Machine Vs Deep Learning Performance for Heart Aattack prediction

Mordechai Mushkin, Dmitry Strizhak, Nathan Schumann

1. Keywords:

ML – Machine Learning

DL – Deep Learning

LR- Logistic Regression

MES- Mean Error squared.

EDA- Exploratory Data Analysis

XGBoost - Extreme Gradient Boosting

ROC - Receiver Operating Characteristic

AUC - Area Under the Curve

SVM- Support Vector Machines

MLP - Multi-layer Perceptrons

NN- Neural Network

RF- Random Forest

2. Abstract:

This project presents a comprehensive study in predictive analytics by leveraging a Kaggle dataset to address a classification problem. The work begins with EDA to understand underlying patterns and data quality issues. A preprocessing is applied with a minor data cleaning, to prepare the dataset for modelling. Initial results are obtained by using a baseline ML algorithms of LS, XGBoost and RF decision tree, evaluated via multiple relevant metrics. A NN model, deployed with default hyperparameters. The project explores the impact of hyperparameter tuning by varying key parameters across different settings. Additional experiments involve refining the network architecture to improve convergence speed and overall performance. The model is assessed the model's by altering data balance levels and applying alternative dimensionality reduction techniques. Overall, the project provides insights of preprocessing, model selection and systematic experimentation contribute to optimizing predictive performance in the given data.

3. Introduction:

Heart disease remains one of the leading causes of mortality in the United States, making early detection and prediction of heart attacks a critical public health objective. This project leverages a heart attack prediction dataset sourced from Kaggle, contributed to develop and evaluate ML models capable of identifying individuals at high risk.

The project begins with an extensive EDA to uncover patterns, assess feature distributions, and identify potential anomalies or missing values. For the modelling phase, a baseline ML algorithm is first deployed to establish performance benchmarks. Advanced methods—including the XGBClassifier and NN architectures are subsequently employed to improve predictive accuracy. The performance of these models is evaluated using a range of metrics. Standard classification metrics

such as accuracy, precision, recall, and F1 score are used to assess the models' ability to correctly predict heart attack risks. Additionally, the ROC and AUC are examined to provide insight into the models' discrimination ability between classes, and confusion matrices are analysed to understand the distribution of prediction errors. Beyond these metrics, further experiments explore the impact of hyperparameter tuning, dataset modifications, and architectural adjustments on model performance. This multi-faceted evaluation approach ensures a thorough understanding of each model's strengths and limitations, ultimately guiding the development of a robust predictive system for heart attack risk assessment.

4. Tables and Figures:

5. Literature review:

5.1 Studies revies on heart disease prediction using ML and DL

5.11 (Kumar, 2021)Researchers used the publicly available Framingham and UCI Heart datasets to predict heart attacks. They applied four classifiers (Gradient Boosting, Decision Tree, RF, and LS) optimized via hyperparameter tuning and feature engineering. Extensive evaluation used a 90:10 train-test split with accuracy and recall metrics. Gradient Boosting achieved about 85% accuracy. High cholesterol, increased heart rate, and chest pain types were key factors. Early prediction and prevention could avert 80% of premature heart attacks; DL remains highly promising.

5.12 (Alshraideh, et al., 2024) The researchers used the JUH Heart Disease dataset from Jordan University Hospital (486 cases with 58 attributes), the study applies multiple machine learning algorithms—including SVM, RF, decision tree, naive Bayes, and KNN—with particle swarm optimization (PSO) for feature selection. Models are rigorously evaluated via 10-fold cross-validation using metrics such as accuracy, precision, recall, F1-score, and ROC AUC. Notably, SVM with PSO achieves a 94.3% accuracy. The study

concludes that optimized machine learning models significantly enhance early heart disease prediction, supporting timely diagnosis and improved patient outcomes.

5.13 (Dritsas & Trigka, 2024)The study employs DL to predict heart attacks using a Cleveland dataset with 14 features and around 300 instances. The researchers developed multiple models—MLP, CNN, RNN, LSTM, GRU—and a Hybrid CNN-GRU model. They evaluated these models on an 80–20 train-test split using accuracy, precision, recall, F1-score, and AUC metrics. By integrating SHAP for explainability, the study enhances the interpretability of the predictions. Notably, the Hybrid model achieved 91% accuracy and a 0.95 AUC, demonstrating its superior performance and potential for clinical application.

5.14 (Rojek, Kotlarz, w Kozielski, Jagodzi ´nski, & Królikowski, 2024) Using a patient dataset capturing diverse clinical parameters, this study develops an AI-based tool to predict heart attack risk for preventive medicine. Multiple machine learning models—Linear SVC, LR, KNN, and RF —were compared to determine personalized risk and identify a minimal feature set (heart rate, age, BMI, cholesterol). Evaluation revealed that LR, while moderately predictive, provided the most accurate results for initial screening. The system offers a rapid, cost-effective, and non-invasive predictive analysis, enabling early intervention and improved preclinical care.

5.15 (Aghamohammadi, Madan, Ki Hong, & Watson, 2019)The paper proposes a novel classification method that combines a Genetic Algorithm (GA) with an Adaptive Neural Fuzzy Inference System (ANFIS) to predict heart attack risk using the Cleveland dataset (297 patients, 14 features). The system categorizes risk from no to very high and is evaluated using sensitivity, specificity, precision, accuracy, RMSE, and 9-fold cross-validation. Training and testing results indicate satisfactory performance, with significant RMSE reduction. Overall, explainable outputs and an Importance Evaluation Function (IEF) robustly reveal key predictive features, demonstrating transparent and effective diagnosis.

5.2 **Previous studies common approaches:** Across these five studies, traditional machine learning methods most frequently used for heart attack prediction include LR, RF, k-nearest neighbours, and SVM. DL approaches commonly involve MLP, convolutional NN, and recurrent NN variants such as RNN, LSTM, and GRU—with hybrid CNN-GRU architectures also popular evaluation metrics across these works typically include accuracy, precision, recall, F1-score and AUC, with additional use of RMSE and k-fold cross-validation to ensure robustness.

6. Methodology:

6.1 Dataset description:

6.1.1 Data set details: This dataset contains information about different health and lifestyle factors that may influence heart attacks in the USA. It includes details like age, cholesterol, blood pressure, and smoking habits, along with outcomes like

whether a heart attack occurred. The goal is to help identify potential risks and trends that can lead to better heart health awareness and prevention. It has 372,974 rows × 32 columns

6.1.2 **Dataset preprocess:** There are no missing values. Most columns are numeric including binary columns. The non-binary categorial predictors were encoded in a one hop encoding. Most of the categorial predictors are binary. The non-binary categorial predictors are: cp, restecg, slope, and thal. We scaled the numeric predictors to [0 1] range. Later on while trying to improve NN performance we changes the balance of the training and validation subsets by removing 50% and 90% of the positive-target observations in order to ensures that each class is sufficiently represented. We reduce dimensions by removing the predictors trestbps, restecg, chol, and fbs, which have the lowest correlation with the target, decreased the metrics.

6.1.3 **Evaluation of the predictor's characteristics:** The evaluation of corelation between predictors is visualized by the "Correlation Heat Map" (Figure 1.) and analytically calculated. The results indicate that exang, cp, oldpeak, and thalach have moderate correlation with the target. ca, slope, thal, sex, and age have weak correlation with the target.

trestbps, restecg, chol, and fbs have no correlation with the target. The results are shown at table 2. slope-oldpeak pair, might have moderate collinearity, all other pairs have low or acceptable collinearity. Checking of the balance of the target predictor, shows that the target is well balanced. Most of the predictors are categorial. At a later stage, when we preformed DL modeling we tried to remove the corelated predictors trestbps, restecg, chol, and fbs from the dataset.

6.1.4 **Data set split:** The dataset is divided into a 60% training subset, 20% Validation subset, and test 20% test subset.

Variable Name	Description	Values/Notes
age	Age of the patient	(years)
sex	Gender	1 = male, 0 = female
cp	Chest pain type	0: Typical angina 1: Atypical angina 2: Non-anginal pain 3: Asymptomatic
trestbps	Resting blood pressure	(mm Hg)
chol	Serum cholesterol	(mg/dl)
fbs	Fasting blood sugar	1 = true, 0 = false
restecg	Resting ECG results	0: Normal 1: ST-T wave abnormality 2: Left ventricular hypertrophy
thalach	Max heart rate achieved	(bpm)
exang	Exercise-induced angina	1 = yes, 0 = no
oldpeak	ST depression (exercise vs. rest)	(numeric)
slope	Slope of peak exercise ST segment	0: Upsloping 1: Flat 2: Downsloping
ca	Major vessels colored by fluoroscopy	(0–3)
thal	Thalassemia	0: Normal 1: Fixed defect 2: Reversible defect
target	Heart disease diagnosis	0: No significant disease 1: Significant disease

Table 1: Predictors List

exang	-0.436757
cp	0.433798
oldpeak	-0.430696
thalach	0.421741
ca	-0.391724
slope	0.345877
thal	-0.344029
sex	-0.280937
age	-0.225439
trestbps	-0.144931
restecg	0.137230
chol	-0.085239
fbs	-0.028046

Table 2: Predictors Target Correlation

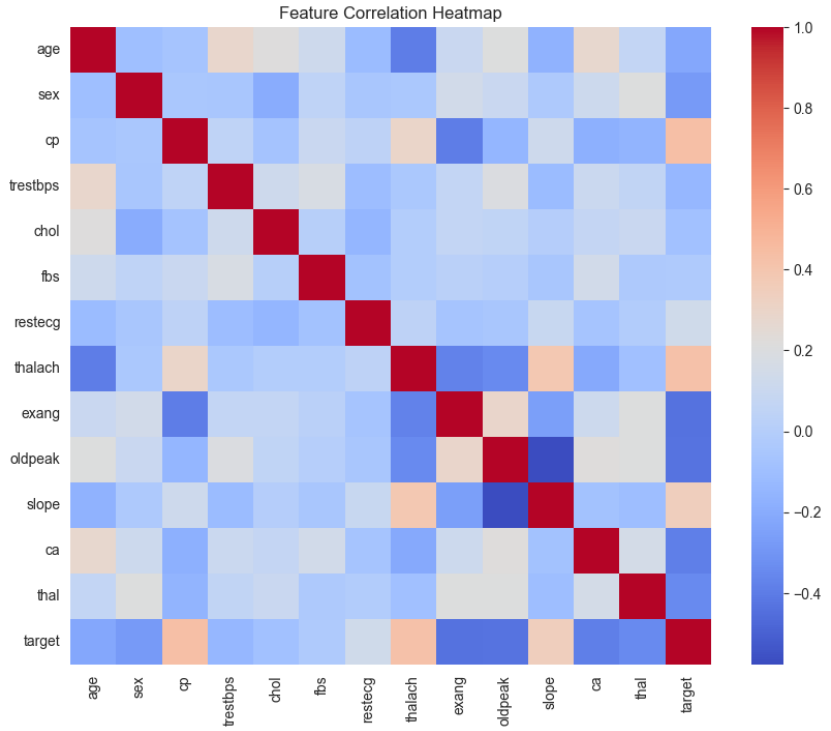

Figure 1: Feature Correlation Heatmap

6.2 ML Modelling:

6.2.1 **Decision Tree:** we used the XGBoost classifier which implements gradient boosting for classification using decision trees. It's used to predict categorical predictors as well as numerical.

6.2.2 **Randon Forest:** A RF with a majority vote over 100 trees.

6.2.3 **Logistic Regression:** The model was calculated with a max iteration of 10,000,000

6.3 **DL Modelling:** Staring with basic NN with a single input layer, a 64 neurons hidden layer with a ReLU activation function and a single output layer with a Sigmoid activation function and a binary-cross-entropy a loss function.

The weights are restored to the epoch with the best validation loss. Early stopping callback is used according to validation set loss values. The best weights from set at epoch 43 with a loss value of 0.4472.

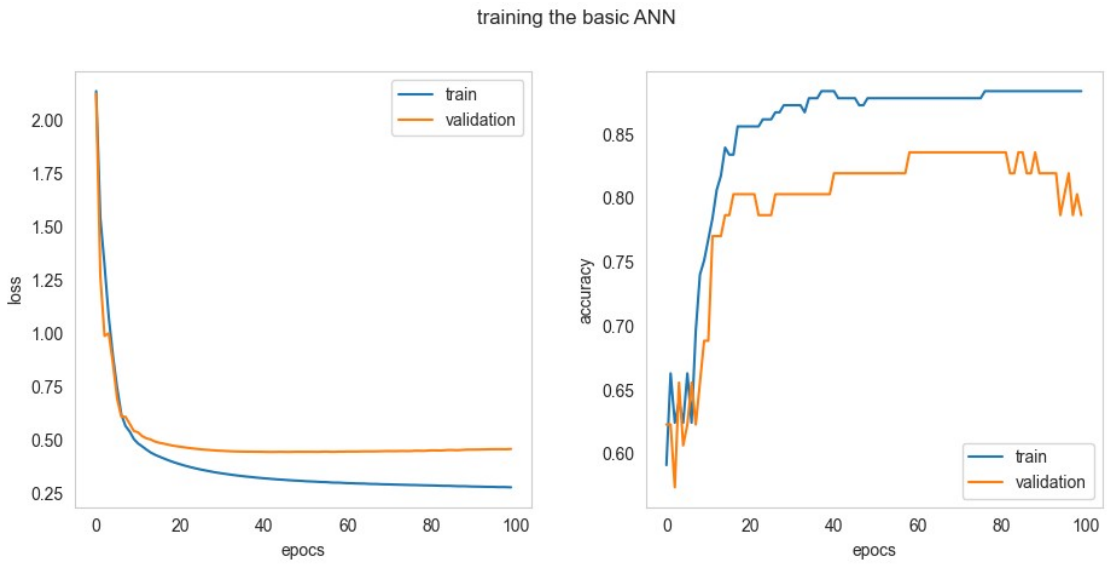

Figure 2: Basic NN training & accuracy & loss per Epoch

The hyperparameters to be checked are the depth, width, and activation functions.

We try to find the optimal hidden layer width by changing the widths to 20,30,40,60,80,100 neurons. We selected 60 as the layer's width. We added a second hidden layer, we tried widths of 4,6,8,10 and 12 neurons. 2nd layer at a width of 10 achieved the best performance. The comparison between the performance of a single layer model and two hidden layers resulted similar performance. Adding a third hidden layer and testing at widths of 16, 20, 24, 28, 32 resulting a width of 28 with the best performance. The metrics of 3 hidden layers model are similar to single and two layers. In order to reduce complexity a simple model is preferred. We tried to improve by removing 55 (30%) of the potential outliers and 6 (3%) of the potential outliners from the training subset. It didn't lead to a more robust training. We continue "dropping" neurons during training, It didn't lead to an improvement. A Hinge loss function results are similar to Cross Entropy.

7. Results:

7.1 **ML Results:** Gradient boost tree-based classifier reached train accuracy of 1.0 and test accuracy of 0.77. RF reached train accuracy of 1.0 and test accuracy: 0.836.

Logistic regression reached train accuracy of 0.867 and test accuracy of 0.803. RF got the best test accuracy. RF metrics are:

Accuracy: 0.8361

Recall: 0.9062

Precision: 0.8056

F1-Score: 0.8529

```
Logistic regression accuracy: 0.803
Random forest accuracy:       0.836
XGBC accuracy:                0.77
```

Table 3: ML Test Accuracy Comparison

7.2 **DL Results:** A basic DL model slightly worse than those of the RF model

metric	RF	basic ANN
Accuracy	0.84	0.82
Recall	0.91	0.91
Precision	0.81	0.78
F1-Score	0.85	0.84

Table 4: Basic NN Vs RF Metrics

The chart below shows the loss per epoch for each tested width.

Best weights at epoch 30 got a loss of 0.4184.

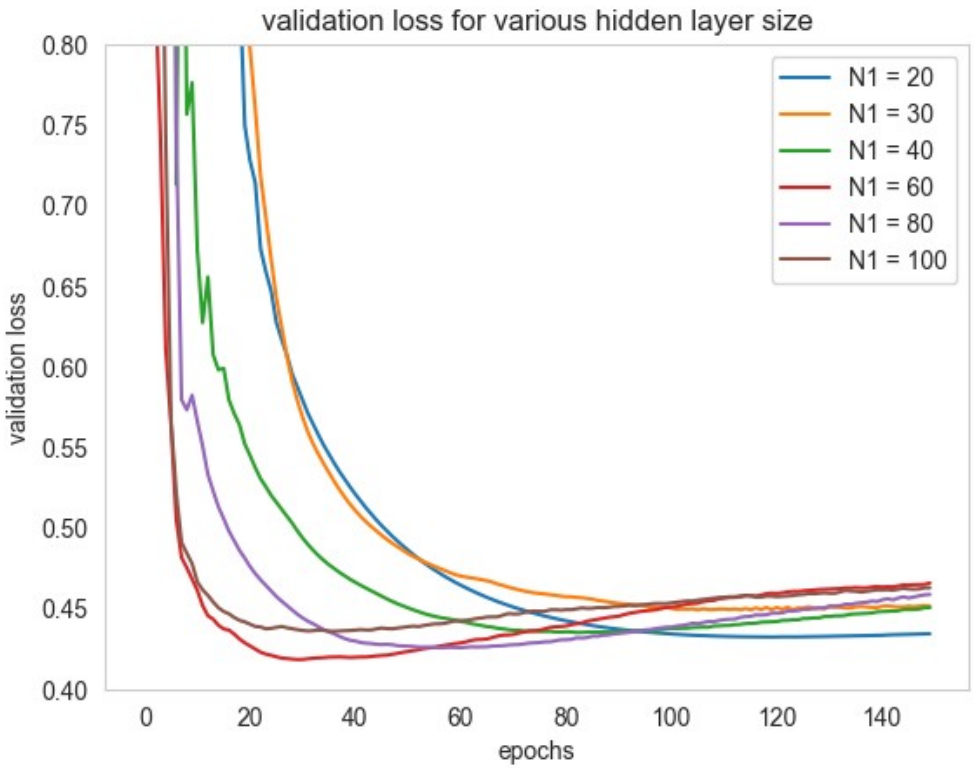

Figure 3: Single layer loss per epoch @ validation subset

best single layer model test preformances

metric	value
Accuracy	0.82
Recall	0.91
Precision	0.78
F1-Score	0.84

Table 5: Best Single layer metrics

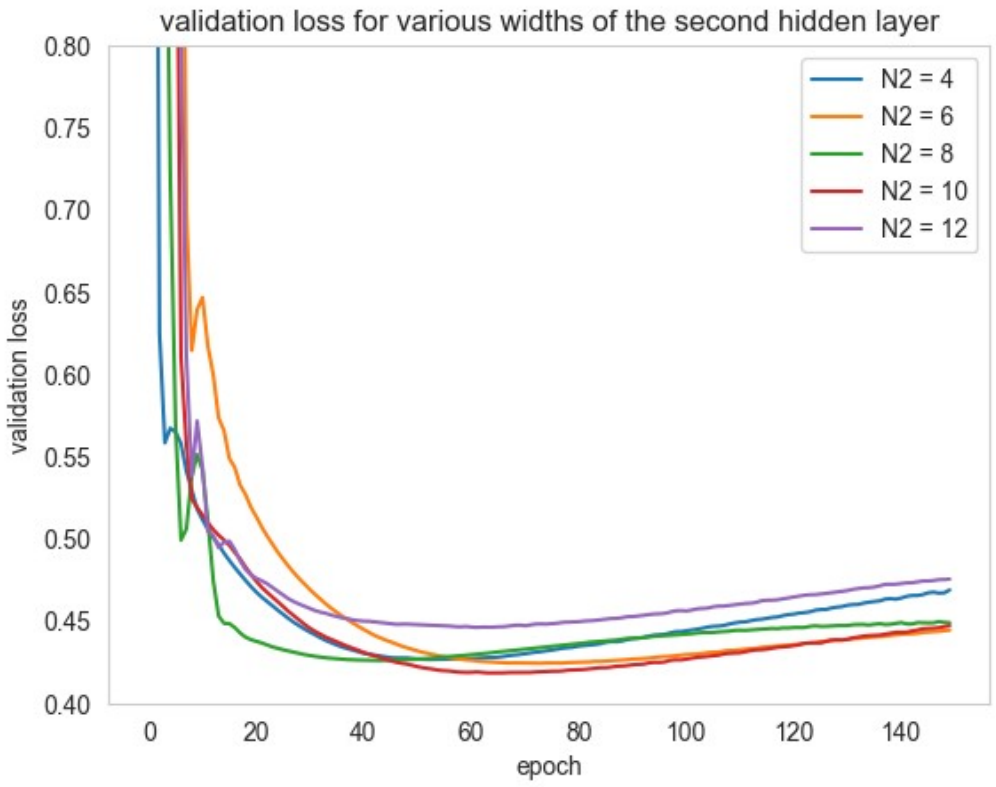

Figure 4: Two layers loss per epoch @ validation subset

The metrics of a single and two hidden layers are the same.

```
best single layer model versus best two layers model test preformances
------------------------------------------------------------------------
metric      single      two
Accuracy    0.82        0.82
Recall      0.91        0.91
Precision   0.78        0.78
F1-Score    0.84        0.84
```

Table 6: Single- and two-layers metrics

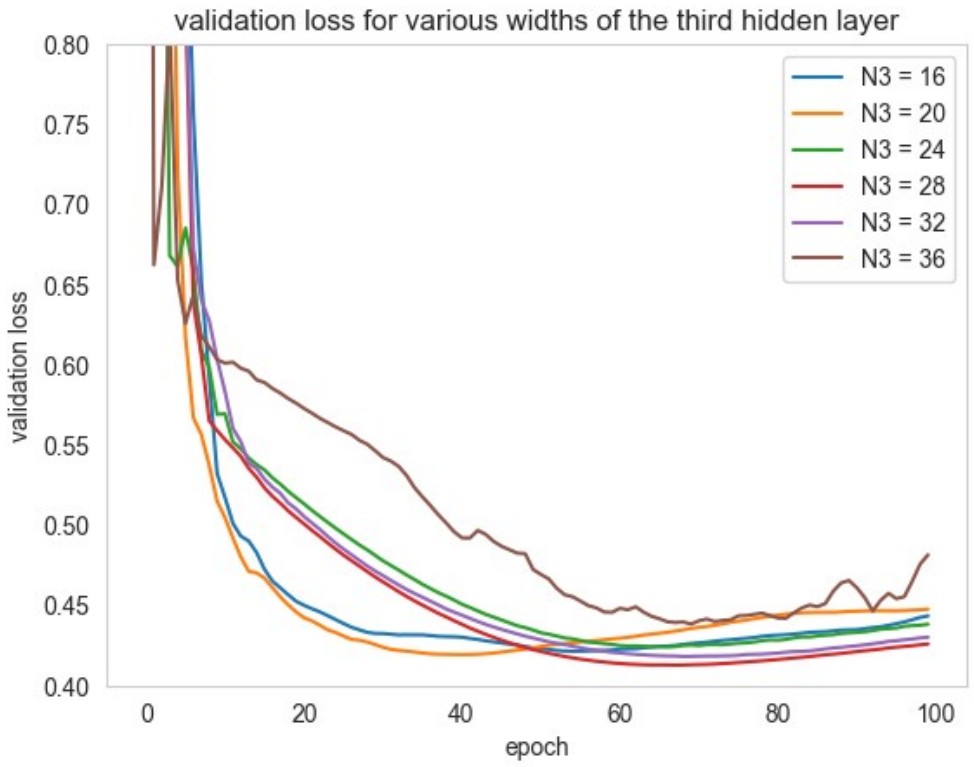

Figure 5: Three layers loss per epoch @ validation subset

best single, two, and three layer models preformances

metric	single	two	three
Accuracy	0.82	0.82	0.80
Recall	0.91	0.91	0.91
Precision	0.78	0.78	0.76
F1-Score	0.84	0.84	0.83

Table 7: Single, two- and three-layers metrics

preformances before and after removing 3% of the potential outliers

metric	before	after
Accuracy	0.82	0.82
Recall	0.91	0.88
Precision	0.78	0.80
F1-Score	0.84	0.84
Precision	0.78	0.72
F1-Score	0.84	0.81

preformances before and after dropout

metric	before	after
Accuracy	0.82	0.80
Recall	0.91	0.88
Precision	0.78	0.78
F1-Score	0.84	0.82

Table 8: metrics comparisons, outliners, dropout

```
Original versus modified data set
---------------------------------
metric      original      modified
Accuracy    0.82       0.64
Recall      0.91       0.44
Precision   0.78       0.78
F1-Score    0.84       0.56
```

Table 9: Balanced train and evaluation subsets

```
Comparing the performance metrics before and after removing non-correlated predictors
--------------------------------------------------------------------------------------
before         after
Accuracy:  0.82       0.8
Recall:    0.91       0.88
Precision: 0.78       0.78
F1-Score:  0.84       0.82
```

Table 10: Dimensions reduction results

8. Discussion and Concoctions:

The data set contains predictors that explain the target with a linear logistic regression. Techniques like, network architecture, balancing the data, dimensions reduction, Loss function modifications didn't improve the performance of the NN model. For this case NN is an "over keel".

9. Further Work:

Heart attacks prediction at this study and previous studies show that it can be done with good results. However, heart attacks prevention is more challenging and more important in terms of public health. Further work should include models that provide Inference. Once we can quantify how much each predictor contribute to the probability to suffers from heart attack, we might be able to prevent it. We should also look at data sets with other predictors which are not biological like environmental, social, and psychological.

10. References:

Bibliography

Aghamohammadi, M., Madan, M., Ki Hong, J., & Watson, I. (2019, December 4). Predicting Heart Attack through Explainable. *EasyChair Preprint*, *2093*. From https://link.springer.com/chapter/10.1007/978-3-030-22741-8_45

Kumar, S. G. (2021, August). A Machine Learning Approach for Heart Attack. *International Journal of Engineering and Advanced Technology (IJEAT)*, *10*(6). doi:10.35940/ijeat.F3043.0810621

Alshraideh, M., Alshraideh, N., Alshraideh, A., a Alkayed, Y., Al Trabsheh, Y., & Alshraideh, B. (2024, March 7). Enhancing Heart Attack Prediction with Machine Learning: A. *2024*. doi:https://onlinelibrary.wiley.com/doi/10.1155/2024/5080332

Dritsas, E., & Trigka, M. (2024, September 24). Application of Deep Learning for Heart Attack Prediction with. *Computers*. doi:https://www.mdpi.com/2073-431X/13/10/244

Rojek, I., Kotlarz, P., w Kozielski, M., Jagodzi ´nski, M., & Królikowski, Z. (2024, January 7). Development of AI-Based Prediction of Heart Attack Risk as an. *electronics*. doi:https://www.mdpi.com/2079-9292/13/2/272