Appendix

Glossary

Fine needle aspirate (FNA)	Tissue or fluid is removed from an abnormal-seeming area of tissue or fluid with a fine needle to be examined under a microscope
Concavity	Severity of concave portions of the contour
Compactness	The quality of being closely packed together. Calculated as ((perimeter)2/ area) −1
Fractal dimension	A ratio that provides a statistical index of complexity, by measuring the rate of change of a pattern relative to its scale. Calculated as “coastline approximation” −1
Kernel trick	The kernel trick is the main strength of SVMs: it avoids the explicit mapping that is generally required to get linear learning algorithms to learn a nonlinear function or decision boundary. Kernel methods represent the data only through a set of pairwise similarity comparisons between original data observations, instead of transforming directly.
Gamma	Gamma is a hyperparameter of SVM with a Gaussian kernel, affecting the size of the kernel function. Gamma is the free parameter of the Gaussian function $K(x_i, x_j) = \exp(-\gamma \|\|x_i - x_j\|\|^2), \gamma > 0$
Regularisation	Regularisation, or C, represents the misclassification cost function for SVM: a small C value creates a softer margin, whereas a large C creates hard boundaries which may make the model prone to overfitting.
Hyperplane	In machine learning, a hyperplane is a decision boundary in n-dimensional space, where data falling on either side would be attributed to a different class. Hyperplanes do not have to be linear, but always has one less dimension than the dimensions of data space.
Radial basis function (RBF)	A function whose value depends on the distance from the origin or some other point. Can only take real values. A Gaussian or RBF kernel transforms the dot product into the Gaussian function of the distance from the centre, creating a normal distribution-shaped hyperplane

Appendix

K-fold cross validation	A resampling method used to improve the reliability of machine learning models. The training data is split into k number of 'folds' of training and validation set, and resampled, producing an average score for the model.
Softmax	The softmax function is an exponential function used as the last activation function in a neural network. It normalises outputs into a probability distribution between 0 and 1 for easy evaluation of neural networks.
Bayesian regularization	Bayesian regularization converts a nonlinear problem into a linear one, similar to a ridge regression. It improves robustness of models, and can be used for early stopping.
Early stopping	Early stopping is a regularisation method which prevents overfitting during training with an iterative method, such as gradient descent.
Generalizability	Refers to the model's ability to adapt to previously unseen data. If a model is unable to adapt to new data, it has overfit the training data and has low generalizability.

Implementation details

Preprocessing

I also tried running both models with significant preprocessing: with and without standardisation, dimensionality reduction (PCA), and log transformations. I found that PCA did not significantly improve the model, in line with the findings of Ghosh et al. This was to be expected as both models are adept at learning the most useful variables. Log transformations also did not improve the model, as both models can cope with differences in magnitude. Standardisation did improve the model and was subsequently used.

Model selection

For MLP, there were several options in implementing the neural network: I used skorch, which is a combination of sklearn and Pytorch. Pytorch is particularly useful in large datasets when parallelisation decreases training time, but since this was a small dataset, it was not particularly useful.

Evaluation

I did not originally stratify when using train_test_split, but this made a huge difference to my model - however, this could be due to overfitting.
I evaluated using AUC, accuracy, precision and recall, in order to ensure comparability.