A solid reference point, the baseline model with two hidden layers of 18 neurons each arrived at a good balance between training and validation accuracy. While adding three layers improved representation learning but caused slight overfitting, reducing the model to one hidden layer resulted in a slight decrease in accuracy, probably because of the limited learning capacity.

There were diminishing returns when the number of neurons per layer was increased to 32 and 64; the latter resulted in overfitting without appreciable increases in accuracy.

Since mean squared error (MSE) is less useful for probability-based classification, using it in place of binary cross-entropy led to decreased accuracy. Tanh activation, which was used in place of ReLU, produced comparable results but had slower convergence because of saturation effects.

Combining L2 regularization and dropout was the most successful regularization method; it decreased overfitting and increased validation accuracy at the expense of somewhat decreased training accuracy. All things considered, the analysis draws attention to the trade-offs in activation functions, regularization, and model complexity, highlighting how crucial it is to strike a balance between generalization and learning capacity for best results.