**BREAST CANCER DIAGNOSIS USING ENHANCED ARTIFICIAL NEURAL NETWORKS: MODEL DEVELOPMENT, CHALLENGES, AND INSIGHTS**

Breast cancer diagnosis is a crucial medical application on the patients where early detection significantly improves patient outcomes or results. This project focuses on developing and improving an Artificial Neural Network (ANN) model to classify breast cancer tumors as benign or malignant in human. The model was enhanced iteratively by incorporating regularization techniques and fine-tuning hyperparameters, resulting in better performance and generalization.

The included dataset is the Breast Cancer Wisconsin (Diagnostic) dataset obtained from Kaggle, made up of 569 records and 30 features that are numeric attributes extracted from a digitized image of a fine needle aspirate **(FNA)** of breast mass cells, depicting cell nuclei radii, textures and smoothness. The target variable is binary (malignant = 1, benign = 0).

**PREPROCESSING STEPS**

**Missing Values:** The dataset did not contain missing values, so no imputation was necessary.

**Feature Scaling:** Standardization was performed to center all the features to have a mean of 0 and standard deviation of 1 to improve model convergence.

**Train-Test Split:** The dataset was divided into training and test set with a ratio of 80:20. The key takeaway is that the model is evaluated on previously unseen data.

**MODEL DEVELOPMENT AND ARCHITECTURE**

Ultimately, we built the initial ANN model, and continued with the experiment iteratively by enhancing the initial model at several stages.

**Baseline Model**: It was composed by two hidden layers (128 neurons and 64 neurons), with ReLU activation and a single output neuron with sigmoid activation for binary classification. The Adam optimizer was used (LR=0.001) and binary cross-entropy the loss function. It was a baseline model which performed relatively well but it presented very mild overfitting tendencies.

**Improvement 1:** After each hidden layer, a dropout layer was added to pin down some of the classical network’s trainable parameters. Specifically, a 0.2 dropout rate was applied, which entails twiddling weights for 1/5, or alternately 80 per cent, of the images seen during training. Since this is done on the fly, it is referred to as ‘dynamic per-minibatch’. PyTorch also offers a gradual version, but that is marginally more complicated to use. The program now became more sensitive to overfitting, a trade-off for increased accuracy. To prevent it from being trained too long (and potentially overfitting), EarlyStopping was added, which causes training to cease when the validation performance has stopped improving.

**Second improvement:** We added L2 regularisation to the weights of each of the hidden layers. This regularisation brought down the weights that lead to large hidden nodes, thus de-emphasising complex features and encouraging the model to fit simpler features better. This, in turn, helped it generalise to unseen data. We kept Dropout and EarlyStopping from the previous version.

**Improvement 3:** Number of neurons in the hidden layers: 256 and 128, to be able to capture more complex patterns from the data. L2 regularisation was maintained, as was the dropout, and learning rate decay was added to tweak the learning process. This model performed best overall with a good learning capacity and generalisation.

**TRAINING AND EVALUATION RESULTS**

For the different version of the model, we trained for up to 100 epoch. We implement using the loss function binary cross-entropy and the optimisation function as Adam using the scikit library.

We decide to use validation accuracy and the loss to checkpoint out model. I decided to use EarlyStopping, this is a fitter function that stop the training when the validation score stops improving.

**Initial model performance:** The basic ANN reached test accuracy of 97.37%, which seemed to be moderately overfit.The results of the confusion matrix in the malignant and benign case categories in the test set (being the most important categories) indicated that benign tumours were predicted more accurately as compared with malignant tumours, but recall has slightly dropped off for malignant cases.

**Improvement 1 (Dropout):** Adding the dropout reduced overfitting but decreased the test accuracy to 95.61 per cent. It was in exchange for a better generalisation. The lower performance in accuracy didn’t harm the precision for the malignant cases as it helped in reducing the number of false positives.

**Improvement 2 (L2 Regularization):** Test accuracy was not changed (95.61 per cent), but the model was more stable during training. Regularisation preserved the balance between learning and generalising, as is evidenced by a decrease in validation loss and validation accuracy.

**Improvement 3 (Neurons + L2):** This third and final model returned a test accuracy score of 98.25% and the best precision and recall for the benign and malignant case. The confusion matrix displays two misclassifications (1 false positive and 1 false negative).

**CHALLENGES AND SOLUTIONS**

One can identify several difficulties with the project, in particular overfitting (model capacity).

The original model still had a very good performance on the training data, but the validation accuracy almost always fell slightly behind. This problem could be resolved with the introduction of dropout, regularisation and learning rate scheduling.

Dropout helped to prevent overfitting by randomly dropping out neurons during training. L2 regularisation penalised large weights thereby discouraging the model from being too dependent of specific features.

Finally, increasing the amount of neurons introduced more capacity into the model, which, in combination with regularisation, allowed it to learn more complex patterns while still being somewhat generalisable.

By making sequential improvements, the final model achieved very high accuracy (98.25 per cent) and generalisation ability. Dropout and regularisation provided a very effective solution to overfitting. Adaptive learning rates improved the optimisation algorithm. The increase in the number of neurons was achieved by increasing the width of the layers. This resulted in a much more expressive model grounded on the complex interactions among the variables, that yet achieves a very good generalisation ability. The high accuracy and robustness of the model renders it suitable for real-world applications. Among them are the following: Breast cancer diagnosis.

.