**Kamaal Bartlett**

**COMP4442**

**Prof. Chan**

**Final Project**

Neural networks are powerful tools for classification problems where the relationships between input features and the target variable are complex and non-linear. This method is particularly useful in medical predictions, such as predicting lung cancer, where multiple factors like demographics and clinical symptoms interact in ways that traditional linear models may not capture effectively. Neural networks can answer research questions such as, "Can we predict the likelihood of lung cancer based on patient data, and how accurately can this prediction be made?"

**Data Requirements & Assumptions:**

* **Input Features:** Requires diverse, normalized inputs (e.g., age, smoking status).
* **Class Balance:** Ideally balanced classes; SMOTE can address imbalances.
* **Assumptions:**
  + Inputs should be relevant and somewhat independent.
  + Works best with large datasets, though tuning can help smaller ones.
* **Assessment:**
  + **Normalization:** Use z-score normalization.
  + **Imbalance Handling:** Apply SMOTE for class balance.
  + **Feature Selection:** Address multicollinearity through data analysis.

**Conceptual Overview & Interpretation:**  
Neural networks consist of layers of neurons processing inputs to produce outputs. The common architecture used is a feedforward network where data moves from input to output.

* **Key Components:**
  + **Input Layer:** Receives normalized data.
  + **Hidden Layers:** Apply weights and activation functions (e.g. Sigmoid) to transform inputs.
  + **Output Layer:** Produces final lung cancer probability.
* **Functions:**
  + **Activation:**
    - **Sigmoid:** Suitable for binary classification, outputs between 0 and
  + **Loss:**
    - **Cross-Entropy:** Measures the difference between actual and predicted probabilities.
    - **MSE:** Calculates the average squared differences (used for regression).
* **Model Performance:**
  + **Accuracy:** SMOTE-balanced nnet model reached ~91.30% accuracy.
  + **AUC:** Tuned nnet model achieved an AUC of ~0.9323, showing strong discrimination ability.
* **Interpretation of Results:**
  + **Confusion Matrix:** Evaluates classification performance through true/false positives/negatives.
  + **ROC Curve:** Illustrates the trade-off between sensitivity and specificity. Threshold adjustments can minimize either false positives or false negatives depending on context. In lung cancer prediction, minimizing false negatives is critical for early detection.

**8. Recommended Reading**

1. **General Neural Networks:**
   * [A beginner’s guide to neural networks and deep learning](https://www.superannotate.com/blog/guide-to-neural-networks-and-deep-learning)
   * [“How Do Neural Networks Work? Your 2024 Guide”](https://www.coursera.org/articles/how-do-neural-networks-work)
2. **Practical Guides in R:**
   * [Building a Simple Neural Network in R Programming](https://www.geeksforgeeks.org/building-a-simple-neural-network-in-r-programming/)
   * [Beginner’s guide to machine learning in R](https://www.r-bloggers.com/2022/02/beginners-guide-to-machine-learning-in-r-with-step-by-step-tutorial/)
3. **SMOTE Methodology:**
   * [Overcoming Class Imbalance with SMOTE: How to Tackled Imbalanced Datasets in Machine Learning](https://www.blog.trainindata.com/overcoming-class-imbalance-with-smote/)
   * [A Comprehensive Guide to Handling Imbalanced Datasets in Classification Problems](https://talent500.co/blog/a-comprehensive-guide-to-handling-imbalanced-datasets-in-classification-problems/)
4. **Machine Learning Fundamentals:**
   * [Machine Learning: Everything You Should Know](https://www.grammarly.com/blog/what-is-machine-learning/)
   * [9 Key Machine Learning Algorithms Explained In Plain English](https://www.bomberbot.com/machine-learning/9-key-machine-learning-algorithms-explained-in-plain-english/)
5. **Model Evaluation Techniques:**
   * [What is Accuracy, Precision, Recall and F1 Score](https://www.labelf.ai/blog/what-is-accuracy-precision-recall-and-f1-score)
   * [Classification: ROC and AUC](https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc)