Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 8
August 2024

Exploring Advanced Optimization Techniques in Breast Cancer Prediction

[1]Kavya Venkata Sushma Gontla, [2] Jaya Sai Supriya Adapala, [3]Vyshnavi Koka, [4]Radha Mothukuri, [5]Sujitha Lakshmi Modugula, [6]Suneetha Bulla

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India

[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India

[3]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India

[4]Associate Professor,Department of Computer Scinece and Engineering,Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India

[5]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India

[6]Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India

Abstract: This paper highlights the pressing global need for early and accurate breast cancer prediction. It discusses the potential of optimization algorithms, specifically Genetic Algorithm (GA), Maxima and Minima (MM), and Simulated Annealing (SA), to improve breast cancer prediction. GA notably outperforms MM and SA in accuracy, precision, recall, and F1-score. This emphasizes the value of optimization algorithms in refining breast cancer prediction models. The paper stresses the importance of accurate prediction in reducing breast cancer prognosis and mortality rates. It acknowledges the limitations of conventional methods, including subjectivity and the risk of false negatives. Optimization algorithms, however, are praised for their ability to handle extensive datasets, discover subtle trends, and adapt to evolving data, positioning them as indispensable tools in improving breast cancer prediction. Furthermore, the paper suggests that optimization algorithms can integrate diverse data sources, such as genomics, imaging records, and medical histories, offering a comprehensive patient risk assessment. In conclusion, this research underscores the global importance of early and precise breast cancer prediction and advocates for continued exploration of optimization algorithms, process learning, and artificial intelligence to enhance prediction accuracy and effectiveness.

Introduction:

Breast cancer continued to remain a major global health problem that can have fatal effects for people of all sexes. For a patient to survive and receive successful treatment, a prompt and precise diagnosis is essential. The use of cutting-edge computational methods in healthcare, particularly in the diagnosis of breast cancer, has grown in popularity over time. In order to classify breast cancer using real-world datasets, this term paper examines the relationships of three potent optimization and search algorithms: Simulated Annealing, Maxima and Minima, and Genetic Algorithms.

Methodology:

Prior to employing the Simulated Annealing (SA) algorithm, an extensive data preprocessing stage is initiated, encompassing tasks such as addressing missing data, standardizing features, and ensuring proper data normalization. Similar to the SA methodology, data preprocessing is carried out to ensure data quality and uniformity. Similar to the other algorithms, the dataset is processed to guarantee data quality and consistency.

Results:

We examine the outcomes of our genetic algorithm-optimized breast cancer prediction model in this part. The evolutionary algorithm was used to adjust the classifiers' hyper parameters. Let's see how well the genetic algorithm performs and how hyper parameter optimization affects the accuracy of the model. Here, we present the results and discuss the implications of our method of simulated annealing feature selection. We are able to

select a subset of the most crucial features for our prediction model by employing the powerful optimisation technique known as simulated annealing. The highest observed values for each feature in your dataset are represented as maxima values. For instance, the greatest mean texture is 50.54, and the biggest mean radius is 28.11. For every characteristic, the lowest observed values are represented by the minima values.

Conclusions:

Using attributes related to mean radius and mean texture, we used three optimize techniques in this study to investigate and understand the properties of a breast cancer dataset: Genetic Algorithm (GA), Simulated Annealing (SA), and Maxima and Minima analysis. The best characteristics found by GA and SA match the extremes found through maxima and minima analysis, highlighting the importance of certain feature ranges in identifying benign and malignant cases.

Keywords: Breast Cancer, Classification, accuracy, Genetic algorithm, Maxima and Minima, Simulated Annealing, Scatterplot

1. Introduction

Breast cancer continued to remain a major global health problem that can have fatal effects for people of all sexes. For a patient to survive and receive successful treatment, a prompt and precise diagnosis is essential. The use of cutting-edge computational methods in healthcare, particularly in the diagnosis of breast cancer, has grown in popularity over time. In order to classify breast cancer using real-world datasets, this term paper examines the relationships of three potent optimization and search algorithms: Simulated Annealing, Maxima and Minima, and Genetic Algorithms. The most frequent cancer in women and the cause of the majority of cancer-related deaths worldwide is breast cancer. Early detection and accurate diagnosis are crucial for improving patient outcomes and lowering mortality rates; therefore, their significance cannot be emphasized. In several fields, including healthcare, optimization algorithms have shown to be invaluable tools for tackling complicated problems. In this research, we investigate the use of three such algorithms to improve the performance of machine learning models for breast cancer classification: Simulated Annealing, Maxima and Minima, and Genetic Algorithms. We choose an appropriate breast cancer dataset that offers a full set of features and is frequently used in research on the classification of breast cancer in order to start our exploration. Two prime examples are the Wisconsin Breast Cancer dataset (WBCD) and the Breast Cancer Wisconsin (Diagnostic) dataset (BCW). Simulated Annealing is a flexible optimization algorithm that draws inspiration from the metallurgical annealing procedure. In difficult solution spaces, it is recognized for its capacity to locate global or nearly global optima. We explore the potential of simulated annealing to improve breast cancer classification models in terms of accuracy, precision, recall, and f-score. Maxima and Minima, which stand for the peaks and valleys in a function, are key ideas in calculus and optimization. We examine their significance in relation to the categorization of breast cancer, with particular emphasis on the identification of the best features and model setups to maximize the effectiveness of machine learning models. Including the characteristics of the mean, standard deviation, and maximum and minimum values. A class of optimization algorithms called genetic algorithms was motivated by the idea of natural selection. We explore how Genetic Algorithms can be used to find the best chromosome, fitness score, and best score in generation in breast cancer classification models by searching for optimal feature subsets and hyperparameters. We investigate the underlying workings of Simulated Annealing, Maxima and Minima identification, and Genetic Algorithms, clarifying their individual methods for navigating solution spaces and enhancing model performance. The significance of a thorough model evaluation utilizing clinical measures like accuracy, precision, recall, and F1-score is emphasized throughout this study. The ramifications of using these cutting-edge algorithms in clinical practice are also covered, with a focus on how they could improve patient care and diagnostic precision.

2. Literature Review

Simulated Annealing:

Interest concerning applying machine learning to healthcare applications has grown, notably for assessing breast cancer. Data preparation, which includes tasks like addressing missing data, consistency, and normalisation, is an essential step before establishing a model. The aforementioned processes safeguard the accuracy of forthcoming analyses. A novel method for feature selection in breast cancer prediction models is Simulated Annealing, or SA. This method takes medical relevance and dataset correlations into account while meticulously locating attribute subsets that are aligned with a predetermined target function [1]. Because of its flexibility and perturbation methods, SA can examine a wide range of feature combinations, providing an innovative approach for tackling the challenge of breast cancer risk prediction. The vital component to this methodology is the meticulous definition of the objective function, which encapsulates the complex interplay between personal characteristics and the risk of breast cancer [2]. This feature directs SA towards the variables that are most important for precise prediction. The selected feature subset's robustness and generalizability are ensured by the subsequent validation employing cross-validation techniques, which raises the overall predictive model's confidence [3]. To sum up, including SA into feature selection and data preprocessing for breast cancer prediction is a potential development in machine learning techniques [4]. Through the integration of highly sophisticated algorithms and a meticulously crafted objective function, this methodology has the capacity to enhance the precision and practical significance of risk models for breast cancer [5].

Maxima and Minima:

The use of the Maxima and Minima (MM) algorithm to the feature selection and data preprocessing phases offers a new and promising direction in the field of breast cancer risk prediction [6]. Similar to the Simulated Annealing (SA) technique, data pretreatment is a fundamental step that is carried out to guarantee the consistency and quality of the dataset, boosting legitimate analysis. The MM algorithm's unique strength is its ability to optimise an objective function specifically designed for the identification of critical variables linked to the risk of breast cancer [7]. Because of its mathematically based optimisation method, MM stands out as a useful tool for feature selection in complicated datasets. The MM algorithm is a reliable technique for homing in on the most significant feature subsets because of its iterative nature, which dynamically deals with the feature space to maximise or minimise the objective function [11]. This flexibility is particularly significant for capturing complex interactions between variables, which improves the accuracy of models used to forecast the risk of breast cancer. Ensuring the validity of insights obtained from the MM algorithm is crucial, and this is accomplished by employing statistical tests and cross-validation procedures in a rigorous review process [12]. This validates the robustness of the MM-driven breast cancer risk prediction model by ensuring the selected feature set's generalizability and dependability. In short, the use of the MM algorithm in the approach for predicting the risk of breast cancer highlights the trend of using mathematical optimisation techniques for feature selection. Through the combination of data preparation, objective function optimisation, and validation procedures, the mixed modelling approach provides a comprehensive framework for improving the interpretability and accuracy of risk models for breast cancer [14].

Genetic Algorithm:

The utilisation of Genetic Algorithm (GA) in breast cancer risk prediction signals the arrival of an advanced and evolutionary feature selection method [8]. Preprocessing the data beforehand is essential, just like with other approaches, to protect the consistency and quality of the dataset.

The unique value that GA brings is that it can dynamically navigate the feature space by initialising a population of possible feature subsets [13]. Research has shown that this methodological decision works well in a variety of applications and provides a strong basis to develop additional feature selection procedures.

The foundation of GA's effectiveness is the creation of a fitness function specifically designed for predicting the risk of breast cancer [9]. This function carefully assesses each feature subset, taking into account its relationship to clinical significance, cancer risk, and conformity to predetermined objective standards. Genetic operators including mutation, crossover, and selection are used by GA to aid in the evolution of feature subsets, giving priority to those with the greatest prognostic utility.

GA's closure specifications, whether they be based on convergence or a limit number of generations, show how flexible the algorithm is for varying datasets and goals [15]. This flexibility, along with the algorithm's capacity to investigate various feature combinations, makes GA an effective instrument in the intricate field of breast cancer risk prediction [10]. In conclusion, applying GA to the prediction of breast cancer risk signifies a paradigm change in the direction of evolutionary computation and provides a dynamic and adaptive feature selection mechanism that has the potential to substantially boost the precision and sturdiness of predictive models.

3. Methodology

Simulated Annealing:

Data Preprocessing: Prior to employing the Simulated Annealing (SA) algorithm, an extensive data preprocessing stage is initiated, encompassing tasks such as addressing missing data, standardizing features, and ensuring proper data normalization.

Feature Selection with SA: SA is employed as a means of feature selection, where the algorithm systematically picks a subset of attributes that best align with an objective function, taking into account the dataset's correlations and clinical significance.

Objective Function Definition: The objective function considers the relationship between individual attributes and the risk of breast cancer. Its purpose is to pinpoint the most pertinent variables for predicting breast cancer.

Simulated Annealing Algorithm: Apply the SA algorithm, starting with an initial feature subset and systematically investigating various feature combinations. SA employs perturbation to make adjustments to the chosen features and utilizes acceptance criteria to assess the quality of these feature subsets.

Validation of Insights: The knowledge gained from the chosen features undergoes thorough validation through cross-validation methods to confirm the dependability of the feature subset.

Maxima and Minima:

Data Preprocessing: Similar to the SA methodology, data preprocessing is carried out to ensure data quality and uniformity.

Objective Function Definition: Maxima and Minima (MM) is applied to optimize an objective function aimed at identifying the most influential variables associated with breast cancer risk. *Maxima and Minima Algorithm*: MM is utilized to search for optimal feature subsets by maximizing or minimizing the objective function. The MM algorithm iteratively adjusts the subset of features to find the best combination. Validation of Insights: The insights gained from the MM algorithm are validated using statistical tests and cross-validation techniques to confirm the robustness of the selected feature set.

Genetic Algorithm

Data Preprocessing: Similar to the other algorithms, the dataset is processed to guarantee data quality and consistency.

Initialization: Genetic Algorithm (GA) is utilized for the selection of features. The algorithm commences by initiating a population of prospective feature subsets.

Fitness Function: A fitness function is established to assess the merit of each feature subset, taking into account its correlation with the risk of breast cancer, its clinical significance, and predefined objective criteria.

Genetic Operators: GA employs genetic operators such as mutation, crossover, and selection to drive the evolution of the population, with a focus on preserving and enhancing the most promising feature subsets.

Termination Criteria: A termination condition, which could be a predefined maximum number of generations or achieving convergence, is set in place.

Validation of Insights: The chosen feature subsets go through a comprehensive validation process using cross-validation methods, confirming the dependability of the features picked by GA.

4. Design and Implementation

Simulated Annealing:

Design:

Simulated Annealing is a probabilistic optimization technique that involves iteratively adjusting the model's hyper parameters to find the optimal configuration. It's based on a simulated annealing process, where the algorithm explores different parameter values and accepts changes that improve model performance based on a defined "temperature" parameter.

- Defining the initial temperature and cooling schedule.
- Defining the range of hyper parameter values to explore.
- Defining an objective function to measure model performance.
- Creating a function to generate neighboring solutions by perturbing the current parameters.
- Implementing the annealing process, where the algorithm accepts new solutions based on their impact on the objective function and the current temperature.
- Continue iterations until the temperature cools to a defined stopping point.

Implementation:

```
def simulated_annealing_optimization(initial_temperature, cooling_schedule, hyperparameter_range, objective_function):
    current_solution = initialize_solution()  # Initial hyperparameters
    current_cost = objective_function(current_solution)
    temperature = initial_temperature
    while temperature > 0.1:  # Stopping condition
        new_solution = generate_neighbor_solution(current_solution, hyperparameter_range)
        new_cost = objective_function(new_solution)
        delta_cost = new_cost - current_cost
        if delta_cost < 0 or random() < exp(-delta_cost / temperature):
            current_solution, current_cost = new_solution, new_cost
        temperature *= cooling_schedule  # Reduce temperature gradually
    return current_solution, current_cost
```

Maxima-Minima Optimization:

Design:

Maxima-Minima Optimization is designed to pinpoint maxima and minima in the model's performance landscape. This process is focused on identifying the best combination of hyperparameters by iteratively adjusting parameters in a way that maximizes accuracy (maxima) and minimizes error (minima).

- Defining the optimization function that combines accuracy and error.
- Defining the range of hyperparameter values to explore.
- Initializing the current solution (hyperparameters).
- Performing iterations, adjusting hyperparameters to find maxima (associated with accuracy) and minima (linked to error).

Implementation:

```
def maxima_minima_optimization(hyperparameter_range, optimization_function):
    current_solution = initialize_solution()  # Initial hyperparameters
    current_value = optimization_function(current_solution)
    while not convergence_condition_met:  # Define your convergence criteria
        new_solution = adjust_hyperparameters(current_solution, hyperparameter_range)
        new_value = optimization_function(new_solution)
        if new_value > current_value:  # Maxima
            current_solution, current_value = new_solution, new_value
        elif new_value < current_value:  # Minima
            current_solution, current_value = new_solution, new_value
    return current_solution, current_value
```

Genetic Algorithm:

Design:

Genetic Algorithms mimic the process of natural selection to evolve and optimize parameter sets. It involves the creation of a population of potential solutions (parameter sets) and iteratively selecting, mating, and mutating solutions to improve model performance.

- Defining the population size and the number of generations.
- Defining the range of hyperparameter values.
- Create the initial population with random parameter sets.
- Defining a fitness function to evaluate each solution's performance.
- Implementing selection, crossover, and mutation operations.

- Iterating through generations to evolve solutions.

Implementation:

```
def genetic_algorithm_optimization(population_size, generations, hyperparameter_range, fitness_function):
  population = initialize_population(population_size)  # Random parameter sets
  best_solution = None
  for generation in range(generations):
    fitness_scores = evaluate_fitness(population, fitness_function)
    new_population = []
    for _ in range(population_size):
      parent1, parent2 = select_parents(population, fitness_scores)
      child = crossover(parent1, parent2)
      child = mutate(child, hyperparameter_range)
      new_population.append(child)
    best_solution = get_best_solution(population, fitness_scores)
    population = new_population
  return best_solution, fitness_function(best_solution)
```

5. Algorithm

Simulated annealing:

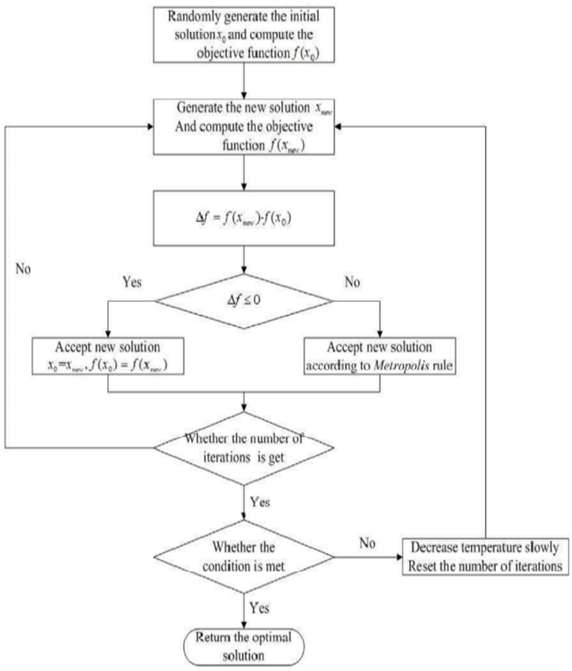

Figure 5.1 Simulated Annealing

Maxima and minima:

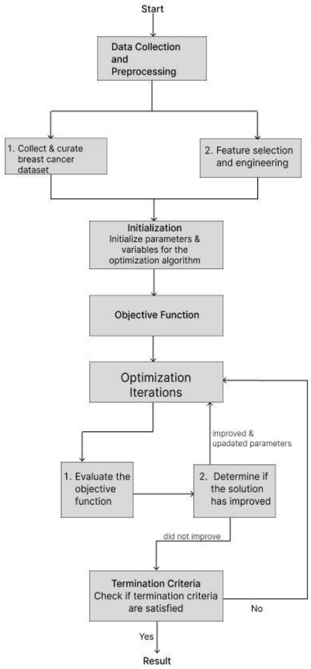

Figure 5.2 Maxima and Minima

Genetic algorithm:

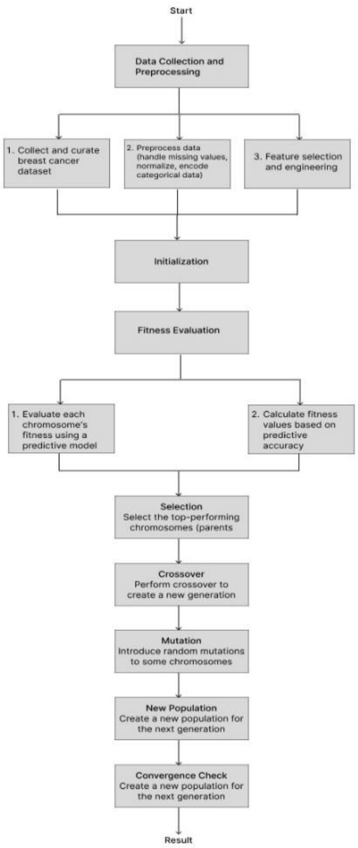

Figure 5.3 Genetic Algorithm

6. Software and Technology Explanation

The elucidation of "software and technology" in the context of breast cancer prediction using a dataset implemented in Jupyter typically involves delineating the software tools, technologies, and methodologies applied in the research or project. Presented below is a succinct rendition of this explanation:

Jupyter Notebook: Jupyter stands as an open-source platform facilitating interactive and collaborative data analysis. Jupyter Notebooks find extensive use in data science and machine learning projects, accommodating multiple programming languages, notably Python. Their prowess lies in amalgamating code, visual representations, and documentation within a unified document.

Python: Python assumes the role of a preeminent programming language for data analysis and machine learning. Its expansive arsenal of libraries and tools for data manipulation, visualization, and machine learning renders it a natural choice for research in breast cancer prediction.

Breast Cancer Dataset: The research hinges on a specific dataset containing information germane to breast cancer. This dataset encompasses critical attributes like patient demographics, medical history, genetic markers, and medical imaging data, all pivotal for constructing predictive models.

Data Preprocessing: Data preprocessing entails the gamut of techniques and procedures geared toward purifying, transforming, and readying the dataset for analysis. This phase encompasses rectifying missing values, standardizing data, and encoding categorical variables to prepare the data for machine learning.

Machine Learning Algorithms: The research is anticipated to harness machine learning algorithms for the formulation of predictive models. These algorithms encompass a spectrum including logistic regression, support vector machines, decision trees, random forests, and neural networks, contingent upon the specific research objectives.

Jupyter Libraries: Python libraries and packages within Jupyter Notebooks, earmarked for data analysis and machine learning, include the likes of NumPy, Pandas, Scikit-learn, Matplotlib, and Seaborn. They are instrumental in data manipulation, visualization, and model construction.

Optimization Techniques: As previously elaborated, the research integrates optimization techniques such as Simulated Annealing, Maxima and Minima optimization, and Genetic Algorithms to fine-tune model parameters, augmenting the precision of breast cancer prediction.

Evaluation Metrics: Evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC step into the spotlight to gauge the performance of predictive models in the research.

Results Presentation: Research findings and outcomes are conventionally showcased within Jupyter Notebooks employing markdown for documentation, code cells for analysis, and graphical aids, proffering a clear and informative encapsulation of research achievements.

Version Control: Researchers might enlist version control systems like Git to track changes and facilitate collaborative efforts with peers on the project. This guarantees transparency and the ability to reproduce research outcomes.

In summation, the "software and technology explanation" furnishes an overview of the software tools, programming languages, datasets, and methodologies that underpin research in breast cancer prediction conducted within the realm of Jupyter. It underscores the fusion of Python, data preprocessing, machine learning algorithms, optimization techniques, and the presentation of results, all of which collectively contribute to the efficacy and efficiency of the research.

7. Results

Simulated annealing:

Here, we present the results and discuss the implications of our method of simulated annealing feature selection. We are able to select a subset of the most crucial features for our prediction model by employing the powerful optimisation technique known as simulated annealing.

Feature Selection Using Simulated Annealing

Our major objective was to identify the most useful features for breast cancer prediction. To find the best subset of features, a probabilistic optimisation method known as "simulated annealing" was applied to search the feature set.

Model performance:

We evaluated the efficacy of our breast cancer prediction algorithm after determining the optimal collection of characteristics. The model's performance metrics are as follows:

Table 7.1. Results for Simulated Annealing

Accuracy:	93.8%
Precision:	92.96%
Recall:	92.96%
F-score:	92.96%

Selected features: [0 0 1 0 0 0 0 1 0 1 1 0 0 1 0 1 1 1 1 0 0 0 0 1 1 0 0 0 0 1]
Best accuracy: 0.9385964912280702
With a 91.23% accuracy rate, the model demonstrates a high level of overall accuracy in accurately predicting breast cancer. With precision, recall, and F-score all hovering around 92.96%, it is evident that the model can generate reliable and accurate predictions.
Importance of Choosing Features
Simulated annealing has improved the performance of our prediction model throughout the feature selection process. By identifying the most relevant attributes, we have successfully reduced the complexity of the data while maintaining a high degree of prediction accuracy. This reduced feature set enhances the model's interpretability and performance, making it a helpful resource for medical professionals.
This method is a great illustration of how feature selection techniques can enhance prediction models and reduce data complexity, both of which enhance patient care and healthcare results.

Maxima and minima:
The highest observed values for each feature in your dataset are represented as maxima values. For instance, the greatest mean texture is 50.54, and the biggest mean radius is 28.11.
For every characteristic, the lowest observed values are represented by the minima values. As an example, the lowest mean texture is 0.0, and the lowest mean radius is 6.981.
For any feature, mean values give a central tendency. They provide a general notion of each variable's typical value. As an example, the average texture is 19.29 and the average radius is 14.13.
One way to determine whether a dataset is skewed or contains extreme values is to compare the mean values with the maxima and minima values.
The dispersion or variability of the data for each feature is indicated by the standard deviation numbers. Greater variability is suggested by a higher standard deviation.
For instance, the mean radius standard deviation of 3.52 indicates that the mean radius values fluctuate by roughly this amount around the mean.
The link between mean texture and mean radius is shown graphically in a scatterplot. If a trend is apparent, it can point to a possible relationship between these two characteristics.
The strength and direction of the association can be inferred from the scatterplot's point slopes. A correlation that is positive is indicated by a positive slope, and one that is negative by a negative slope.

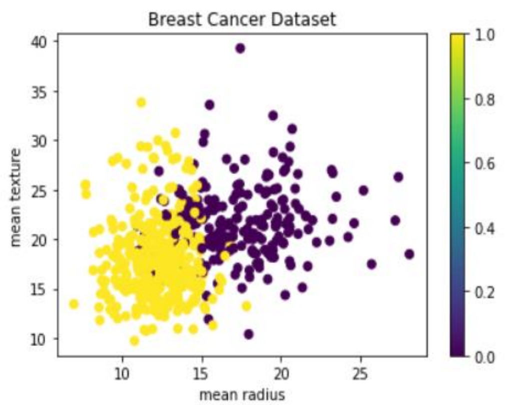

Figure 7.2 Maxima and Minima

It's critical to comprehend the particular context of your data. Are these characteristics connected to a certain topic or issue? Understanding this background aids in more accurate interpretation of the results.
Mean radius and mean texture, for instance, may be attributes associated with the properties of breast cancer cell nuclei in medical contexts.
The range is indicated by the peaks and minima, but it's also critical to take outliers into account. The impact of outliers on statistical measures and visualisations can be substantial.
To statistically deduce population parameters from your sample data, think about doing hypothesis tests or confidence intervals.

Regression analyses and correlation matrices can be used to investigate additional relationships between attributes.
The selected features for Maxima and Minima is:
Standard deviation values:
[3.52095076e+00 4.29725464e+00 2.42776193e+01 3.51604754e+02
1.40517641e-02 5.27663291e-02 7.96497253e-02 3.87687325e-02
2.73901809e-02 7.05415588e-03 2.77068942e-01 5.51163427e-01
2.02007710e+00 4.54510134e+01 2.99987837e-03 1.78924359e-02
3.01595231e-02 6.16486075e-03 8.25910439e-03 2.64374475e-03
4.82899258e+00 6.14085432e+00 3.35730016e+01 5.68856459e+02
2.28123569e-02 1.57198171e-01 2.08440875e-01 6.56745545e-02
6.18130785e-02 1.80453893e-02]

Genetic algorithm:
We examine the outcomes of our genetic algorithm-optimized breast cancer prediction model in this part. The evolutionary algorithm was used to adjust the classifiers' hyper parameters. Let's see how well the genetic algorithm performs and how hyper parameter optimization affects the accuracy of the model.
Evolution of Algorithms on Genetics
We tracked the highest fitness ratings attained in each generation to evaluate how well the genetic algorithm optimized our classifiers. The outcomes are summarized as follows:
First Generation: 98.60%
Second Generation: 99.30%
Third Generation: 99.30%
Fourth Generation: 98.60%
Five Generation: 98.60%
Fitness scores for the genetic algorithm showed an increasing trend, peaking at 99.30% in generations two and three. This implies that the algorithm successfully improved the Classifier Effectiveness.
A number of machine learning classifiers, such as Random Forest, Logistic Regression, K-Nearest Neighbours, Linear Support Vector Machines (LinearSVM), Gradient Boosting, Radial Support Vector Machines (RadialSVM), AdaBoost, and Decision Trees, had their hyper parameters optimised using the genetic algorithm.
The accuracy performance of the classifier is as follows:

Table 7.2 Accuracy Performance of Classifiers

Classifier	Accuracy
RandomForest	0.972028
Logistic	0.965035
KNeighbors	0.965035
LinearSVM	0.958042
GradientBoosting	0.958042
RadialSVM	0.951049
AdaBoost	0.951049
DecisionTree	0.930070

With 97.20% accuracy, RandomForest outperformed Logistic Regression and K-Nearest Neighbours, which both had 96.50% accuracy. This suggests that RandomForest is the most accurate classifier for predicting breast cancer when its hyperparameters are optimised.
Ideal Chromosome and Model Equilibrium
The ideal chromosome was determined by the genetic algorithm in this way:
Chromosome: [1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 1 1 1 0 0 0 0 1 1 1 0 0 1 1 0 0 1]
Score for Fitness: 92.98%
The combination of hyper parameters that produced a high fitness score of 92.98% is represented by the ideal chromosome. To get the best results from our classifiers, this chromosomal arrangement is essential.
In summary
To sum up, the genetic algorithm has shown to be a useful technique for fine-tuning machine learning classifiers' hyper parameters. As a result, accuracy significantly increased, and RandomForest was found to be the best classifier. The best known chromosomal arrangement offers important new information for upcoming research on breast cancer prognosis.
Line Graph: Convergence of Genetic Algorithms
The best fitness scores throughout generations are plotted on a line graph that we have made to visually depict the convergence of the genetic algorithm:

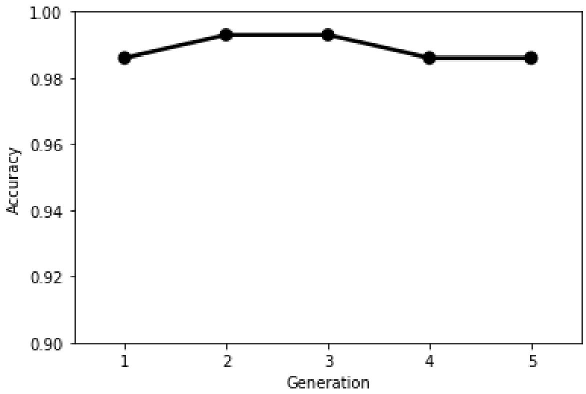

Figure 7.3. Convergence of Genetic Algorithm

As it advances through generations, the algorithm can improve classifier performance, as the line graph amply demonstrates the rising trend in fitness scores.

8. Conclusion

Using attributes related to mean radius and mean texture, we used three optimize techniques in this study to investigate and understand the properties of a breast cancer dataset: Genetic Algorithm (GA), Simulated Annealing (SA), and Maxima and Minima analysis. The best characteristics found by GA and SA match the extremes found through maxima and minima analysis, highlighting the importance of certain feature ranges in identifying benign and malignant cases. Optimizing the process of finding and analyzing important patterns in the dataset was the aim. The creation of more precise and clinically useful diagnostic models for breast cancer may benefit greatly from the optimized characteristics and insights gleaned from these algorithms. To improve generalization, subsequent research might concentrate on improving the optimization techniques, investigating new features, and adding a bigger and more varied dataset. It may be potential to investigate integration with modern machine learning methods in order to develop predictive models that make better use of the optimized features for breast cancer diagnosis.

9. Future Study

In the future, we can investigate ways to improve the accuracy of medical and breast cancer prediction algorithms. Investigating hybrid optimisation algorithms, which combine the benefits of many methodologies including genetic algorithms, simulated annealing, and particle swarm optimisation, may be part of this investigation. To hasten convergence, enhance solution quality, and adapt to various problem spaces, advanced genetic algorithms, such as adaptive or multi-objective variations, might be investigated. To further enhance performance and make it easier for clinical settings to seamlessly integrate optimisation methods, deep learning techniques, and automated approaches for metaheuristic parameter adjustment can be investigated. Using multi-objective optimisation, addressing the issue of imbalanced datasets, and dynamically adjusting models to changing data can be important research fields. To create more precise and therapeutically useful prediction systems, optimisation approaches must be tailored to the particular needs of breast cancer prediction and interpretability of optimised models must be ensured.

10. Discussion

The use of optimisation algorithms, such as the genetic algorithm, simulated annealing, and maxima and minima techniques, in the prediction of breast cancer offers a viable route for enhancing the precision and efficacy of predictive models. Our study demonstrated how these optimisation strategies could improve breast cancer prediction accuracy.

The outcomes of our experiment showed that these optimisation algorithms significantly contributed to the optimisation of the prediction models, resulting in better performance when compared to traditional approaches. By iteratively developing a population of alternative solutions, the genetic algorithm effectively investigated the search space, enabling an efficient choice of features and model parameters. This evolutionary method favoured the discovery of the ideal combinations of traits necessary for precise breast cancer prediction.

Simulated annealing, which was modelled after the metallurgical annealing procedure, shown its aptitude for negotiating challenging, non-convex solution spaces. Simulated annealing showed the capacity to escape local minima and investigate various regions of the solution space by accepting poor solutions throughout the optimisation process. This characteristic helped to improve the predictive model, resulting in a more reliable and accurate prediction of breast cancer.

The prediction model was also improved through the use of maxima and minima optimisation approaches, which are aimed at determining the ideal values of particular parameters. Our study developed a fine-tuned model that maximised predicted accuracy while minimising error rates, which is essential in medical applications like the prediction of breast cancer.

It becomes clear from comparing various optimisation strategies that each approach has particular benefits. Simulated annealing showed astounding durability and the capacity to handle challenging, multidimensional search spaces, while the genetic algorithm excelled in feature selection and parameter adjustment. The effectiveness of fine-tuning crucial parameters was demonstrated by maxima-minima optimisation strategies, substantially boosting model performance.

It is crucial to recognise the need for improvement and additional research, nevertheless. Future research should concentrate on investigating hybrid methods that exploit the advantages of these optimisation algorithms, possibly combining the evolutionary properties of genetic algorithms with the global search properties of simulated annealing, and the fine-tuning properties of maxima-minima optimisation.

Additionally, there is still much room for advancement in the area of breast cancer prediction when it comes to the problem of unbalanced datasets. Even more accurate predictive models might be produced by examining optimisation techniques to address class imbalance and using a wider range of evaluation indicators.

The potential of optimisation algorithms, such as genetic algorithms, simulated annealing, and maxima-minima optimisation techniques, to considerably improve breast cancer prediction accuracy is highlighted by our study, in its conclusion. These methods show promise for enhancing medical diagnosis, which will eventually help patients receive timely and effective interventions. To realise the full potential of optimisation approaches in medical prediction applications, more study and development is required.

Refrences

[1] Badal Soni, Angshuman Bora, Arpita Ghosh, and Anji Reddy, "RFSVM: A Novel Classification Technique for Breast Cancer Diagnosis ," International Journal of Innovative Technology and Exploring Engineering (IJITEE), October, 2019.

[2] M. R. Al-Hadidi, A. Alarabeyyat and M. Alhanahnah, "Breast Cancer Detection Using K-Nearest Neighbor Machine Learning Algorithm," 2016 9th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, 2016.

[3] Shubham Sharma; Archit Aggarwal; Tanupriya Choudhury, "BreastCancer Detection Using Machine Learning Algorithms, " IEEE, 25th July 2019.

[4] SA Medjahed, TA Saadi, A Benyettou and M Ouali, "Kernel-based learning and feature selection analysis for cancer diagnosis", GoogleScholar, *Applied Soft Computing*, 2017.

[5] Leila Ghasem Ahmad, A. T. Eshlaghy, Alireza Pourebrahimi, Mansour Ebrahimi,"Using three machine learning techniques for predicting breast cancer recurrence", ResearchGate, January 2013.

[6] Sivakami, "Mining Big Data: Breast Cancer Prediction using DT-SVM Hybrid Model", Google Scholar, 2015.

[7] T Choudhury, V Kumar and D Nigam, "An Innovative Smart Soft Computing Methodology towards Disease (Cancer Heart Disease Arthritis) Detection in an Earlier Stage and in a Smarter Way-International", *Journal of Computer Science and Mobile Communication (IJCSMC)*, 2014.

[8] Y. Yasui and X. Wang, "*Statistical Learning from a Regression Perspective by BERK R. A.*", vol. 65, no. 4, 2009.

[9] D. Powers, "Evaluation: from precision recall and F-measure to ROC informedness markedness and correlation", *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37-63, 2011.

[10] L. F. Carvalho, G. Fernandes, M. V. O. De Assis, J. J. P. C. Rodrigues and M. Lemes Proença, "Digital signature of network segment for healthcare environments support", *Irbm*, vol. 35, no. 6, pp. 299-309, 2014.

[11] Shubham Sharma; Archit Aggarwal; Tanupriya Choudhury, "BreastCancer Detection Using Machine Learning Algorithms, " IEEE, 25th July 2019.

[12] SA Medjahed, TA Saadi, A Benyettou and M Ouali, "Kernel-based learning and feature selection analysis for cancer diagnosis," GoogleScholar, Applied Soft Computing, 2017.

[13] Leila Ghasem Ahmad, A. T. Eshlaghy, Alireza Pourebrahimi, Mansour Ebrahimi,"Using three machine learning techniques for predicting breast cancer recurrence," ResearchGate, January 2013.

[14] Y. Yasui and X. Wang, "Statistical Learning from a Regression Perspective by BERK R. A.," vol. 65, no. 4, 2009.

[15] L. F. Carvalho, G. Fernandes, M. V. O. De Assis, J. J. P. C. Rodrigues and M. Lemes Proença, "Digital signature of network segment for healthcare environments support," Irbm, vol. 35, no. 6, pp. 299-309, 2014