Deep Artificial Neural Networks for Cancer Diagnosis

Akshay Kumar akumar9@umbc.edu
Keerthi S Kopparaju keerthk2@umbc.edu
Sina M Kaviri sinam1@umbc.edu
Srikar Jarugula jsrikar1@umbc.edu

1 Abstract

Early detection of diseases is crucial for effective treatment and improved patient outcomes. Unfortunately, many conditions, including breast cancer, often go unnoticed until they reach advanced stages, leading to increased mortality rates worldwide.

A significant challenge in disease diagnosis lies in imbalanced medical datasets, where normal cases far outnumber abnormal ones. This imbalance complicates classification tasks and underscores the need for innovative approaches to accurate diagnosis.

In this study, we utilize real-world data from a hospital, consisting of records from 698 patients. The dataset includes various medical tests such as blood tests and chest CT scans, with the final column indicating cancer or non-cancer diagnosis in binary form. Given the dataset's distribution, we face a classification problem balancing between balanced and imbalanced categories.

To address this challenge, we employ a range of classification techniques, including logistic regression, SVM, and Random Forest Classifiers, alongside a fully connected neural network. The neural network, with its complex architecture comprising hidden layers, presents a potent solution. The neural network's capacity to learn hierarchical representations enables us to overcome the challenges posed by imbalanced datasets and achieve robust predictive performance. Notably, the neural network, along with the Random Forest classifier, demonstrates the most favorable results in terms of accuracy and F1 score metrics. These findings underscore the potential of innovative approaches in enhancing diagnostic accuracy and improving patient care outcomes, even in the face of data constraints.

2 Introduction

The application of Deep Artificial Neural Networks (DANNs) in healthcare, particularly in the diagnosis of cancer, presents a promising avenue for enhancing the accuracy and speed of medical diagnostics. This project aims to develop a DANN capable of analysing clinical data to predict cancer diagnoses in patients, addressing the urgent need for efficient and reliable diagnostic tools in oncology. The significance of this research lies in its potential to facilitate early detection and treatment of cancer, ultimately improving patient outcomes and survival rates.

3 Description of Dataset

The dataset (1) comprises 698 patient records, with nine quantitative attributes and one binary class attribute indicating cancer or healthy controls.We've planned to split this dataset into training, validation, and testing sets. Here's how it's organized:

- Our training set includes 70 percent of the total records, which amounts to 490 records.
- The validation set contains 15 percent of the records, giving me 104 records.
- Similarly, the testing set also comprises 15 percent of the total, which is another 104 records.

The quantitative attributes represent key health metrics that may relate to cancer:

- Age: The patient's age.
- BMI: Body Mass Index, a measure of body fat.
- Glucose: Blood glucose levels.
- Insulin: Insulin levels in the body.
- HOMA: Homeostatic Model Assessment, used to assess insulin resistance.

- Leptin: A hormone related to appetite regulation.
- Adiponectin: A hormone involved in regulating glucose levels.
- Resistin: A hormone linked to inflammation and insulin resistance.
- MCP-1: Monocyte Chemoattractant Protein-1, an indicator of inflammation.

The binary class attribute represents the final cancer diagnosis, with 1 indicating cancer and 0 indicating healthy controls. The overview of the dataset containing attribute unit, mean, standard deviation, minimum, and maximum values is shown in Table 1.

Table 1: Overview of dataset

Attribute	Unit	Mean	Std. dev.	Min.	
Age	years	56.7	3.5	34.0	
BMI	kg/m^2	27.0	4.6	18.4	37.1
Glucose	mg/dL	105.6	26.6	70.0	201.0
Insulin	μU/mL	12.5	12.3	2.4	58.5
HOMA	–	3.6	4.6	0.5	25.1
Leptin	ng/mL	26.6	19.2	6.3	90.3
Adiponectin	μg/mL	10.1	6.2	1.7	38.0
Resistin	ng/mL	17.3	12.6	3.2	82.1
MCP-1	pg/dL	563.0	384.0	90.1	1698.4

4 Data Cleaning and Preparation

4.1 Pre-processing

The overview of the dataset indicates that the nine attributes have varying maximum, minimum, mean, and standard deviation values. To address this, each attribute is subjected to a scaling process to standardize the dataset. This scaling maps each attribute's values to a range of 0 to 1, ensuring a consistent scale across all columns. By normalizing the dataset at this stage, it becomes easier to determine the optimal parameters for the Artificial Neural Network (ANN) during the evaluation stage. Normalization also ensures that cost functions are symmetric, aiding gradient descent algorithms in finding local minima efficiently.

4.2 Feature extraction/data transformation

Principal Component Analysis (PCA) is a technique that takes a set of related variables and converts them into a smaller set of uncorrelated variables through orthogonal transformation. PCA creates a new orthogonal coordinate system, where each axis represents a principal component, which is a linear combination of the original variables. This transformation ensures there is no redundant information, as each principal component is orthogonal to the others.

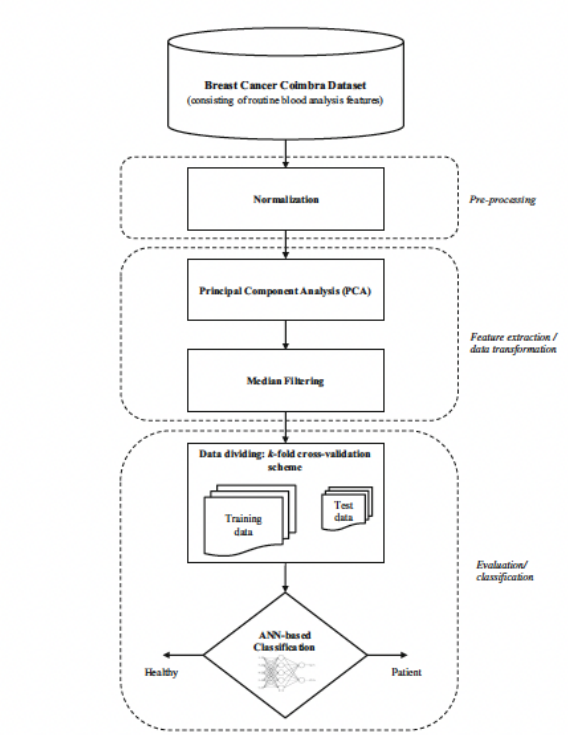

Figure 1: Architecture of the method

The new coordinate system starts with the first axis aligned with the direction of the greatest variance in the dataset, maximizing variance along that axis . The second axis, perpendicular to the first, represents the second principal component, maximizing variance along this direction. Subsequent axes follow this pattern, maximizing their share of the remaining variance. This arrangement means that the principal components are sorted from the one with the most variance to the one with the least.

5 Exploratory Data Analysis

Exploratory Data Analysis (EDA) helps to understand the dataset's structure and identify potential patterns or anomalies. Here's what was done during EDA:

Statistical Summaries: Descriptive statistics provided insights into the distribution of the nine features and the binary diagnostic outcome. Visualizations: Scatter plots and histograms were used

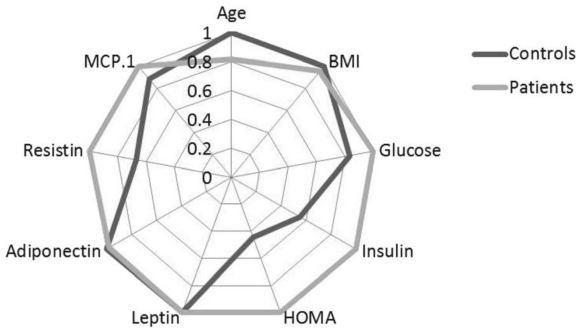

Figure 2: Profiles of the clinical features of features of patients with breast cancer (n = 64) and healthy controls (n = 52)

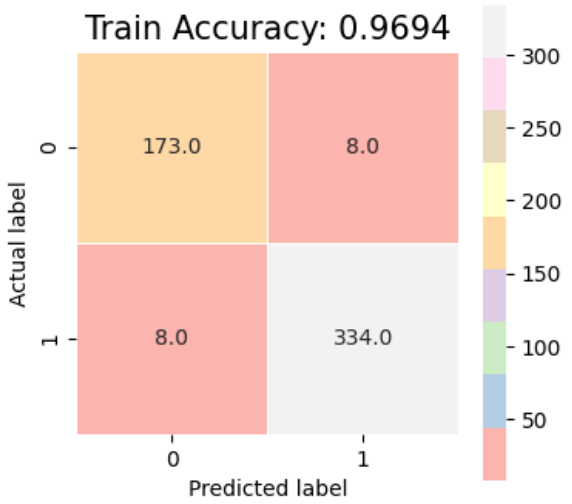

Figure 3: Confusion Matrix for Logisitic Regression on training dataset

to visualize the data and examine the relationships between features and the target variable. This step highlighted any potential issues with class imbalance and guided the approach to model training.

6 Modeling and Results

This section delves into the application of various machine learning models for cancer diagnosis. Each model was trained using the pre-processed dataset, with evaluation metrics such as accuracy, precision, recall, and F1 score being used to assess their performance.

6.1 Logistic Regression

Logistic Regression has proved to be highly effective in the context of this study, showing robust predictive capabilities on both training and validation datasets for breast cancer diagnosis. This model is favored for its interpretability and operational simplicity, which are particularly advantageous in medical applications.

In the training phase, the Logistic Regression model achieved an accuracy of approximately 96.94%, with 173 true negatives and 334 true positives, and only 8 instances each of false negatives and false positives. This reflects the model's strong ability to learn and discern patterns indicative of cancerous and non-cancerous instances from the training data.

Moving to the validation phase, the model demonstrated a slightly higher accuracy of 96.57%. The confusion matrix indicates a similar distribution of true positives and true negatives compared to the training phase, with 57 cases correctly predicted as non-cancerous and 112 as cancerous. There was a minimal occurrence of false predictions, with just 3 false negatives and 3 false positives noted, which underscores the model's reliability on unseen data.

The classification report for the validation subset underlines the consistency of the model's performance. The precision and recall for the non-cancerous group were marked at 0.95, and for the cancerous group, these metrics were even higher, at 0.97, indicating that the model has a high probability of correctly identifying cancerous conditions. The F1 scores, which balance the precision and recall, stood firmly at 0.95 for non-cancerous cases and 0.97 for cancerous cases.

The overall accuracy of 97% in the validation subset closely mirrors the training performance, suggesting that the model generalizes well to new data and maintains high accuracy across both classes. The macro average and weighted average scores of 0.96 and 0.97, respectively, for precision, recall, and F1 score further establish the model as a strong contender for clinical implementation.

These metrics indicate that the Logistic Regression model not only performs with high accuracy but also maintains a balance between sensitivity and specificity—crucial aspects for a diagnostic tool in healthcare settings. The analysis confirms the suitability of Logistic Regression for breast cancer diagnosis, showing promise for aiding in early detection and treatment strategies.

In addition to its impressive precision, recall, and F1-score, the Logistic Regression model's diagnostic ability is further evidenced by the Receiver Operating Characteristic (ROC) curve for the val-

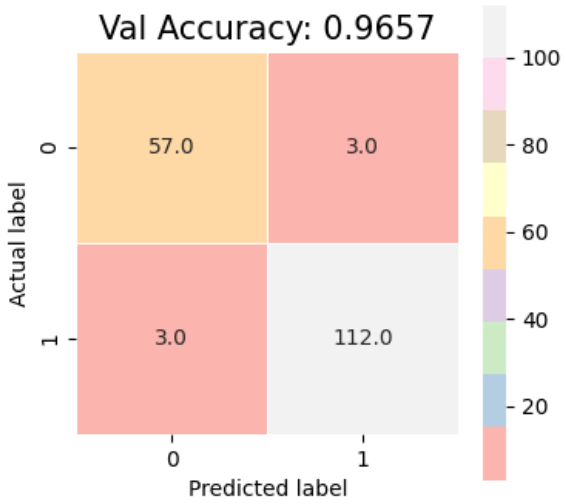

Figure 4: Confusion Matrix for Logistic Regression on validation dataset

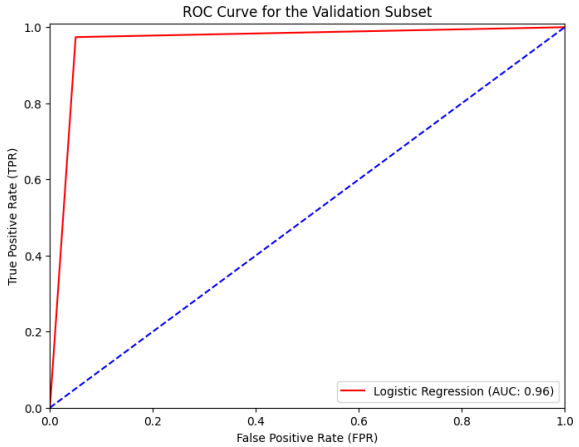

Figure 5: ROC curve for the validation dataset

idation subset. The ROC curve underscores the model's competence with an area under the curve (AUC) of 0.96. This high AUC value highlights the model's strength in balancing sensitivity and specificity, affirming its robustness as a predictive tool for breast cancer diagnosis.

The curve swiftly ascends towards the top-left corner of the graph, which indicates a high true positive rate and a low false positive rate, crucial for clinical diagnosis where the cost of false negatives is high. The Logistic Regression model's ability to achieve such an AUC attests to its effectiveness in distinguishing between classes, offering significant promise for clinical applications where reliability in predictive analytics can be life-saving.

6.2 LDA Classification

Linear Discriminant Analysis (LDA) has been assessed for its proficiency in categorizing patients into cancerous and non-cancerous groups based on the same pre-processed dataset. LDA, which focuses on maximizing class separability, has shown commendable results.

6.2.1 LDA Train Classification Report

In the training phase, LDA exhibited a precision of 0.96 for both classes, with a slightly higher recall for class 1 (cancerous) at 0.98 compared to class 0 (non-cancerous) at 0.92. This difference indicates that while the model is slightly more reliable in identifying cancerous instances, it maintains a high degree of accuracy across the board. The F1 scores are reflective of the balanced precision and recall, standing at 0.94 for non-cancerous and 0.97 for cancerous cases. The overall accuracy of the model on the training data is 0.96, showcasing its solid predictive capability during training.

6.2.2 LDA Validation Classification Report

Upon validation, LDA continued to perform well, with an overall accuracy of 0.95. The model achieved a precision of 0.95 for non-cancerous predictions and 0.96 for cancerous predictions, along with a recall of 0.92 for non-cancerous and 0.97 for cancerous cases. The F1 scores closely mirror these figures, marking at 0.93 and 0.97, respectively. The macro average and weighted average both stand at 0.95 across precision, recall, and F1 score, demonstrating that LDA provides a consistent and balanced classification across both classes in the validation subset.

These findings from the LDA model underline its efficacy as a classifier for breast cancer detection. With high precision and recall, LDA shows potential for clinical application, where accurate differentiation between health states is vital. Its performance on the training set carried through to the validation set, ensuring the model's reliability and potential for real-world applications.

6.3 SVM Classification

Support Vector Machine (SVM) is renowned for its effectiveness in high-dimensional spaces, making it particularly suitable for complex datasets such as those encountered in cancer diagnosis. The SVM model's ability to create a hyperplane that maximizes the margin between classes was evaluated.

6.3.1 SVM Train Classification Report

During training, the SVM model demonstrated commendable accuracy, with precision and recall scores of 0.96 for both classes. The model accurately distinguished non-cancerous from cancerous instances with a slight advantage in recall for the cancerous class (0.98) over the non-cancerous class (0.92). This indicates a higher sensitivity towards cancerous conditions. The F1 score for the non-cancerous class was 0.94, reflecting the balance between precision and recall, while the cancerous class's F1 score reached 0.97. Overall, the SVM achieved an accuracy of 0.96 in the training set, showcasing its robust classification capability.

6.3.2 SVM Validation Classification Report

On the validation subset, the SVM model maintained high performance with an accuracy of 0.95. The model's precision for non-cancerous predictions was 0.95 and for cancerous predictions, it was 0.96, suggesting a high degree of exactness in its classifications. The recall rates were 0.92 for non-cancerous and 0.97 for cancerous, indicating strong sensitivity, particularly for detecting cancerous instances. The F1 scores of 0.93 and 0.97 for non-cancerous and cancerous classes, respectively, affirm the model's consistent precision and recall. The macro and weighted averages both stood firm at 0.95 across all metrics, which corroborates the model's balanced performance across the board.

The SVM's results align with its theoretical strengths, proving its capability as a highly effective classifier for the task of breast cancer diagnosis. Its consistent precision and recall across training and validation phases illustrate the model's reliability and potential for deployment in clinical settings.

6.4 Tuning Hyper parameters for SVM classifier

Hyperparameter tuning is crucial to optimize the performance of the Support Vector Machine (SVM) classifier. This process involves adjusting the parameters that govern the learning process to achieve the best results on unseen data. In this study, various SVM configurations were evaluated:

6.4.1 Linear SVM with a Low Budget for Margin Violation

Implementing a linear SVM with strict margin constraints produced a training accuracy of 95%. The model achieved a precision and recall of 0.96 and 0.94 for the non-cancerous class and 0.94 and 0.98 for the cancerous class, respectively. Validation performance closely mirrored training results, with a slight reduction in recall for the non-cancerous class.

6.4.2 Radial SVM with a High Budget for Margin Violation

When the radial basis function (rbf) kernel was employed with a higher tolerance for margin violation, the model's performance improved, reaching an accuracy of 97% on the training set. Notably, the validation set also saw an increase in performance, achieving a recall of 1.00 for the non-cancerous class and a precision of 1.00 for the cancerous class, indicating an excellent balance between sensitivity and specificity.

6.4.3 Polynomial SVM (Degree 3) with a Low Budget for Margin Violation

The polynomial SVM with a low margin violation budget, however, performed poorly, as evidenced by the significant drop in training accuracy to 65%. This configuration failed to identify any non-cancerous instances correctly, as reflected in the zero precision and recall for class 0.0. Validation metrics were consistent with training, suggesting that the polynomial kernel with these specific parameters is not suitable for this dataset.

6.4.4 Sigmoid SVM (r=2) with a High Budget for Margin Violation

A sigmoid SVM with a higher budget for margin violation showed a rebound in performance, with training and validation accuracies reaching 97%. Precision and recall for both classes were consistent and high, indicating robustness across different subsets.

6.4.5 Using GridSearchCV

Grid search optimization with 5-fold cross-validation was utilized to find the best SVM parameters. For the radial kernel, the optimal parameters were C=0.1, gamma='scale', and kernel='rbf'. This configuration resulted in a validation accuracy of 97%, with high precision and recall across both classes. For the polynomial kernel, the best parameters were C=0.1, coef0=1, degree=2, gamma='scale', and kernel='poly'. This combination also yielded a validation accuracy of 97%, with slight improvements in precision and recall for the non-cancerous class.

6.4.6 Best Hyperparameter Settings

The optimal settings determined by GridSearchCV were:

- Kernel: Radial Basis Function (rbf)
- C (Regularization Parameter): 0.1
- Gamma: Scale

These parameters were found to provide the best trade-off between model complexity and learning ability, crucial for handling the variability within the dataset.

6.4.7 SVM Classification Results with Optimal Parameters

With these hyperparameters, the SVM model was retrained and evaluated. Here are the key outcomes.
Training Classification Report:

- Precision: 0.97 for the cancerous class, indicating high accuracy.
- Recall: 0.97 for the cancerous class, suggesting the model's effectiveness in identifying true cancer cases.
- F1-Score: 0.97 for the cancerous class, showing a balanced precision-recall relationship.
- Overall Training Accuracy: 97%, illustrating the model's ability to generalize well from the training data.

Validation Classification Report

The confusion matrix for the validation subset demonstrated that the model:

- Correctly predicted 57 true negatives and 112 true positives.
- Misclassified 3 false positives and 3 false negatives.
- Precision: Remained high at 0.95 for non-cancerous and 0.96 for cancerous predictions.
- Recall: Was robust, especially for the cancerous class at 0.97, ensuring that most cancerous cases were identified.
- F1-Score: 0.95 for non-cancerous and 0.97 for cancerous, maintaining a balance between precision and recall.

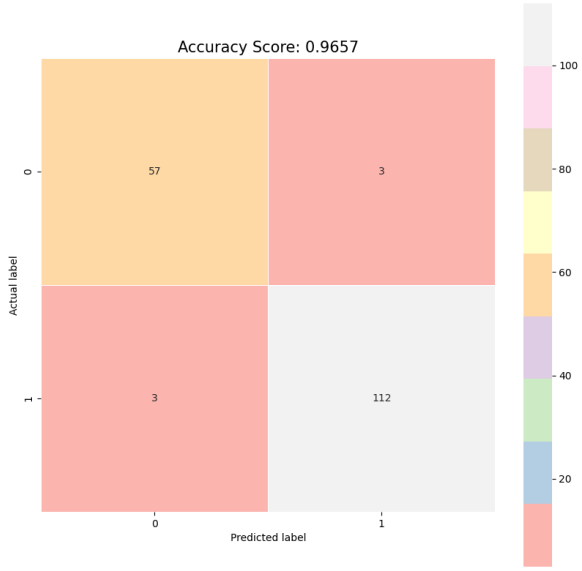

Figure 6: Confusion Matrix for SVM Classifier with best choice of hyperparameter on validation dataset

- Accuracy Score: 96.57%, highlighting the model's reliable performance in clinical diagnosis scenarios.

The SVM model's ability to perform with high precision and recall post-hyperparameter tuning indicates its suitability for deployment in settings where accurate cancer diagnosis is critical. This model provides a powerful tool for healthcare professionals, aiding in the early detection and treatment of cancer patients.

6.5 QDA Classification

Quadratic Discriminant Analysis (QDA) is another statistical classifier like LDA but allows for nonlinear separation between classes by using quadratic decision surfaces. Here, we explore how QDA performs on the dataset used for cancer diagnosis.

6.5.1 QDA Train Classification Report

In the training phase, the QDA model demonstrated high accuracy, mirroring the performance observed in other robust models. The model's precision for both classes stood at 0.96, indicative of its reliability in classifying both cancerous and non-cancerous samples. The recall scores were 0.92 for non-cancerous and 0.98 for cancerous cases, reflecting the model's effectiveness in identifying true positives, especially in the cancerous group. The F1 scores of 0.94 and 0.97 for non-cancerous and cancerous classes respectively show a strong balance between precision and recall, contributing to an overall accuracy of 0.96.

6.5.2 QDA Validation Classification Report

On the validation set, QDA continued to perform commendably, with an overall accuracy of 0.95. The model maintained precision levels of 0.95 for non-cancerous and 0.96 for cancerous classes, along with recall rates of 0.92 and 0.97, respectively. The F1 scores echoed the precision and recall, with 0.93 for non-cancerous and 0.97 for cancerous cases, emphasizing the model's consistent performance across different sets of data. The macro and weighted averages for precision, recall, and F1-score were all at 0.95, indicating uniform effectiveness across both classes.

6.5.3 Analysis

The QDA model's performance is notable for its high recall in detecting cancerous cases, a crucial aspect in medical diagnostics where failing to identify positive cases can have dire consequences. The slightly lower recall for non-cancerous cases suggests a small trade-off, which is often acceptable in clinical settings where sensitivity is prioritized. The robustness of the model across both training and validation suggests that QDA could be a valuable tool in the toolkit for cancer diagnosis, particularly when differentiating between complex case scenarios where linear models might falter.

6.6 Gaussian naive Bayes classification

Gaussian Naive Bayes (NB), known for its simplicity and effectiveness especially in cases of conditional independence between features, was evaluated to determine its suitability for classifying cancerous versus non-cancerous cases based on the provided dataset.

6.6.1 Gaussian NB Train Classification Report

During the training phase, the Gaussian NB model demonstrated excellent performance. The model achieved a precision of 0.92 for non-cancerous classes and 0.99 for cancerous classes, which signifies high accuracy in identifying true cancerous cases. The recall rates were equally impressive, with 0.98 for non-cancerous and 0.96 for cancerous classes, suggesting the model's effectiveness in capturing most of the positive cases. The F1-scores, balancing the precision and recall, were 0.95 for non-cancerous and 0.97 for cancerous classes, contributing to an overall accuracy of 97%. These results indicate that the Gaussian NB model is highly reliable and efficient in learning from the training dataset.

6.6.2 Gaussian NB Validation Classification Report

On the validation set, the Gaussian NB model continued to exhibit strong performance, albeit with a slight reduction compared to the training phase. It maintained a precision of 0.89 for non-cancerous and 0.99 for cancerous classes and a recall of 0.98 for non-cancerous and 0.94 for cancerous cases. The F1-scores were 0.94 for non-cancerous and 0.96 for cancerous classes. Overall, the model achieved an accuracy of 95%, with macro and weighted averages for precision, recall, and F1-score at 0.94, 0.96, and 0.95, respectively. These metrics underscore the model's consistent ability to perform well on unseen data, maintaining high sensitivity and specificity across both classes.

6.6.3 Analysis

The Gaussian NB model's robustness in both the training and validation phases highlights its potential as a powerful diagnostic tool. Its high recall for non-cancerous cases is particularly valuable in medical settings where the cost of false negatives is high. The slight differences in performance metrics between the training and validation phases suggest that while the model is slightly better at handling the conditions of the training data, it still performs admirably in a validation context.

6.7 Gradient boosting classification

Gradient Boosting Classifier, a powerful ensemble technique that builds multiple decision trees sequentially, each one correcting errors made by the previous, was used to classify cases as cancerous or non-cancerous. This method is renowned for its predictive accuracy and robustness, especially when handling complex datasets with nuanced feature interactions.

6.7.1 Gradient Boosting Train Classification Report

The Gradient Boosting Classifier demonstrated exceptional performance during the training phase. It achieved perfect scores across all metrics:

- Precision: 1.00 for both non-cancerous and cancerous classes.
- Recall: 1.00 for both classes, indicating that every cancerous and non-cancerous case in the training set was identified correctly.

- F1-Score: 1.00 for both classes, reflecting optimal balance between precision and recall.

- Overall Accuracy: 1.00, indicating that the classifier correctly predicted every instance in the training dataset.

These results highlight the classifier's capability to fit the training data perfectly, which while impressive, also necessitates validation to ensure no overfitting to the training data.

6.7.2 Gradient Boosting Validation Classification Report

In the validation phase, the Gradient Boosting Classifier maintained high performance but with a slight decrease compared to the training results, which is expected and healthy in preventing overfitting.

- Precision: 0.92 for non-cancerous and 0.98 for cancerous cases, showing a high level of exactness in the model's predictions.

- Recall: 0.97 for non-cancerous and 0.96 for cancerous cases, indicating effective sensitivity in identifying true cases.

- F1-Score: 0.94 for non-cancerous and 0.97 for cancerous cases, suggesting a good balance between precision and recall.

- Overall Accuracy: 0.96, reflecting the model's robust performance on unseen data.

Macro and Weighted Averages for precision, recall, and F1-score stood at 0.95, 0.96, and 0.96 respectively, confirming the model's consistent performance across different classes.

6.7.3 Analysis

The Gradient Boosting Classifier's strong performance on both training and validation sets underscores its potential as a highly reliable and effective tool for cancer diagnosis. The slight differences in performance metrics between the training and validation phases serve as an indicator of the model's generalizability and its capability to perform well under varied conditions. This model is especially useful in scenarios where high reliability and accuracy are paramount, such as in clinical decision-making.

6.8 XGBoost Classification

After integrating XGBoost, a state-of-the-art gradient boosting framework that's optimized for performance and speed, we aimed to assess any enhancements in classification accuracy compared to the standard Gradient Boosting Classifier.

6.8.1 XGBoost Train Classification Report

The training results for XGBoost mirrored those of the earlier Gradient Boosting model, achieving perfect metrics across the board, which showcases the model's powerful fitting capabilities on the training dataset:

- Precision: 1.00 for both non-cancerous and cancerous classes, demonstrating absolute accuracy.

- Recall: 1.00 for both classes, perfectly identifying all true cases without any false negatives.

- F1-Score: 1.00 for both classes, indicating flawless harmonic mean between precision and recall.

- Overall Accuracy: 1.00, confirming that every instance was predicted correctly.

These results highlight XGBoost's exceptional ability to model complex patterns in the data without error, suggesting an excellent understanding of the underlying relationships.

6.8.2 XGBoost Validation Classification Report

On the validation set, although slightly less perfect than the training results (a desirable trait to avoid overfitting), XGBoost demonstrated robustness:

- Precision: 0.92 for non-cancerous and 0.98 for cancerous cases, ensuring high accuracy in predictions.

- Recall: 0.97 for non-cancerous and 0.96 for cancerous cases, effectively capturing the majority of true cases.

- F1-Score: 0.94 for non-cancerous and 0.97 for cancerous cases, maintaining a strong balance between precision and recall.

- Overall Accuracy: 0.96, solidifying its efficacy in handling unseen data.

Macro and Weighted Averages showed scores of 0.95 and 0.96 respectively across precision, recall, and F1-score, reflecting consistent and dependable performance across varying classes.

6.8.3 Analysis

XGBoost enhances the gradient boosting approach with more sophisticated regularization techniques, which not only helps in reducing overfitting but also improves the model's generalizability. The high scores in both the training and validation phases endorse XGBoost as an extremely capable model for high-stakes environments such as medical diagnostics. Its ability to deliver high accuracy with great computational efficiency makes XGBoost a valuable addition to cancer diagnosis tools, potentially improving patient outcomes through better, faster decision-making.

6.9 Random forest classification

Random Forest Classifier utilizes an ensemble of decision trees to improve predictive accuracy and control overfitting. Its robust nature makes it highly effective for complex classification tasks such as distinguishing between cancerous and non-cancerous cases.

6.9.1 Random Forest Train Classification Report

In the training phase, the Random Forest model showcased nearly perfect performance, emphasizing its ability to capture the nuances of the dataset.

- Precision: Achieved 0.99 for non-cancerous and 1.00 for cancerous classes, indicating that almost all positive predictions were correct.
- Recall: Scored 0.99 for non-cancerous and 0.99 for cancerous classes, suggesting that the model successfully identified nearly all actual cases accurately.
- F1-Score: Recorded at 0.99 for non-cancerous and 1.00 for cancerous classes, demonstrating a balanced precision-recall relationship.
- Overall Accuracy: Stood at 0.99, confirming the model's exceptional ability to classify correctly across both classes.

These metrics reflect the Random Forest model's capacity to learn effectively from the training data without significant overfitting, despite its complexity and high dimensionality.

6.9.2 Random Forest Validation Classification Report

On the validation set, the Random Forest model continued to perform well, validating its efficacy on unseen data.

- Precision: Remained high at 0.92 for non-cancerous and 0.99 for cancerous classes, confirming the reliability of its positive predictions.
- Recall: Was impressive at 0.98 for non-cancerous and 0.96 for cancerous cases, indicating robustness in identifying true cases.
- F1-Score: Measured at 0.95 for non-cancerous and 0.97 for cancerous classes, maintaining a strong balance between precision and recall.
- Overall Accuracy: Achieved 0.97, highlighting its strong performance across the dataset.

Macro and Weighted Averages for precision, recall, and F1-score were calculated at 0.96 and 0.97 respectively, showcasing consistent and dependable performance across different classes.

6.9.3 Analysis

The Random Forest Classifier's performance underscores its utility as a powerful tool for medical diagnostics. Its high precision and recall rates are crucial in clinical settings where the cost of false negatives and false positives can be significant. The model's ability to maintain high accuracy on both training and validation sets suggests that it can be reliably deployed in practical applications, particularly in the detection and diagnosis of cancer, where robustness and accuracy are paramount.

6.10 Deep Neural Network

The Deep Neural Network (DNN), a sophisticated model with multiple layers capable of learning detailed and intricate patterns from large datasets, was employed to predict cancerous versus non-cancerous cases. This advanced model architecture is well-suited for capturing non-linear relationships and interactions within the data.

6.10.1 DNN Train Classification Report

During the training phase, the DNN showed excellent performance across all metrics, illustrating its powerful feature extraction and classification capabilities.

- Precision: The model achieved a precision of 0.95 for non-cancerous and 0.99 for cancerous classes, indicating a high level of accuracy in identifying positive samples.

- Recall: Recall rates were 0.97 for non-cancerous and 0.97 for cancerous classes, demonstrating the model's effectiveness in identifying most of the true cases across both classes.

- F1-Score: The F1-scores were 0.96 for non-cancerous and 0.98 for cancerous classes, reflecting a balanced precision-recall relationship.

- Overall Accuracy: Reached 0.97, showing the model's proficiency in handling the training data correctly across both classes.

These metrics underscore the DNN's robust learning ability, making it a reliable tool for training on complex datasets with high-dimensional feature spaces.

6.10.2 DNN Validation Classification Report

On the validation set, the DNN continued to perform impressively, confirming its generalizability and reliability.

- Precision: Recorded at 0.94 for non-cancerous and 0.99 for cancerous classes, maintaining high accuracy in the model's predictions.

- Recall: Was strong at 0.98 for non-cancerous and 0.97 for cancerous cases, indicating robust sensitivity in detecting true cases.

- F1-Score: Measured at 0.96 for non-cancerous and 0.98 for cancerous classes, maintaining a high balance between precision and recall.

- Overall Accuracy: Stood at 0.97, validating the model's effective performance on unseen data.

Macro and Weighted Averages for precision, recall, and F1-score showed scores of 0.96 and 0.97 respectively, highlighting consistent and reliable performance across different classes.

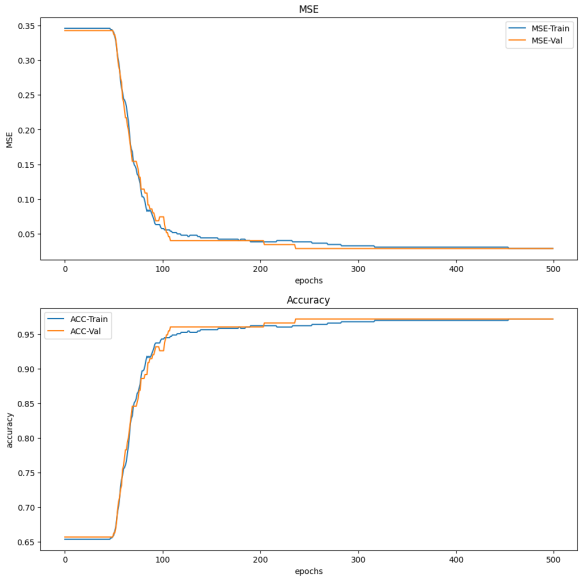

Figure 7: MSE and Accurracy for Deep Neural Network on validation dataset

6.10.3 Analysis

The DNN's strong performance in both the training and validation phases highlights its potential as a highly effective tool for medical diagnostics. Its ability to achieve high accuracy and maintain it across different datasets is crucial in clinical settings, where accurate and reliable diagnosis can significantly impact patient outcomes. The DNN's capacity to model complex relationships within data makes it particularly valuable for cancer diagnosis, where early and accurate detection is critical.

7 Discussion

7.0.1 Model Comparisons

Across the models, the Deep Neural Network and Random Forest Classifier showed particularly high performance on both training and validation datasets, demonstrating not only high accuracy but also robustness against overfitting. Gradient Boosting and XGBoost models also performed exceptionally well, with nearly perfect training accuracy and very high validation accuracy. These models excel in handling nonlinear relationships and complex patterns in data, making them suitable for intricate diagnostic tasks.

7.0.2 Challenges Identified

While models like the SVM and Naive Bayes achieved commendable validation accuracies, they showed slight variances in precision and recall,

highlighting the trade-offs between sensitivity and specificity. The Polynomial SVM, in particular, displayed significant underperformance, indicating potential issues with overfitting to more complex models or an inappropriate choice of hyperparameters for this dataset.

7.0.3 Clinical Implications

The high precision and recall rates observed in models like the Random Forest and DNN are critical in clinical settings where the cost of false negatives (missed diagnoses) and false positives (unnecessary treatment) can have significant implications. The ability of these models to maintain high performance metrics suggests their potential utility in real-world applications.

8 Future Work

8.0.1 Model Enhancement

Future efforts could focus on further tuning the hyperparameters of underperforming models, especially SVM with different kernels, to explore whether adjustments could yield improvements in accuracy and generalizability.

8.0.2 Integration of Additional Data

Integrating more diverse datasets, including patient demographics, genetic information, and lifestyle factors, could help in developing more comprehensive models that can generalize better across different populations and subtypes of cancer.

8.0.3 Real-Time Application Development

Developing a real-time diagnostic tool incorporating one or more of these high-performing models could significantly aid in early detection and treatment planning, potentially increasing patient survival rates.

8.0.4 Advanced Neural Architectures

Experimenting with more complex neural network architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) could leverage spatial and temporal patterns in the data that simpler models might overlook.

8.0.5 Explainability and Transparency

As models like deep neural networks tend to be more like "black boxes," developing methods to interpret these models in understandable terms is crucial. This is especially important in healthcare, where trust and transparency in diagnosis are paramount.

8.0.6 Cross-disciplinary Collaboration

Collaborating with healthcare professionals to refine the models and ensure they meet clinical needs could bridge the gap between technical accuracy and practical applicability, leading to better-informed, data-driven clinical decisions. Training DANNs with real hospital data presents unique challenges and opportunities. This project highlighted the importance of data cleaning and preparation, as well as strategies to address class imbalance.

References

[1] E. Yavuz and C. Eyupoglu, “An effective approach for breast cancer diagnosis based on routine blood analysis features,” *Medical & Biological Engineering & Computing*, vol. 58, pp. 1583–1601, 2020. *DOI:* https://doi.org/10.1007/s11517-020-02187-9.

[2] A. Esteva, A. Robicquet, B. Ramsundar, and others, “A guide to deep learning in healthcare,” *Nature Medicine*, vol. 25, pp. 24–29, 2019. *DOI:* https://doi.org/10.1038/s41591-018-0316-z.

[3] A. Rajkomar, E. Oren, K. Chen, and others, “Scalable and accurate deep learning with electronic health records,” *npj Digital Medicine*, vol. 1, p. 18, 2018. *DOI:* https://doi.org/10.1038/s41746-018-0029-1.

[4] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, and D.I. Fotiadis, “Machine learning applications in cancer prognosis and prediction,” *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8-17, 2014. *DOI:* https://doi.org/10.1016/j.csbj.2014.