1. **Introduction**

Heart disease remains a critical health concern globally, demanding precise diagnostic approaches. The Heart Disease dataset from the UCI Machine Learning Repository offers a promising avenue for predictive modelling. With more than 300 instances and features of varied types, this dataset sourced from four databases allows for the purpose to determine the presence or absence of heart disease in patients. Therefore, the goal is to construct predictive models that aid in early detection and risk assessment, ultimately enhancing patient care and outcomes in cardiovascular health.

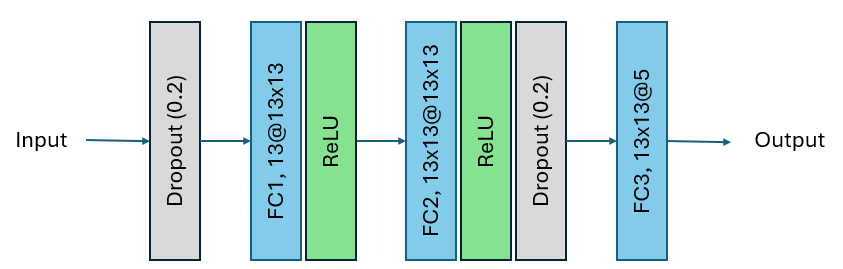
1. **Description of Data**

The Heart Disease dataset comprises 303 instances and 13 features, encompassing a mix of categorical, and integer attributes. These attributes provide information relevant to the diagnosis and prognosis of heart-related conditions. While the dataset originally includes 76 attributes, most experiments focus on a subset of 14 features. The primary target variable indicates the presence or absence of heart disease, with integer values ranging from 0 (no presence) to 4 (presence). This dataset, sourced from databases including Cleveland, Hungary, Switzerland, and VA Long Beach, serves as a valuable resource for classification tasks aimed at predicting heart disease status based on clinical characteristics.

*Table 1. Variables table*

| **Variable name** | **Description** | **Role** | **Type** | **Note** |
| --- | --- | --- | --- | --- |
| age | Age | Feature | Integer |  |
| sex | Sex | Feature | Categorical |  |
| cp | Chest pain type | Feature | Categorical | 1: typical angina  2: atypical angina  3: non-anginal pain  4: asymptomatic |
| trestbps | Resting blood pressure | Feature | Integer | mm/Hg |
| chol | Serum cholesterol | Feature | Integer | mg/dl |
| fbs | Fasting blood sugar | Feature | Categorical | >120 mg/dl ? 1(true) : 0(false) |
| restecg | Resting electrocardiographic results | Feature | Categorical | 0: normal  1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) |
| thalach | Maximum heart rate achieved | Feature | Integer |  |
| exang | Exercise induced angina | Feature | Categorical | 0: no  1: yes |
| oldpeak | ST depression induced by exercise relative to rest | Feature | Integer |  |
| slope | The slope of the peak exercise ST segment | Feature | Categorical | 1: upsloping  2: flat  3: downsloping |
| ca | number of major vessels (0-3) coloured by fluoroscopy | Feature | Integer |  |
| thal |  | Feature | Categorical | 3: normal  6: fixed defect  7: reversible defect |
| num | diagnosis of heart disease | Target | Integer |  |

1. **Model Structure**



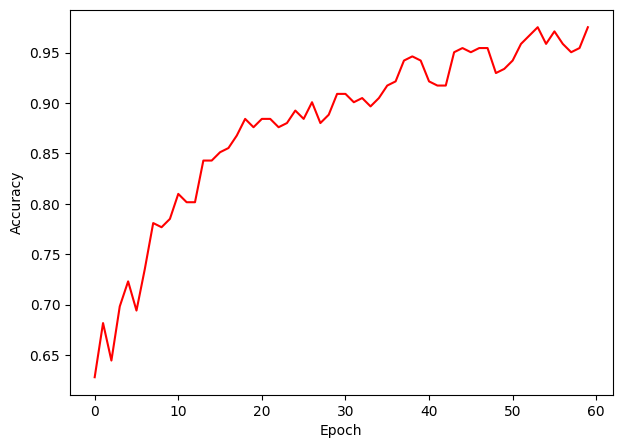
*Fig. 1. Models Structure*

The blocks in the diagram labelled FC1, FC2, and FC3 likely represent fully-connected layers which are responsible for combining features extracted from the previous layers to form more complex representations.

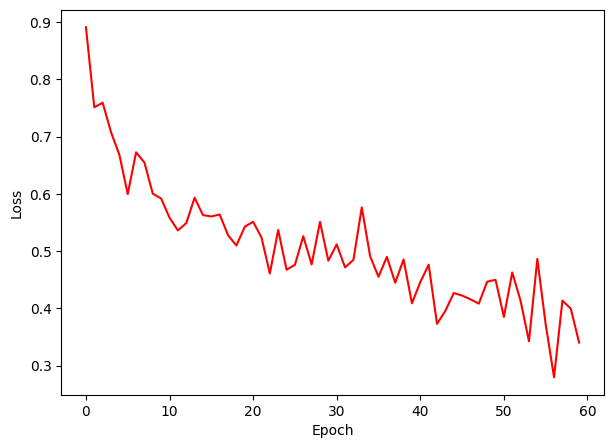
The dropout layer is also important as it masks partial feature map’s information during training to prevent the overfitting as well as increase the model’s generalisation

ReLU is an activation function that sets all negative values in the output to zero, while leaving positive values unchanged.

1. **Results**
2. **Training**The dataset is splitted 75% for training and 25% for testing. The model is trained with 60 epochs, Cross Entropy Loss, a learning rate of 0.01 and Adam optimizer.



*Fig.2 Accuracy Curve*

**

*Fig. 3 Loss curve*

1. **Testing**The test accuracy is 55.26% which is a really low and unreliable model.

- Initialised parameter:

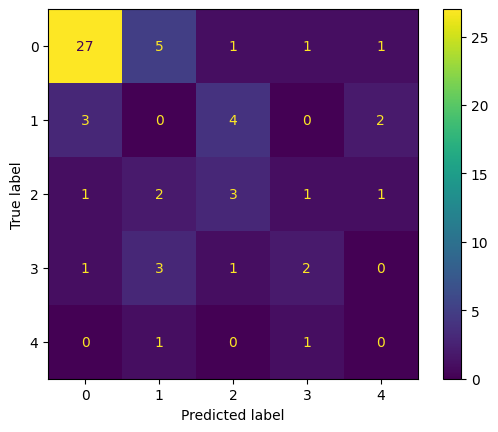
Learning rate: 0.01

Epoch: 60

Data split: 75:25

| **Training parameters** | **Testing Accuracy** |
| --- | --- |
| 0.01, 50, 80:20 | 44.3% |
| 0.001, 50, 80:20 | 59% |
| 0.001, 60, 80:20 | 54.1% |
| Add dropout layer after first ReLU  0.001, 50, 80:20 | 55.7% |
| Remove the first dropout layer  Add dropout layer after first ReLU  0.001, 50, 80:20 | 63.9% |

1. **Evaluation**It appears that 27 samples labelled as 0 are correctly classified. However, samples labelled 1 and 4 are incorrectly classified.

****

*Fig. 4 Confusion matrix*

1. **Discussion**

In conclusion, the model must be adjusted more in order to achieve better accuracy by altering the number of layers, initialised parameters or modify the block in the model (add, remove…)