

Preprocessing D1NAMO dataset

1. Download the dataset from <https://zenodo.org/record/1421616/files/D1NAMO.tgz>
2. Extract the dataset
3. Loop through the dataset and find the files ending with “ECG.csv” which contain the ECG data for all subjects. **Important!** – Each subject has 4 separate ECG recordings. Due to the lack of data each recording was treated as a different subject.
4. For each ECG recording extract 1 min of data starting from the 15th minute (to allow time for stabilization). From this data, the following parameters were calculated.
 1. Standard deviation of all NN intervals (SDNN) in seconds,
 2. Square root of the mean of the sum of the squares of differences between adjacent NN interval (RMSSD) in milliseconds (ms),
 3. Number of adjacent NN intervals differing more than 50 ms. (NN50 count),
 4. Percentage of difference between adjacent NN intervals differing more than 50 ms. (pNN50%),
 5. Integral of sample density distribution of RR intervals divided by the maximum of the density distribution (R- R triangular index)
 6. Baseline width of the minimum square difference triangular interpolation of the maximum of the sample density distribution of NN intervals in seconds (TINN)
 7. Mean R-R interval in seconds
 8. Mean heart rate
 9. Standard deviation (STD) of the mean heart rate (per minute)
5. These parameters were used by Ahamed Seyd, P.T; Paul K. Joseph & Jeevamma Jacob in their paper “Automated Diagnosis of Diabetes Using Heart Rate Variability Signals”
6. The processed data was saved to a csv file named “processed_data.csv”. This file contains the columns: 'filename', 'SDNN', 'RMSSD', 'nn_50', 'pNN_50', 'meanHR', 'SDHR', 'meanRR', 'TINN', 'HRVTriIndex'

Processing SWELL Dataset

The model trained from the D1NAMO dataset had insufficient accuracy (~65%) therefore the SWELL dataset was used. SWELL dataset contains various Heart rate variability parameters which are used for detecting stress.

Out of the 36 different parameters, the 9 parameters were considered. These were chosen to be similar to the ones used in the D1NAMO model. **SDNN, RMSSD, pNN_50, meanHR, mean_RR, MEDIAN_RR, LF, HF, HF_LF** was chosen. The parameters **MEDIAN_RR, LF, HF, HF_LF** was chosen instead of nn_50, SDHR, TINN, HRVIndex, And as raw ECG was not included in the dataset the required parameters could not be calculated.

The data was labelled as “no stress”, “time pressure” and “interruption”. These were relabelled as “diabetic”(time pressure, interruption) and “healthy”(no stress). The SWELL dataset contained two files train.csv and test.csv these files were processed and was saved as “swell_training_data.csv” and “swell_testing_data.csv” respectively.

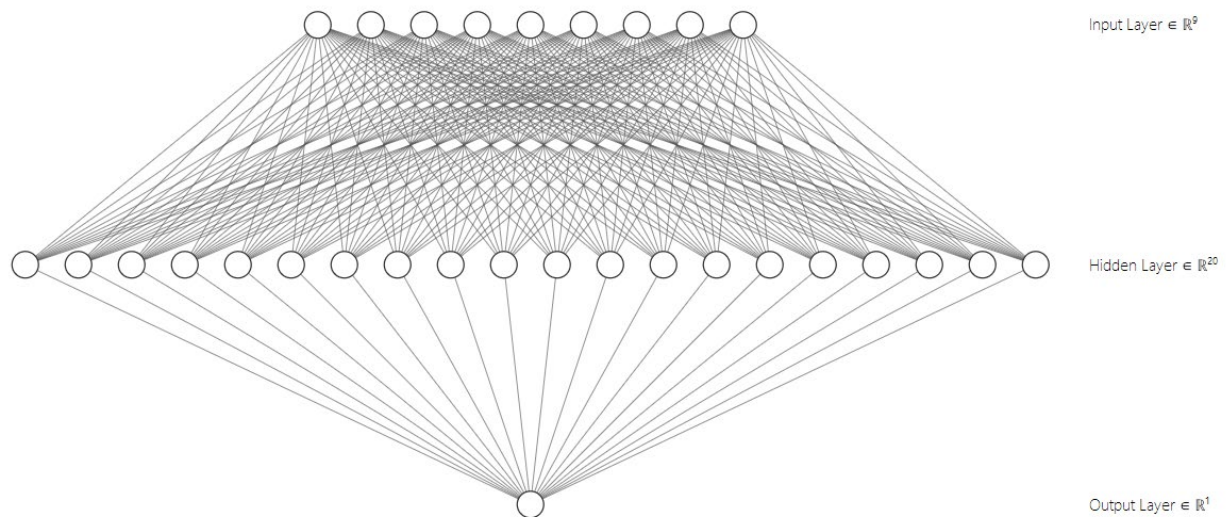
Technologies

- Tensorflow 2.20 and Keras was used for model creation
- Pandas , numpy libraries were used for data manipulation
- Jupyter lab with Python 3.8 kernel was used for programming.

Algorithm

1. Remove rows containing NaN values from the dataset
2. Normalize the data and binary encode the labels as healthy – 1, diabetic – 0
3. Train the neural network for 20 epochs.
4. Export the model as diabetesHRV.h5

The neural network architecture used in the “Automated Diagnosis of Diabetes Using Heart Rate Variability Signals” was used with minor modifications.



- Adam optimiser was used instead of backward propagation with momentum.
- RELU activation function was used instead of sigmoid for the hidden layer
- Binary cross entropy loss function was used

Evaluation

1. Load the model “diabetesHRV.h5”
2. Read the test data in swell_testing_data.csv to a pandas dataframe
3. Normalize the data and encode labels. Normalizing is done using statistical parameters of the training dataset (model_stats.csv)
4. Use `model.evaluate` function to obtain accuracy.