Justifying DCNN architecture

Layer 1 (Conv1D):
- The first layer consists of a convolutional base, aimed at finding general patterns amongst my input features
- Keeping a small kernel size allows the model to look for patterns amongst neighbouring features
- The use of a L2 regularization penalizes the model for having excessive large weights, helping to prevent any particular feature from having biasedness

BatchNormalization & MaxPooling between layer 1 and 2:
- Batch normalization helps to normalize feature scales, helping to prevent overfitting
- Max pooling helps the model focus only on the most important features, helping to reduce overfitting

Layer 2 (Conv1D):
- By doubling the number of filters from layer 1 to 2, this architecture is in line with the more common design pattern for CNNs, where layer 1 is aimed at finding simple patterns, while the subsequent layers will combine such features to extract deeper patterns and relationships

Dropout:
- Allows the model to turn off X% (dropout_rate) of neurons during training will help to prevent my model from memorizing data which could lead to overfitting

Decision layer:
- A dense layer acts as the final reasoning layer, combining patterns found in the model
- A final node which has a sigmoid activation function acts as the output node, classifying the results of the model

Justifying model parameters

Learning Rate (0.001):
- A lower learning rate was chosen, so as to allow the model to have a steady rate of learning towards the optimal weights
- A higher learning rate (0.01) might cause the model to learn at too fast a rate, overshooting the optimal solution

Dropout Rate (0.5):
- Selecting a higher dropout rate allows the model to reduce over dependence on specific features, forcing the model to find multiple patterns to form the correct prediction

- It is observed during the parameter tuning stage, when a lower dropout_rate (0.2) was chosen, the model performed worse (peaked at 75%) as compared to the first trained model with dropout_rate at 0.5 (accuracy at 82%)

Filters (32):

- For the dataset of 16 features, selecting 32 filters allow sufficient space for the model to start finding patterns
- Selecting too high a filter number may result in the model memorising the data and overfitting

Kernel Size (2):

- We can observe that some of the features in the dataset are related to each other (for e.g. Blood pressure and Glucose, or BMI and Physical activity), as such choosing a small kernel size is helpful for the model to start finding patterns between pairs of neighbouring features, and to capture local feature interactions