Jewel Journal of Scientific Research (JJSR) 9(1&2): 160–172, 2024
©Federal University of Kashere-Nigeria (Print ISSN: 2384 – 6267, Online ISSN 2756-651X)
https://journals.fukashere.edu.ng/index.php/jjsr

A Neural Network-Based Prediction for Time-to-Metastasis in Breast Cancer of Later Occurrence

***Muhammed, B. J.[1];Yusuf, M. M. [2]; Jeremiah, I.[1] and Mohammed, A.[3]**

[1]Department of Computer Science, Federal University of Kashere, Nigeria
[2]Department of Computer Science, Modibbo Adama University, Yola, Nigeria
[3]Department of Computer Science, Ahmadu Bello University, Zaria, Nigeria
***Corresponding Author**: bashmjibrin@gmail.com; +2348069604862

Abstract

Metastatic breast cancer (MBC) poses a significant global health concern, with increasing incidence rates, notably in regions like Nigeria. Despite advancements in diagnosis and prognosis, challenges persist in accurately predicting MBC, particularly when dealing with imbalanced datasets. This study introduces a neural network-based model for MBC prediction, integrating feature importance measures (FIM) such as the chi-square filter, Jaya algorithm wrapper, and gini-index random forest embedded, alongside data imbalance handling techniques including oversampling (ROS), under sampling (RUS), and synthetic minority oversampling technique (SMOTE). Initially, the three FIM methods were used on the original unbalanced dataset to independently select the top 10 features from a pool of 24 features in a 5-year MBC dataset. The selected features from each FIM method were then passed to the neural network classifier. Among these methods, the chi-square consistently demonstrated superior performance in accuracy, F1-score, and sensitivity metrics. Subsequently, RUS, ROS, and SMOTE were applied to balance the selected dataset subsets, including all features. Extensive experiments revealed that utilizing all 24 features with SMOTE consistently yielded superior performance across all metrics with significant margins, highlighting the importance of comprehensive FIM strategies and holistic data imbalance handling methods for enhancing BCM prediction.

Keywords: Neural network, breast cancer metastasis, imbalanced data, Jaya algorithm, feature importance measure.

Received: 10th Oct, 2024 *Accepted:* 16th Dec, 2024 *Published Online:* 30th Dec, 2024

Introduction

Cancer is a formidable medical condition characterized by the uncontrolled proliferation of cells within the body (Botlagunta *et al.*, 2023). Jiang and Xu (2022) emphasize the profound impact of cancer disease, with lung, skin, and breast cancers (BC) emerging as the most prevalent among various types. Their findings reveal BC as the foremost concern, surpassing all others with over 2.3 million new cases reported globally in 2020 alone. Moreover, BC retains its serious distinction as the leading cause of cancer-related deaths among women, claiming over 685,000 lives in the same year (Jiang and Xu, 2022). In Africa, specifically in Nigeria, the rising number of BC cases has become a significant concern among breast oncologists (Azubuike *et al.*, 2022). An estimate of over 26,310 cases were reported in the year 2018, with projections indicating an annual increase of approximately 4,000 cases over the next 10 years (Azubuike *et al.*, 2022).

Despite its lethal nature, BC confined solely to the breast holds a higher likelihood of successful treatment compared to when it has metastasized (Jiang and Xu, 2022). Metastatic breast cancer (MBC) occurs when cancer spreads to vital organs or tissues, serving as a primary cause of breast cancer-related fatalities (Nassar, 2023). While not all diagnosed cases of BC metastasize, there is always a risk of future development (Nassar, 2023). Therefore, it is crucial not only to diagnose current cases of BC but also to predict the probability of future recurrences (Botlagunta *et al.*, 2023; Nassar, 2023; Walsh and Tardy, 2023).

Traditional methods of diagnosing BC typically involve a combination of mammography, ultrasound, and biopsy (Olayide *et al.*, 2023). Ultrasound is often used in conjunction with mammography to further evaluate suspicious findings, providing detailed images of the breast tissue (Abdollahi *et al.*, 2022; Marti *et al.*, 2022). If abnormalities are detected, an image-guided core needle biopsy is performed to obtain a tissue sample for microscopic examination, confirming the presence of cancer cells [8]. Nonetheless, Jiang and Xu, (2022) reported that the use of an image-guided core needle biopsy can return non-definitive results in 5 - 15% of patients, making it difficult for breast oncologists to decide on an effective treatment decision.

However, modern technology has introduced more reliable and automated methods for diagnosing BC and its chances of metastasis, notably through machine learning (ML) and deep learning (DL) techniques (Tran *et al.*, 2022). ML and DL algorithms such as support vector machines (SVM), neural networks (NN), and random forest (RF), among others, utilize extensive datasets containing either image or text data to build predictive models (Isuwa *et al.*, 2023; Xu, Coen-Pirani and Jiang, 2023). While the integration of ML and DL techniques has significantly advanced the diagnosis and prognosis of BCM, the effectiveness of these models heavily relies on the quality of the datasets used for training. Often, these datasets may contain redundant or irrelevant variables or features, leading to challenges such as increased computational complexity, reduced model interpretability, and ultimately, lower accuracy (Isuwa *et al.*, 2022).

To address these issues, it is essential to utilize only the most significant features that contribute to BCM prediction, a process known as feature importance measure (FIM). Moreover, the abundance of features in these datasets can worsen data imbalance problems, a common occurrence in medical datasets due to the rareness of certain health conditions and ethical constraints among others. This imbalance in datasets leads to challenges such as an increased risk of model over fitting (Nassar, 2023). Consequently, it becomes difficult to generalize (Nassar, 2023). FIM methods encompass various approaches. Filter methods are one category, such as Chi-Square for individual feature assessment (Magboo and Magboo, 2021). Another category is the wrapper method, which evaluates feature subsets using learning algorithms. Examples are Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995), and the Jaya Algorithm (JA) (Gunduz and Aslan, 2021). Additionally, there is the embedded method, which assesses feature relevance directly within the model training process. An example is the Gini Index Random Forest (GI-RF) (Algehyne *et al.*, 2022).

As a result, researchers have extensively explored effective strategies to tackle the challenges of BCM prediction such as the utilization of various ML and DL algorithms, along with employing the filter, wrapper, embedded, or a combination for FIM (Marti *et al.*, 2022; Nassar, 2023). Additionally, techniques such as ROS, RUS, and SMOTE have been investigated for handling imbalanced data (Magboo and Magboo, 2021; Nassar, 2023). Despite the abundance of studies in the literature focusing on BCM prediction using diverse FIM and addressing data imbalance, there is a noticeable absence of detailed comparative analyses among different classes of FIM. Similarly, in the case of data imbalance handling, studies that attempt to leverage all three methods for BCM prediction are lacking. The lack of comprehensive comparative assessments of

FIM methods and holistic utilization of data imbalance handling methods highlight the ongoing challenge of achieving higher accuracy levels in BCM prediction.

Hence, the focus of this study lies in developing a neural network-based model for predicting the later occurrence of BCM in an imbalanced data setting. We employed one example of each of the three fundamental methods of FIM: the Chi-square filter method, the JA wrapper method, the GI-RF embedded method, and also all features for feature selection from the 5-year BCM dataset. Subsequently, we applied the three data imbalance handling techniques to balance the subset of the dataset selected by each of the utilized FIM methods. Specifically, we undertake the following actions:

- Conduct a comprehensive and up-to-date review of ML-based approaches for BCM prediction, encompassing both FIM and strategies for handling data imbalance.
- Develop multiple methodologies by combining the Chi-square, JA, GI-RF, and all features with the ROS, RUS, and SMOTE imbalance data handling techniques.
- Finally, an extensive experimental investigation, comparing the developed methodologies against each order, utilizing the 5-year BCM datasets.

This paper is structured as follows: Section 2 offers a background and literature review. Section 3 details the proposed approach. Section 4 covers experiments, comparisons, and result analysis. Section 5 concludes with insights from experiments and future research directions.

Background

This section highlights key terms essential for understanding the subject. It provides a detailed explanation of feature importance measures, including their various types: filter, wrapper, and embedded methods. Additionally, it explores the concept of imbalanced datasets and neural networks, concluding with a review of current literature related to the subject.

Feature Importance Measure (FIM)

Filter Methods of FIM. Filter methods utilize statistical techniques to assess individual features in a dataset and rank them based on their scores. Subsequently, a cutoff point is established, below which lower-performing features are discarded, while those above the cutoff are retained. Filter methods offer advantages of computational efficiency, however, they may overlook interactions between features and can be less effective when dealing with complex datasets or non-linear relationships (Jeremiah *et al.*, 2022).

Chi-Square. The chi-square test, described byAlrefai and Ibrahim, (2022)measures the interdependence between two categorical variables. It is widely favored owing to its notable strengths such as its ease of interpretability, nonparametric nature, and resilience to outliers (Alrefai and Ibrahim, 2022). Ghosh et al., (2020) introduced a mathematical formulation of the chi-square test, dividing the expected range into intervals. The calculation of the chi-square (X^2) value for feature f is as shown in Equation 1:

$$X^2 f = \sum_{j=1}^{r} \sum_{s=1}^{c} \frac{(n_{js} - \mu_{js})^2}{\mu_{js}} \quad (1)$$

Here, r indicates the count of distinct values in the feature, c signifies the number of distinct values within a class, n_{js} represents the frequency of the j^{th} element within the s^{th} class, and μ_{js} is calculated as $\frac{n_{*s} n_{j*}}{n}$, where n_{j*} denotes the frequency of the j^{th} element and n_{*s} denotes the total number of elements within the s^{th} class.

Wrapper Methods of FIM. This approach assesses a set of features simultaneously. It employs a learning algorithm to assess the quality of selected features. While offering superior classification performance it can be computationally intensive due to the multiple evaluations required (Isuwa, Abdullahi and Abdulrahim, 2022). Metaheuristic Algorithms (MA) such as JA are widely adopted for this purpose.

Jaya Algorithm (JA). The JA, proposed by Venkata Rao, (2016) is a population-based optimization algorithm inspired by the concept of continuous improvement and optimization. Originally designed to tackle both constrained and unconstrained mathematical optimization problems, JA has also been adapted for discrete optimization tasks, such as feature selection (Noshad and Fallahi, 2023). While other MAs such as PSO require tuning multiple parameters for optimal performance, JA is characterized by its simplicity with only the population size and number of iterations as adjustable parameters. During each iteration, JA identifies both the best and worst solutions, ensuring that the algorithm progresses toward the optimal solution while avoiding poorer outcomes (Chaudhuri and Sahu, 2021). The position update of each solution in JA is governed by Equation 2, as described by Chaudhuri and Sahu, (2021):

$$X_{k,j}^{new} = X_{k,j}^{current} + r_1 * \left(Best_j - \left|X_{k,j}^{current}\right|\right) - r_2 * \left(Worst_j - X_{k,j}^{current}\big|\right) \quad (2)$$

where k represents the number of solutions ranging from $1\ to\ N$, and D represents the dimension of the problem ranging from $1\ to\ D$. $X_{k,j}^{current}$ denotes the value of the k^{th} solution in the j^{th} dimension. $Best_j$ and $Worst_j$ denote the best and worst solutions in terms of fitness for the j^{th} dimension, respectively. r_1 and r_2 are random numbers in the range [0,1] introduced to inject randomness into the search process. Finally, $X_{k,j}^{new}$represents the updated solution.

Embedded Methods of FIM. Embedded methods of FIM evaluate feature relevance directly within the learning algorithm's training process (Algehyne *et al.*, 2022). This approach allows the model to identify the most important features while optimizing its performance on the task at hand (Algehyne *et al.*, 2022). A Common example of an embedded method is the Gini Index Random Forest (Dritsas and Trigka, 2022).

Gini-Index Random Forest (GI-RF). GI-RF relies on the Gini Impurity (GI) measure, commonly used in decision trees to assess split purity (Algehyne *et al.*, 2022). GI measures the level of misclassification at a decision tree node, where lower values suggest higher purity, indicating nodes primarily comprise samples from a single class (Algehyne *et al.*, 2022). In RF, GI is utilized to assess the importance of features across individual trees within the forest. Feature importance is determined by examining the decrease in GI when a particular feature is used for data split at each node (Algehyne *et al.*, 2022). The impurity importance measure is defined by the Gini Index algorithm, which is consistent with Gini theory. According to this theory, a set of samples S comprising k classes $(C_i, i = 1,2,3,\ldots,k)$ can be partitioned into k subsets based on class distinctions. Here, S_i represents a subset containing samples belonging to the class C_i, and s_i denotes the subset's sample count. The Gini Index of S can be computed using Equation 3 (Algehyne *et al.*, 2022).

$$Gini(S) = 1 - \sum_{i=1}^{m} P_i^2 \quad (3)$$

where P_i is the estimated probability as $\frac{s_i}{s}$ for all samples belonging to C_i.

Imbalance Data

Imbalanced datasets are characterized by highly skewed class distributions where certain classes significantly outnumber others (Magboo and Magboo, 2021). ML algorithms often struggle to accurately predict minority classes of such datasets due to a bias towards the majority class (Aruna and Nandakishore, 2022). Strategies for handling unbalanced data according to Aruna and Nandakishore, (2022) can be broadly categorized into algorithm-driven and data-driven approaches. Techniques such as SMOTE, RUS, and ROS are the commonly employed data-driven approaches to tackling data imbalance.

ROS Technique. According to Viloria, Lezama and Mercado-Caruzo, (2020), this technique involves randomly increasing the occurrences in the minority class to conform to the majority class. ROS is an easy technique to use, however, it might lead to overfitting and missing opportunities to introduce new information.

RUS Technique. Undersampling, on the other hand, randomly eliminates the instances in the majority class until a balance is reached (Viloria, Lezama and Mercado-Caruzo, 2020). RUS may result in the loss of important information present in the majority class, which could lead to underfitting and decreased model performance, even if it can reduce computational complexity (Aruna and Nandakishore, 2022).

SMOTE. SMOTE is a popular oversampling approach designed to address the shortcomings of simpler oversampling algorithms like ROS (Aruna and Nandakishore, 2022). SMOTE generates synthetic instances by interpolating between minority class samples that already exist, as opposed to randomly replicating instances from the minority class (Aruna and Nandakishore, 2022).

Neural Network (NN) Learning Algorithm

NN are computational models inspired by the human brain's structure and function. They consist of interconnected layers of artificial neurons that process input data to produce output (Xu, Coen-Pirani and Jiang, 2023). Through training, NNs learn to recognize patterns and relationships in data, making them powerful tools for tasks such as classification (Xu, Coen-Pirani and Jiang, 2023). NN is composed of three layers as shown in Figure 1 where each layer consists of weighted interconnected nodes (Jiang and Xu, 2022). During training, the network adjusts these weights based on the input data and the desired output, to minimize a loss function (Jiang and Xu, 2022).

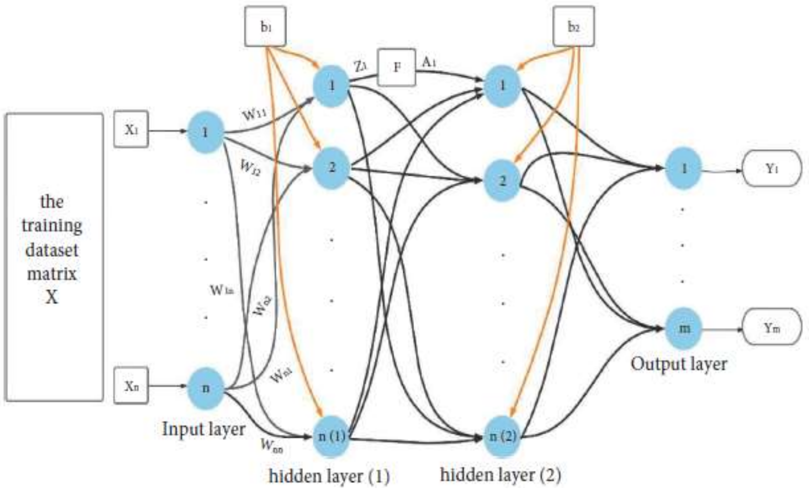

Figure. 1. Conventional structure of a NN with two hidden layers (Beghriche *et al.*, 2021)

Related Works

Over the years, many studies have explored the application of ML and DL techniques for predicting the 5, 10, and 15-year BCM. Some studies, however, have employed the Wisconsin BC dataset sourced from the University of California in Irvine (UCI) repository. Moreover, these studies have also incorporated FIM and techniques for handling data imbalance. For instance, Botlagunta *et al*. (2023) utilized nine ML algorithms to predict BCM. The Welch Unpaired t-test was employed to determine the statistical significance of the datasets. Results from the experiment indicate that the decision tree classifier outperforms the ensemble and other ML algorithms. However, the absence of a clear description of the dataset utilized by the authors creates uncertainty about its balance. This is essential as it can notably affect model performance, particularly in reducing overfitting. Without this vital information, assessing the generalizability and reliability of their findings becomes difficult. Walsh and Tardy (2023) employ the ResNet-22 architecture to classify four distinct full-field imbalanced digital image mammography datasets. They applied class weighting, ROS, RUS, and a synthetic lesion generation approach to augment the number of malignant samples. Findings from the experiment revealed the effectiveness of synthetic malignant lesions in balancing classes, mitigating bias towards the majority class, and enhancing AUC-ROC performance. However, given the extensive number of experiments and the lengthy training duration (1–8 days) for each run, conducting thorough hyperparameter tuning was not feasible. Nassar (2023) introduced eight ML algorithms for analyzing an imbalanced BCM dataset from Khartoum State Hospital. Preprocessing techniques, including resampling and feature selection, were applied to the data. Results indicate that the ensemble methods, Boosting and J48, yielded the most favorable outcomes. However, ROS was utilized to tackle data imbalance. Nonetheless, this approach proves ineffective as it fails to introduce new information and may result in heightened computational time. Marti et al., (2022) employed the Markov Blanket and Interactive Risk Factor Learner to pinpoint key factors contributing to the late recurrence of BCM at 5, 10, and 15 years. Their experiments revealed that factors such as ER status, smoking, race, and alcohol usage exerted the most significant influence on recurrence. Jiang and Xu (2022) compare the performance of a Deep Neural Network (DNN) with nine other ML algorithms in predicting BCM at 5, 10, or 15-year intervals. Results show DNN ranking 6th, 4th, and 3rd for 5, 10, and 15-year predictions respectively, with XGBoost, RF, and KNN performing best for 5-year BCM, XGBoost, RF, and Naïve Bayes for 10-year BCM, and SVM, LR, and LASSO for 15-year BCM. Nonetheless, the study overlooked the imbalanced datasets and neglected to perform feature selection, potentially compromising the robustness and generalizability of the findings.

Methodology

This section provides a comprehensive overview of the methodology employed in this study. It covers the dataset description, the design of the proposed method, the architectural framework, and the parameter values and settings used.

Dataset Description

The study utilized a 5-year BCM dataset, referenced in Xu, Coen-Pirani and Jiang, (2023) and Marti *et al.*, (2022). Comprising 6726 samples from both malignant and benign cases. The dataset encompasses 26 clinical features as described in the work of Xu et al., (2023) and a binary target class denoting whether a patient experienced metastasis within 5 years of initial treatment. Notably, 92% of the samples originate from benign cases, while only 8% pertain to malignant cases. This striking class imbalance highlights the necessity for careful management of class imbalance during model training.

Proposed Method Design

Figure 2 illustrates the architecture of the proposed method. The design comprises two main segments: predictions utilizing the original imbalanced dataset and predictions employing the balanced dataset. Specifically, the design encompasses the following essential elements:

Data Cleaning. The utilized dataset includes three obvious irrelevant columns: patient ID and two unnamed columns with over 70% missing values. Therefore, these columns are removed during the initial data-cleaning phase. Moreover, the dataset primarily comprises categorical data, with all values being non-numeric. Consequently, the LabelEncoder is utilized to convert categorical variables into a numerical format. Additional data preprocessing steps involve normalization using the MinMaxScaler to standardize the features within a defined range, typically between 0 and 1.

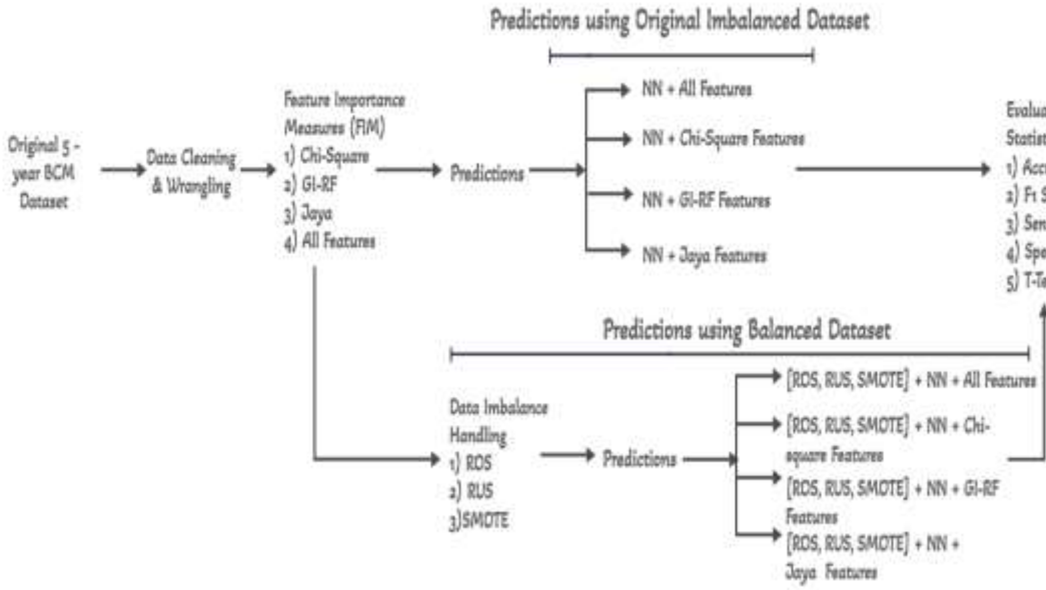

Figure. 2. Architecture of the proposed method

Feature Importance Measure (FIM). Three FIM methods were employed i.e., the Chi-square filter, JA Wrapper, and GI-RF embedded methods. Furthermore, all features were included in the experiment primarily to reveal the importance of FIM in ML training.

Chi-Square. The Chi-Square method was used to identify 10 top-performing features. The decision to select only 10 features out of the available 24 was based solely on intuition. Figure 3 depicts the feature ranking using Chi-Square, focusing on the top 10 features with the highest scores.

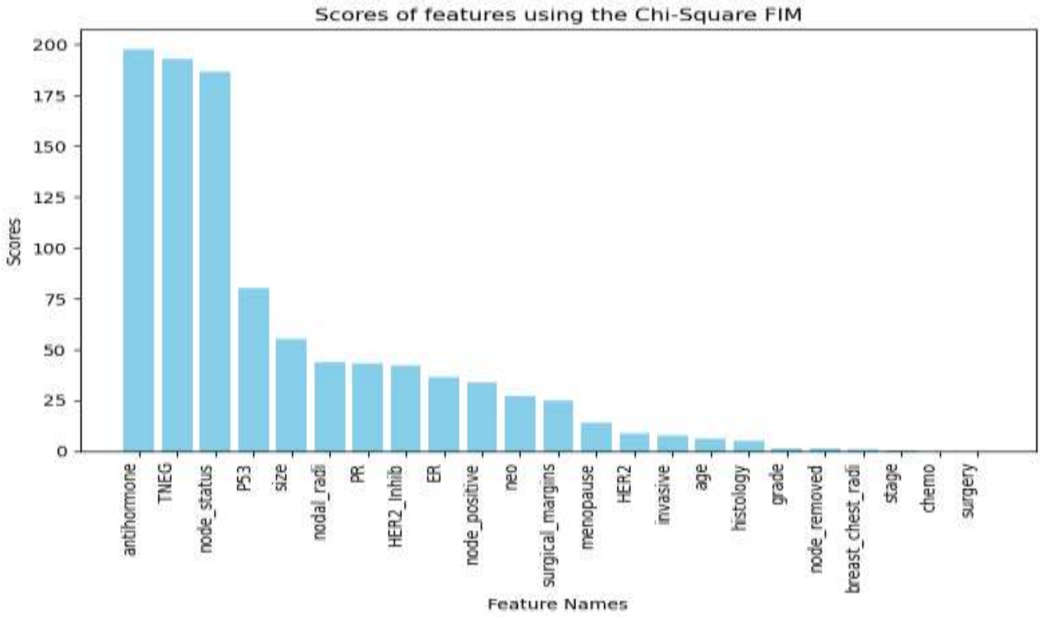

Figure. 3. Scores of features using Chi-Square

GI-RF. We similarly chose only the top 10 features using GI-RF. Figure 4 showcases the feature ranking achieved through GI-RF.

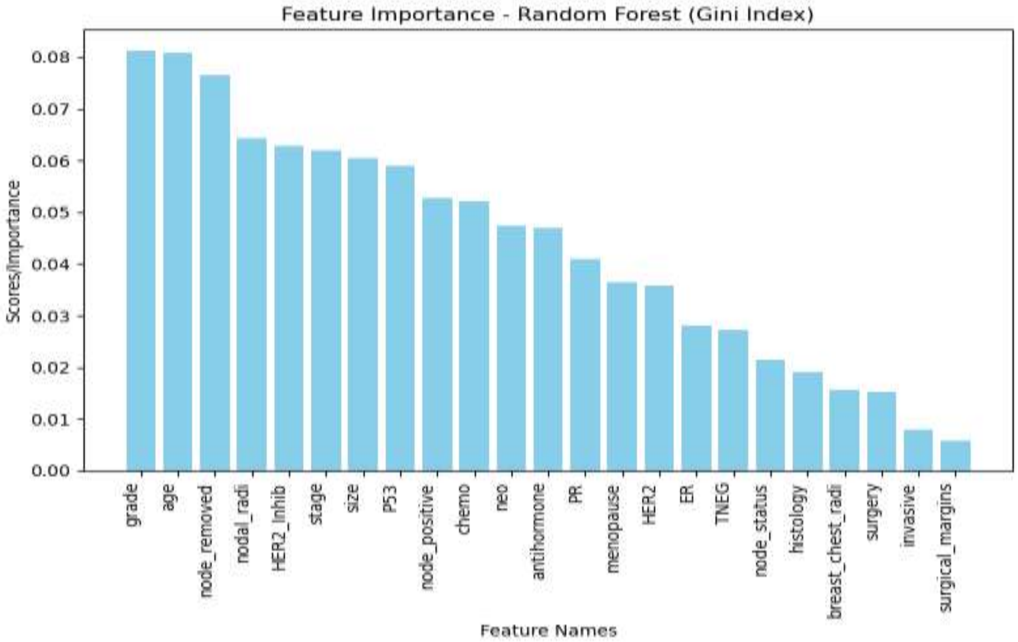

Figure. 4. Scores of features using GI-RF

Jaya Algorithm. JA is a stochastic algorithm, thus feature selection may vary between iterations. However, over 10 runs, an average of 8 features were selected, with the 10 most frequent being 'age', 'menopause', 'P53', 'stage', 'node_positive', 'size', 'surgical_margins', 'surgery', 'breast_chest_radi', and 'nodal_radi'.

Predictions Utilizing the Original Imbalanced Dataset. In the initial phase of the experiment, we employed each of the three subsets of selected features, along with using all features, to train the NN model on the original imbalanced dataset. Hence, at this stage, we built 4 distinct methods, as depicted in Figure 2.

Predictions Employing the Balanced Dataset. During this stage, we applied the three discussed techniques for handling imbalanced data to balance the datasets. Each dataset comprised features selected by one of the three FIM methods, along with a dataset containing all features. Consequently, we developed 12 unique methods at this stage, as depicted in Figure 2.

Evaluation and Statistical Tests. All 16 methods, built from the two stages, are evaluated based on accuracy (Acc), sensitivity (Sens), specificity (Specs), F1-score (F1), and their statistical significance using the T-Test. Method stability is assessed through Standard Deviation (SD).

Parameter Values and Settings

The parameters utilized in the study are outlined in Table 1. Settings in JA such as population initialization and fitness function as well as other non-JA configurations were intuitively selected and adhered to the standard practices in the literature.

Table 1. Parameter values and settings utilized in the proposed method

Jaya Algorithm	k value for KNN = 5, population size = 20, number of iterations = 20
Neural Network	Number of neurons (input layer) = number of dataset features, activation function (2 hidden layers) = Relu, activation function (output layer) = sigmoid, number of neurons (first hidden layer) = 64, number of neurons (second hidden layer) = 32, number of neurons (output layer) = 1, optimizer = adam, loss = binary_crossentropy, epoch = 10, batch_size = 32, validation_split = 0.1, dropout (2 layers) = 0.2
Other Settings	train_test_spilt = 70:30, random_state = 42, number of algorithm run = 10

Experimental Results And Discussions

In this section, we present the outcomes of the experiments alongside their corresponding discussions. All experimental procedures were conducted using Jupyter Notebook with Python version 3.10, executed on a computer equipped with an Intel(R) Core (TM) i7-6600U processor running at 2.60GHz, and possessing 8GB of RAM. The SD of each result is shown alongside it using the notation $x \pm y$. Where x represents the average result (in %), and y represents the SD value of the associated result, all across ten independent runs. Boldface results represent the best from each category of metric i.e., column-wise.

Results Discussion: Original Imbalanced Dataset Predictions

Table 2 displays the outcomes of the experiment conducted with the original imbalanced dataset. Notably, the chi-square filter method outperformed others in terms of Acc, F1, and Sens. The superiority of chi-square over its peer could be due to its independence assumption between features, which can be advantageous in scenarios where this assumption holds. Furthermore, it is robust to noise since it evaluates the statistical significance of the relationship between each feature and the target variable independently.

Table 2. Results from experiments using the original imbalanced BCM dataset

FIM	Accuracy (Acc)	F1-Score (F1)	Sensitivity (Sens)	Specificity (Spec)
Chi-Square	**0.9139±0.0009**	**0.0588±0.0413**	**0.0317±0.0232**	0.9977±0.0016
RF-Gini Index	0.9130±0.0006	0.0012±0.0037	0.0006±0.0019	**0.9996±0.0007**
Jaya Algorithm	0.9132±0.0006	0.0073±0.0153	0.0038±0.0081	**0.9996±0.0013**
All Features	0.9137±0.0009	0.0503±0.0289	0.0267±0.0157	0.9979±0.0009

However, the results also highlight the detrimental impact of the dataset's imbalanced nature. While Acc and Spec exhibit commendable performance, both F1 and Sens display severely inferior results. This phenomenon is attributed to the inherent challenges posed by the imbalanced state of the dataset, where the dominance of the majority class leads models to favor predictions that align with this majority. Consequently, Acc and Spec, which measure overall correctness and true negative rates respectively, tend to be inflated due to the prevalence of the majority class. Conversely, the F1, representing the harmonic mean of precision and recall, and Sens, suffer as they are sensitive to the correct identification of minority class instances.

Results Discussion: Balanced Dataset Predictions

Table 3 displays the results obtained from the experiment conducted on balanced datasets. It is evident that regardless of the data-imbalanced handling method employed, utilizing all features consistently yields superior performance across nearly all metrics. This could be due to several factors. It could be that the dataset contains no noisy features, and each feature contributes uniquely to the classification task. Consequently, the three FIM techniques may have inadvertently discarded informative features, leading to a degradation in performance. By using all features, the model has access to the complete set of information available, potentially improving its ability to discern patterns and make accurate predictions. Another factor may be attributed to the fact that the top 10 features selected from chi-square and GI-RF may not be enough to capture the full complexity and information present in the dataset.

Table 3. Results from exp s using balanced BCM datasets

	FIM methods	Accuracy (Acc)	F1-Score (F1)	Sensitivity (Sens)	Specificity (Spec)
Up-Sampling	Chi-Square	0.7467±0.0023	0.7499±0.0059	0.7661±0.0258	0.7275±0.0278
	RF-Gini Index	0.7499±0.0050	0.7557±0.0105	0.7807±0.0328	0.7196±0.0282
	Jaya Algorithm	0.6928±0.0461	0.7058±0.0310	0.7418±0.0636	0.6445±0.1234
	All Features	**0.7894±0.0058**	**0.7949±0.0112**	**0.8241±0.0352**	**0.7553±0.0258**
Down-Sampling	Chi-Square	0.7045±0.0059	0.6989±0.0085	0.7270±0.0192	0.6844±0.0151
	RF-Gini Index	0.6951±0.0063	0.6926±0.0212	**0.7314±0.0664**	0.6627±0.0583
	Jaya Algorithm	0.6656±0.0311	0.6643±0.0334	0.7057±0.0869	0.6299±0.1030
	All Features	**0.7135±0.0080**	**0.7055±0.0064**	0.7270±0.0168	**0.7015±0.0242**
SMOTE	Chi-Square	0.7363±0.0031	0.7268±0.0078	0.7004±0.0250	0.7723±0.0246
	RF-Gini Index	0.7732±0.0029	0.7785±0.0072	0.7961±0.0343	0.7503±0.0361
	Jaya Algorithm	0.6831±0.0439	0.6968±0.0309	0.7270±0.0707	0.6390±0.1270
	All Features	**0.8013±0.0039**	**0.8049±0.0070**	**0.8184±0.0275**	**0.7841±0.0261**

These FIM techniques prioritize features based on their relevance to the target variable, but they may overlook interactions between features that are crucial for accurate classification, leading to suboptimal performance compared to using all available features. Thus, in future research, we aim to investigate the selection of a larger set of top-performing features, exceeding the limit of 10, for subsequent experiments. Moreover, the outcomes presented in Table 3 highlight the significant influence of tackling the imbalanced nature of a dataset. Although there is a notable enhancement in the F1 and Sens metrics following the application of balancing techniques, there is generally a reduction in performance in Acc and Spec which we aim to investigate in future research.

Also from Table 3, it is seen that using the SMOTE technique in handling the data imbalance problem results in superior performance in Acc, F1, and Spec. The observed superiority of the SMOTE can be attributed to its effectiveness in generating synthetic samples for the minority class. SMOTE addresses the class imbalance by creating synthetic instances of the minority class, thereby increasing its representation in the dataset. This approach helps to alleviate the bias towards the majority class and enables the classifier to learn more effectively from the minority class instances. Overall, utilizing all features from the dataset along with the SMOTE technique to address the imbalanced dataset consistently yields the highest performance.

Moreover, we conducted a t-test to evaluate potential statistically significant differences among the 12 methods developed in the second phase of our experiment, focusing solely on Acc. Table 4 displays the comparison results, indicating significant differences if the p-value is equal to or less than 0.05. Notably, all methods demonstrate statistically significant differences except for the JA algorithm (highlighted in bold).

Table 4. T-test result on accuracy (Up-S = oversampling, D-S = undersampling)

Chi-Square	**Up-S Vs. D-S**	**Up-S Vs SMOTE**	**D-S Vs SMOTE**
	1.15561E-09	8.23011E-07	5.63021E-09
RF-Gini Index	**Up-S Vs. D-S**	**Up-S Vs SMOTE**	**D-S Vs SMOTE**
	1.59501E-12	2.38993E-08	1.03074E-12
Jaya Algorithm	**Up-S Vs. D-S**	**Up-S Vs SMOTE**	**D-S Vs SMOTE**
	0.164917263	**0.654064797**	**0.346468898**
All Features	**Up-S Vs. D-S**	**Up-S Vs SMOTE**	**D-S Vs SMOTE**
	6.18771E-13	0.000164038	2.54442E-12

This consistent lack of significance may stem from the algorithm's instability, resulting in varying performance across the 10 iterations. Such instability could arise from factors like parameter sensitivity, convergence issues, or inadequate adaptation to dataset characteristics. To address these concerns in future research, we plan to conduct more rigorous parameter-tuning efforts.

Conclusıon And Future Works

This study focuses on developing a neural network-based model for predicting later occurrences of BCM in an imbalanced data setting. Initially, three feature importance measures including Chi-Square, Jaya Algorithm, Gini-Index Random Forest, and using all features were employed to independently select the top 10 features from a total of 24 in the 5-year BCM datasets. While Chi-Square consistently outperformed others in accuracy, F1-score, and sensitivity metrics, the results indicate poor performance of the methods in terms of F1-Score and sensitivity metrics, attributed to the effect of data imbalance. Subsequently, RUS, ROS, and SMOTE imbalanced data handling techniques were applied to balance the selected dataset subsets. Extensive experiments and comparisons were conducted, revealing that utilizing all features along with SMOTE consistently yielded superior performance across all metrics including accuracy, F1-Score, Sensitivity, and Specificity. Consequently, applying SMOTE may have enhanced the method's robustness to variability and noise in the data resulting in better performance. These findings further strengthen the significant impact of addressing dataset imbalance, as evidenced by notable improvements in F1-Score and Sensitivity metrics before and after employing this approach.

For future research, the impact of varying sizes of top-performing features selected from these methods could be explored. Additionally, in response to the instability observed in the Jaya algorithm as indicated by its P-values, primarily attributed to parameter sensitivity, we plan to conduct more thorough hyperparameter tuning as well as for the neural network. Furthermore, we aim to investigate factors such as the initialization and search strategies of the Jaya algorithm to potentially yield a more promising feature selection.

References

Abdollahi, S., Acero, F., Baldini, L. and Ballet, J. (2022) 'Detection of Metastatic Breast Cancer from Whole-Slide Pathology Images Using an Ensemble Deep-Learning Method', *Archives of Breast Cancer*, 9(3): 364–376. Available at: https://doi.org/10.32768/abc.202293364-376.

Algehyne E.A, Muhammad L.J, Naseh A. A, Osama A.A and Abdullah K. A (2022) 'Fuzzy Neural Network Expert System with an Improved Gini Index Random Forest-Based Feature Importance Measure Algorithm for Early Diagnosis of Breast Cancer in Saudi Arabia', *Big Data and Cognitive Computing*, 6:13. Available at: https://doi.org/10.3390/bdcc6010013.

Alrefai, N. and Ibrahim, O. (2022) 'Optimized feature selection method using particle swarm intelligence with ensemble learning for cancer classification based on microarray datasets', *Neural Computing and Applications* 34:13513-13528 Available at: https://doi.org/10.1007/S00521-022-07147-Y.

Aruna, S. and Nandakishore, L. V. (2022) 'Empirical Analysis of the Effect of Resampling on Supervised Learning Algorithms in Predicting the Types of Lung Cancer on Multiclass Imbalanced Microarray Gene Expression Data', *EAI/Springer Innovations in Communication and Computing*, pp. 15–27. Available at: https://doi.org/10.1007/978-3-030-86165-0_2/COVER.

Azubuike S. O., Hayes, L., Sharp, L., Alabi, A., Oyesegun R. and McNally, R. (2022) 'Socio-economic status and the risk of breast cancer among Nigerian women : a case control study' *Pan African Medical Journal*. 41: 175

Beghriche T, Mohamed D, Youcef B, Bilal A, and Samir B. B. (2021). An Efficient Prediction System for Diabetes Disease Based on Deep Neural Network', *Complexity*, 2021. Available at: https://doi.org/10.1155/2021/6053824.

Botlagunta, Madhu B.M, D.Lakshmi, Anand N (2023). Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms', *Scientific Reports*, 13(1): 1–17. Available at: https://doi.org/10.1038/s41598-023-27548-w.

Chaudhuri, A. and Sahu, T.P. (2021). A hybrid feature selection method based on Binary Jaya algorithm for micro-array data classification', *Computers and Electrical Engineering*, 90: 106963. Available at: https://doi.org/10.1016/j.compeleceng.2020.106963.

Dritsas, E. and Trigka, M. (2022). Data-Driven Machine-Learning Methods for Diabetes Risk Prediction', *Circulation*, 134(19): 1441–1443. Available at: https://doi.org/10.1161/circulationaha.116.024941.

Ghosh, K. K., Begum, S., Sardar, A., Adhikary, S., Ghosh, M., Kumar, M. and Sarkar, R. K. (2020). Theoretical and Empirical Analysis of Filter Ranking Methods : Experimental Study on Benchmark DNA Microarray Data', *Expert Systems With Applications*, 169: 114485. Available at: https://doi.org/10.1016/j.eswa.2020.114485.

Gunduz, M. and Aslan, M. (2021). DJAYA : A discrete Jaya algorithm for solving traveling salesman problem', *Applied Soft Computing*, 105: 107275. Available at: https://doi.org/10.1016/j.asoc.2021.107275.

Isuwa J. Mohammed A, Yusuf S. A, Ibrahim H.H and Olaide N. O (2023). Optimizing Microarray Cancer Gene Selection using Swarm Intelligence : Recent Developments and An Exploratory Study Optimizing Microarray Cancer Gene Selection using Swarm Intelligence : Recent', *Egyptian Informatics Journal*, 24(4): 100416.

Isuwa, J., Abdullahi, M. and Abdulrahim, A. (2022). Hybrid particle swarm optimization with sequential one point flipping algorithm for feature selection', Concurrency and Computation: Practice and Experience 34(25): 1–18. Available at: https://doi.org/10.1002/cpe.7239.

Jeremiah I, Mohammed A, S. A. Yusuf, M. N. Idris, and Baffa S.G (2022). Integration of Specific Local Search Methods in Metaheuristic Algorithms for Optimizing the Feature Selection Process : A Survey', *Sule Lamido University Journal of Science and Technology*. 4(1): 34–48.

Jiang, X. and Xu, C. (2022). Deep Learning and Machine Learning with Grid Search to Predict Later Occurrence of Breast Cancer Metastasis Using Clinical Data', *Journal of Clinical Medicine*, 11(19): 5772. Available at: https://doi.org/10.3390/jcm11195772.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization', in *Proceedings of ICNN'95 - International Conference on Neural Networks*. IEEE, pp. 1942–1948. Available at: https://doi.org/10.1109/ICNN.1995.488968.

Magboo, V.P.C. and Magboo, M.S. (2021). Machine learning classifiers on breast cancer recurrences', *Procedia Computer Science*, 192: 2742–2752. Available at: https://doi.org/10.1016/j.procs.2021.09.044.

Marti J L, Adam B, Alan W, and Xia J. (2022). Machine Learning to Discern Interactive Clusters of Risk Factors for Late Recurrence of Metastatic Breast Cancer', *Cancers*, 14(1): 253. Available at: https://doi.org/10.3390/cancers14010253.

Nassar, H.B. (2023). Classification for Imbalanced Breast Cancer Dataset Using Resampling Methods', International Journal of Computer Science and Network Security (IJCSNS). Available at: https://doi.org/10.22937/IJCSNS.2023.23.1.12.

Noshad, A. and Fallahi, S. (2023). A new hybrid framework based on deep neural networks and JAYA optimization algorithm for feature selection using SVM applied to classification of acute lymphoblastic Leukaemia', *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(4): 1549–1566. Available at: https://doi.org/10.1080/21681163.2022.2157748.

Olayide A, Aremu I, Rahman G and Olatoke S. (2023). Breast Cancer Treatment and Outcomes in Nigeria : A Systematic Review and Meta-analysis', 8(3): 591–598. Available at: https://doi.org/10.31557/APJCC.2023.8.3.591.

Tran, T., Le, U. and Shi, Y. (2022). An effective up-sampling approach for breast cancer prediction with imbalanced data: A machine learning model-based comparative analysis', *PLoS ONE*, 17(5): Available at: https://doi.org/10.1371/journal.pone.0269135.

Venkata Rao, R. (2016). Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems', *International Journal of Industrial Engineering Computations*, 7(1): 19–34. Available at: https://doi.org/10.5267/j.ijiec.2015.8.004.

Viloria, A., Lezama, O.B.P. and Mercado-Caruzo, N. (2020). Unbalanced data processing using oversampling: Machine Learning', *Procedia Computer Science*, 175: 108–113. Available at: https://doi.org/10.1016/J.PROCS.2020.07.018.

Walsh, R. and Tardy, M. (2023). A Comparison of Techniques for Class Imbalance in Deep Learning Classification of Breast Cancer', *Diagnostics*, 13(1): 1–19. Available at: https://doi.org/10.3390/diagnostics13010067

Xu, C., Coen-Pirani, P. and Jiang, X. (2023). Empirical Study of Over fitting in Deep Learning for Predicting Breast Cancer Metastasis', *Cancers*, 15(7) Available at: https://doi.org/10.3390/cancers15071969.