DATA SCIENCE SEMINAR (DASC 6810)

WINTER 2025

PROJECT 1

INSTRUCTOR: Mohamed Tawhid

GROUP G:

Arpitha Thippeswamy (T00749833)

Taiwo Ogunkeye (T00751495)

Affiliation

Faculty of Science, Thompson Rivers University,Kamloops

Date: 15th February 2025

Breast Cancer Predictions using Classification Models and Neural Networks.

Table Of Content

1	Table Of Content	2
2	Abstract	3
3	Background and Motivation	3
4	Objectives	4
5	Literature Review *a.* Breast Cancer Detection Using Machine Learning Algorithms (Sharma et al., 2018) *b.* Breast Cancer Classification Using Machine Learning(Amrane et al..,2018) *c.* Machine Learning Techniques to Diagnose Breast Cancer(Osareh & Shadgar,2010)	4 4 5 5
6	Limitations Of Existing Research	6-7
7	Approaches	8
8	Methodology 8.1. Datasets 8.1.1.Wisconsin Dataset 8.1.2. Breast-Cancer-Dataset 8.1.3. BRCA Dataset 8.1.4. German BS Dataset 8.1.5. Seer Dataset 8.2.Data Quality 8.3Feature Correlations 8.4.Preprocessing 8.5.Models 8.6.Evaluation Metrics	8 8 9 9 9 9 9 10-11 12-13 14 15 16
9	Results	17-19
10	Limitations	19
11	Future Work	20
12	Conclusion	20
13	Github Repository	20
14	References	21 - 22

2. Abstract

The objective of this project is to train machine learning models to predict breast cancer using classification techniques and neural networks. Five Kaggle datasets were preprocessed to handle class imbalance, normalization of numerical features, and one-hot encoding for categorical features. Three approaches were utilized: individual classification models (SVM, Logistic Regression, Random Forest, XGBoost), a weighted fusion model of these classification models, and a feedforward neural network (MLP). Model performances were evaluated using accuracy, F1-score, and ROC-AUC. Results were good on organized datasets such as Wisconsin but struggled with complicated datasets such as BRCA. Areas of future research are sophisticated oversampling methods, hyperparameter tuning, and ensemble techniques to enhance generalization and robustness.

3. Background and Motivation

Breast cancer is the most prevalent cancer in women all over the world and is also the cause of a high proportion of cancer-related deaths. It is estimated by the World Health Organization (WHO) that over 2 million women are diagnosed with breast cancer every year, and approximately 600,000 died in 2018 alone (Sharma et al., 2018). Late diagnosis is most frequently given as the cause of the high mortality rate, as early-stage breast cancer is asymptomatic. Early diagnosis, therefore, becomes extremely crucial to improve survival and treatment outcomes.

Mammography, ultrasound, and magnetic resonance imaging (MRI) are conventional techniques of detecting breast cancer and rely heavily on the expertise of radiologists. The techniques are prone to human error, since studies have revealed that radiologists miss as many as 15% of breast cancer (Yassin et al., 2018). Besides, radiological image interpretation is subjective and greatly reliant on the practitioner. To address these limitations, more emphasis has been placed on developing computer-aided diagnosis (CAD) systems using machine learning and artificial intelligence to assist radiologists in detecting and diagnosing breast cancer more accurately.

Support Vector Machine (SVM), k-Nearest Neighbors (kNN), and Random Forest machine learning algorithms have also been proven to detect breast cancer with great accuracy. These models are capable of handling vast data sets, identifying patterns, and making accurate predictions.Feature selection and extraction techniques, such as Principal Component Analysis (PCA) and Signal-to-Noise Ratio (SNR), have also been employed to improve the performance of these models by reducing dimensionality and eliminating irrelevant features (Osareh & Shadgar, 2010).

The impetus for this project is the need to develop more and more effective models of forecasting breast cancer so that they can inform medical practitioners to make enhanced

decisions. In drawing from existing knowledge in earlier research, the project aims to assist the present effort towards a better diagnosis of breast cancer as well as avoidance of death caused by it.

4. Objectives:

The overall objective of this project is to develop a machine learning-based breast cancer prediction model that can effectively classify malignant and benign tumors. The specific problems to be addressed are:

- Literature Review: Thoroughly review the existing literature on the application of machine learning methods for breast cancer prediction, including techniques, datasets, and evaluation metrics adopted in previous research.

- Data Preprocessing: Preprocessed the datasets, dealt with the class imbalances, though there were no missing values, scaling with standard scaler and one hot encoding.

- Feature Extraction and Selection

- Modelling phase: Use of 2 approaches using classification models and neural networks.

- Model Evaluation with some metrics like accuracy, F1 score.

5. Literature Review

The project literature review is based on three important papers that address various machine learning approaches and methods to breast cancer diagnosis and prediction. The papers provide perspective on the possibilities and limitations in the field, in addition to the performance of various algorithms and feature selection methods**.**

1. Detection of Breast Cancer Using Machine Learning Algorithms (Sharma et al., 2018)

Sharma et al. (2018) provided a comparison of three machine learning models, namely Random Forest, k-Nearest Neighbors (kNN), and Naïve Bayes, for the classification of breast cancer. The Wisconsin Diagnosis Breast Cancer (WDBC) database, consisting of 569 instances and 32 features, was employed by the authors. The authors compared the performance of algorithms based on accuracy, precision, recall, and F1 score. The result

showed kNN having the highest accuracy level (95.90%) by Random Forest with 94.74% accuracy and Naïve Bayes with 94.47%. The study laid emphasis on how feature selection becomes important as well as conducting additional research in the future to be able to perform better in detecting breast cancer for machine learning algorithms

2.**Breast Cancer Classification Using Machine Learning (Amrane et al., 2018)**
Amrane et al. (2018) proposed a comparison between kNN and Naïve Bayes classifiers for breast cancer classification. The work utilized the Wisconsin Breast Cancer Database (WBCD) and cross-validated. kNN performed best at 97.51% and Naïve Bayes at 96.19%. The authors mentioned that the significance of cross-validation is utilized in measuring the performance of machine learning classifiers and proposed that kNN is a more appropriate classifier for addressing breast cancer classification problems.

3. Machine Learning Techniques to Diagnose Breast Cancer (Osareh & Shadgar, 2010)
Osareh and Shadgar (2010) studied the usage of Support Vector Machines (SVM), kNN, and Probabilistic Neural Networks (PNN) to identify breast cancer.Two datasets were employed in the study: fine-needle aspirate dataset and gene microarray dataset. Feature selection techniques such as Signal-to-Noise Ratio (SNR) and Sequential Forward Selection (SFS) were employed by the authors in an effort to improve model performance. The results showed that SVM with RBF kernel generated the highest accuracy (98.80%) on the fine-needle aspirate dataset and 96.33% on the gene microarray dataset. The study demonstrated the effectiveness of feature selection and the effectiveness of SVM in breast cancer diagnosis.

6. Limitations of Existing Research:

The three articles on machine learning algorithms used in breast cancer detection and diagnosis are enlightening regarding how different approaches can be utilized to enhance diagnosis accuracy. However, certain gaps and shortcomings in the available research must be filled in future research:

1. Restricted Dataset Diversity and Dimension
    - Paper 1 (Breast Cancer Detection using Machine Learning Approaches): This paper utilizes the Wisconsin Diagnosis Breast Cancer (WDBC) database, which just so happens to be not too big in size (569 instances). This data might likely not represent all the categories of cases of breast cancer, especially in different groups or populations.

    - Paper 2 (Breast Cancer Classification through Machine Learning): The study also uses the Wisconsin Breast Cancer database, which contains 699 data points. The data set is widely used and might not be appropriate for all breast cancers, particularly geographically or clinico-pathologically.

- Paper 3 (Machine Learning Techniques to Diagnose Breast Cancer): Paper 3 employs two data sets, one consisting of 692 cases and the other of 295 gene microarrays. The strength of using gene microarray data is undermined by the relative smallness of the data set and the fact that the paper does not inform us about how generalizable the outcomes are to other, larger, more heterogeneous populations.

2. Absence of External Validation
   All three studies employ cross-validation on the same dataset to make model comparisons. While cross-validation is a robust technique, it does not replace the need for external validation on other datasets. None of the studies carry out external validation on alternate datasets, and this creates questions regarding the generalizability of the results to other clinics or populations.

3. Inadequate Exploration of Feature Importance
   - Paper 1: Paper 1 doesn't explore as much the impact of the individual features in the data. Even where it does mention significant variables, it won't perform the full research on how the input of each characteristic affects model performance.
   - Paper 2: Yet this paper only uses methods for ranking features by the signal-to-noise ratio and does not touch upon the more sophisticated methods of selection or extraction of features that would enhance the quality of the model performance.
   - Paper 3: While the methods of feature ranking and selection (SNR, PCA, and SFS) are ascribed to this paper, there is one aspect that is barely attended to: how do these features interact with one another and how do they affect the classification processes?

4. Model Interpretability That Is Unexplored
   In all three works, the authors focus on high accuracy and not on interpretability of models. In healthcare, interpretability of models is very crucial for clinicians to trust and comprehend the predictions. Methods such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) could be utilized in modeling how decisions are made.

5. Limited Excursion into Imbalanced Data Handling
   - Both papers explore the domain of imbalanced classification (malignant vs. benign), although neither paper addresses the apparent issue of class imbalance itself. These methods for increasing minority

class performance could be oversampling, undersampling, or the utilization of class-weighted models.
- Paper 3 uses the Matthews Correlation Coefficient to estimate imbalanced datasets but does not discuss advanced techniques dealing with class imbalance, like SMOTE or ADASYN.

6. Limited Research on Advanced Learning Techniques
    - Paper 1: Some research involved the use of traditional machine learning methods: Naïve Bayes, Random Forest, kNN. While these processes do all right in theory, there may have been even more sophisticated ways of learning over convolutional neural networks or kitchen oven ensemble-like techniques like XGBoost, LightGBM that could also have done well.
    - Paper 2: Like paper 1, paper 2 looks at the shared classifiers (Naïve Bayes and kNN) but does not tend to examine more advanced methodologies.
    - Paper 3: SVM, kNN, and PNN are employed by the authors of paper 3 but are not attempted to follow deep learning and other upcoming strategies that are now beneficial in the context of medical image classification and analysis.

7. Limited Discussion on Model Robustness
   The reports are not giving enough information about the robustness of their models to noise, missing values, or attacks. Real-world data is generally noisy or missing; hence, any good model should be immune to its occurrence.

The literature review focuses on the potential of machine learning techniques for the improvement of breast cancer diagnosis and prediction. Sharma et al. (2018), Amrane et al. (2018), and Osareh and Shadgar (2010) are among the papers that provide significant information on how different algorithms work, the importance of feature selection, and the need for powerful evaluation methods. These findings will inform the development of the proposed breast cancer prediction model in such a way that it will be highly accurate and reliable in the prediction of benign and malignant tumors. Building on the strengths of previous studies, this project will contribute to enhancing computer-aided diagnosis systems for breast cancer.

While the three papers are valuable additions to the topic of breast cancer detection through machine learning, there are certain areas that can be addressed in further research. These are the calls for larger, more diverse data, external evaluation, improved feature selection and interpretation, real clinical deployment, imbalance data management, examination of high-end machine learning approaches, multiple data modality fusion, ethics, longitudinal models, model interpretability, cost-benefit

calculations. Closing such gaps could provide more accurate, interpretable, and clinically salient models to diagnose and classify breast cancer.

7. Approaches

We implemented and tested 3 different Machine Learning based approaches to predict breast cancer using 5 different datasets from Kaggle. These approaches are used to compare the performance of individual models , weighted fusion of these models and a feed forward neural network for classification tasks. The methods are the following:

1. Use of 4 different types of classification models:
    - Support Vector Machines (SVM)
    - Logistic Regression (LR)
    - Random Forest (RF)
    - XGBoost
2. Use of a weighted fused model of all the classification models run over all 5 datasets
3. Use of Neural Networks: A feedforward neural network (also known as a multi-layer perceptron, MLP).

8.Methodology

8.1 Datasets

We used 5 different datasets, publicly available in Kaggle, which are related to the prediction of breast cancer. Each dataset provides a different combination of numerical and categorical features, as well as a target variable that indicates the presence of a benign or malignant breast condition for the patient.

The datasets vary in size, characteristics, and data quality, offering the variety needed to undertake a full evaluation of the various approaches to this problem. This ensures that the results are consistent and can be extended to analogous cases within the context of the same problem.

8.1.1.1 Wisconsin Dataset

- Target: Diagnosis (B or M. Benign or Malignant)
- Numerical Features: radius_mean, texture_mean,

perimeter_mean, area_mean, smoothness_mean, compactness_mean, concavity_mean, concave points_mean, symmetry_mean, fractal_dimension_mean, radius_se, texture_se, perimeter_se, area_se, smoothness_se, compactness_se, concavity_se, concave points_se, symmetry_se, fractal_dimension_se, radius_worst, texture_worst, perimeter_worst, area_worst, smoothness_worst, compactness_worst, concavity_worst, concave points_worst, symmetry_worst, fractal_dimension_worst

- Categorical Features: none
- Number of Samples: 570
- Link: Wisconsin Dataset

8.1.1 Seer Dataset

- Target: Status (Alive or dead)
- Numerical Features: Age, Survival Months, Regional Node Examined

- Categorical Features Race, Marital Status, T Stage, N Stage, 6th Stage, differentiate, Grade, A Stage, Tumor Size, Estrogen Status, Progesterone Status, Reginol Node Positive
- Number of Samples: 4025
- Link: Seer Dataset

8.1.2 German BS Dataset

- Target: status(0 or 1 , dead or alive)
- Numerical Features: age, size, grade, nodes, pgr, er, rfstime
- Categorical Features meno, hormon
- Number of Samples: 686
- Link: German BS Dataset

8.1.3 BRCA Dataset

- Target: Patient_Status (Alive or dead)
- Numerical Features: Age, Protein1, Protein2, Protein3, Protein4
- Categorical Features: Gender, Tumour_Stage, Histology, ER status, PR status, HER2 status, Surgery_type
- Number of Samples: 335
- Link: BRCA Dataset

8.1.4 Breast-Cancer-Dataset

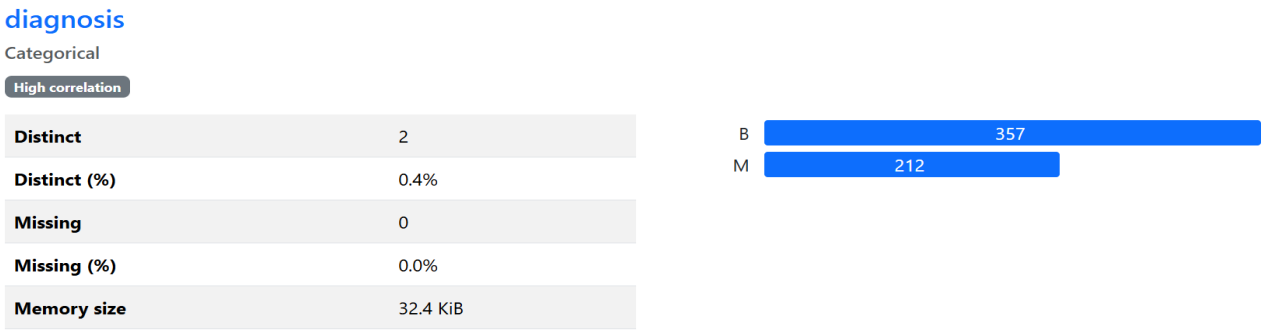
- Target: Diagnosis Result (B or M. Benign or Malignant)

- Numerical Features: Year, Age, Tumor Size (cm), Inv-Nodes
- Categorical Features: Menopause, Breast, Metastasis, Breast Quadrant, History
- Number of Samples: 214
- Link: [breast-cancer-prediction]

8.2 Data Quality

All datasets need cleaning and preprocessing. A data profiling package was run in all of them, and there were no signs of missing values and outliers. But on the other hand, all the 5 datasets Wisconsin dataset, BRCA dataset, german bs dataset, breast-cancer-dataset and seer dataset have class imbalance issue.

1. Wisconsin Dataset.



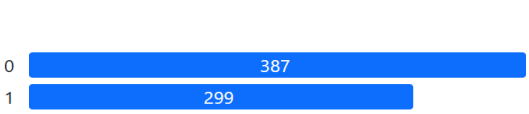
2. Seer Dataset

3. German BS Dataset.

status

Categorical

Distinct	2
Distinct (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory size	39.0 KiB



4.BRCA Dataset

Patient_Status

Categorical

Distinct	2
Distinct (%)	0.6%
Missing	0
Missing (%)	0.0%
Memory size	21.6 KiB

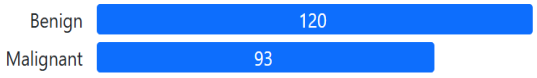
5.Breast – Cancer - Dataset

Diagnosis Result

Categorical

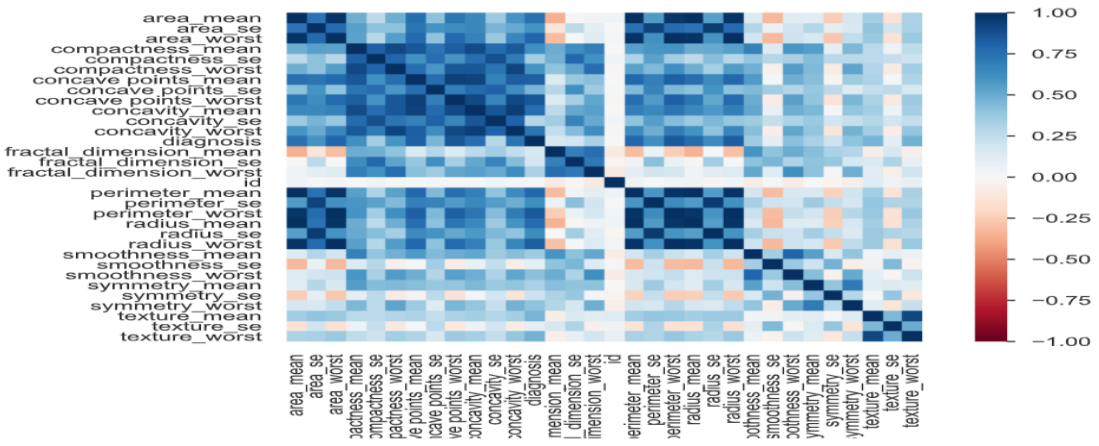
High correlation

Distinct	2
Distinct (%)	0.9%
Missing	0
Missing (%)	0.0%
Memory size	13.5 KiB

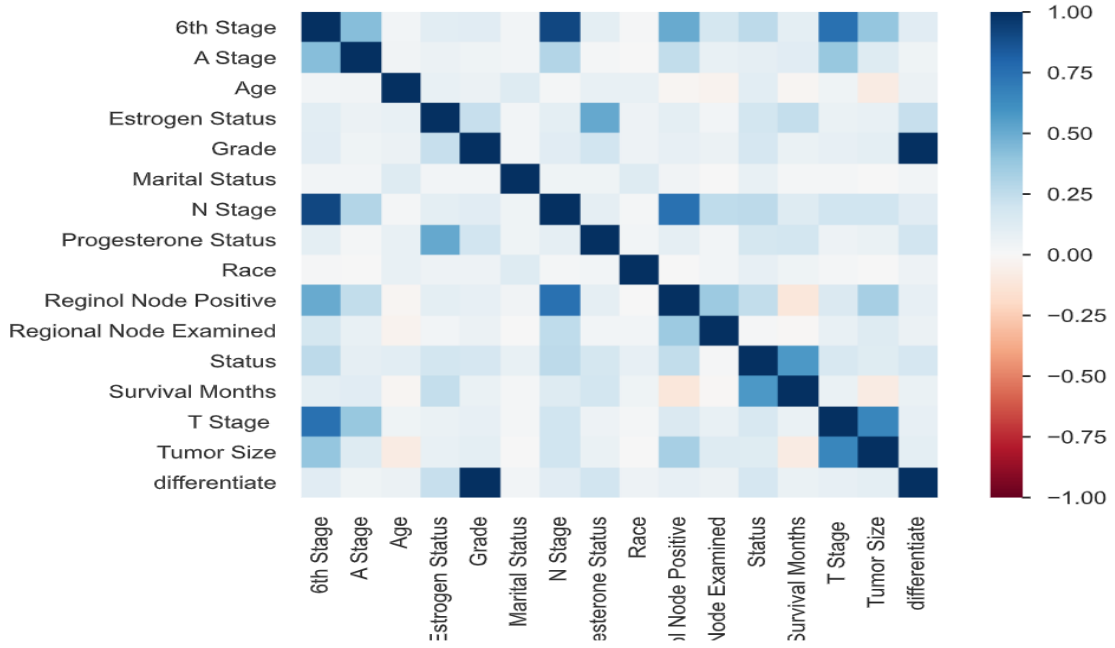


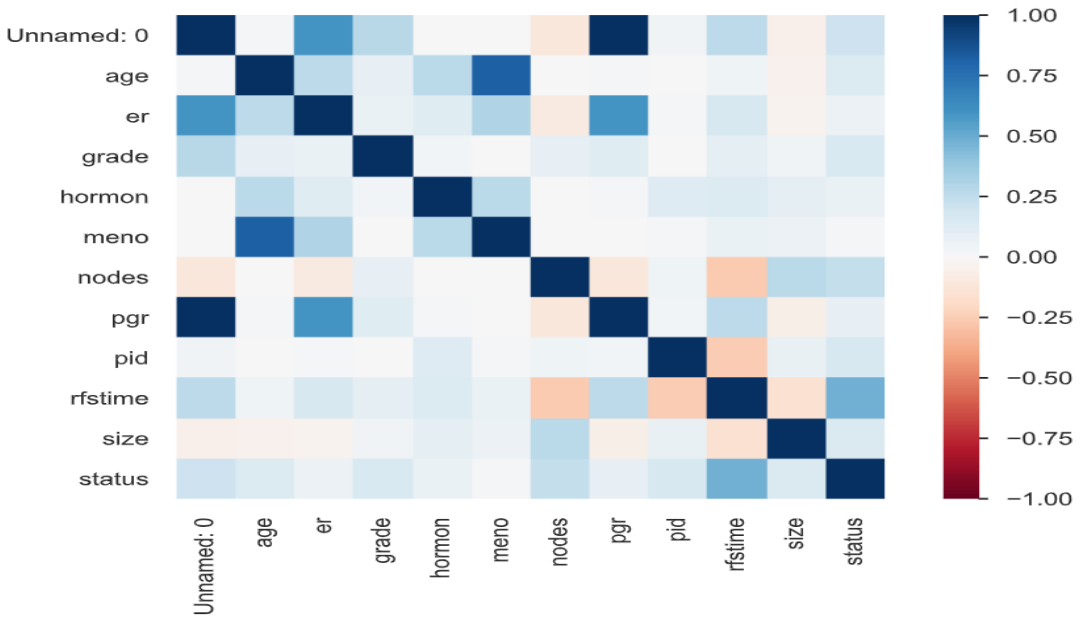
8.3 Feature Correlations.

1.Wisconsin Dataset - The highly correlated features with diagnosis are radius_mean` (0.73), radius_worst` (0.78), perimeter_mean` (0.74), `perimeter_worst` (0.78), `area_mean` (0.71), `area_worst` (0.74), concavity_mean` (0.70),`concavity_worst` (0.73), `concave points_mean` (0.69),and `concave points_worst` (0.72).



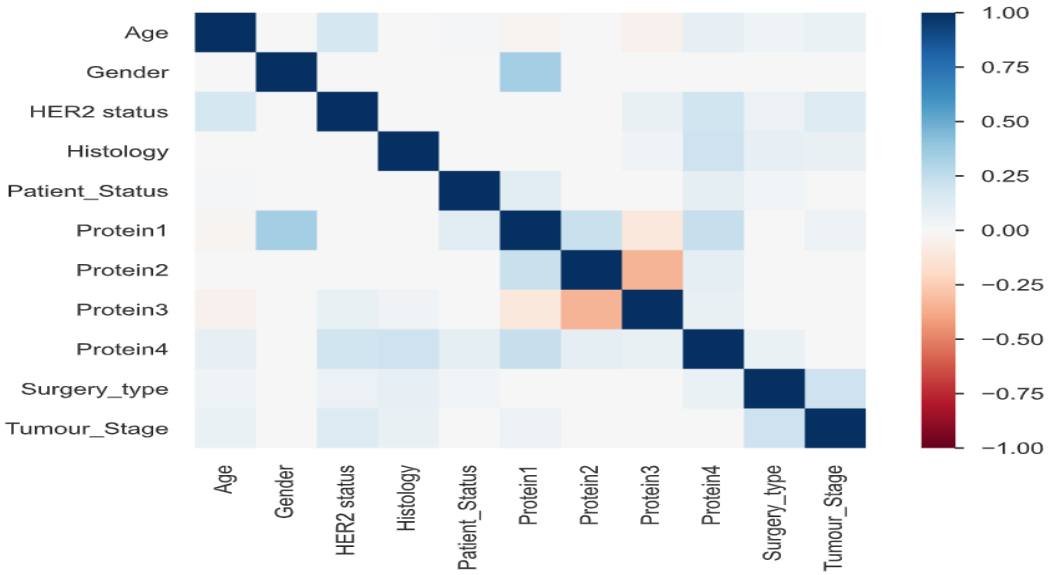
2.Seer Dataset - The features highly correlated with Status are Survival Months (+), T Stage (+), 6th Stage (+), and Tumor Size (+) .



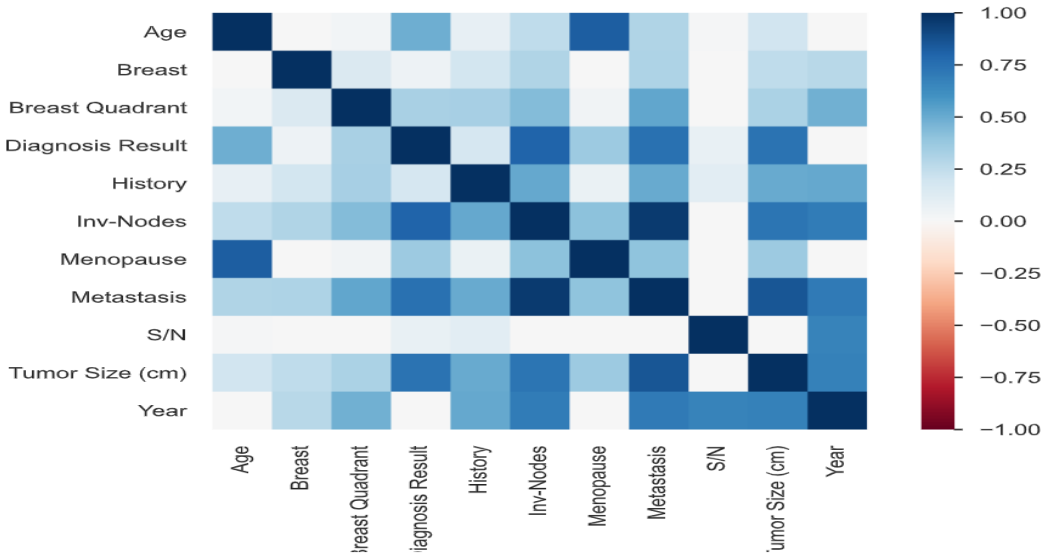


3.German BS Dataset – except rfstime (0.481) none of the other features are close to strong correlation with the target variable (status)

4.BRCA Dataset – none of the features show strong correlation with target variable – Patient_Status



5.Breast-Cancer-Dataset – Inv-Nodes(0.802),Metastasis(0.745),Tumor size(0.738)



show strong correlation with the target variable (diagnosis result).

8.4 Preprocessing

The preprocessing pipeline was implemented as an orchestrator in Scikit-Learn, using common steps for all datasets and models. The differences are mainly in how each of the individual numerical features, categorical features, and target variables are handled dynamically, based on the dataset.

The implementation involves a modular structure so that different versions of the same pipeline may be produced, using sub-pipelines for handling numerical and categorical features.

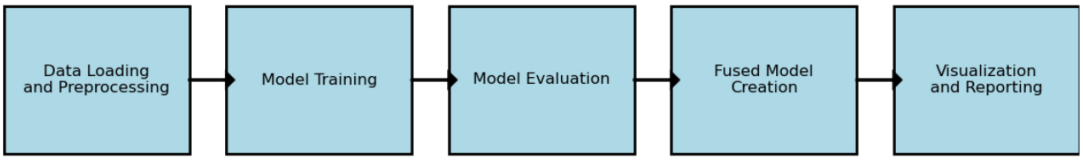
The Numerical Features: StandardScaler was used to normalize the numeric features, meaning that these features will have a mean of 0 and a standard deviation of 1. Initially, filling in the missing value in the numeric features would be done using a SimpleImputer; however, this was rectified upon confirming that no missing values are actually present within the numeric features.

The Categorical Features: For categorical features, One-Hot Encoding was used,

generating binary columns. No significant cardinality or sparsity was seen for the categorical features. A SimpleImputer was part of the original plan to handle any problems of missing values within categorical features, but this was dismissed after checking that no missing values are actually within the categorical columns.

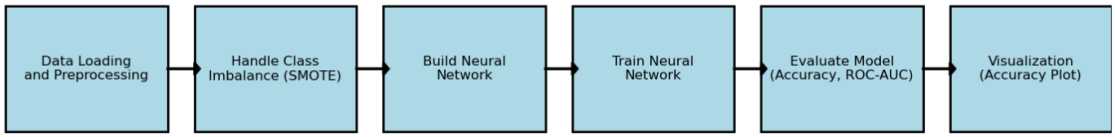
8.5 Models

Approach 1 – Architecture – Pipeline



The preprocessing phase consisted of correction of imbalances in classes, scaling of numerical features, one-hot encoding of categorical features, and no missing values or outliers found. The modeling phase applied Approach 1 by implementing different classification models including logistic regression factor, XGBoost, Random Forest, and SVM while Approach 2 worked by a weighted fusion of these models. Approach 3 introduced feed-forward neural networks (MLP). The models are evaluated on F1-score, accuracy, and ROC-AUC terms. In the fusion model, the predictions of all models were combined through a weighted average. Various performance metrics, including accuracy, F1-score, and AUC curves across datasets, were visualized in the results.

Approach 2 – Architecture Pipeline



The missing values were filled in, the categorical features were converted to numeric format, and numeric attributes were scaled. Class imbalance was addressed by applying SMOTE oversampling of the minority class. The neural network model architecture consisted of an input, hidden, and output layer, trained on a balanced-class dataset with class weights. The accuracy and ROC-AUC metrics were used to evaluate the model performance for further analysis. Also, a bar graph was drawn to show the model accuracy for the datasets, which clearly indicated the

efficiency and stability of its performance.

8.6 Evaluation Metrics

Accuracy:

This measure is the ratio of accurately classified instances to the total instances. It is simple and intuitive, but only effective if the dataset is balanced.

$$\text{Accuracy} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{Total Samples}}$$

F1 Score:

The F1 score is the harmonic mean of Precision and Recall and provides a single score that is a trade-off between the two.It's particularly helpful when working with imbalanced datasets where accuracy is not always an accurate measure of performance.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

AUC (Area Under the ROC Curve):

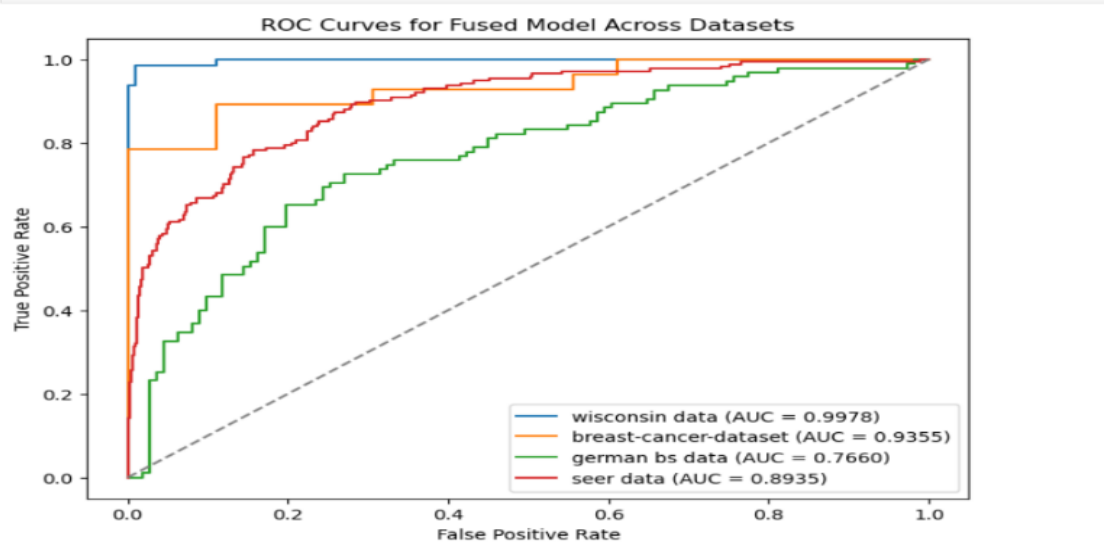
AUC estimates the model's performance across all possible classification thresholds by computing the area under the Receiver Operating Characteristic (ROC) curve. In our study, all are class-imbalanced. Since we did not apply oversampling or subsampling, F1 score may be low for some models in these datasets. However, AUC is a good measure because it assesses how well the model can distinguish between classes at various thresholds, thus the best measure for our study

$$\text{AUC} = \int_0^1 \text{TPR}\, d(\text{FPR})$$

9. Results

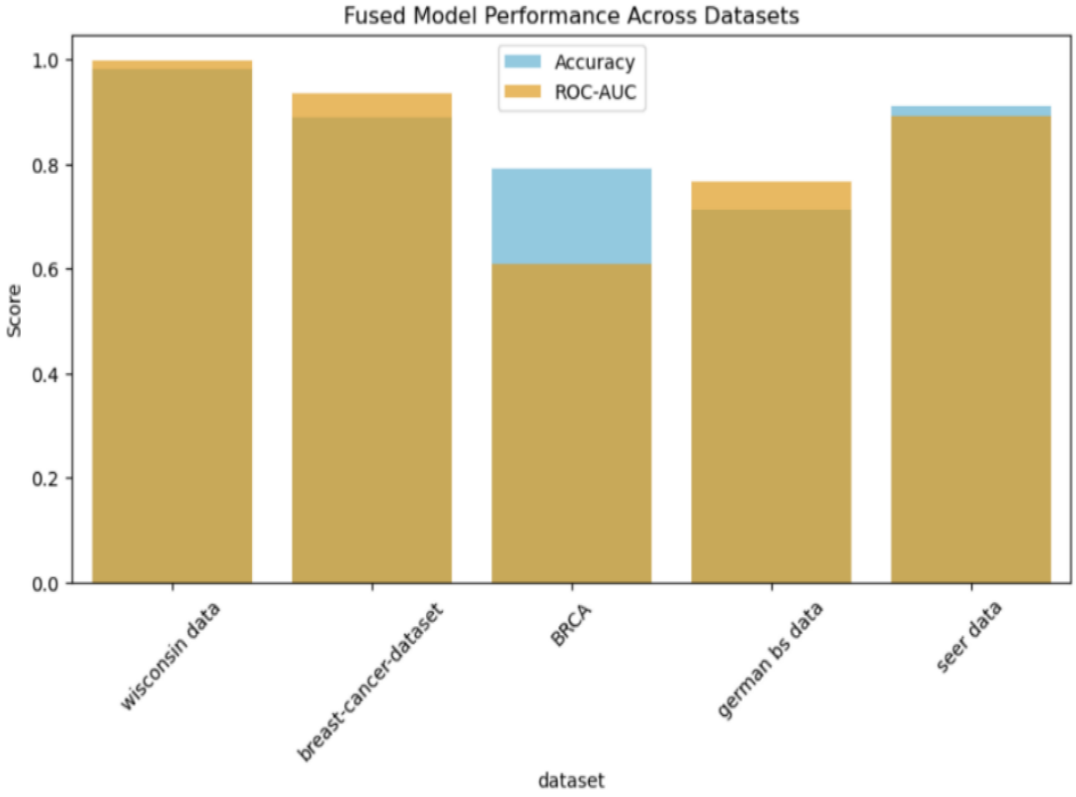
Approach 1 – Classification Models

ROC curve comparison was made with the True Positive Rate (TPR) versus False Positive Rate (FPR) trade-off at different values of thresholds across different performance levels of a model. Ratings for Area Under Curve (AUC) on data sets indicate performance levels for each modeling case.



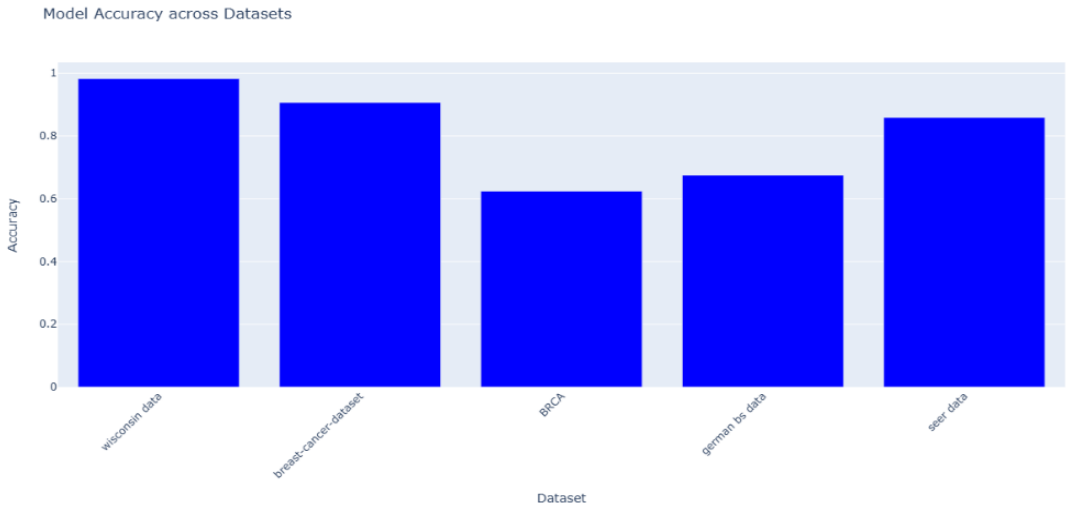
Summary :

- The model works like a charm on the Wisconsin dataset (AUC ≈ 0.9978).
- Scores decently on Breast Cancer and SEER datasets (AUC >0.89).
- Although the performance on the German IS dataset was lower (AUC = 0.7660), there is room for improvement.



The accuracy and ROC-AUC of a fused model were tested on five datasets (Wisconsin, breast cancer, BRCA, German BS, and SEER). The plot shows consistently high ROC-AUC across all datasets, which is a sign of excellent class separation. Accuracy, while generally good, varied somewhat between datasets. The Wisconsin dataset performed best overall. The combined model effectively combines data, with great discriminatory ability, though dataset-specific characteristics influence accuracy of classification.

Approach 2 – Neural Networks



The bar graph represents the accuracy achieved on five datasets: Wisconsin, breast-cancer, BRCA, German BS, and SEER. Accuracy scores lie between 0.6 and just under 1.0. The highest accuracy here is found for the Wisconsin dataset, which approaches 1.0. BRCA and German BS are about 0.6 and 0.7 respectively, which would mean they are the least accurate.Breast-cancer and SEER data pretty much sit in the middle. In actuality, the graph efficiently reflects the model's varying performance levels across diverse datasets while portraying the possible dataset-related problems that influence the model's accuracy. Further research needs to be carried out on the poorer-performing datasets.

10. Limitations

Firstly . Related to performance across datasets.
Approach 1 works perfectly well providing 98.2% accuracy with 99.35% of ROC AUC on the Wisconsin dataset while it failed to give a good performance on the BRCA dataset with just 62.4% on accuracy versus 53.1% ROC AUC that indicates that the model is dependent on the dataset. Approach 2 performed very well on German BS Data, recording an accuracy of 67.5% with 74.1% ROC AUC but did not perform well on BRCA itself.

Secondly. Imbalanced data issues.Nonetheless, the model proved to induce a nearly 50% ROC AUC for the BRCA, thus implying that SMOTE-offered synthetic sampling failed to optimize learning for the minority class distribution.
Such a first method provided slightly higher recall scores than the other method that possessed higher precision depicting different operational aspects of bias against variability

11. Future Work

- Next, Dataset-Specific Changes.

Replace SMOTE with something like ADASYN or the use of weighted loss functions for BRCA as AR-AUC was low even after oversampling, at 53.1%.

Imply the feature selection techniques such as the SHAP values on the diagnosis factor from SEER data in order to minimize noise of non-promising variables.

- Then, Hyperparameter tuning.

The Random Forest and SVM in approach 1 could be improved by grid search or Bayesian optimization for generalization.

And in approch 2, a neural network might benefit from dropout inputs and batch normalization in order to address overfitting yet again, especially when dealing with datasets coming from Wisconsin and SEER.

12. Conclusion.

Classic ML performed exceedingly well on structured datasets but was not as brittle on complex datasets like BRCA and SEER. In contrast, Method 2 (Neural Networks), well generalized, was nevertheless deficient by overfitting and imbalance problems during evaluation on BRCA and SEER.

Future improvements must include dataset-specific tuning to switch SMOTE with more advanced oversampling techniques and to improve the use of transfer learning in boosting deep learning models further. Hyperparameter tuning and the improvement of ensemble models, like stacking XGBoost with a neural net, could further cement performance on heterogeneous datasets.

13. Github Repository

Breast- Cancer - Prediction: https://github.com/SanfoCodes/Breast-Cancer-Prediction.git

14. References

Amrane, M., Oukid, S., Gagaoua, I., & Ensari, T. (2018). Breast cancer classification using machine learning. *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, 78–82. https://doi.org/10.1109/CTEMS.2018.8769187.

Osareh, A., & Shadgar, B. (2010). Machine learning techniques to diagnose breast cancer. *2010 5th International Symposium on Health Informatics and Bioinformatics*, 114–120. https://doi.org/10.1109/HIBIT.2010.5478899

Sharma, S., Aggarwal, A., & Choudhury, T. (2018). Breast cancer detection using machine learning algorithms. *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, 114–118. https://doi.org/10.1109/CTEMS.2018.8769185

Yassin, N. I. R., Omran, S., El Houby, E. M. F., & Allam, H. (2018). Machine learning techniques for breast cancer computer-aided diagnosis using different image modalities: A systematic review. *Computer Methods and Programs in Biomedicine, 156*, 25–45. https://doi.org/10.1016/j.cmpb.2017.12.012

Data Sources

Kaggle. (n.d.). Data for analysis and machine learning competitions. Retrieved from https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data?resource=download

Gitignore Creation

Toptal Developers. (n.d.). Gitignore templates. Retrieved from https://www.toptal.com/developers/gitignore

ChatGPT Contributions

OpenAI. (n.d.). ChatGPT contributions: *format_results function docstring and warning removal*.

Scikit-learn Documentation

Scikit-learn developers. (n.d.). *Stacking Classifier. Scikit-learn 1.5 documentation.* Retrieved from https://scikitlearn.org/1.5/modules/generated/sklearn.ensemble.StackingClassifier.html#sklearn.ensemble.StackingClassifier

Scikit-learn. (n.d.-a). *StandardScaler*. In *scikit-learn: Machine learning in Python*. Retrieved from https://scikit-learn.org/1.5/modules/generated/sklearn.preprocessing.StandardScaler.html

Scikit-learn. (n.d.-b). *OneHotEncoder*. In *scikit-learn: Machine learning in Python*. Retrieved from https://scikit-learn.org/1.5/modules/generated/sklearn.preprocessing.OneHotEncoder.html

Scikit-learn. (n.d.-c). *SimpleImputer*. In *scikit-learn: Machine learning in Python*. Retrieved from https://scikit-learn.org/1.5/modules/generated/sklearn.impute.SimpleImputer.html

Scikit-learn. (n.d.-d). *Pipeline*. In *scikit-learn: Machine learning in Python*. Retrieved from https://scikit-learn.org/1.5/modules/generated/sklearn.pipeline.Pipeline.html#sklearn.pipeline.Pipeline

Scikit-learn. (n.d.-e). *Column transformer with mixed types*. In *scikit-learn: Machine learning in Python*. Retrieved from https://scikitlearn.org/1.5/auto_examples/compose/plot_column_transformer_mixed_types.html