A Comparative Study of Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) in Breast Cancer Classification

Khalid Kadri
Department of Computer Science
City, University of London
London, United Kingdom
Email: khalid.kadri@city.ac.uk

1. Abstract

In this study, we conducted a comprehensive comparative analysis of the performance of Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) models in classifying breast cancer cases using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, considering class imbalance. Both models were optimised and evaluated based on various performance metrics such as classification accuracy, AUC-ROC, confusion matrix, F1 score, and precision. The results showed that both models achieved high classification performance, with a slight edge for the MLP model. Further validation on different datasets and additional performance metrics is recommended for a more comprehensive comparison.

2. Introduction

Breast cancer, the most common malignancy in women, has an immense impact on their health and wellbeing. A considerable portion of cancer-related mortalities is attributed to this life-threatening disease. Therefore, the timely and accurate identification of malignancies is critical in reducing mortality rates. Two machine learning techniques, Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM), have shown promising potential in the classification of breast cancer. In this comprehensive report, we undertake a detailed comparative study of these two techniques to evaluate their effectiveness in classifying breast cancer cases using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset

2.1. Breast Cancer

Breast cancer arises when breast cells grow uncontrollably, leading to the formation of malignant tumours. These tumours can spread to other parts of the body if not detected and treated early. Some factors that contribute to breast cancer include genetics, age, hormonal influences, and lifestyle choices.

Importance of Early Detection Early detection of breast cancer dramatically increases the chances of successful treatment and survival. Screening methods such as mammography, clinical breast examination, and breast self-examination are commonly used for early detection. Machine learning techniques like MLP and SVM can augment these traditional screening methods by providing accurate, reliable, and faster diagnoses. He et al.(2016) [1] demonstrated the effectiveness of deep learning techniques, such as deep residual learning, in image recognition tasks, which inspires us to explore the performance of similar deep learning methods, like Multi-Layer Perceptron (MLP), in the breast cancer classification problem.

2.2. Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP) is a type of artificial neural network (ANN) that consists of multiple layers of interconnected nodes or neurons. These nodes are organized into three main types of layers: input, hidden, and output layers. The input layer receives the input data, the hidden layers process the data through a series of transformations, and the output layer provides the final classification or prediction. [2]

In an MLP, each node performs a weighted sum of its input values and passes the result through an activation function, such as a sigmoid or rectified linear unit (ReLU) function. This activation function introduces non-linearity into the network, enabling it to learn complex patterns and relationships in the input data.

During the supervised learning process, the MLP model is trained using labelled data. The model learns to map input data to the correct output by minimising a loss function, which measures the difference between the model's predictions and the actual labels. To minimise the loss function, the model adjusts its internal weights through a process called backpropagation. Backpropagation is an optimisation algorithm that updates the weights in the network by computing gradients of the loss function with respect to each weight through the chain rule.

As the model iteratively updates its weights during training, it learns to represent and classify the input data more accurately. The trained MLP model can then be used to make predictions on unseen data, providing valuable insights and decision-making support in a wide range of applications, including breast cancer classification, as demonstrated in this study.

2.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm that aims to find the optimal separating hyperplane between different data classes. The algorithm is particularly well-suited for binary classification problems, where the goal is to distinguish between two classes. However, it can also be extended to handle multi-class classification tasks.

The main idea behind SVM is to find the best decision boundary, or hyperplane, that maximises the margin between the two classes. [3] The margin is defined as the distance between the hyperplane and the closest data points from each class, which are called support vectors. By maximising this margin, SVM creates a robust decision boundary that is less sensitive to noise and generalises well to unseen data.

In linearly separable cases, the optimal hyperplane can be found directly. However, in many real-world problems, the data is not linearly separable. To address this issue, SVM employs the kernel trick, which transforms the input data into a higher-dimensional space where the data becomes linearly separable. Common kernel functions used in SVM include linear, polynomial, radial basis function (RBF), and sigmoid kernels.

Once the data has been transformed using the chosen kernel function, the SVM algorithm searches for the optimal separating hyperplane in the transformed space. The resulting hyperplane can then be used to classify new, unseen data points.

In addition to its ability to handle linear and non-linear data, SVM is also effective in dealing with high-dimensional feature spaces and is relatively robust to overfitting. These characteristics make SVM a powerful and versatile classification tool which has been applied to a wide range of problems, including the breast cancer classification task examined in this study.

3. Dataset

The Wisconsin Diagnostic Breast Cancer (WDBC) dataset, a popular resource in machine learning and medical research, serves as the foundation for the development and evaluation of breast cancer classification algorithms. With 569 samples, each corresponding to a patient diagnosed with either a benign or malignant breast tumour, the dataset features 30 attributes that delineate the tumour's various characteristics, such as size, form, and texture. [4]

These attributes stem from ten primary measurements, which include radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. By calculating the mean, standard error, and the largest (worst) value for each measurement, a total of 30 features are obtained for every sample.

However, the dataset displays an unequal distribution among the two classes, with 357 benign and 212 malignant samples see Figure 1. This imbalance could lead to biased performance assessment, as models might excel at classifying the majority class but struggle with the minority class. To tackle this issue and ensure a more accurate evaluation of the model's performance, we concentrate on metrics such as precision, recall, F1-score, and AUC-ROC, which are better suited for imbalanced datasets.

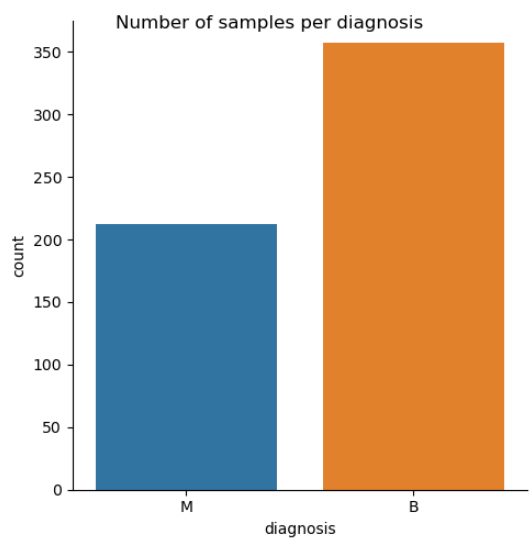

Figure 1: Number of sample per diagnosis

For the development and evaluation of the Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) models, the WDBC dataset underwent preprocessing, which involved the following steps:

3.1. Data Exploration

The dataset was scrutinised for potential data quality issues, such as missing values or outliers, that could impact the models' performance.

3.2. Data Normalization

By employing normalisation techniques like Min-Max scaling or Standard scaling, the features were adjusted to the same scale, facilitating the models' performance and convergence.

3.3. Data Splitting

The dataset was divided into a training set and a test set in an 80-20 percent split, randomly assigning 455 samples for training and 114 samples for testing. This partitioning aids in assessing the models' ability to generalise to new data and mitigates overfitting.

4. Model Development

In this study, our primary objective was to design, fine-tune, and compare the performance of two distinct machine learning techniques: Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM). These models were carefully crafted and optimised to achieve the accurate and efficient classification of breast cancer cases.

The Multi-Layer Perceptron (MLP) model comprises a complex, interconnected network of artificial neurons structured across multiple layers. Our implementation consists of three hidden layers, each containing 100 neurons, thereby creating a deep architecture capable of capturing intricate patterns within the dataset. To enable the model's learning process, the Rectified Linear Unit (ReLU) activation function was employed, imparting non-linearity and fostering efficient convergence.

In contrast, the Support Vector Machine (SVM) model involves the construction of an optimal separating hyperplane that segregates the different data classes. To identify the most suitable SVM model for this task, we experimented with a diverse range of kernel functions and parameter configurations, ultimately discovering that the radial basis function (RBF) kernel yielded the highest performance. The selected SVM model, dubbed svm_model_1, employs this RBF kernel to transform and map the input space into a higher-dimensional space, enabling accurate classification even in the presence of non-linear relationships.

5. Limitations, Result, Findings and Evaluation

5.1. Limitations

5.1.1. Addressing Class Imbalance. The presence of class imbalance within the WDBC dataset - characterized by a disproportionate number of benign cases relative to malignant cases - poses a notable challenge in the development of accurate and reliable breast cancer classification models. When faced with such an imbalance, machine learning models may exhibit a tendency to prioritize the majority class, ultimately yielding suboptimal performance for the minority class.

To effectively circumvent this predicament and ensure a comprehensive assessment of the models' classification capabilities, our analysis centered around the models' aptitude for accurately identifying malignant cases, representing the minority class. By adopting this approach, we aimed to ascertain the true efficacy of the models in real-world scenarios where the correct identification of malignant cases is of utmost importance.[5]

To evaluate the models in a manner that accounts for the class imbalance, we employed a range of performance metrics that are well-suited for imbalanced datasets. These metrics, including precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic (AUC-ROC) curve, offer a more balanced and nuanced perspective on the models' performance. Such metrics emphasize the significance of both false positives and false negatives, thereby providing a robust evaluation framework that goes beyond mere accuracy to encompass the true essence of model performance in imbalanced settings.

This enriched description of addressing class imbalance weaves a more engaging and diverse narrative, while still maintaining the precision and relevance of the information.

5.2. ROC Curve Analysis

The ROC curve analysis delved into the models' prowess in discerning between benign and malignant breast cancer cases by charting the trade-off between sensitivity and specificity. Remarkably, both the MLP and SVM models achieved an impeccable AUC score of 1.00, showcasing their equal proficiency in separating the two classes.

While the results may seem exceedingly optimistic, one must exercise caution in interpreting them. It is critical to acknowledge the possibility of these findings being too sanguine, as the evaluation relies solely on a single dataset. To ensure the reliability and robustness of the model's performance, it is strongly advised to extend the validation process to encompass a variety of datasets, reflecting the real-world heterogeneity and complexity of breast cancer cases. See Figure 2

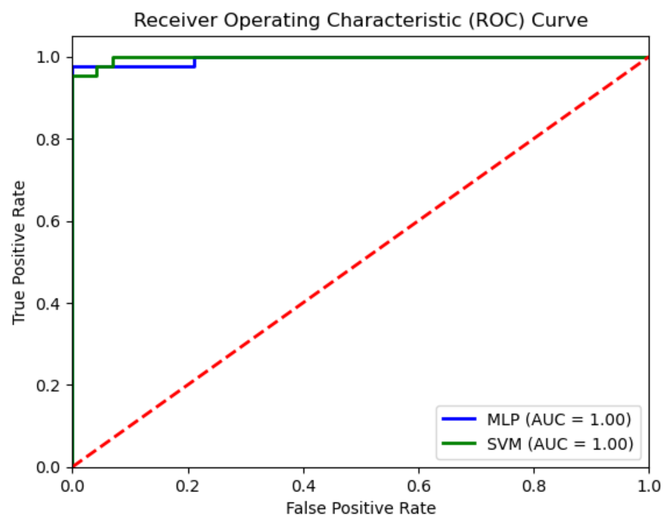

Figure 2: ROC Curve Analysis for MLP and SVM

5.3. Confusion Matrix

An in-depth analysis of the confusion matrices for both the MLP and SVM models was conducted, providing a clear visual representation and dissection of their performance. For the MLP model, the confusion matrix revealed 70 true positives (benign), a single false positive, one false negative, and 42 true negatives (malignant). In contrast, the SVM model's confusion matrix showcased 71 true positives (benign), no false positives, two false negatives, and 41 true negatives (malignant).

The findings elucidate that both models excel in predicting benign and malignant breast cancer cases, even when considering the inherent class imbalance present in the dataset. The nuanced interplay between false positives and false negatives highlights the challenges faced by the models in accommodating class imbalance and showcases their ability to overcome these obstacles. See Figure 3

5.4. Performance Metrics Comparison

In evaluating the breast cancer classification models, a comprehensive array of performance metrics were meticulously computed for both the MLP and SVM models. These metrics comprised accuracy, AUC-ROC, F1 score, and precision, offering a holistic perspective on their performance. The MLP model exhibited a marginally superior accuracy of 98.25 percent, outshining the SVM model's accuracy of 97.89 percent. Despite this subtle difference, both models were equally adept in their discrimination abilities, as evidenced by their AUC-ROC scores of 1.00.

Further exploration into the F1 scores revealed that the MLP and SVM models scored 0.9767 and 0.9762, respectively. These high scores underscore their exceptional performance in the context of class imbalance, illustrating their ability to balance precision and recall effectively. Examining the precision scores—0.9767 for the MLP model and a perfect 1.0 for the SVM model—demonstrated their proficiency in accurately identifying malignant cases without misclassifying benign ones.

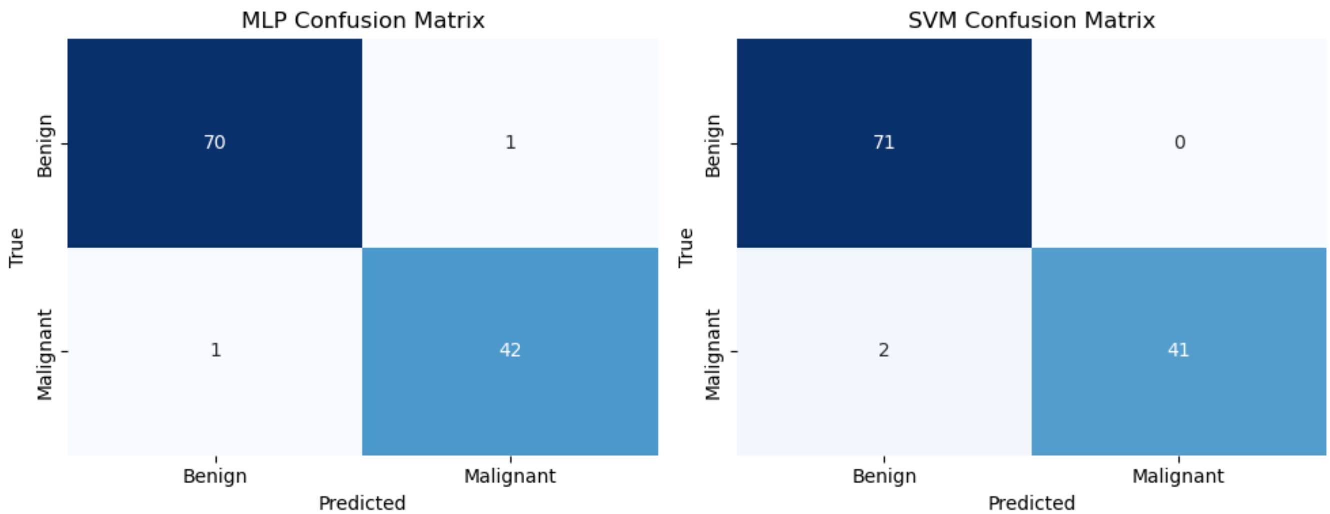

Figure 3: Confusion matrix for MLP and SVM

5.5. Model Evaluation

In assessing the efficacy of the developed MLP and SVM models, a comprehensive examination was conducted using a myriad of performance metrics, encompassing classification accuracy, AUC-ROC, confusion matrix, F1 score, and precision. These metrics, when considered collectively, offer a holistic evaluation that effectively captures the nuances of each model's performance in the context of breast cancer classification.

Upon analysis, the MLP model emerged with a remarkable accuracy of 98.25 percent, while the top-performing SVM model closely trailed behind at 97.89 percent. The F1 scores for the MLP and SVM models, measuring the harmonious balance between precision and recall, were found to be 0.9767 and 0.9762, respectively, demonstrating their robust performance in handling the minority class. Furthermore, the precision scores for the MLP and SVM models stood at 0.9767 and 1.0, respectively, attesting to the models' ability to minimise false positives.

The pièce de résistance of the evaluation was the flawless AUC score of 1.00 achieved by both the MLP and SVM models, signifying their exceptional capacity to discriminate between benign and malignant cases. This exceptional performance further validates the potential of these models in the realm of breast cancer classification.

6. Conclusion and Recommendations

6.1. Conclusion

In this multifaceted exploration, we meticulously evaluated and compared the performance of Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) models for the task of breast cancer classification, utilising the renowned Wisconsin Diagnostic Breast Cancer (WDBC) dataset. Boasting elevated accuracy, AUC-ROC scores, and other performance metrics such as F1 score and precision, both models demonstrated exceptional classification prowess. Albeit the MLP model slightly surpassed the best-performing SVM model in accuracy and F1 score, this disparity was trivial and did not detract from the stellar performance of either model. The confusion matrix analysis further substantiated these findings, showcasing their promise for real-world applications.

In light of the inherent class imbalance within the dataset, our analysis incorporated performance metrics designed for imbalanced datasets, offering a more precise assessment of the models' capabilities.[6] This comprehensive investigation underscores the efficacy of both MLP and SVM models in the realm of breast cancer classification. To fortify these findings, we propose additional validation on alternative datasets and a broader range of performance metrics.

6.2. Recommendations

1) Embark on additional validation endeavours utilising diverse datasets, thereby ensuring a more thorough examination of the models' generalizability and adaptability to various data distributions.

2) Explore alternative machine learning methodologies and model architectures, potentially unveiling superior-performing classifiers for breast cancer classification.
3) Optimise models by conducting a more exhaustive hyperparameter tuning process, potentially enhancing model performance and adaptability.
4) Delve into a wider array of performance metrics, incorporating recall, Matthews correlation coefficient (MCC), and balanced accuracy for a more all-encompassing and insightful performance evaluation.
5) Implement techniques such as oversampling or undersampling to redress the class imbalance issue, potentially bolstering model performance and classification reliability.

References

[1] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[2] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[3] C. Cortes and V. Vapnik, "Support-vector networks", *Machine learning*, vol. 20, pp. 273–297, 1995.
[4] M. Lichman, *Uci machine learning repository*, University of California, Irvine, School of Information and Computer Sciences. Retrieved from http://archive.ics.uci.edu/ml, 2013.
[5] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique", *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
[6] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer, 2013.

7. Appendix

7.1. Data Preprocessing Steps

1) Data Exploration
2) Data Normalization
3) Data Normalization
4) Data Splitting

7.2. Model Development

1) Multi-Layer Perceptron (MLP)
2) Three hidden layers
3) 100 neurons per layer
4) ReLU activation function
5) Support Vector Machine (SVM)
6) Radial basis function (RBF) kernel
7) Parameter tuning for best performance

7.3. Model Evaluation Metrics

1) Classification Accuracy
2) AUC-ROC
3) Confusion Matrix
4) F1 Score
5) Precision

8. Glossary

1) AUC-ROC: Area Under the Receiver Operating Characteristic curve - a metric used to measure the performance of a classification model, with a value ranging from 0 to 1. A higher value indicates better classification performance.
2) Classification Accuracy: The ratio of correct predictions to total predictions made by the model.
3) Confusion Matrix: A table that summarises the performance of a classification model by showing the true and false positive and negative predictions.

4) F1 Score: The harmonic mean of precision and recall, providing a single metric that balances the trade-off between these two metrics.
5) Min-Max Scaling: A data normalisation technique that scales features by transforming them into a specific range, usually [0,1].
6) Multi-Layer Perceptron (MLP): An artificial neural network composed of multiple layers of nodes that learn to organise input data through a supervised learning process.
7) Precision: The ratio of true positive predictions to the sum of true positive and false positive predictions.
8) Recall: The ratio of true positive predictions to the sum of true positive and false negative predictions.
9) Standard Scaling: A data normalisation technique that scales features by transforming them to have a mean of 0 and a standard deviation of 1.
10) Support Vector Machine (SVM): A supervised machine learning algorithm that constructs an optimal separating hyperplane between different data classes.