https://afrjmis.net

Performance Comparison of Selected Swarm Intelligence Algorithms on Breast Cancer Diagnosis

Babatunde Julius Olorunsola[1], Tinuke Omolewa Oladele[2], Taye Oladele Aro[3], Hakeem Babalola. Akande[4] and Oyenike Adunni Olukiran[5]
[1, 2]Department of Computer Science, University of Ilorin, Ilorin, Nigeria
[3,5] Department of Mathematical and Computing Sciences, KolaDaisi University, Ibadan, Oyo State, Nigeria
[4]Department of Telecommunication Science, University of Ilorin, Ilorin, Nigeria
Email: [1]juliusbabatunde140@gmail.com [2]tinuoladele@gmail.com [3]taiwo774@gmail.com [4]hakeemobabs@gmail.com
* Corresponding author (Phone Number: +234-803-617-6349)

ABSTRACT

Techniques of data mining have a growing reputation in healthcare for the diagnosis of breast cancer due to its robust diagnostic capability and better classification. Analyzing huge features generally require large memory and computation. If features are carefully chosen, it is expected that the features set will remove the relevant information from the input data to perform the desired task using this reduced features representation. This paper conducted a comparative analysis of the performance of selected swarm intelligence algorithms on breast cancer diagnosis. Genetic Algorithm (GA), Particle Swarm Optimization (PSO), HarmonySearch (HS) and Tabu Search were employed to select the most relevant from the Wisconsin breast cancer (original) dataset. The selected features were passed to seven different machine learning algorithms: Support Vector Machine (SVM), Decision Tree (C4.5), Naïve Bayes (NB), K Nearest Neighhood (KNN), Neural Network (NN), Logistic Regression (LR) and Random Forest (RF). The earlier study applied PSO, Imperialist Competitive Algorithm (ICA), Firefly Algorithm (FA) and Invasive Weed Optimization (IWO) on the original Wisconsin breast cancer dataset. The breast cancer diagnostic model was evaluated based on accuracy, precision, recall and F1-measure. The results showed that the best accuracy of 97.1388% was obtained in the PSO and Tabu-search applying RF classifier, the best precision value of 0.9720 was recorded in Tabu-search applying RF classifier, the best kappa statistic value of 0.9372 was obtained in Tabu-search algorithm applying RF classifier, the best recall value of 0.9750 was achieved in PSO using RF classifier, the best F1-measure value of 0.9700 was obtained in HarmonySearch Algorithm applying SVM classifier. Finally, it was revealed that the RF classifier performed efficiently with swarm intelligence algorithms.

.

Keywords: *Breast cancer, Diagnosis, Data mining, Feature selection, Swarm intelligence*

Reference Format:

Olorunsola, Babatunde Julius; Oladele, Tinuke Omolewa; Aro, Taye Oladele; Akande, Hakeem Babalola and Olukiran, Oyenike Adunni (2021), A Performance Comparison of Selected Swarm Intelligence Algorithms on Breast Cancer Diagnosis *Afr. J. MIS*, Vol. 3, Issue 1, pp. 5 - 21.

1. INTRODUCTION

Breast cancer is a serious ailment that has been discovered to be the second cause of death among women in society and also identified to be a popular type of cancer in females that is affecting approximately 10% of all women globally [1][2]. Among the medical applications, breast cancer diagnosis and prognosis pose a major challenge to researchers in getting desired diagnostic model in healthcare [3]. The medical experts are faced with various problems in the diagnosis of some diseases in which breast cancer is included. Several issues that often come across during the diagnosis of breast cancer prediction include inadequate understanding of symptoms, risk factors, lack of quality diagnostic measure for the patient and no adequate information to properly predict the case of breast cancer [4].

In the data mining approach for diagnosis of diseases such as breast cancer diagnosis, the data used most times always contains more features needed or sometimes the wrong kind of features (irrelevant features) to build a diagnostic model [5]. The use of irrelevant features in building a model may result in more memory space used during the training process and classification phase. The selection of features has been identified to be a dynamic area of research in machine learning and data mining techniques [6]. Several conventional techniques for the selection of features: filter, wrapper and hybrid have been applied in the machine learning domain and data mining technique [7]. The rationale behind the use of all the techniques is to remove inappropriate or redundant features from the database. The filter approach chooses the subset of a feature based on essential characteristics of the data, independent of mining algorithm. The filter technique is usually applied to data with high dimensionality due to its generality and high computation efficiency [8]. The wrapper method chooses features for accuracy calculation by the target learning algorithm, applying a definite learning algorithm, wrapper finds the feature space by neglecting some features and analyzing the effect of feature omission on the metrics for prediction [9].

Among the feature selection techniques, the swarm intelligence techniques have been identified to outperform other techniques such as filter, wrapper and hybrid methods [10]. The selection of features is a general problem common with large datasets. To improve the accuracy and speed of a model, it is necessary to choose the subset of features that are discriminative using swarm intelligence algorithms [11] such as Genetic algorithms, Swarm intelligence optimization, Elephant Algorithm, HarmonySearch Algorithm, Lion Optimization Algorithm, firefly Algorithm, Gravitational Search Algorithm, Ant colony optimization and so on. These methods can be effective for this problem, which require less amount of computation time and memory.

This paper conducted a comparative analysis on some selected swarm intelligence algorithms for feature selection techniques in breast cancer diagnosis.

2. RELATED WORK

In data mining, several studies have been carried on breast cancer diagnosis. These include:
Several studies were reviewed concerning diagnosis, prediction and classification of cancers [12], The main focus was on the application of diverse algorithms for the prediction of cancer using data mining. Considering the analysis of various outcomes by the study, it was understandable that the compact of multidimensional diverse data, connected with the applicant of different techniques for feature selection and classification can produce better tools for inference in the cancer domain. The application of these techniques by researchers has helped physicians to create concepts that contribute to achieving effective results. [13] used ensemble-based on homogenous approach employing data mining for diagnosis of breast cancer. The study applied three classification algorithms: K-Nearest Neighbour (KNN), Decision Tree (C4.5) and Support Vector Machines (SVM) with their homogenous ensembles of Bagging and Boosting. The experimental result showed that the individual-based approach of SVM recorded the best accuracy of 97.14% compared with the homogenous ensembles of bagging and boosting methods.

Development of a predictive model for the survivability of breast cancer by [14]. Three commonly used machine learning algorithms: Naıve Bayes, RBF Network and J48 were used to build predictive models using a large dataset (683 breast cancer cases). A 10-fold cross-validation approach was applied to measure the unbiased estimate of the three prediction models for performance comparison purposes. Results revealed that the Naıve Bayes recorded the best accuracy of 97.36% on the holdout sample, RBF Network produced 96.77% of accuracy and J48 gave an accuracy of 93.41%. A predictive model for breast cancer at an early stage was developed [15]. The study involved four main modules: a collection of data, pre-processing of data, feature selection and classification were considered. In the pre-processing stage, Global histogram equalization (GHE) was used to find a uniform histogram for the output image. Discrete Wavelet Transform was applied for feature extraction. Two machine learning algorithms: SVM and NN were used

for classification. The evaluation of the model was done with MIAS digital mammography database. The result showed that the Neural Network produced a higher accuracy of 95.15%.

Breast cancer diagnostic model was built using an ensemble approach [16]. Different learning techniques: C4.5 Decision Tree (DT), Support Vector Machine (SVM), Artificial Neural Network (ANN) and also the ensemble of these techniques were used. SPSS Clementine software was used to carry out the experimental analysis. Results showed that the accuracy of 98.97% was obtained in both SVM and C4.5, while the accuracy of 97.54% was recorded in ANN. The ensemble model gave the best accuracy of 98.77%. Application of diverse classification techniques of data mining to the Breast Cancer-Wisconsin dataset was performed by [17] The comparison of the accuracy of models was achieved using dual data partitions. The BayesNet gave 97.13% of accuracy in the case of 10-fold data partitions. When applied with feature selection applying info-gain technique employing BayesNet and SVM. The highest accuracy of 97.28% was obtained in BayesNet with only six feature subsets.

Data mining techniques were employed to build a cancer predictive model [18], the study classified the cancer of colon microarray dataset using five diverse machine learning algorithms: Naïve Bayesian, KNN, SVM, RF and NN. The performance of these algorithms was calculated based on accuracy, precision and recall. The experimental result showed that the highest accuracy was recorded in both KNN and NN classifiers in all the algorithms. Direct classification without and with feature selection techniques on the breast cancer dataset was performed [19], the accuracy of the classifier was improved upon through feature selection because it removed irrelevant attributes. The result showed that the feature selection increased the accuracy of all three different classifiers, reduced the Mean Standard Error (MSE) and increased Receiver Operating Characteristics (ROC).

[20] used swarm intelligence algorithms for feature selection techniques to improve the accuracy of a breast cancer diagnosis. Particle Swarm Optimisation (PSO), Imperialist Competitive Algorithm (ICA), Firefly Algorithm (FA) and Invasive Weed Optimization (IWO) were applied to the original Wisconsin breast cancer dataset. The study combined ANN with the four swarm intelligence algorithms to further enhance the performance of breast cancer diagnostic model. The experimental results showed that FA algorithm recorded the highest accuracy of 98.54% in all the algorithms applied.

[21] employed particle swarm optimization, non-dominating sorting and multi-classifier techniques, namely, k-nearest neighbour method, fast decision tree and kernel density estimation for breast cancer diagnosis. Bayes' theorem was implemented for revising the results to achieve optimum accuracy in the breast cancer prediction. The proposed particle swarm optimization and non-domination sorting with classifier technique model selected the most significant features relevant to breast cancer predictions. The selected features design the objective of the problem model. The proposed model is implemented on the WBCD and WDBC breast cancer data sets publicly available from the UCI machine learning data repository. The metrics considered are sensitivity, specificity, accuracy and time complexity. The experimental results using measures such as sensitivity, specificity, accuracy and time complexity. The experimental results of the study are evaluated against the state-of-the-art algorithms, namely, genetic algorithm kernel density estimation and particle swarm optimization kernel density estimation wherein the results justify the superiority of the proposed model

3. METHODOLOGY

3.1 The Developed Breast Cancer Diagnostic Model

The developed breast cancer diagnostic model was simulated using Weka 3.9.4 data mining tool. The study used one dataset, the Breast cancer Wisconsin dataset was employed to evaluate the breast cancer diagnosis model. The dataset was obtained from the University of Wisconsin Hospitals, Madison, Wisconsin, USA. Data was converted into Attribute Relation File Format (arff) and then later loaded into the system. The selection of relevant features was done using swarm intelligence algorithms: HarmonySearch, Genetic Algorithm, Particle Swarm Optimization and Tabu Search Algorithm selector to rank the features according to their relevance. Seven learning algorithms: C4.5, SVM, KNN, NN, LR, RF and NB were used for classification. The data from the breast cancer dataset was cleaned during the preprocessing stage. After the input data normalization, classification was performed using the seven machine learning techniques. The framework of the breast cancer diagnostic model is shown in Figure 1.

3.2 Weka Data Mining Tool

The WEKA data mining tool was developed at the University of Waikato, New Zealand. The software was implemented using the JAVA programming language. WEKA involves the collection of different machine learning algorithms for data mining tasks, it

is an open-source software issued under the General Public License. Weka normally makes use of Attribute Relation File Format (ARFF), which contains special tags to indicate different things in the data file. It implements algorithms for data pre-processing, classification, regression, clustering and association rules; it also includes a visualization tool. The main interface in Weka is the Explorer. It has a set of panels, each of which can be employed to carry out a definite task.

3.3 Acquisition of Breast Cancer Dataset

This study used the Wisconsin breast cancer dataset to evaluate the developed breast cancer diagnosis model using data mining classification techniques. The details of the attributes found in the Wisconsin Breast Cancer Dataset (WDBC) dataset: ID number, Diagnosis (M = malignant, B = benign) and ten real-valued features are computed for each cell nucleus: Radius, Texture, Perimeter, Area, Smoothness, Compactness, Concaity, Concave points, Symmetry and Fractal dimension. The sample of the Wisconsin breast cancer dataset with different attributes is shown in Figure 2.

3.4 Swarm Intelligence Algorithms

The breast cancer diagnosis model used different swarm intelligence algorithms to perform the selection of the most relevant and discriminant features before the classification. Tabu-Search, Particle Swarm Optimization, HarmonySearch Algorithm and Genetic Algorithm were employed. The following figures give algorithms of the swarm intelligence algorithms used in this study:

3.5 Evaluation Metrics

It is an integral phase of the model development process. The following evaluation metrics commonly used in data mining techniques:

(i) Accuracy: The percentage of correctly classified instances is often called accuracy and the percentage of incorrectly classified instances is gotten by subtracting the correctly classified instances from 100 as shown in Equation (1).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \times 100\% \qquad (1)$$

Where
TP = True positive (It is an outcome where the model correctly predicts the positive class)
TN = True negative (It is an outcome where the model correctly predicts the negative class)
FP = False positive (It is an outcome where the model incorrectly predicts the positive class)
FN = False negative (It is an outcome where the model incorrectly predicts the negative class)

(ii) Recall: This is the proportion of actual positives that have been correctly identified as shown in Equation (2)

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \times 100\% \qquad (2)$$

(iii) Precision: It can also be defined as the number of true positives divided by the number of true positives plus the number of false positives as shown in Equation (3).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \times 100\% \qquad (3)$$

(iv) F1-Score: It is a measure of the test's accuracy**.** This is calculated by considering the weighted average of precision and recall (harmonic mean of precision and recall) as shown in Equation (4).

$$\text{F1} = 2 * \frac{\text{precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (4)$$

4. RESULTS AND DISCUSSION

The experimental analysis was conducted in WEKA data mining which consists of a Graphic User Interface (GUI) Chooser which enables users to navigate to different interfaces for diverse operations. The applications are with five buttons Explorer, Experimenter, Knowledge Flow, Workbench and Simple CLI, as illustrated in Figure 7.

From the interface, the Explorer button is clicked to navigate to the Explorer Interface. This interface is required for this study to achieve its purpose. The pictorial view is shown in Figure 8.

The interface shown in Figure 8 supports operations such as data pre-processing, classification, Clustering, Association, Selection of attributes and Visualization of data.

4.1 Dataset Loading

The breast cancer Wisconsin (original) dataset was loaded into the WEKA data mining tool as presented in Figure 9.

4.2 Results of Developed Breast Cancer Diagnosis model

4.2.1 Results of Optimised Features for Swarm Intelligence Algorithms

The swarm intelligence algorithms: Particle Swarm Optimization, Genetic Algorithm, Tabu Search and Ant Colony Optimization obtained the ten most discriminant features which are Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses, and class from eleven features of original breast cancer dataset as shown in Figure 10, 11, 12 and 13.

4.2.2 Results of Performance of the Breast Cancer Diagnostic

The performance of the breast cancer diagnostic model was evaluated using different metrics such as accuracy, precision, recall, F-measure, Kappa-statistic and Time taken to build a model. Samples of results are represented in Figure 14, 15, 16, 17 and 18.

The results are illustrated in the following tables.

In Table 1, the highest accuracy of 96.7096% was obtained in SVM and RF classifier, the highest precision value of 0.9680 was recorded in RF classifier, the highest recall value of 0.9670 was recorded in SVM and RF classifier, the highest F1-measure value of 0.9670 was obtained in both SVM and RF classifier, the highest kappa statistics of 0.9278 was obtained in RF classifier. The lowest time of 0s taken to build a model was achieved in KNN classifier.

In Table 2, the highest accuracy of 97.1388% was obtained in RF classifier, the highest precision value of 0.9700 in SVM classifier, the highest recall value of 0.9750 was recorded in RF classifier, the highest F1-measure value of 0.9700 was obtained in SVM classifier, the highest kappa statistics of 0.9370 was obtained in RF classifier. The lowest time of 0s taken to build a model was achieved in KNN and NB classifier.

In Table 3, the highest accuracy of 96.9957% was obtained in SVM, the highest precision value of 0.9700 was obtained in SVM, the highest recall value of 0.9700 was recorded in SVM, the highest F1-measure value of 0.9700 was recorded in SVM, the highest kappa statistics value of 0.9337 was obtained in SVM. The lowest time of 0s taken to build a model was achieved in KNN.

In Table 4, the highest accuracy of 96.7096% was obtained in RF classifier, the highest precision value of 0.9700 was obtained in SVM classifier, the highest recall value of 0.9700 was recorded in SVM classifier, the highest F1-measure value of 0.9700 was recorded in SVM classifier, the highest kappa statistics of 0.9338 was obtained in SVM classifier. The lowest time of 0.08s taken to build a model was achieved in KNN and NB.

In Table 5, the highest accuracy of 97.1388% was obtained in RF, the highest precision value of 0.9720 was recorded in RF, the highest recall value of 0.9710 was recorded in RF, the highest F1 measure of 0.9710 was obtained in RF, the highest kappa statistics of 0.9372 was obtained in RF. The lowest time of 0s taken to build a model was achieved in C4.5, KNN, SVM and NB.

4.3 Results of Comparative Analysis

The comparative analysis was conducted for the different evaluation metrics used. Results are comprehensively compared in the following tables.

From Table 6, the highest accuracy of 97.1388% was obtained in feature selection using PSO and Tabu Search algorithm applying RF classifier, while the lowest accuracy of 94.5637% applying was recorded in without feature selection, with feature selection using PSO, Genetic Algorithm and Tabu search applying C4.5 classifier.

From Table 7, the highest precision value of 0.9720 was obtained in feature selection using Tabu-search algorithm applying RF classifier, while the lowest precision value of 0.9440 was recorded in feature selection PSO applying RF classifier.

From Table 8, the highest recall value of 0.9750 was obtained in feature selection using PSO algorithm applying RF classifier, while the lowest recall value of 0.9460 was recorded without feature selection, with feature selection using PSO, Genetic Algorithm and Tabu-search applying C4.5.

From Table 9, the highest kappa statistic value of 0.9372 was obtained in feature selection using Tabu-Search algorithm applying RF classifier, while the lowest kappa statistic value of 0.8799 was recorded without feature selection, with feature selection using PSO, Genetic Algorithm and Tabu-search applying C4.5 classifier.

From Table 10, the highest F1-measure value of 0.9700 was obtained in feature selection using HarmonySearch Algorithm applying SVM classifier, while the lowest F1-measure value of 0.8799 was recorded without feature selection, with feature selection using PSO, Genetic Algorithm and Tabu-search applying C4.5 classifier.

5. CONCLUSION

In data mining, data may sometimes contain more features needed or wrong features (irrelevant features) to build a predictive or diagnostic model. The use of unneeded or irrelevant features employ in building a model may result in more memory space used during the training process and classification phase. This paper comparatively analysed the performance of swarm intelligence-based feature selection techniques on breast cancer disease diagnosis. The experimental results showed the effectiveness of swarm intelligence algorithms with respect to the learning algorithms employed for classification. The best accuracy of 97.1338% was obtained in PSO and Tabu-search applying RF classifier. The best precision value of 0.9720 was recorded in Tabu-search algorithm, while also the best recall value of 0.9750 was recorded in PSO applying RF. The best kappa statistic value of 0.9372 was obtained in feature selection using Tabu-search algorithm applying RF classifier. The best F1-measure value of 0.9700 was obtained in HS applying RF classifier. It was concluded that RF outperformed other machine learning algorithms in general performance of the breast cancer diagnostic model.

REFERENCES

1. I. O. Kehinde, W., Peter, A. I., Jeremiah, A. B. and Adeniran (2015). Breast Cancer Risk Prediction Using Data Mining, Transmission. Networks Communication., 3(2). 1–12.

2. J. Majali, R. Niranjan, V. Phatak, and O. Tadakhe (2015). Data Mining Techniques For Diagnosis And Prognosis of Cancer, International Journal of Advanced Research Computer and Communication Engineering. 4(3). 613–616.

3. B. Padmapriya (2014). A survey on Breast Cancer Analysis using Data Mining Techniques, IEEE Int. Conf. Comput. Intell. Computing Res., 4–5.

4. M. & Deneshkumar, V., Manoprabha and K. k Senthamarai (2019). Comparison of data mining techniques for prediction of breast cancer, International Journal of Science and Technology. Res., 8(8) 1331–1337.

5. K. Sheena, Krishan, K and Gulshan (2016). "Analysis of Feature Selection Techniques : A Data Mining Approach," in International Conference on Engineering & Technology, 17–21.

6. M. B. Kursa (2010). Feature Selection with the Boruta Package, Journal of Statistical. Software. 36(11) 1–13.

7. A. Kumar, P. Prabhat and K. L. Jaiswal (2014). Classification Model for the Heart Disease," Global Journal. of Medical. Research. 14(1).

8. K. Sutha, (2015). A Review of Feature Selection Algorithms for Data Mining Techniques, International Journal of Computer Science and Engineering., 7(6) 63–67.

9. P. Yildirim, Filter Based Feature Selection Methods for Prediction of Risks in Hepatitis Disease (2015). International Journal of Machine. Learning and. Computing.5(4), 258–263.

10. B. Kumari and T. Swarnkar (2011). Filter versus Wrapper Feature Subset Selection in Large Dimensionality Microarray : A Review, International Journal of Computer Science and Information Technology. 2(3) 1048–1053..

11. S. I. Adi and M. Aldasht (2018). Parallel Evolutionary Algorithms for Feature Selection in High Dimensional Datasets. American. Journal of Computer. Science. Engineering. Survey. 6(1) 13–21.

12. Z. N. S. Weli, (2020). Data Mining in Cancer Diagnosis and Prediction: Review about Latest Ten Years. Current Journal of Applied Science and Technology. 39(6) 11–32.

13. Aro, T. O., Akande, H. B., Jibrin, B. M., and U. A Jauro (2019). "Homogenous Ensembles on Data Mining Techniques for Breast Cancer Diagnosis," Daffodil International University Journal of Science and Technology.14(1) 9–12.

14. V. Chaurasia, S. Pal, and B. B. Tiwari (2018). Prediction of Benign and Malignant Breast Cancer using Data Mining Techniques, Journal of Algorithm and ComputerTechnology. 12(2) 119–126.

15. B. Raufi, J. Ajdari, X. Zenuni, and F. Ismail (2017). Enhancing Breast Cancer Detection using Data Mining Classification Techniques, in 2nd World Conference on Technology, Innovation and Entrepreneurship, 605, 310–316.

16. G. Zorluoglu and M. Agaoglu (2017). "Diagnosis of Breast Cancer Using Ensemble of Data Mining Classification Methods " International Journal Oncology Cancer Therapy. 2(4) 24–27.

17. A. K. Shrivas and A. Singh (2016). Classification of breast cancer diseases using data mining techniques. International Journal of Engineering and Science Inventory. 5(12) 62–65

18. G. Porkodi, R and Suganya (2015). A

Comparative Study of Different Deployment Models in a Cloud," International Journal of Advanced. Research. Computer Science and Software Engineering. 3(5) 512–515.

19. A. Lebba, A., Sayeth, S., Elankovan, S., and Azuraliza (2014). Comparative Study on Different Classification Techniques for Breast Cancer Dataset," International Journal of Computer Science and Mobile Computing. 3(10) 185–191,

20. M. Zamani, H., and Nadimi-Shahraki (2016). Swarm Intelligence Approach for Breast Cancer Diagnosis," International Journal of Computer Application. 151(1) 40–44.

21. S. Vijayalakshmi, A. John, R. Sunder, and S. Mohan (2020), "Multi-modal Prediction of Breast Cancer using Particle Swarm Optimization with non-dominating sorting. International Journal of Distributed. Sensor Networks. 16(11) 3-12

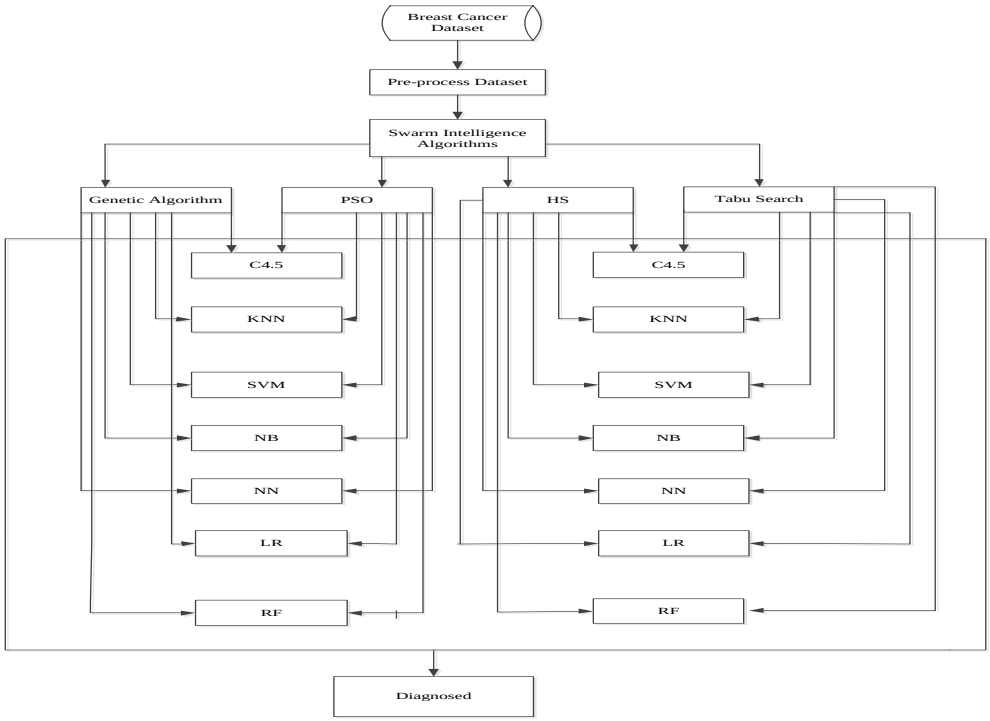

Figure 1: Framework of the Breast Cancer Diagnostic Model

```
@attribute Cell_Shape_Uniformity integer [1,10]
@attribute Marginal Adhesion integer [1,10]
@attribute Single_Epi_Cell_Size integer [1,10]
@attribute Bare_Nuclei integer [1,10]
@attribute Bland_Chromatin integer [1,10]
@attribute Normal Nucleoli integer [1,10]
@attribute Mitoses integer [1,10]
@attribute Class {benign, malignant}
@data
5,1,1,1,2,1,3,1,1,benign
5,4,4,5,7,10,3,2,1,benign
3,1,1,1,2,2,3,1,1,benign
6,8,8,1,3,4,3,7,1,benign
4,1,1,3,2,1,3,1,1,benign
8,10,10,8,7,10,9,7,1,malignant
1,1,1,1,2,10,3,1,1,benign
```

Figure 2: Sample of Wisconsin Breast Cancer Dataset

Step 1: Start
Step 2: Generate an initial solution set
Step 3: Create a list of candidate movement
Step 4: Choose the best candidate
Step 5: Update the tabu search list
Step 6: Stopping criterion met else if Goto Step 3
Step 7: End

Figure 3: Tabu-Search Algorithm

Step 1: Start
Step 2: Generate an initial population
Step 3: Perform selection (objective function for each chromosome) fitness function
Step 4: Generate operation crossover
Step 5: Perform mutation (replacement of gene)
Step 6: Terminal condition met else if Goto Step 2
Step 7: End

Figure 4: Genetic Algorithm

Step 1 : Initialize the problem and algorithm parameters.
Step 2: Initialize the harmony memory.
Step 3: Improvise a new harmony.
Step 4: Update the harmony memory.
Step 5: Repeat Step 2, 3, 4 until the stopping criterion is met.
Step 6: The best harmony stored in HM is returned as the found optimum solution
Step 7: End

Figure 5: HarmonySearch Algorithm

Step 1: Start
Step 2: Initialize group of Particles
Step 3: Evaluate pBest for each particle
Step 4: If current position is better than pbest then Update pBest Else assign pBest to gBest
Step 5: Compute velocity
Step 6: Update particle position
Step 7: Target reach Else Goto Step 3
Step 8: End

Figure 6: Particle Swarm Optimization Algorithm

Figure 7: WEKA Tool Applications Interface (GUI Chooser)

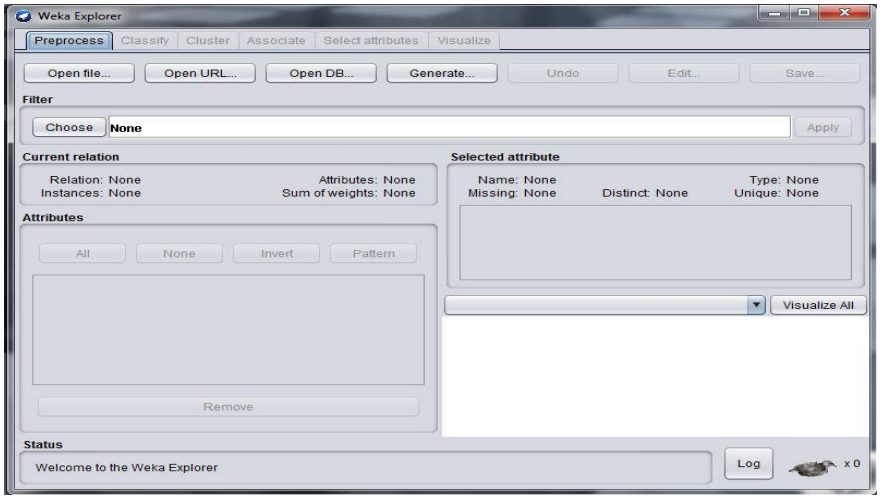

Figure 8: The WEKA Explorer Interface.

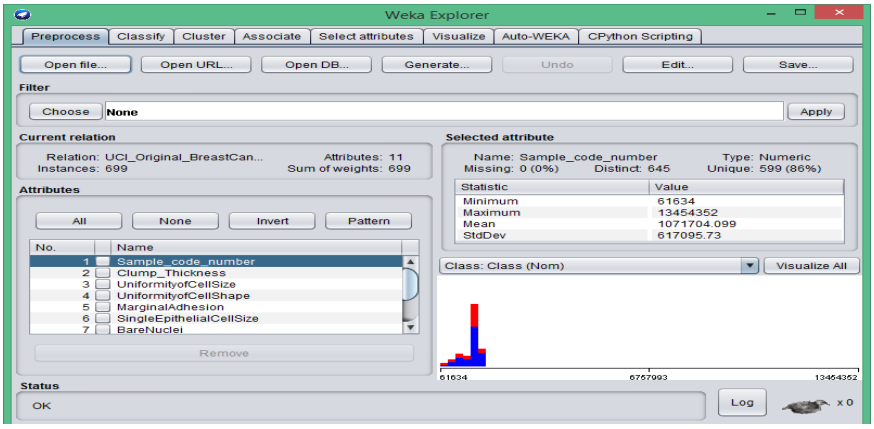

Figure 9: Interface for Loading of Dataset

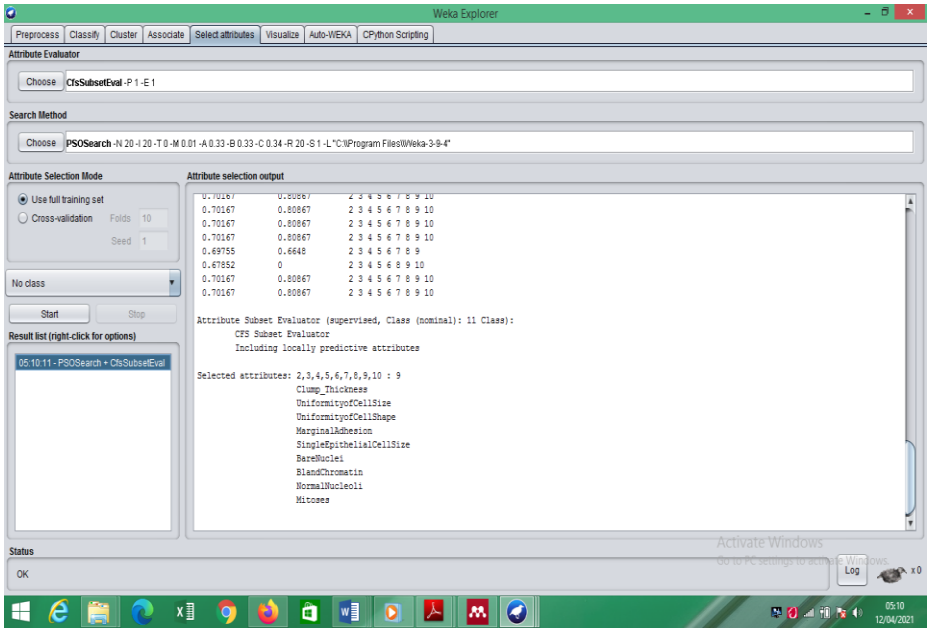

Figure 10: Sample of selected features by PSO Algorithm

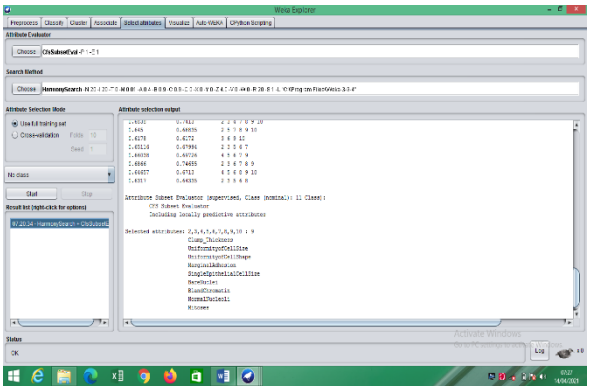

Figure 11: Sample of selected features by HarmonySearch Algorithm

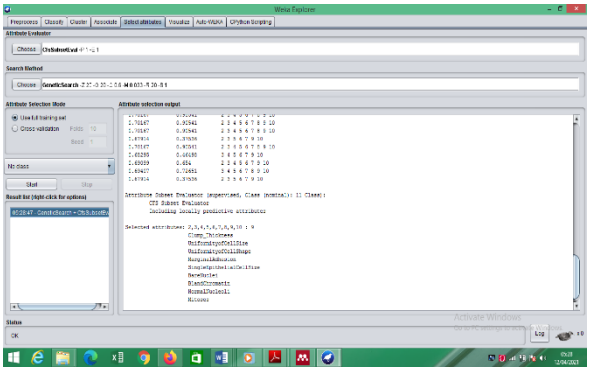

Figure 13: Sample of selected features by Genetic Algorithm

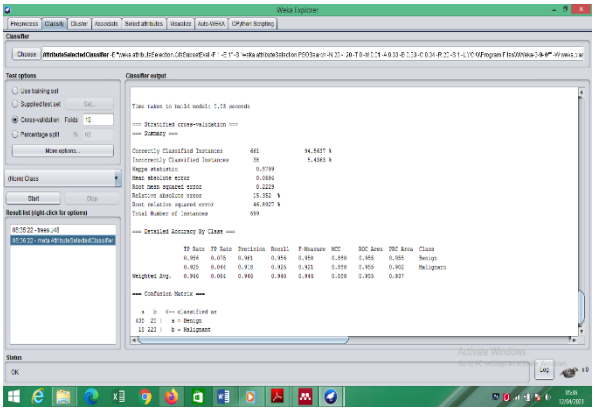

Figure 12: Sample of selected features by Tabu-Search Algorithm

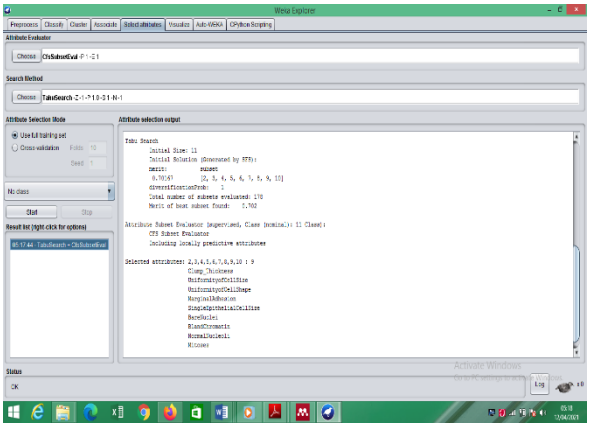

Figure 15: Sample Result for PSO Algorithm Using C4.5 Decision Tree (J48

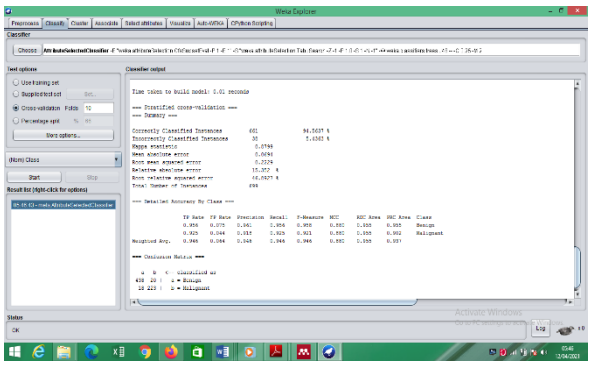

Figure 17: Sample Result for Tabu-Search Algorithm Using C4.5 Decision Tree (J48)

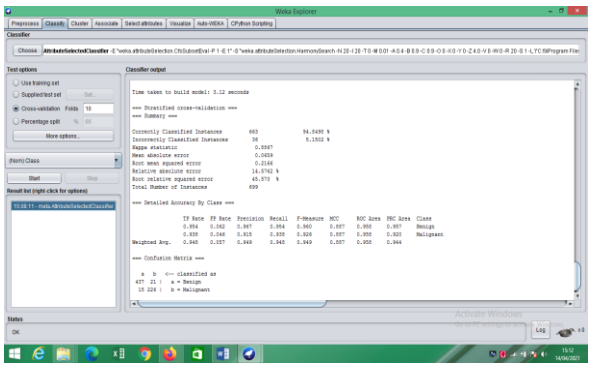

Figure 18: Sample Result for HarmonySearch Algorithm Using C4.5 Decision Tree (J48)

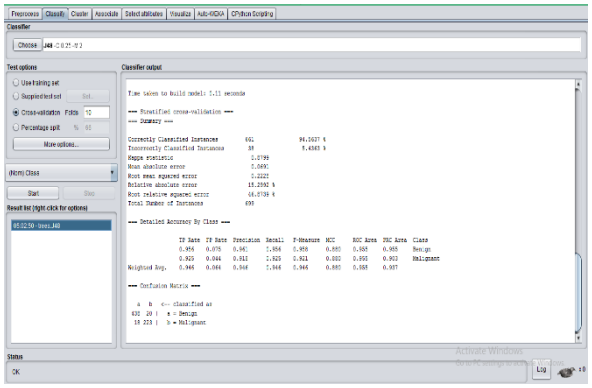

Figure 14: Sample Result for no Feature Selection Algorithm Using C4.5 Decision Tree (J48)

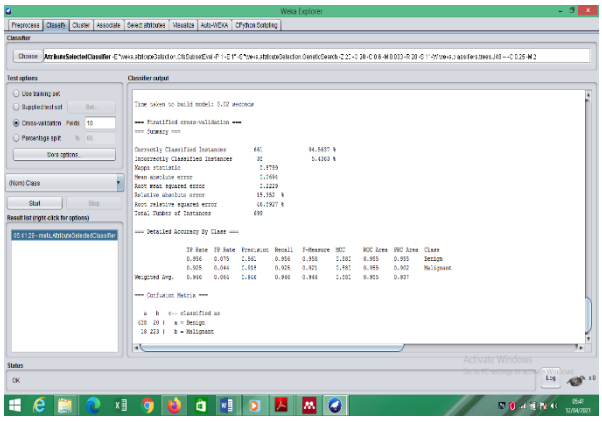

Figure 16: Sample Result for Genetic Algorithm Using C4.5 Decision Tree (J48)

Table 1: Performance Evaluation Metrics (With no Feature Selection Algorithm)

Classifier	Accuracy (%)	Precision	Recall	F1-Measure
C4.5	94.5637	0.9460	0.9460	0.9460
KNN	95.1359	0.9510	0.9510	0.9510
SVM	96.7096	0.9670	0.9670	0.9670
NB	95.9943	0.9620	0.9600	0.9600
NN	95.8512	0.9590	0.9590	0.9590
LR	96.5665	0.9660	0.9660	0.9660
RF	96.7096	0.9680	0.9670	0.9670

Table 2: Performance Evaluation Metrics (PSO Algorithm)

Classifier	Accuracy (%)	Precision	Recall	F1-Measure
C4.5	94.5637	0.9460	0.9460	0.9460
KNN	95.1359	0.9510	0.9510	0.9510
SVM	96.9957	0.9700	0.9700	0.9700
NB	95.9943	0.9620	0.9600	0.9600
NN	95.2790	0.9530	0.9530	0.9530
LR	96.5665	0.9660	0.9660	0.9660
RF	97.1388	0.9440	0.9750	0.9590

Table 3: Performance Evaluation Metrics (Genetic Algorithm)

Classifier	Accuracy (%)	Precision	Recall	F1-Measure	Kappa-Statistic	Time-Taken (s)
C4.5	94.5637	0.9460	0.9460	0.9460	0.8799	0.01
KNN	95.2790	0.9530	0.9530	0.9530	0.8952	0
SVM	96.9957	0.9700	0.9700	0.9700	0.9337	0.06
NB	95.9943	0.9620	0.9600	0.9600	0.9127	0.02
NN	95.2790	0.9530	0.9530	0.9530	0.8958	1.17
LR	96.5665	0.9660	0.9660	0.9660	0.9240	0.2

Table 4: Performance Evaluation Metrics (HarmonySearch Algorithm)

Classifier	Accuracy (%)	Precision	Recall	F1-Measure	Kappa-Statistic	Time-Taken (s)
C4.5	94.8498	0.9490	0.9480	0.9490	0.8867	3.12
KNN	95.4220	0.9540	0.9540	0.9540	0.8985	0.08
SVM	96.9957	0.9700	0.9700	0.9700	0.9338	1.19
NB	95.9943	0.9620	0.9600	0.9600	0.9127	0.08
NN	95.1359	0.9510	0.9510	0.9510	0.8926	1.1
LR	96.5665	0.9660	0.9660	0.9660	0.9240	0.21
RF	96.7096	0.9680	0.9670	0.9670	0.9278	0.5

Table 5: Table 4: Performance Evaluation Metrics (Tabu Search Algorithm)

Classifier	Accuracy (%)	Precision	Recall	F1-Measure	Kappa-Statistic	Time-Taken (s)
C4.5	94.5637	0.9460	0.9460	0.9460	0.8799	0
KNN	95.1359	0.9510	0.9510	0.9510	0.8919	0
SVM	96.9957	0.9700	0.9700	0.9700	0.9337	0
NB	95.9943	0.9620	0.9600	0.9600	0.9127	0
NN	95.2790	0.9600	0.9600	0.9600	0.9240	1.13
LR	96.5665	0.9660	0.9660	0.9660	0.9240	2.18
RF	97.1388	0.9720	0.9710	0.9710	0.9372	0.16

Table 6: Result of Comparative Analysis (Accuracy)

Classifier	Without Feature Selection	PSO	Genetic Algorithm	HS	Tabu-Search
C4.5	94.5637%	94.5637%	94.5637%	94.8498	94.5637%
KNN	95.1359%	95.1359%	95.2790%	95.4220	95.1359%
SVM	96.7076%	96.9957%	96.9957%	96.9957	96.9957%
NB	95.9943%	95.9943%	95.9943%	95.9943	95.9943%
NN	95.8512%	95.2700%	95.2700%	95.1359	95.2700%
LR	96.5665%	96.5665%	96.5665%	96.5665	96.5665%
RF	96.7076%	97.1388%	96.2804%	96.7096	97.1388%

Table 7: Result of Comparative Analysis (Precision)

Classifier	Without Feature Selection	PSO	Genetic Algorithm	HS	Tabu-Search
C4.5	0.9460	0.9460	0.9460	0.9490	0.9460
KNN	0.9510	0.9510	0.9530	0.9540	0.9510
SVM	0.9670	0.9700	0.9700	0.9700	0.9700
NB	0.9620	0.9620	0.9620	0.9620	0.9620
NN	0.9590	0.9530	0.9530	0.9510	0.9530
LR	0.9660	0.9660	0.9660	0.9660	0.9660
RF	0.9680	0.9440	0.9630	0.9680	0.9720

Table 8: Result of Comparative Analysis (Recall)

Classifier	Without Feature Selection	PSO	Genetic Algorithm	HS	Tabu-Search
C4.5	0.9460	0.9460	0.9460	0.9480	0.9460
KNN	0.9510	0.9510	0.9530	0.9540	0.9510
SVM	0.9670	0.9700	0.9700	0.9700	0.9700
NB	0.9600	0.9600	0.9600	0.9600	0.9600
NN	0.9590	0.9530	0.9530	0.9510	0.9530
LR	0.9660	0.9660	0.9660	0.9660	0.9660
RF	0.9670	0.9750	0.9630	0.9670	0.9710

Table 9: Result of Comparative Analysis (Kappa Statistic)

Classifier	Without Feature Selection	PSO	Genetic Algorithm	HS	Tabu-Search
C4.5	0.8799	0.8799	0.8799	0.8867	0.8799
KNN	0.8919	0.8919	0.8952	0.8985	0.8919
SVM	0.9274	0.9337	0.9337	0.9338	0.9337
NB	0.9127	0.9127	0.9127	0.9127	0.9127
NN	0.9086	0.8958	0.8958	0.8926	0.9240
LR	0.9240	0.9240	0.9240	0.9240	0.9240
RF	0.9278	0.9370	0.9127	0.9278	0.9372

Table 10: Result of Comparative Analysis (F1-Measure)

Classifier	Without Feature Selection	PSO	Genetic Algorithm	HS	Tabu-Search
C4.5	0.8799	0.8799	0.8799	0.9490	0.8799
KNN	0.8919	0.8919	0.8952	0.9540	0.8919
SVM	0.9274	0.9337	0.9337	0.9700	0.9337
NB	0.9127	0.9127	0.9127	0.9600	0.9127
NN	0.9086	0.8958	0.8958	0.9510	0.8958
LR	0.9240	0.9240	0.9240	0.9660	0.9240
RF	0.9278	0.9372	0.9177	0.9670	0.9372