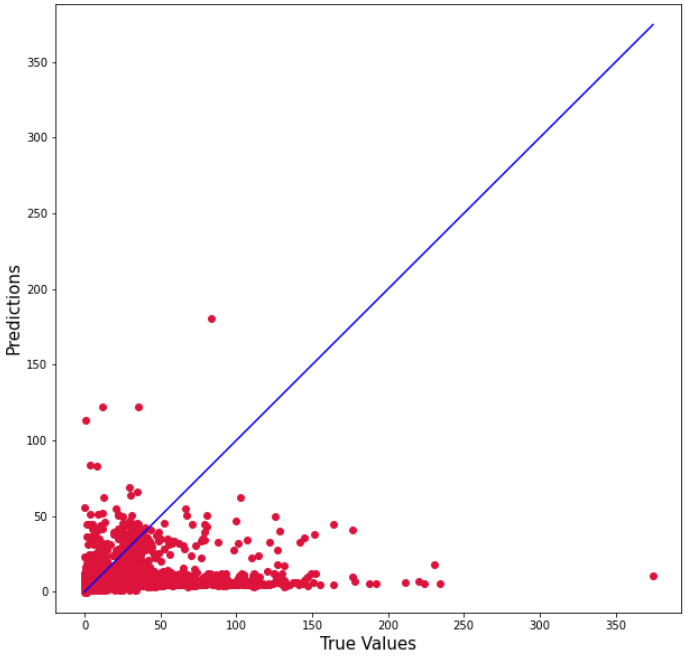
PCA+ LnR – Chronic condition dataset with Outliers :



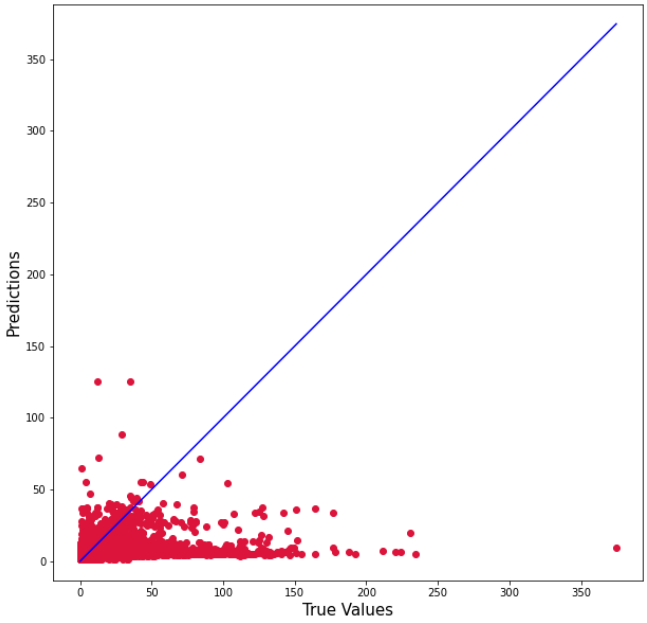
Non-PCA- Chronic condition dataset with Outliers:



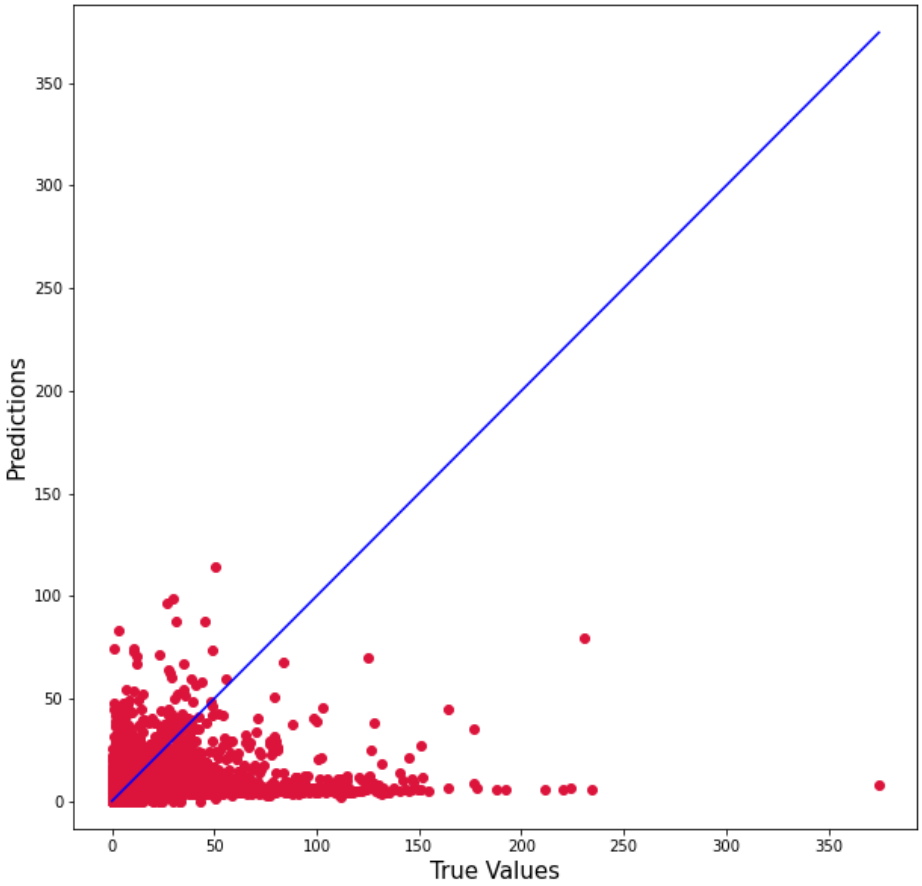
PCA+ NN- Chronic condition dataset with Outliers :



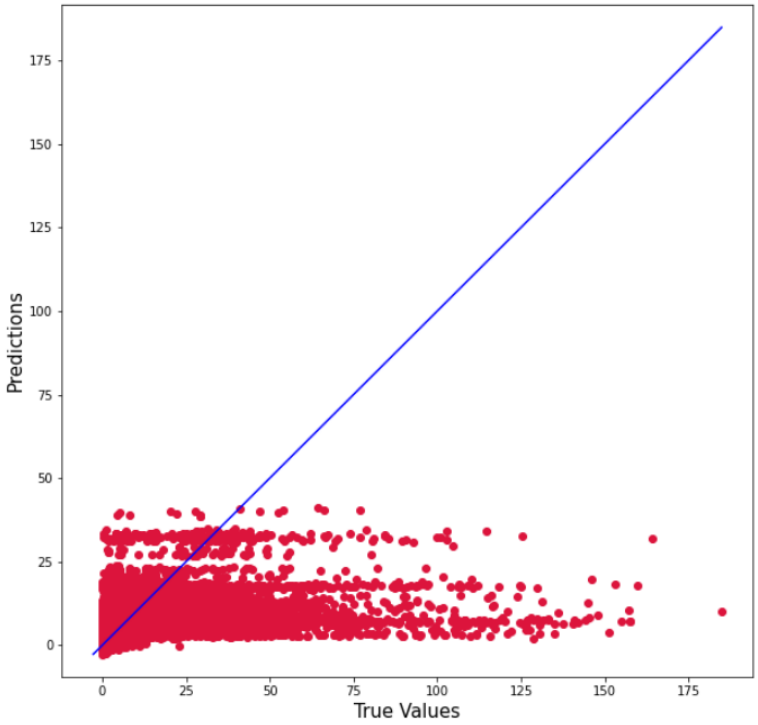
Non PCA+ NN- Chronic condition dataset with Outliers:



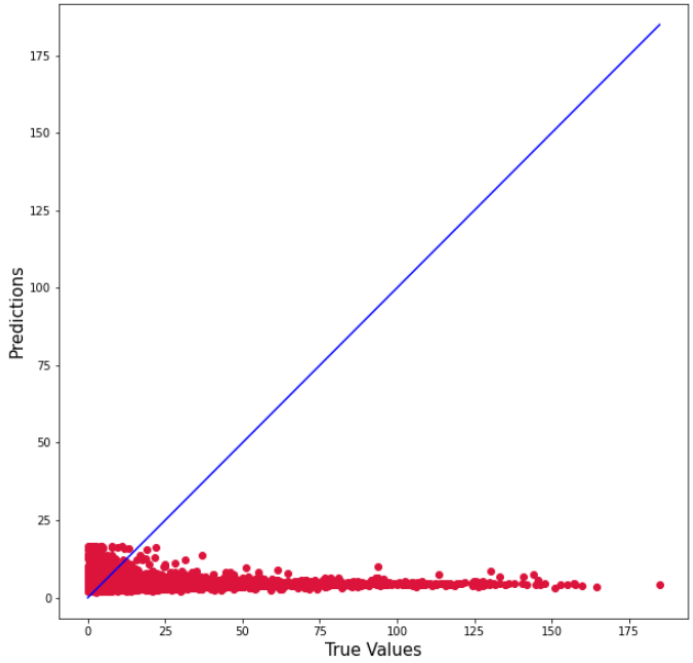
MAER - Chronic condition dataset with Outliers :



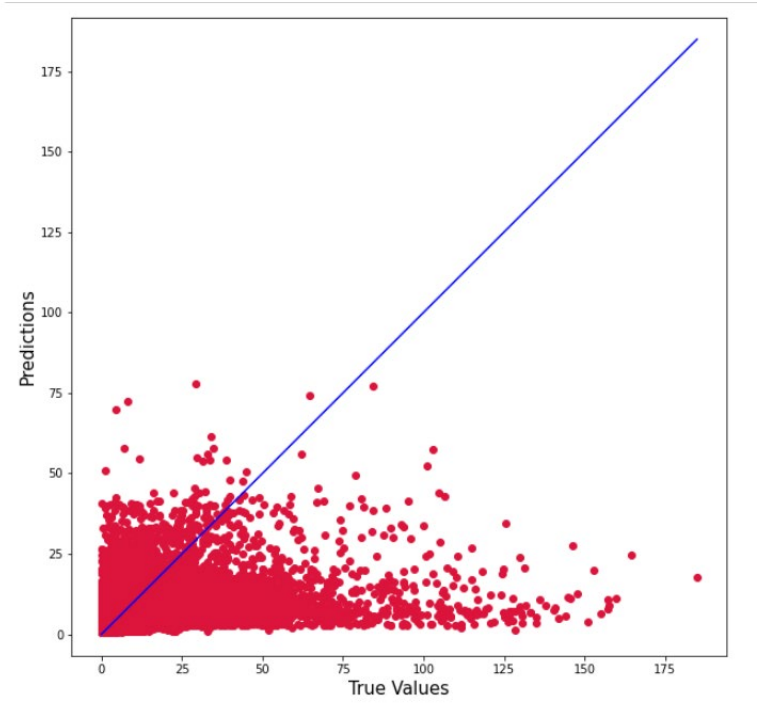
PCA+ LnR - Full dataset without Outliers :



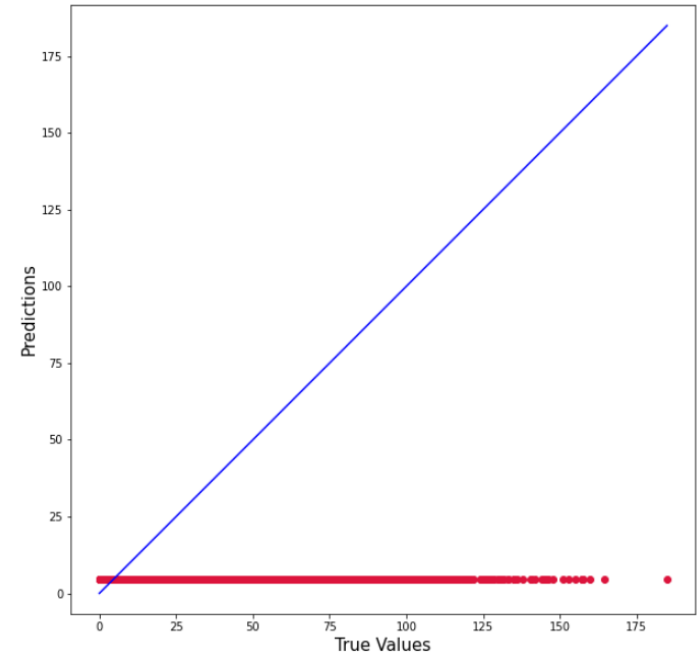
Non-PCA- Full dataset without Outliers:



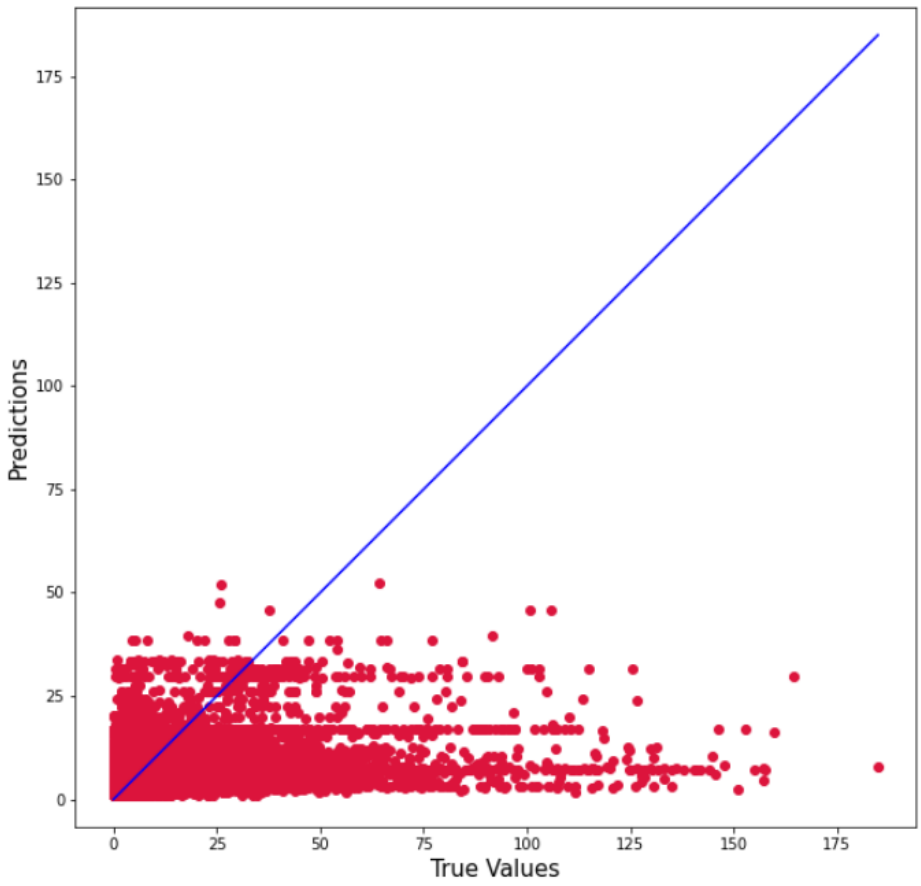
PCA+ NN- Full dataset without Outliers :



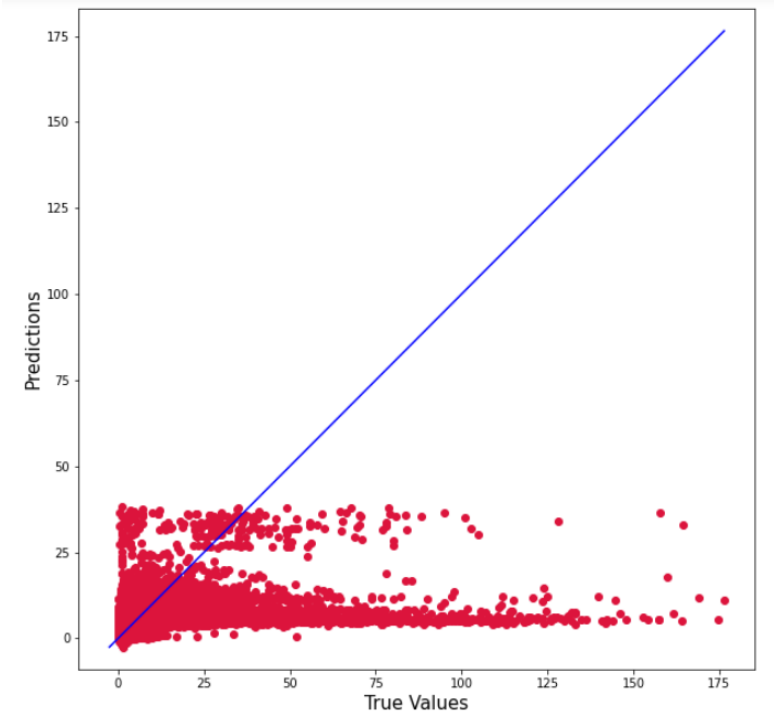
Non PCA+ NN- Full dataset without Outliers:



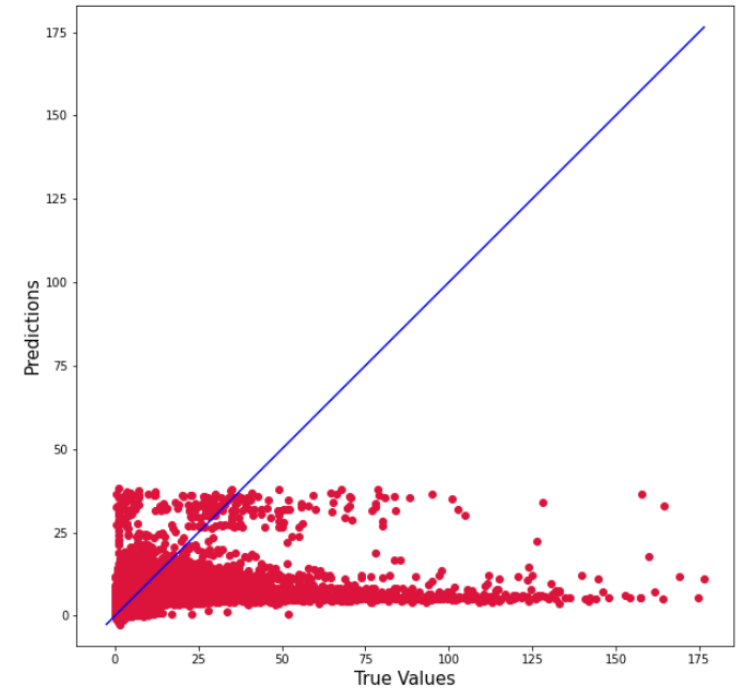
MAER - Full dataset without Outliers :


Predictions
True Values
0
25
50
75
100
125
150
175

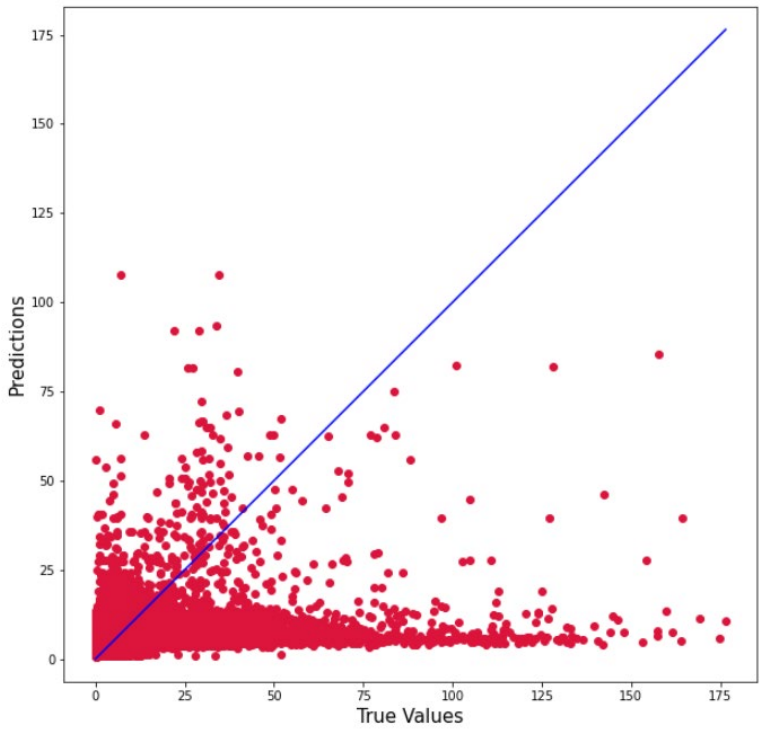
PCA+ LnR – Chronic condition dataset without Outliers :



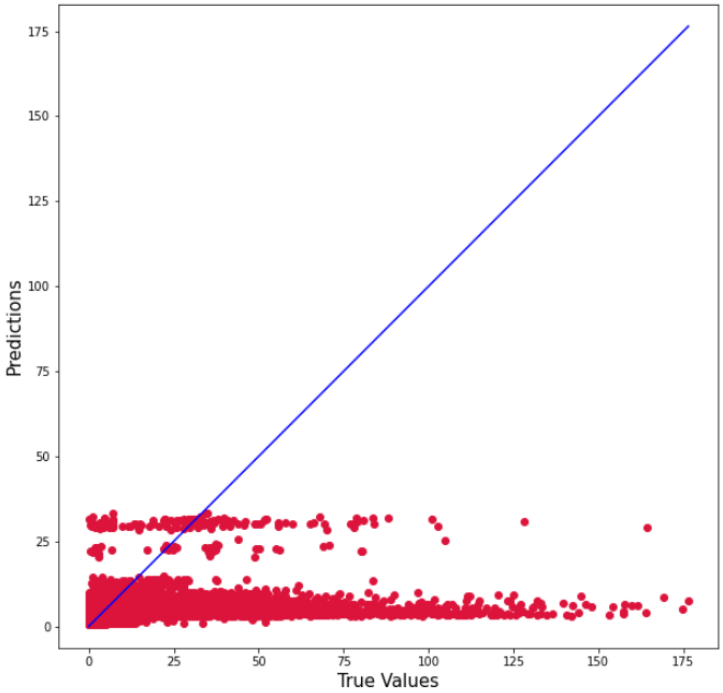
Non-PCA- Chronic condition dataset without Outliers:



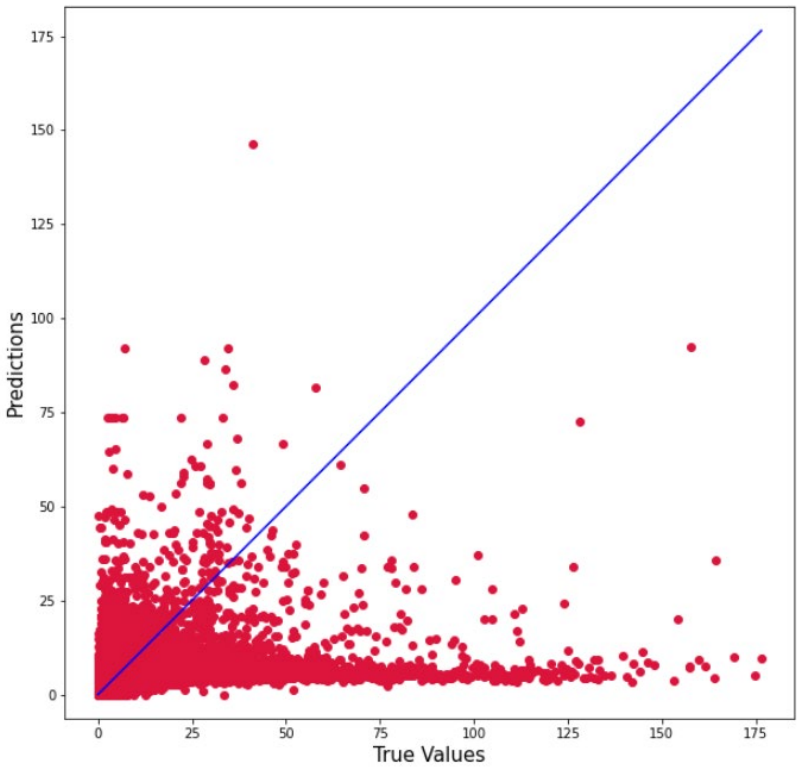
PCA+ NN- Chronic condition dataset without Outliers :



Non PCA+ NN- Chronic condition dataset without Outliers:



MAER - Chronic condition dataset without Outliers :



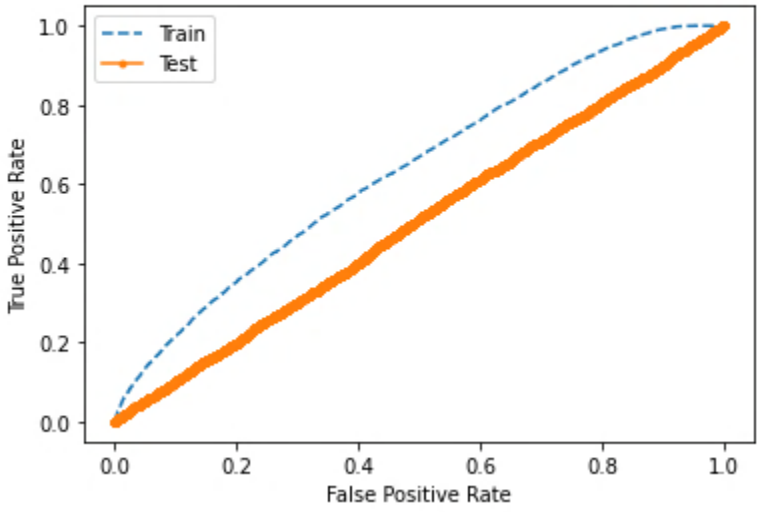
Classification Test Results:

Across the entire dataset, none of the models performed well in predicting mortality. While some models displayed high AUROC values, they exhibited lower precision, recall, and F1 scores, indicating significant class imbalance.

Here are the results:

PCA+ Logistic Regression- Full dataset:

```
Train: ROC AUC=0.636
Test: ROC AUC=0.504
```



train report

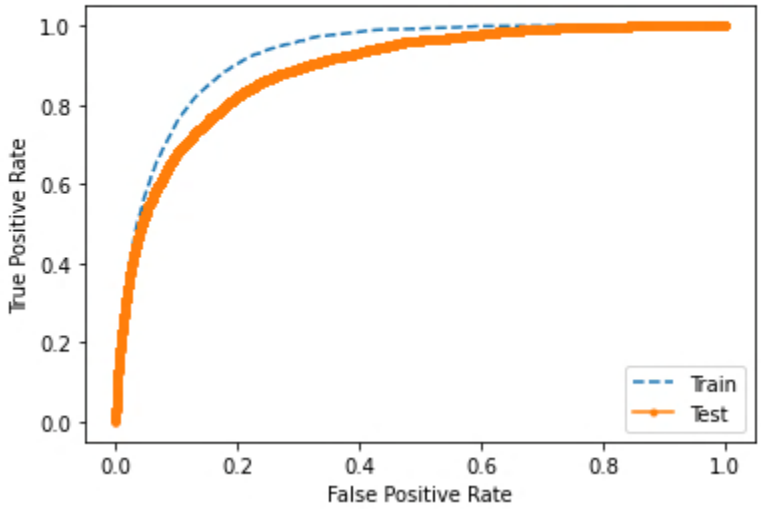
	precision	recall	f1-score	support
No_death	0.89	0.41	0.56	358211
Death	0.20	0.75	0.32	71642
accuracy			0.47	429853
macro avg	0.55	0.58	0.44	429853
weighted avg	0.78	0.47	0.52	429853

test report

	precision	recall	f1-score	support
No_death	0.98	0.41	0.58	153519
Death	0.02	0.60	0.04	2764
accuracy			0.41	156283
macro avg	0.50	0.51	0.31	156283
weighted avg	0.97	0.41	0.57	156283

Non-PCA-Logistic regression- Full dataset :

```
Train: ROC AUC=0.927
Test: ROC AUC=0.891
```



train report

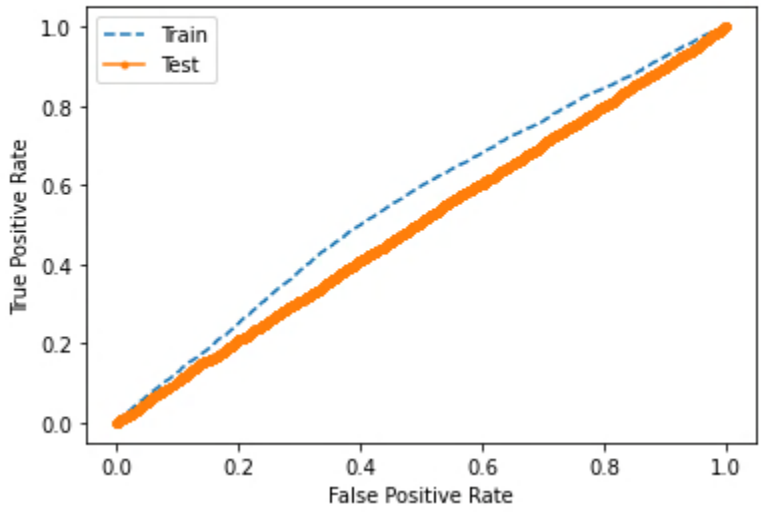
	precision	recall	f1-score	support
No_death	0.74	0.93	0.82	179105
Death	0.90	0.67	0.77	179105
accuracy			0.80	358210
macro avg	0.82	0.80	0.80	358210
weighted avg	0.82	0.80	0.80	358210

test report

	precision	recall	f1-score	support
No_death	0.99	0.93	0.96	153519
Death	0.13	0.61	0.21	2764
accuracy			0.92	156283
macro avg	0.56	0.77	0.58	156283
weighted avg	0.98	0.92	0.94	156283

PCA+ Naivebayes- Full dataset:

```
Train: ROC AUC=0.558
Test: ROC AUC=0.503
```



train report

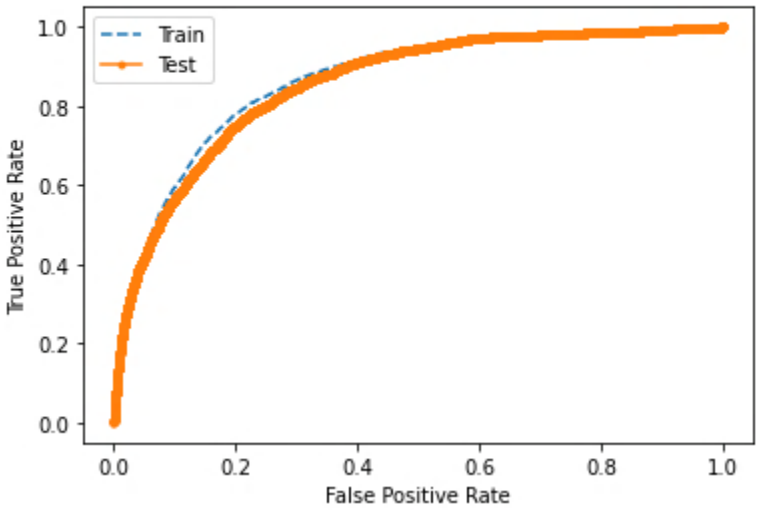
	precision	recall	f1-score	support
No_death	0.85	0.66	0.74	358211
Death	0.20	0.44	0.28	71642
accuracy			0.62	429853
macro avg	0.53	0.55	0.51	429853
weighted avg	0.75	0.62	0.67	429853

test report

	precision	recall	f1-score	support
No_death	0.98	0.66	0.79	153519
Death	0.02	0.35	0.03	2764
accuracy			0.65	156283
macro avg	0.50	0.50	0.41	156283
weighted avg	0.97	0.65	0.78	156283

NON-PCA+ Naivebayes- Full dataset:

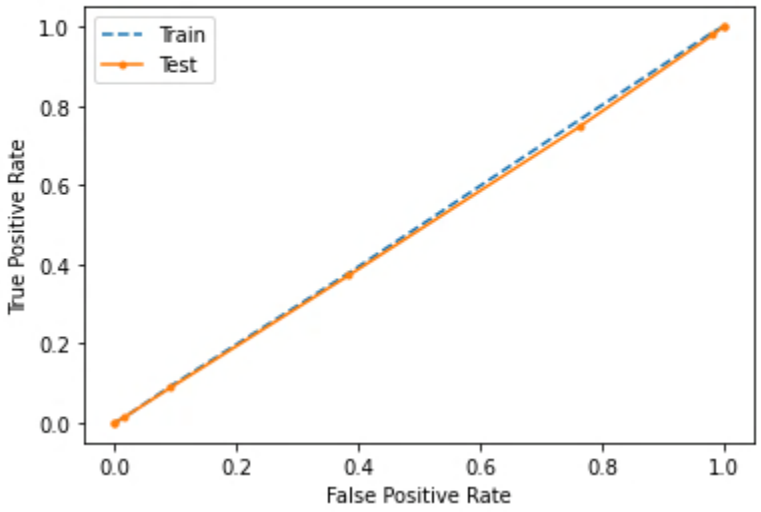
```
Train: ROC AUC=0.865
Test: ROC AUC=0.856
```




```
train report
              precision    recall  f1-score   support

    No_death       0.87      0.62      0.72    179105
       Death       0.70      0.91      0.79    179105

    accuracy                           0.76    358210
   macro avg       0.79      0.76      0.76    358210
weighted avg       0.79      0.76      0.76    358210

test report
              precision    recall  f1-score   support

    No_death       1.00      0.62      0.76    153519
       Death       0.04      0.90      0.08      2764

    accuracy                           0.62    156283
   macro avg       0.52      0.76      0.42    156283
weighted avg       0.98      0.62      0.75    156283
```

PCA+ NNC(neural nets)- Full dataset:

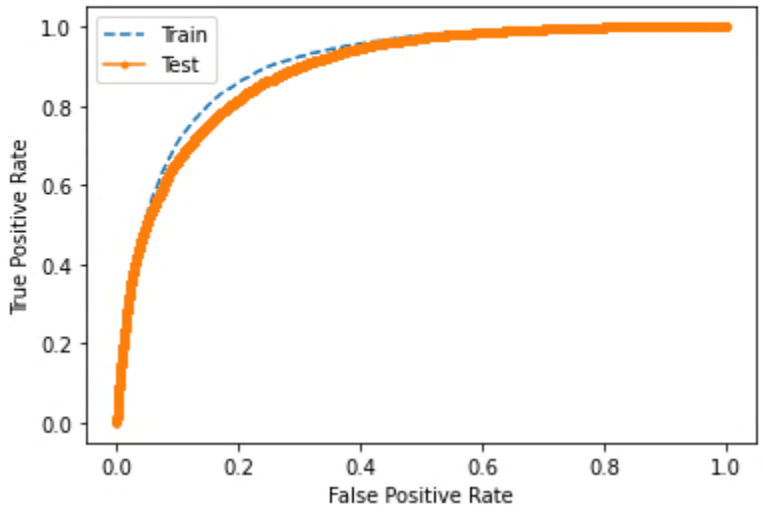
```
Train: ROC AUC=0.499
Test: ROC AUC=0.490
```




```
train report
              precision    recall  f1-score   support

    No_death       0.83      1.00      0.91    358210
       Death       0.00      0.00      0.00     71640

    accuracy                           0.83    429850
   macro avg       0.42      0.50      0.45    429850
weighted avg       0.69      0.83      0.76    429850

test report
              precision    recall  f1-score   support

    No_death       0.98      1.00      0.99    153519
       Death       0.00      0.00      0.00      2764

    accuracy                           0.98    156283
   macro avg       0.49      0.50      0.50    156283
weighted avg       0.96      0.98      0.97    156283
```

Non-PCA+ NNC(neural nets)- Full dataset:

```
Train: ROC AUC=0.906
Test: ROC AUC=0.892
```



train report

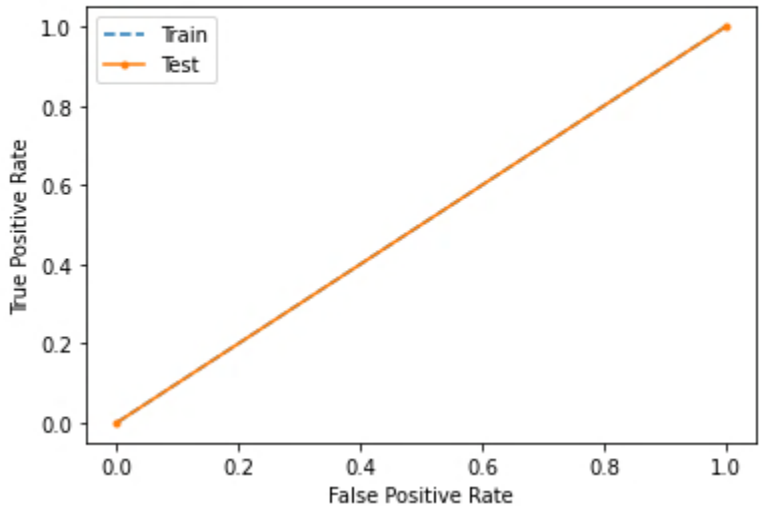
	precision	recall	f1-score	support
No_death	0.75	0.90	0.82	179105
Death	0.88	0.70	0.78	179105
accuracy			0.80	358210
macro avg	0.81	0.80	0.80	358210
weighted avg	0.81	0.80	0.80	358210

test report

	precision	recall	f1-score	support
No_death	0.99	0.90	0.95	153519
Death	0.11	0.66	0.18	2764
accuracy			0.90	156283
macro avg	0.55	0.78	0.56	156283
weighted avg	0.98	0.90	0.93	156283

MAEC- Full dataset:

```
Train: ROC AUC=0.500
Test: ROC AUC=0.500
```



train report

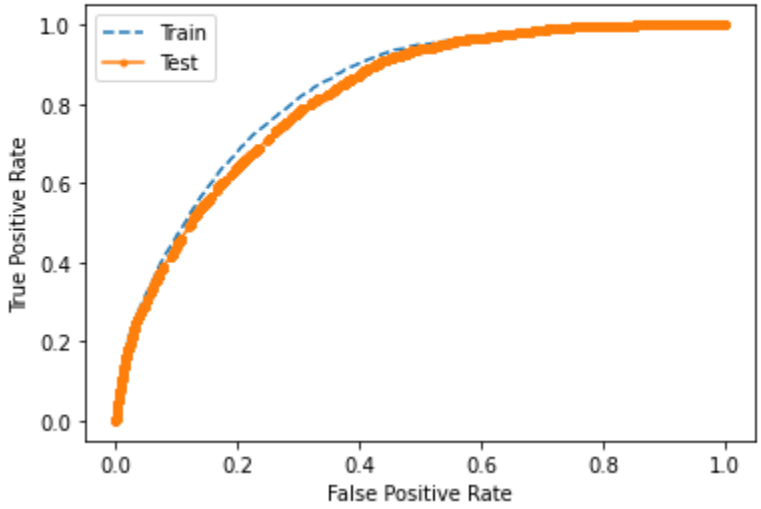
	precision	recall	f1-score	support
No_death	0.50	1.00	0.67	179105
Death	0.00	0.00	0.00	179105
accuracy			0.50	358210
macro avg	0.25	0.50	0.33	358210
weighted avg	0.25	0.50	0.33	358210

test report

	precision	recall	f1-score	support
No_death	0.98	1.00	0.99	153519
Death	0.00	0.00	0.00	2764
accuracy			0.98	156283
macro avg	0.49	0.50	0.50	156283
weighted avg	0.96	0.98	0.97	156283

PCA+ Logistic Regression- Chronic condition dataset:

```
Train: ROC AUC=0.836
Test: ROC AUC=0.821
```



train report

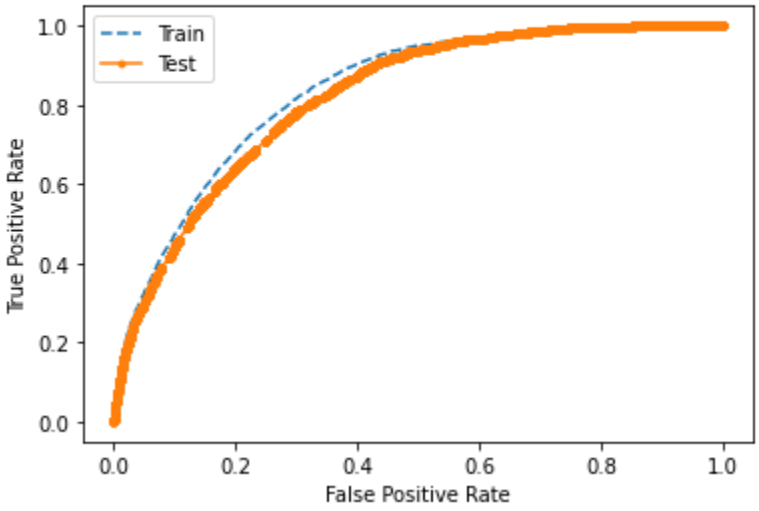
	precision	recall	f1-score	support
No_death	0.86	0.98	0.92	358211
Death	0.65	0.22	0.33	71642
accuracy			0.85	429853
macro avg	0.76	0.60	0.62	429853
weighted avg	0.83	0.85	0.82	429853

test report

	precision	recall	f1-score	support
No_death	0.99	0.98	0.98	153519
Death	0.13	0.20	0.16	2764
accuracy			0.96	156283
macro avg	0.56	0.59	0.57	156283
weighted avg	0.97	0.96	0.97	156283

Non-PCA-Logistic regression- Chronic condition dataset:

```
Train: ROC AUC=0.837
Test: ROC AUC=0.821
```



train report

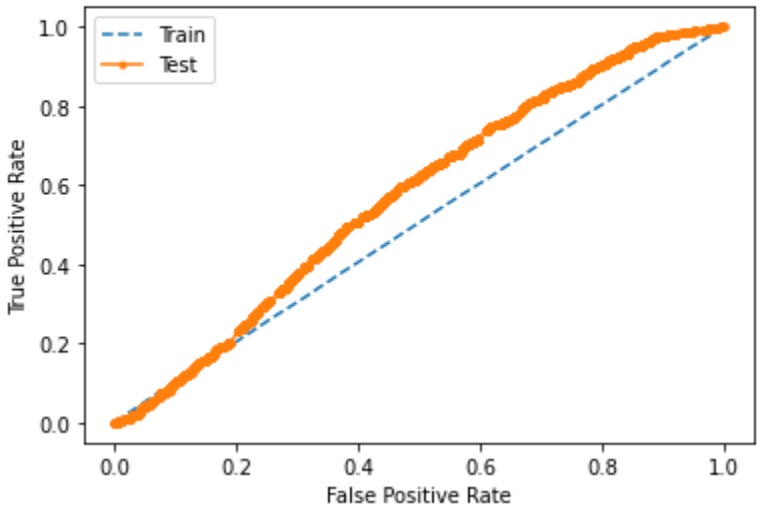
	precision	recall	f1-score	support
No_death	0.87	0.97	0.91	358211
Death	0.61	0.27	0.38	71642
accuracy			0.85	429853
macro avg	0.74	0.62	0.65	429853
weighted avg	0.83	0.85	0.83	429853

test report

	precision	recall	f1-score	support
No_death	0.99	0.96	0.98	153519
Death	0.12	0.25	0.16	2764
accuracy			0.95	156283
macro avg	0.55	0.61	0.57	156283
weighted avg	0.97	0.95	0.96	156283

PCA+ Naivebayes- Chronic condition dataset:

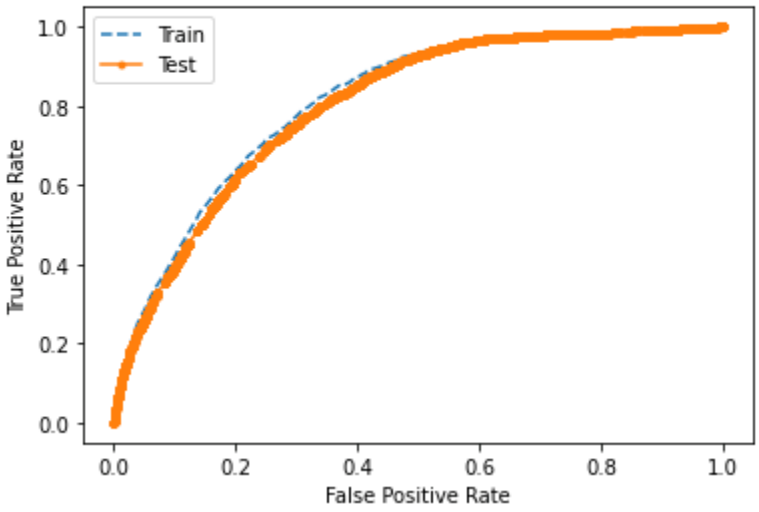
```
Train: ROC AUC=0.504
Test: ROC AUC=0.575
```




```
train report
              precision    recall  f1-score   support

    No_death       0.83      1.00      0.91    358211
       Death       0.00      0.00      0.00     71642

    accuracy                           0.83    429853
   macro avg       0.42      0.50      0.45    429853
weighted avg       0.69      0.83      0.76    429853

test report
              precision    recall  f1-score   support

    No_death       0.98      1.00      0.99    153519
       Death       0.00      0.00      0.00      2764

    accuracy                           0.98    156283
   macro avg       0.49      0.50      0.50    156283
weighted avg       0.96      0.98      0.97    156283
```

NON-PCA+ Naivebayes- Chronic condition dataset :

```
Train: ROC AUC=0.813
Test: ROC AUC=0.803
```



train report

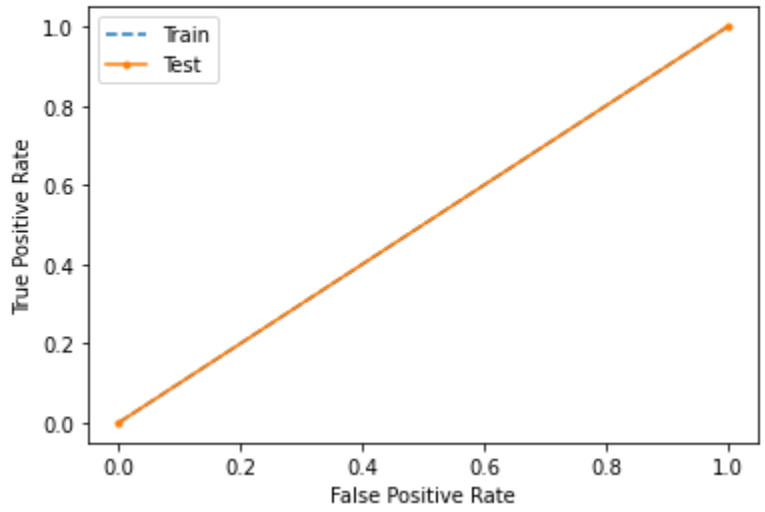
	precision	recall	f1-score	support
No_death	0.89	0.90	0.89	358211
Death	0.46	0.42	0.44	71642
accuracy			0.82	429853
macro avg	0.67	0.66	0.66	429853
weighted avg	0.81	0.82	0.82	429853

test report

	precision	recall	f1-score	support
No_death	0.99	0.90	0.94	153519
Death	0.07	0.39	0.11	2764
accuracy			0.89	156283
macro avg	0.53	0.65	0.53	156283
weighted avg	0.97	0.89	0.93	156283

PCA+ NNC(neural nets)- Chronic condition dataset :

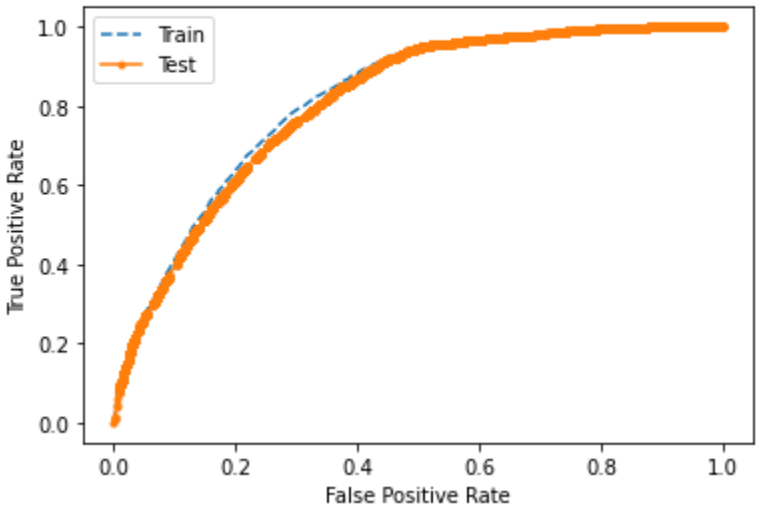
```
Train: ROC AUC=0.500
Test: ROC AUC=0.500
```




```
train report
              precision    recall  f1-score   support

    No_death       0.83      1.00      0.91    358211
       Death       0.00      0.00      0.00     71642

    accuracy                           0.83    429853
   macro avg       0.42      0.50      0.45    429853
weighted avg       0.69      0.83      0.76    429853

test report
              precision    recall  f1-score   support

    No_death       0.98      1.00      0.99    153519
       Death       0.00      0.00      0.00      2764

    accuracy                           0.98    156283
   macro avg       0.49      0.50      0.50    156283
weighted avg       0.96      0.98      0.97    156283
```

Non-PCA+ NNC(neural nets)- Chronic condition dataset:

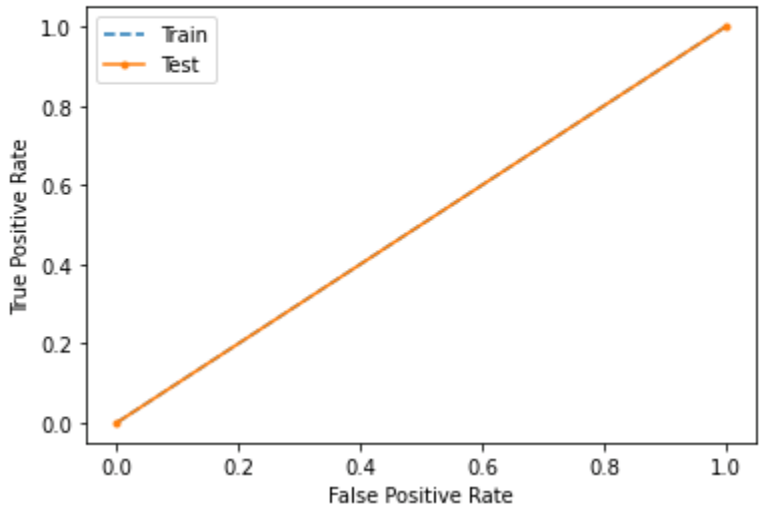
```
Train: ROC AUC=0.818
Test: ROC AUC=0.809
```




```
train report
              precision    recall  f1-score   support

    No_death       0.86      0.97      0.91    358211
       Death       0.57      0.20      0.29     71642

    accuracy                           0.84    429853
   macro avg       0.71      0.58      0.60    429853
weighted avg       0.81      0.84      0.81    429853

test report
              precision    recall  f1-score   support

    No_death       0.99      0.97      0.98    153519
       Death       0.10      0.20      0.14      2764

    accuracy                           0.96    156283
   macro avg       0.55      0.58      0.56    156283
weighted avg       0.97      0.96      0.96    156283
```

MAEC- Chronic condition dataset :

```
Train: ROC AUC=0.500
Test: ROC AUC=0.500
```



train report

	precision	recall	f1-score	support
No_death	0.83	1.00	0.91	358211
Death	0.00	0.00	0.00	71642
accuracy			0.83	429853
macro avg	0.42	0.50	0.45	429853
weighted avg	0.69	0.83	0.76	429853

test report

	precision	recall	f1-score	support
No_death	0.98	1.00	0.99	153519
Death	0.00	0.00	0.00	2764
accuracy			0.98	156283
macro avg	0.49	0.50	0.50	156283
weighted avg	0.96	0.98	0.97	156283

In addition to the above-mentioned planned experiments, the IIAI team explored two more questions related to length of stay.

1. Focusing solely on hospitalizations with a length of stay equal to or less than 31 days due to the high variability in longer LOS (Regression Problem).
2. Converting length of stay (LOS) into a classification problem by categorizing hospitalizations with more than 3 days as longer stays and those with 3 days or less as short stays (Classification Problem).

Here are the results of both the experiments

EXPERIMENT-1 – 31days regression:

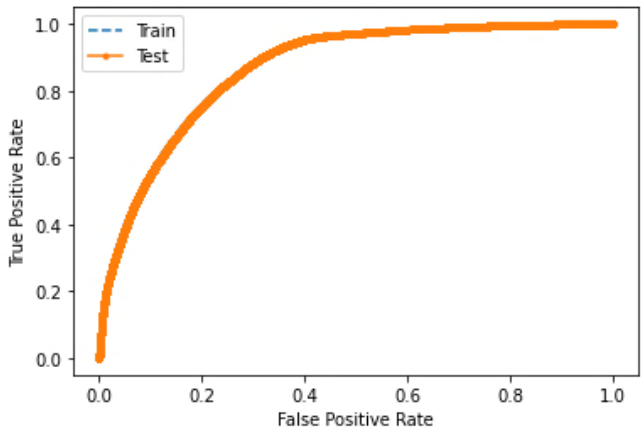
Model Name	Full dataset	Chronic Condition Dataset
PCA+ LnR	0.295	0.217
Non PCA+ LnR	-0.009	0.218
PCA+ NNR	0.311	0.206
Non PCA+ NNR	-.000014	0.2075
MAER	0.21	0.214

Experiments 1 shows that the models perform slightly better on lower LOS hospitalizations. However, the performance is still not significant.

EXPERIMENT-2 – Longstay vs Short stay:

PCA+ Logistic Regression - Fulldataset:

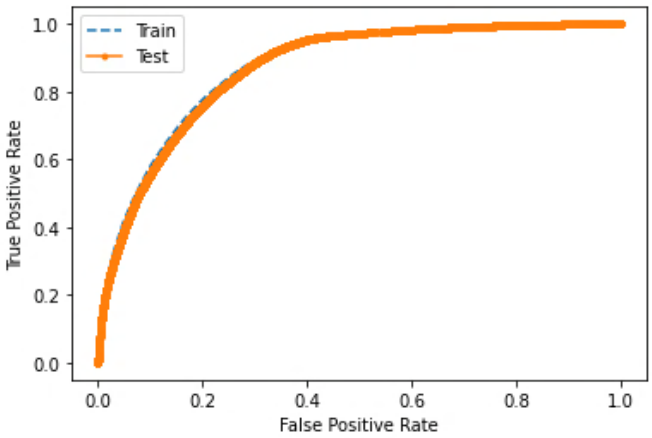
```
Train: ROC AUC=0.872
Test: ROC AUC=0.869
```




```
train report
              precision    recall  f1-score   support

  Short Stay       0.85      0.74      0.79    201860
   Long Stay       0.72      0.84      0.78    162800

    accuracy                           0.79    364660
   macro avg       0.79      0.79      0.79    364660
weighted avg       0.80      0.79      0.79    364660

test report
              precision    recall  f1-score   support

  Short Stay       0.85      0.74      0.79     86512
   Long Stay       0.72      0.84      0.78     69771

    accuracy                           0.78    156283
   macro avg       0.79      0.79      0.78    156283
weighted avg       0.79      0.78      0.78    156283
```

Non-PCA-Logistic regression - Fulldataset:

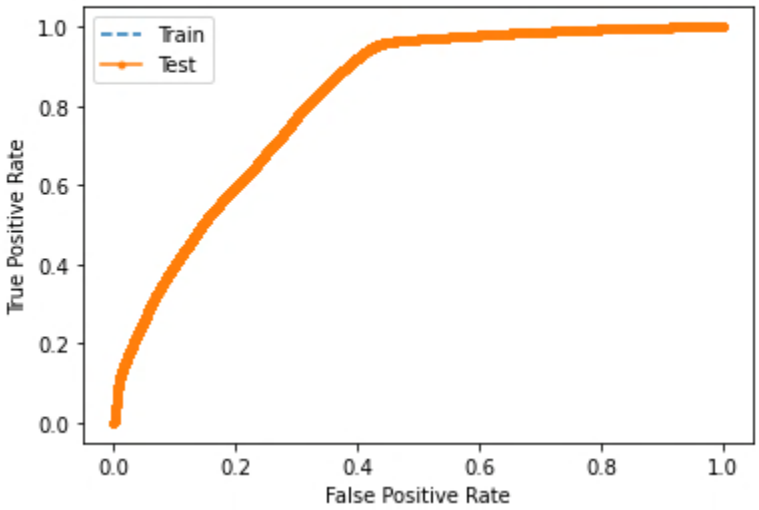
```
Train: ROC AUC=0.876
Test: ROC AUC=0.870
```




```
train report
              precision    recall  f1-score   support

  Short Stay       0.85      0.75      0.80    201860
   Long Stay       0.73      0.84      0.78    162800

    accuracy                           0.79    364660
   macro avg       0.79      0.80      0.79    364660
weighted avg       0.80      0.79      0.79    364660

test report
              precision    recall  f1-score   support

  Short Stay       0.85      0.75      0.79     86512
   Long Stay       0.72      0.83      0.77     69771

    accuracy                           0.78    156283
   macro avg       0.79      0.79      0.78    156283
weighted avg       0.79      0.78      0.78    156283
```

PCA+ Naivebayes - Fulldataset:

```
Train: ROC AUC=0.819
Test: ROC AUC=0.818
```



train report

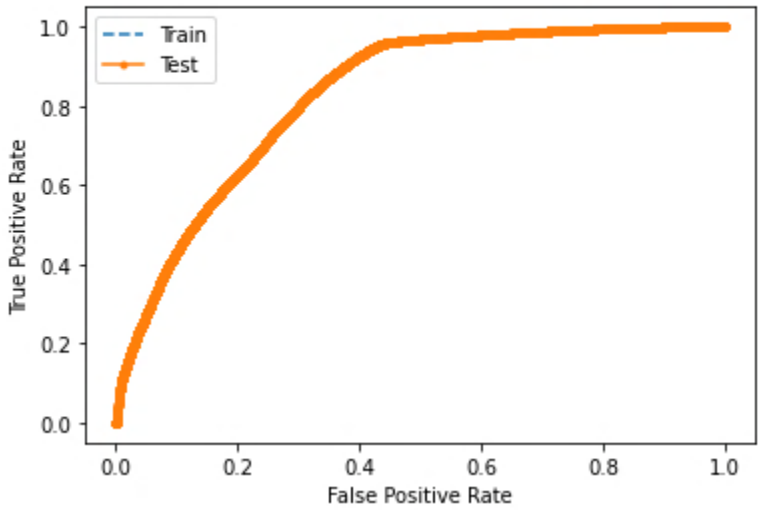
	precision	recall	f1-score	support
Short Stay	0.75	0.74	0.75	201860
Long Stay	0.69	0.69	0.69	162800
accuracy			0.72	364660
macro avg	0.72	0.72	0.72	364660
weighted avg	0.72	0.72	0.72	364660

test report

	precision	recall	f1-score	support
Short Stay	0.75	0.74	0.75	86512
Long Stay	0.68	0.69	0.69	69771
accuracy			0.72	156283
macro avg	0.72	0.72	0.72	156283
weighted avg	0.72	0.72	0.72	156283

NON-PCA+ Naivebayes - Fulldataset:

```
Train: ROC AUC=0.828
Test: ROC AUC=0.828
```



train report

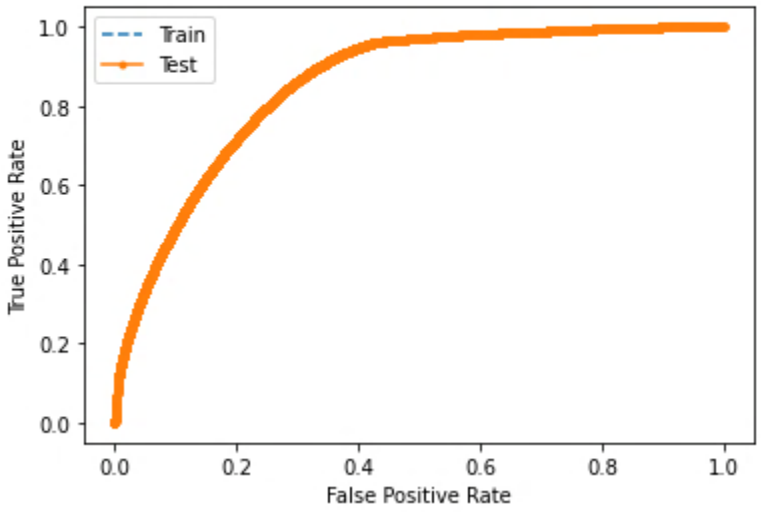
	precision	recall	f1-score	support
Short Stay	0.76	0.76	0.76	201860
Long Stay	0.70	0.70	0.70	162800
accuracy			0.73	364660
macro avg	0.73	0.73	0.73	364660
weighted avg	0.73	0.73	0.73	364660

test report

	precision	recall	f1-score	support
Short Stay	0.76	0.76	0.76	86512
Long Stay	0.70	0.70	0.70	69771
accuracy			0.73	156283
macro avg	0.73	0.73	0.73	156283
weighted avg	0.73	0.73	0.73	156283

PCA+ NNC(neural nets) - Fulldataset:

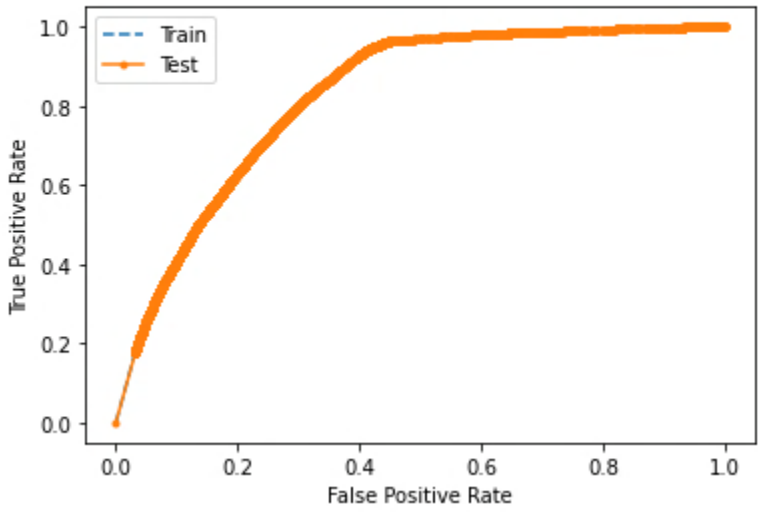
```
Train: ROC AUC=0.856
Test: ROC AUC=0.855
```




```
train report
              precision    recall  f1-score   support

  Short Stay       0.86      0.70      0.77    201860
   Long Stay       0.70      0.86      0.77    162800

    accuracy                           0.77    364660
   macro avg       0.78      0.78      0.77    364660
weighted avg       0.79      0.77      0.77    364660

test report
              precision    recall  f1-score   support

  Short Stay       0.86      0.70      0.77     86512
   Long Stay       0.70      0.86      0.77     69771

    accuracy                           0.77    156283
   macro avg       0.78      0.78      0.77    156283
weighted avg       0.79      0.77      0.77    156283
```

Non-PCA+ NNC(neural nets) - Fulldataset:

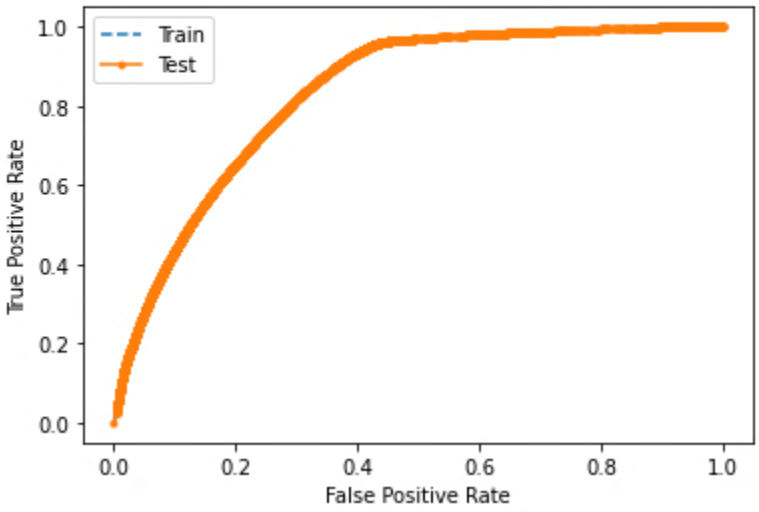
```
Train: ROC AUC=0.825
Test: ROC AUC=0.825
```




```
train report
              precision    recall  f1-score   support

  Short Stay       0.93      0.58      0.71    201860
   Long Stay       0.64      0.94      0.77    162800

    accuracy                           0.74    364660
   macro avg       0.79      0.76      0.74    364660
weighted avg       0.80      0.74      0.74    364660

test report
              precision    recall  f1-score   support

  Short Stay       0.93      0.58      0.71     86512
   Long Stay       0.64      0.95      0.77     69771

    accuracy                           0.74    156283
   macro avg       0.79      0.76      0.74    156283
weighted avg       0.80      0.74      0.74    156283
```

MAEC- Full dataset:

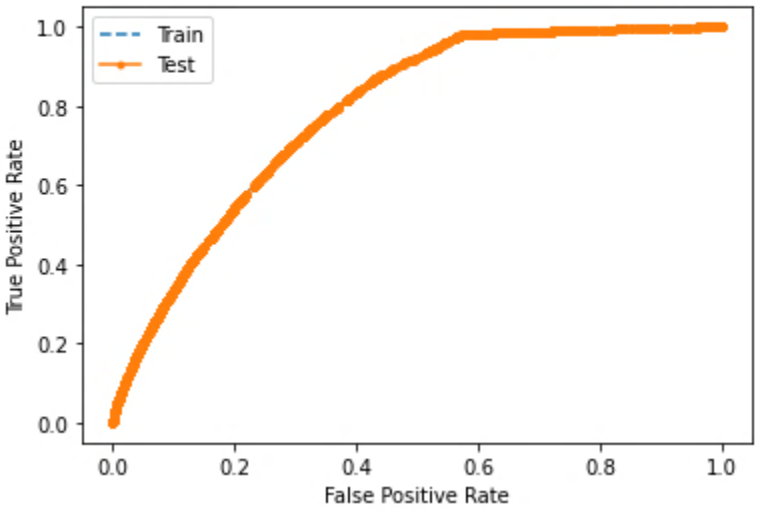
```
Train: ROC AUC=0.833
Test: ROC AUC=0.833
```




```
train report
              precision    recall  f1-score   support

  Short Stay       0.90      0.62      0.74    201860
   Long Stay       0.66      0.91      0.77    162800

    accuracy                           0.75    364660
   macro avg       0.78      0.77      0.75    364660
weighted avg       0.79      0.75      0.75    364660

test report
              precision    recall  f1-score   support

  Short Stay       0.90      0.62      0.73     86512
   Long Stay       0.66      0.92      0.77     69771

    accuracy                           0.75    156283
   macro avg       0.78      0.77      0.75    156283
weighted avg       0.79      0.75      0.75    156283
```

PCA+ Logistic Regression - Chronic condition dataset:

```
Train: ROC AUC=0.785
Test: ROC AUC=0.785
```



train report

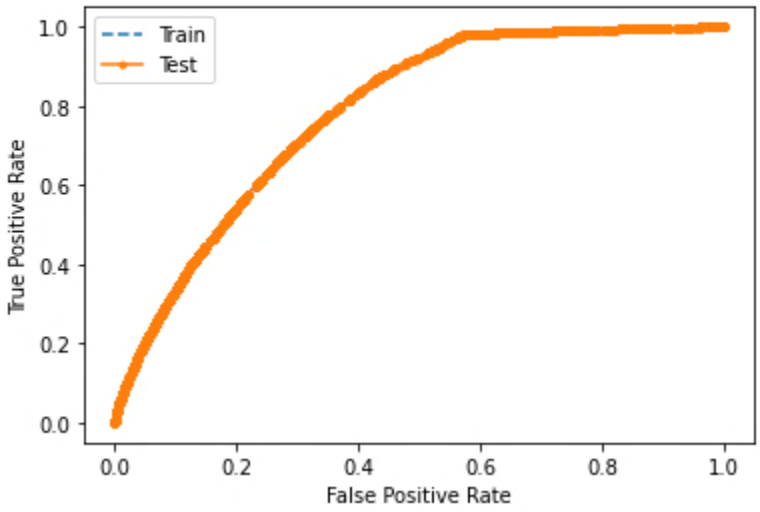
	precision	recall	f1-score	support
Short Stay	0.67	0.82	0.74	201860
Long Stay	0.69	0.50	0.58	162800
accuracy			0.68	364660
macro avg	0.68	0.66	0.66	364660
weighted avg	0.68	0.68	0.67	364660

test report

	precision	recall	f1-score	support
Short Stay	0.67	0.82	0.74	86512
Long Stay	0.69	0.50	0.58	69771
accuracy			0.68	156283
macro avg	0.68	0.66	0.66	156283
weighted avg	0.68	0.68	0.67	156283

Non-PCA-Logistic regression - Chronic condition dataset:

```
Train: ROC AUC=0.785
Test: ROC AUC=0.785
```



train report

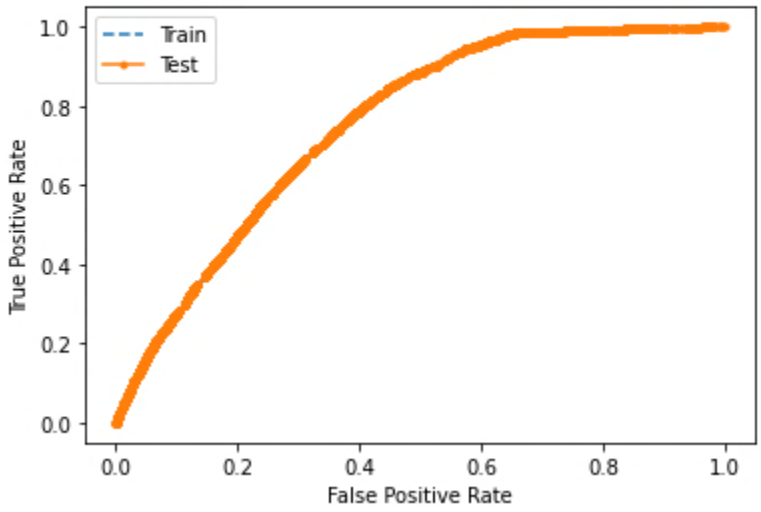
	precision	recall	f1-score	support
Short Stay	0.67	0.82	0.74	201860
Long Stay	0.70	0.50	0.58	162800
accuracy			0.68	364660
macro avg	0.68	0.66	0.66	364660
weighted avg	0.68	0.68	0.67	364660

test report

	precision	recall	f1-score	support
Short Stay	0.67	0.82	0.74	86512
Long Stay	0.69	0.50	0.58	69771
accuracy			0.68	156283
macro avg	0.68	0.66	0.66	156283
weighted avg	0.68	0.68	0.67	156283

PCA+ Naivebayes - Chronic condition dataset:

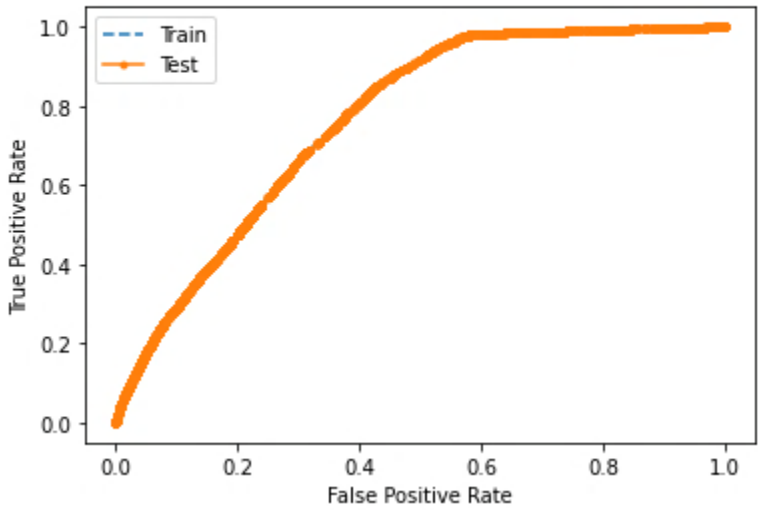
```
Train: ROC AUC=0.754
Test: ROC AUC=0.754
```




```
train report
              precision    recall  f1-score   support

  Short Stay       0.67      0.77      0.72    201860
   Long Stay       0.65      0.52      0.58    162800

    accuracy                           0.66    364660
   macro avg       0.66      0.65      0.65    364660
weighted avg       0.66      0.66      0.66    364660

test report
              precision    recall  f1-score   support

  Short Stay       0.67      0.77      0.72     86512
   Long Stay       0.65      0.53      0.58     69771

    accuracy                           0.66    156283
   macro avg       0.66      0.65      0.65    156283
weighted avg       0.66      0.66      0.66    156283
```

NON-PCA+ Naivebayes - Chronic condition dataset:

```
Train: ROC AUC=0.765
Test: ROC AUC=0.765
```



train report

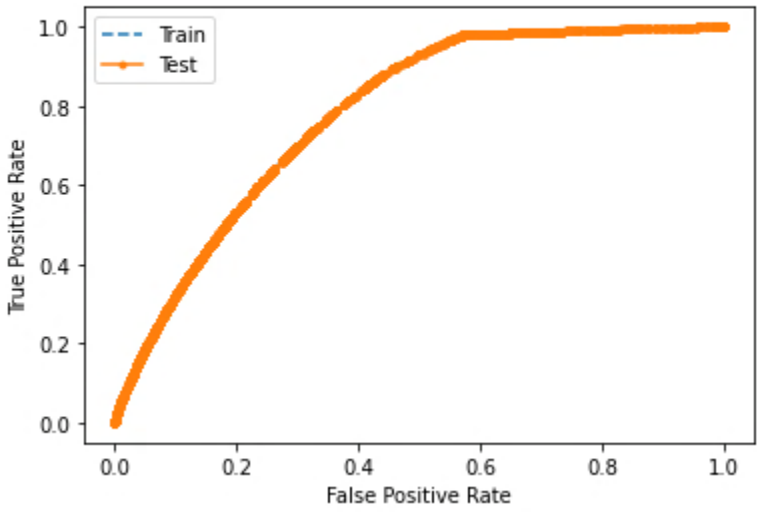
	precision	recall	f1-score	support
Short Stay	0.66	0.79	0.72	201860
Long Stay	0.66	0.50	0.56	162800
accuracy			0.66	364660
macro avg	0.66	0.64	0.64	364660
weighted avg	0.66	0.66	0.65	364660

test report

	precision	recall	f1-score	support
Short Stay	0.66	0.79	0.72	86512
Long Stay	0.65	0.50	0.57	69771
accuracy			0.66	156283
macro avg	0.66	0.64	0.64	156283
weighted avg	0.66	0.66	0.65	156283

PCA+ NNC(neural nets) - Chronic condition dataset:

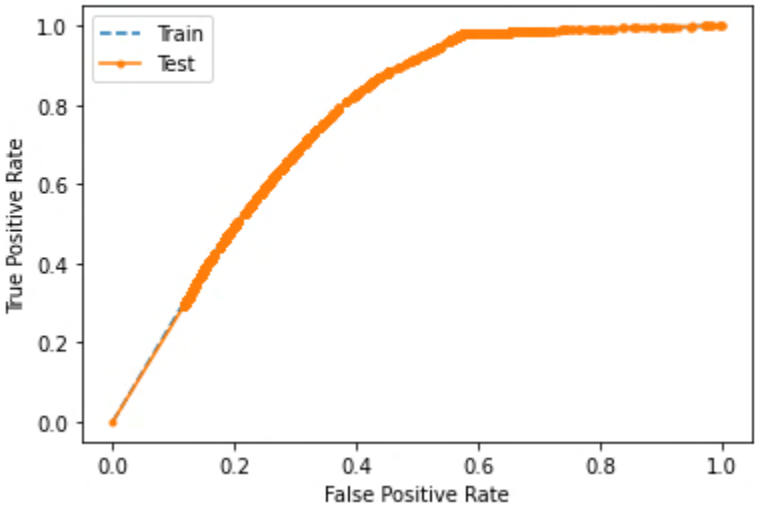
```
Train: ROC AUC=0.781
Test: ROC AUC=0.781
```




```
train report
              precision    recall  f1-score   support

  Short Stay       0.84      0.58      0.68    201860
   Long Stay       0.62      0.86      0.72    162800

    accuracy                           0.70    364660
   macro avg       0.73      0.72      0.70    364660
weighted avg       0.74      0.70      0.70    364660

test report
              precision    recall  f1-score   support

  Short Stay       0.84      0.58      0.68     86512
   Long Stay       0.62      0.86      0.72     69771

    accuracy                           0.70    156283
   macro avg       0.73      0.72      0.70    156283
weighted avg       0.74      0.70      0.70    156283
```

Non-PCA+ NNC(neural nets) - Chronic condition dataset:

```
Train: ROC AUC=0.766
Test: ROC AUC=0.765
```



train report

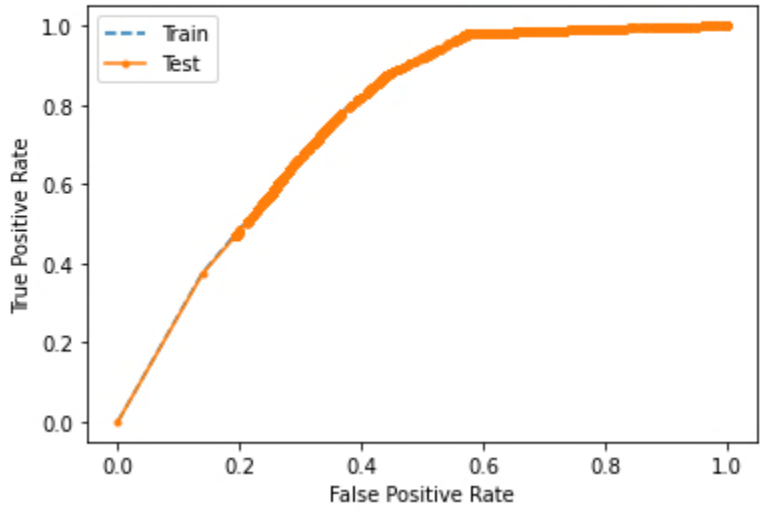
	precision	recall	f1-score	support
Short Stay	0.84	0.57	0.68	201860
Long Stay	0.62	0.87	0.72	162800
accuracy			0.70	364660
macro avg	0.73	0.72	0.70	364660
weighted avg	0.74	0.70	0.70	364660

test report

	precision	recall	f1-score	support
Short Stay	0.84	0.57	0.68	86512
Long Stay	0.62	0.87	0.72	69771
accuracy			0.70	156283
macro avg	0.73	0.72	0.70	156283
weighted avg	0.74	0.70	0.70	156283

MAEC - Chronic condition dataset:

```
Train: ROC AUC=0.765
Test: ROC AUC=0.764
```



train report

	precision	recall	f1-score	support
Short Stay	0.84	0.57	0.68	201860
Long Stay	0.62	0.87	0.72	162800
accuracy			0.70	364660
macro avg	0.73	0.72	0.70	364660
weighted avg	0.74	0.70	0.70	364660

test report

	precision	recall	f1-score	support
Short Stay	0.84	0.57	0.68	86512
Long Stay	0.62	0.87	0.72	69771
accuracy			0.70	156283
macro avg	0.73	0.72	0.70	156283
weighted avg	0.74	0.70	0.70	156283

Ideas for future research directions:

a) Consider dividing data analysis in 2 stages
    a. Short/long stay classification
    b. Regression separately for each such class
    c. Reason – high imbalance of data in these 2 groups
    d. Focus on ICU stays, as the data may be more accurate.

b) Consider converting the presence of ICD codes into text and using transformers like BERT to embed the text before conducting any analysis.