Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 2551 - 2562
Received 15 April 2021; Accepted 05 May 2021.

Breast Cancer Detection Using Machine Learning Algorithms

Basker.N[1*], Theetchenya.S[1] , Vidyabharathi.D[1] , Dhaynithi.J[1],Mohanraj.G[1]
Marimuthu.M[1], Vidhya.G[1]

[1]Sona College of Technology, Department of CSE, Salem, Tamil Nadu
*bas2k9@gmail.com, theetchenya@gmail.com, dvbharathi77@gmail.com,
dhaya.j@gmail.commohanraj134@gmail.com,mari.btech@gmail.com, gvidhyacse@gmail.com

ABSTRACT

Breast cancer (BC) is one among the disease occur in women through the globe. Early Diagnosis of the cancer, on the other hand, will save lives. Radiologists can tell whether the mammography scans show cancer or not, but they can fail 15% of the time. We suggest a new approach for detecting breast cancer with high precision in this article. Data mining techniques had a major role to play in the initialstage diagnosis of breast cancer. We suggest an approach in this paper for improving the accuracy and efficiency of the classifiers Decision Tree (J48), Naive Bayes (NB), and Sequential Minimal Optimization (SMO).The proposed approach uses two benchmark datasets to test and compare the classifiers: Wisconsin Breast Cancer (WBC) and Breast Cancer dataset. Considering that, the probability of instances belonging to the majority class is significantly high; algorithms are far more likely to assign unique findings to the majority class during the classification process. In this paper, we discuss such a dilemma. We use the data-level methodology that involves data resampling to minimize the impact of class imbalance. 10fold cross-validation is used to assess the results. The outcome of the models such as Precision, Recall, ROC curve, Standard Deviation (STD), and accuracy are used to evaluate the performance. Experimentations reveal that applying a resample filter improves the accuracy of the classifier, with SMO outperforms other classifiers in the WBC dataset whereas J48 outperforms the rest in the Breast Cancer dataset.

Keywords

Breast Cancer(BC), Accuracy Measure, Naïve Bayes, J48, Sequential Minimal Optimization (SMO)

Introduction

BC is the primary reason of death among women in the world even with lot of advancements in technology. Women in the United States are predictable to be diagnosed with 268,600 new aggressive cases of BC and 62,930 new non-aggressive cases of BC. The easiest way to improve the chances of recovery and survival is to catch cancer early. With encouraging results, data mining has become a mainstream tool for information discovery in all domains like marketing, social science, economics, and medicine.Numerous machine learning techniques for BC classification and its prediction have been evolved over the last few decades [5–7].The process of classification isbasically categorized into three phases: pre-processing, Feature extraction, and classification. Preprocessing mammography films improves illumination of peripheral areas and strength distribution, which helps with perception and examination [8, 9]. Many approaches have been published to aid with this phase. Feature Extraction helps in the distinction of benign and malignant tumours,which is an important step in the diagnosis of breast cancer. Later, the properties of image such as unevenness, smoothness, regularity and depth are removed using segmentation [10].

The images are turned in to new form by using the pixel intensity differences and several transform-based texture analysis techniques .Wavelet transforms [11], FFT(Fast Fourier transform)[12], GT(Gabor Transforms) [13], and SVD(Singular Value Decomposition) [14] is some of the most commonly used techniques. PCA(Principal Component Analysis)[15] is used for dimensionality reduction of feature representations.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 2551 - 2562
Received 15 April 2021; Accepted 05 May 2021.

Many studies have tried to use machine learning algorithms (Maarlin&Marimuthu et al.)to automate breast cancer detection. Malek et al. [16], for example, suggested a system that combines fuzzy logic and wavelet feature extraction . Sun et al. [17] explored the issue by contrasting the function selection approaches. Zheng et al. [18] used a K-means and SVM to diagnose breast cancer. Several studies were done on clustering and grouping [7]. Alikovi and Subasi [19] proposed a genetic algorithm for feature extraction and classifier.Bannaie performed research [20] using the dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) method, the related data are extracted. Also they focused more on preprocessing. Hyperparameters are those that cannot be planned straight from the data, according to Kuhn and Johnson [21]. To get the optimal value from an algorithm, certain model parameters must usually be tweaked. Since there is no statistical method for determining the appropriate learning rate in a neural network, and certain SVM parameters must be specified manually. As a result, any potential model's final tuning parameters are yet to be determined.

Machine Learning (ML) is now in such high demand that it is being offered as a commodity. Unfortunately, machine learning remains a high-barrier environment that always necessitates expert expertise. The phases preprocessing, feature identification, and classification requires good skills and experience to design an efficient ML model. In any of the proposed model, the methods and parameters used in the pre-processing and classification stages are spontaneously specified. The specialist in ML selects the best methodology for the current problem area. Non-machine learning researchers, on the other hand, devote a significant amount of time optimizing their proposed models and achieving the desired results.Multiple classifier algorithms have recently been introduced to medical datasets in order to do statistical processing on patients and their medical diagnoses. For instance, machine learning methods may be used to determine tumors activity in patients with breast cancer. One issue is that the training data has a class imbalance, with the likelihood of not developing this disorder being greater than the probability of having it. This paper compares the precision of three distinct classifiers: J48, NB, and SMO when it comes to detecting breast cancer. Our goal is to improve the classifier's output by preparing the dataset by recommending a suitable approach for managing the imbalanced dataset and missing values.

The main goal of this paper is to suggest a simple approach for detecting BC. This paper examines existing cancer detection models in depth and reports on the exceptionally reliable and effective outcomes. The paper is structured into four parts. Section 2 presents the literature and recent works. The suggested approach is detailed in Section 3. Section 4 presents the findings and discussions. When compared to other models, the presented findings have proven to be reliable and effective.

Literature Review

Several researchers have used ML(Machine Learning) algorithms on various healthcare databases to identify BC in recent years. The outcome of the algorithms is good, which encouraged many researchersto use them to solve difficult problems. With an accuracy of nearly 88%, a CNN was used to predict and diagnose the invasive ductal carcinoma in BC photos. Furthermore, it is commonly used in the health care community to forecast and diagnose abnormal occurrences in order to get a greater understanding of conditions that are incurable, like cancer. Table 1 contains a set of several studies relevant to this procedure.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 2551 - 2562
Received 15 April 2021; Accepted 05 May 2021.

Table 1. Comparison of the various ML algorithms used for BC

Title of the paper	**Datasets**	**Algorithm Used**	**Observation**
Silva J et al [3]	Breast Cancer(BC)	GRNN,J48 ,NB, SVM classifiers	Accuracy for GRNN , J48 is 91% NB & SVM: 89%
Ojha Uet al[4]	WPBC	Classification: KNN, SVM, NB and C5.0, Clustering: K-means, EM, PAM and Fuzzymeans	Classification Accuracy is superior than clustering,
A. J. Cruz et al[5]	WPBM	NB, C4.5, SVM	Accuracy for NB is 67.17%, C4.5 is 73.73%, and SVM is 75.75%
G. Valvanoet al[6]	WBC	KNN,NB,SVM and C4.5 classifiers	SVM performs better than the other classifiers and the accuracy is 97.13% SVM is superior to NB and Ensemble methods.
M. F. Akayet al[7]	WDBC	NB, SVM and Ensemble methods	Accuracy for SVM is98.5%, Accuracy for NB and Ensemble is 97.3%
D. NarainPonrajet al[8]	WDBC	NB, J48	Accuracy for NB is 97.51%, Accuracy for J48is 96.5%
A. P. Charateet al[9]	Breast Cancer(BC)	J48, MLP and Rough set	Accuracy for J48 is 79.97%, Accuracy for MLP is 75.35%, Accuracy for Rough set is 71.36%
P. Salembieret al[10]	WBC	SMO, IBK and BF Tree	Accuracy for SMO is 96.19%, Accuracy for IBK: 95.90%, Accuracy for BF Tree is 95.46%

In all the above comparison of algorithms only the methods are used with the parameter specifications. It gives better results. To improve the BCprediction we use resample filters

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 2551 - 2562
Received 15 April 2021; Accepted 05 May 2021.

repeatedly in our proposed work.

Methodology

The datasets consideredin this study are prone to the missing values and imbalanced data, a significant portion of the analysis is done in pre-processing to improve the efficiency of the classifier. During preprocessing, the missing values and imbalanced data will be managed.Instances with missing values are omitted to handle the missing attributes. The training data balance must be adjusted to solve the imbalance issue. The data is rebalanced artificially by using the resample filter.After that, 10fold cross validation is used, followed by a comparison of these three classifiers. The detailed explanation of the steps involved in the training phase is given in the next subsections and the process is depicted in the Fig.1.

Preprocessing Phase

Initially the discretize filter is used to discretize the data and then it removes the missing values in the dataset. In order to sustain the distribution of class in the subsample and bias to attain uniform distribution, the resample filter is used to resample the instances. After that 10fold cross validation was applied. And then experimentation is done by using the NB, SMO and J48 classifiers that is explained in Figure.1

Training & Classification phase

After the preprocessing step, 10fold cross validation is used to reduce the bias occurred during random sampling in the training outcomes. The dataset is uniformly divided into K equivalent subsets and K times the model is learned and validated, by using k-fold cross validation. For everyiteration, one subset is considered as validation data for evaluation of model while the left behind k1 subgroups are considered as training data. The algorithms used are: a DT based J48 algorithm, SMO andNB. The NB is completely a Bayes rule-based probabilistic classifier. It works by calculating the probability of each and every class for which it checks whether a given instance is a member or not. The formula is

$$P(c|x) = \left(\frac{P(x|c)P(c)}{P(x)}\right) \qquad (1)$$

whereP(c|x) represents posterior probability, P(c) represents the class prior probability, P(x|c) represents the likelihood and P(x) represents the predictor prior probability.

The J48 algorithm operates by dividing every single data attribute into reduced datasets mainly to analyze differences in entropy. It's a better and more advanced clone of C4.5.Consider, X denotes attribute, P denotes the element and j denotes the position of element X. Then the entropy is calculated using the formula given below.

$$Y(X) = \sum_{j=1}^{k} P_j \log_2 \frac{1}{P_j} \qquad (2)$$

If the obtained value Y(X) is larger, it means that the X is more random and smaller means less random.For training a help vector classifier, the SMO uses Platt's sequential minimal optimization algorithm. This implementation removes all the missing values entirely and the

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 2551 - 2562
Received 15 April 2021; Accepted 05 May 2021.

nominal attributes are converted to binary attributes.

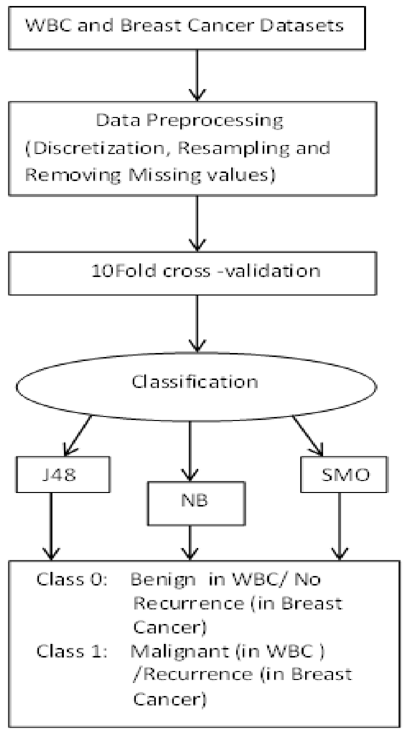

Figure 1. Steps involved in processing

Basically, it also normalizes all attributes. Let us assume the binary classification problem,

$$(x_1, y_1), ..., (x_n, y_n),$$

where x_i - input vector , $y_i \in \{-1, 1\}$ is a binary label. The quadratic programming problem is solved using SVM, is given as:

$$\max_{\propto} \sum_{i=1}^{n} \propto_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j K(x_i, x_j) \propto_i \propto_j \quad (3)$$

With respect to :

$$0 \leq \alpha_i \leq C \text{ fori-=}\{1,2.3.4\ldots\ldots n\}$$

$$\sum_{i=1}^{n} y_i \alpha_i = 0 \quad (4)$$

where C denotes SVM parameter and $K(x_i, x_j)$ denotes the kernel function.

Performance Evaluation Criteria

To test all of the classifiers in this analysis, we used five output measures:ROC curve, Standard Deviation(SD), and precision, Recall and F-measure. The formula for Accuracy, Precision , Recall and F1 measure is:

$$Accuracy = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (5)$$

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 2551 - 2562
Received 15 April 2021; Accepted 05 May 2021.

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1\ measure = 2 \times \frac{precision \times recall}{precision + recall} \quad (8)$$

where TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative. The performance measure values are shown for the Breast cancer dataset and WBC dataset by using the three ML algorithms J48, Naïve Bayes and SMO are given in table 2 and table 4.

Experimental Results

First, without using any pre-processing methods, the classification algorithms are tested on the WBC and BC datasets. The highest results were obtained in the BC dataset tested by J48 algorithm is 75.52%, and in WBC dataset tested by SMO algorithm is 96.99 %.Following that, accuracy improves to 98.20 % in J48 with respect to BC dataset and 99.56 % in SMO with respect to WBC dataset after applying pre-processing techniques.

Datasets

The datasets experimented in the proposed work are available at the University of California, Irvine (UCI) Machine Learning Repository.

WBC Dataset

There are 699 instances and 11 attributes in the WBC dataset, including 458 benign and 241 malignant occurrences. For almost 16 records present in the WBC, the significance of the attribute (Bare Nuclei) status is not there. As a result, data pre-processing is critical for this dataset, as it requires us to handle both uneven data and lost values.

Breast Cancer Dataset

A graphical image of a BCis taken to create the function in this dataset. The prognosis is recorded in the goal feature (Malignant or Benign). Itconsistsof 286 instances and 10 attributes.Also it has 201 non-recurrence events and 85 recurrence events. The values of some of the attribute status are missing in the eight documents of BC dataset.

Experimental Results with BC Dataset

At First, the proposed classifiers are put to the test on real-world results (without any pre-processing). The findings reveal that J48 has the highest accuracy of 75.52 %, while NB and SMO have accuracy of 71.67 % and 69.58%. Following that, a discretization filter is used to delete records with missing values, and the following is how the performance changed with the classifier. For J48, the accuracyis 74.82 %, NB is 75.53 %, and SMOis 72.66 %. The resample filter was then used seven times. As seen in Table 2, the classifiers' performance has increased and been strengthened.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 2551 - 2562
Received 15 April 2021; Accepted 05 May 2021.

Table 2. Accuracy for BC Dataset

Experiment Steps	J48 %	Naive Bayes %	SMO in %
Original without pre-processing	75.52	71.67	69.58
After discretization &eliminating missing values	74.82	75.53	72.66
Apply resample filter (1st iteration)	79.49	77.33	80.93
Apply resample filter (2nd iteration)	81.65	78.05	80.57
Apply resample filter (3rd iteration)	87.41	78.41	82.73
Apply resample filter (4th iteration)	92.08	77.69	88.84
Apply resample filter (5th iteration)	95.68	79.13	91.72
Apply resample filter (6th iteration)	97.48	79.85	95.68
Apply resample filter (7th iteration)	98.20	76.61	95.32

From Table 2 that the more resample filters we use, the better the accuracy. This is because of the unbalanced data and there by applying filter helps to sustain the class distribution. With a score of 98.20%, J48 outperforms the competition in the BC dataset. The accuracy metrics of J48 classifier along with the Roc curve are represented in the Figure 2.

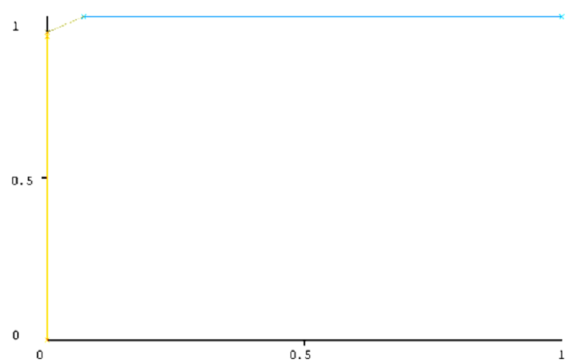

Figure 2.ROC curve for J48 classifier (Breast Cancer Dataset)

We equate the obtained findings with the analysis proposed in [9] to assess the success of the proposed model. The model's output is evaluated using the J48 algorithm. According to the findings, the proposed model attains better precision as compared to other classifiers. It is due to the use of resample filter in preprocessing rather than the feature selection strategy employed in [9], as seen in Table 3.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 2551 - 2562
Received 15 April 2021; Accepted 05 May 2021.

Table 3. Performance measure values for J48, Naïve Bayes and SMO for the BC dataset

ML Classifier	Precision	Recall	F-Measure	ROC Curve	STD
J48	0.9358	0.9572	0.9611	0.986	0.2220
Naïve Bayes	0.8924	0.9011	0.9411	0.936	0.3542
SMO	0.9134	0.9281	0.9562	0.976	0.1254

Experiment Using the WBC Dataset

The WBC dataset was subjected to the same tests. Both algorithms have better classification accuracy when pre-processing techniques are used. The usage of resample filter many times increases the classification performance. The SMO classifier performance is measured as 99.56% when compared to NB which is 99.12 % and J48 which is 99.24 % .With 99.56% in the WBC dataset, SMO outperformed the competition. Table 6 shows the accuracy tests for the SMO classifier.

Table 4. Accuracy for WBC Dataset

Experiment Steps	J48 %	Naive Bayes %	SMO in %
Original without pre-processing	71.68	82.71	86.52
After discretization & eliminating missing values	73.11	85.72	84.55
Apply resample filter (1st iteration)	79.56	87.66	89.79
Apply resample filter (2nd iteration)	80.57	88.15	88.45
Apply resample filter (3rd iteration)	83.69	88.41	89.78
Apply resample filter (4th iteration)	89.33	87.34	94.58
Apply resample filter (5th iteration)	90.35	89.67	96.88
Apply resample filter (6th iteration)	94.38	89.91	97.98
Apply resample filter (7th iteration)	96.32	86.52	99.56

Table 5. Performance measure values for J48, Naïve Bayes and SMO for the WBC dataset

ML Classifier	Precision	Recall	F-Measure	ROC Curve	STD
J48	0.9358	0.9572	0.9611	0.986	0.2220

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 2551 - 2562
Received 15 April 2021; Accepted 05 May 2021.

Naïve Bayes	0.8924	0.9011	0.9411	0.936	0.3542
SMO	0.9134	0.9281	0.9562	0.976	0.1254

The performance measure for the WBC dataset is shown in Table 5.The SMO performs well for the dataset and has good precision and recall. The ROC curve for the WBC dataset by using the SMO classifier is depicted in Figure 3. Since our model uses pre-processing and resampling techniques, the efficiency of the SMO classifier is higher.

Figure 3.ROC curve for SOM classifier(WBC dataset)

Thus, similar to the other techniques in [6, 10], preprocessing and resampling techniques has major contribution in increasing SMO precision.

Conclusion

Breast cancer is one of the most prominent diseases which affect the livelihood of the women. So earlier detection helps to treat the disease and thereby increase the lifespan of the women. Various machine learning algorithms are used in BC diagnosis. In this article, we look at how to use resampling strategies to improve the classification accuracy by dealing with imbalanced data and missing values. The following algorithms J48, NB, and SMO are tested on BC and WBC datasets. The observation from the results shows that the usage of resample filter, at the time of preprocessing improves the accuracy of the classifier. In future, we can experiment the dataset with different classifiers and also different classifier on the same dataset to improve the performance significantly.

References

[1] U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2008 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and Human Services, Centers for Disease ControlGoogle Scholar

[2] http://www.breastcancer.org/symptoms/understand_bc/statistics

[3] Silva J., Lezama O.B.P., Varela N., Borrero L.A.: Integration of data mining classification techniques and ensemble learning for predicting the type of breast cancer recurrence. In:

Miani, R., Camargos, L., Zarpelão, B., Rosas, E., Pasquini, R. (eds.) GPC 2019. LNCS, vol. 11484, pp. 18–30. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-19223-5_2CrossRefGoogle Scholar

[4] Maarlin.R, Marimuthu.M, Dr.Sathyamoorthi.V, Theetchenya.S, Vidhya.G. (2020). A Combınatıon of BI-Clusterıng and Hybrıd Boostıng Algorıthm for Breast Tumor Classıfıcatıon. International Journal of Advanced Science and Technology, 29(7), 12175 - 12184.

[5] Ojha U., Goel, S.: A study on prediction of breast cancer recurrence using data mining techniques. In: 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, IEEE, pp. 527–530, 2017Google Scholar

[6] A. J. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," Cancer Informatics, vol. 2, pp. 59–77, 2006.View at: Publisher Site | Google Scholar

[7] G. Valvano, G. Santini, N. Martini et al., "Convolutional neural networks for the segmentation of microcalcification in mammography imaging," Journal of Healthcare Engineering, vol. 2019, Article ID 9360941, 9 pages, 2019.View at: Publisher Site | Google Scholar

[8] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," Expert Systems with Applications, vol. 36, no. 2, pp. 3240–3247, 2009.View at: Publisher Site | Google Scholar

[9] D. NarainPonraj, M. Evangelin Jenifer, P. Poongodi, and J. Samuel Manoharan, "A survey of the preprocessing techniques of mammogram for the detection of breast cancer," Journal of Emerging Trends in Computing and Information Sciences, vol. 2, no. 12, pp. 656–664, 2011.View at: Google Scholar

[10] A. P. Charate and S. B. Jamge, "The preprocessing methods of mammogram images for breast cancer detection," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 5, no. 1, pp. 261–264, 2017.View at: Google Scholar

[11] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," IEEE Transactions on Image Processing, vol. 9, no. 4, pp. 561–576, 2000.View at: Publisher Site | Google Scholar

[12] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," IEEE Transactions on Image Processing, vol. 1, no. 2, pp. 205–220, 1992.View at: Publisher Site | Google Scholar

[13] G. Carter, C. Knapp, and A. Nuttall, "Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing," IEEE Transactions on Audio and Electroacoustics, vol. 21, no. 4, pp. 337–344, 1973.View at: Publisher Site | Google Scholar

[14] A. Teuner and B. J. Hosticka, "Adaptive Gabor transformation for image processing," IEEE Transactions on Image Processing, vol. 2, no. 1, pp. 112–117, 1993.View at: Publisher Site | Google Scholar

[15] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Borjesson, “OFDM channel estimation by singular value decomposition,” IEEE Transactions on Communications, vol. 46, no. 7, pp. 931–939, 1998.View at: Publisher Site | Google Scholar

[16] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, “Two-dimensional PCA: a new approach to appearance-based face representation and recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 1, pp. 131–137, 2004.View at: Publisher Site | Google Scholar

[17] J. Malek, A. Sebri, S. Mabrouk, K. Torki, and R. Tourki, “Automated breast cancer diagnosis based on GVF-snake segmentation, wavelet features extraction and fuzzy classification,” Journal of Signal Processing Systems, vol. 55, no. 1–3, pp. 49–66, 2009.View at: Publisher Site | Google Scholar

[18] Y. Sun, C. F. Babbs, and E. J. Delp, “A comparison of feature selection methods for the etection of breast cancers in mammograms: adaptive sequential floating search vs. genetic algorithm,” in Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, pp. 6532–6535, Shanghai, China, September 2005.View at: Publisher Site | Google Scholar

[19] B. Zheng, S. W. Yoon, and S. S. Lam, “Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms,” Expert Systems with Applications, vol. 41, no. 4, pp. 1476–1482, 2014.View at: Publisher Site | Google Scholar

[20] E. Aličković and A. Subasi, “Breast cancer diagnosis using GA feature selection and Rotation Forest,” Neural Computing and Applications, vol. 28, no. 4, pp. 753–763, 2017.View at: Publisher Site | Google Scholar

[21] M. Banaie, H. Soltanian-Zadeh, H.-R. Saligheh-Rad, and M. Gity, “Spatiotemporal features of DCE-MRI for breast cancer diagnosis,” Computer Methods and Programs in Biomedicine, vol. 155, pp. 153–164, 2018.View at: Publisher Site | Google Scholar

[22] M. Kuhn and K. Johnson, Applied Predictive Modeling, Springer, New York, NY, USA, 2013.

[23] S. Bouaziz, H. Dhahri, A. M. Alimi, and A. Abraham, “Evolving flexible beta basis function neural tree using extended genetic programming & hybrid artificial bee colony,” Applied Soft Computing, vol. 47, pp. 653–668, 2016.View at: Publisher Site | Google Scholar

[24] K. P. Bennett, “Decision tree construction via linear programming,” in Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society, pp. 97–101, Utica, IL, USA, 1992.View at: Google Scholar

[25] J. M. Dixon, T. J. Anderson, J. Lamb, S. J. Nixon, and A. P. M. Forrest, “Fine needle aspiration cytology, in relationships to clinical examination and mammography in the diagnosis of a solid breast mass,” British Journal of Surgery, vol. 71, no. 8, pp. 593–596, 1984.View at: Publisher Site | Google Scholar

[26] Pritom, A.I., Munshi, M.A.R., Sabab, S.A., Shihab, S.: Predicting breast cancer

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 2551 - 2562
Received 15 April 2021; Accepted 05 May 2021.

recurrence using effective classification and feature selection technique. In: 19th International Conference on Computer and Information Technology (ICCIT), pp. 310–314. IEEE (2016)Google Scholar

[27] Asri, H., Mousannif, H., Al, M.H., Noel, T.: Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Comput. Sci. 83, 1064–1069 (2016)CrossRefGoogle Scholar

[28] Hazra, A., Mandal, S.K., Gupta, A.: Study and analysis of breast cancer cell detection using Naïve Bayes, SVM and ensemble Algorithms. Int. J. Comput. Appl. 145, 0975–8887 (2016)Google Scholar

[29] Rodrigues, B.L.: Analysis of the Wisconsin Breast Cancer dataset and machine learning for breast cancer detection. In: Proceedings of XI Workshop de Visão Computational, pp. 15–19 (2015)Google Scholar

[30] Saabith, A.L.S., Sundararajan, E., Bakar, A.A.: Comparative study on different classification techniques for breast cancer dataset. Int. J. Comput. Sc. Mob. Comput. 3(10), 185–191 (2014)Google Scholar

[31] Chaurasia, V., Pal, S.: A novel approach for breast cancer detection using data mining techniques. Int. J. Innovative Res. Comput. Commun. Eng. 2 (2017). (An ISO 3297: 2007 Certified Organization)Google Scholar

[32] Asraf Yasmin, B., Latha, R., & Manikandan, R. (2019). Implementation of Affective Knowledge for any Geo Location Based on Emotional Intelligence using GPS. International Journal of Innovative Technology and Exploring Engineering, 8(11S), 764–769. https://doi.org/10.35940/ijitee.k1134.09811s19

[33] Muruganantham Ponnusamy, Dr. A. Senthilkumar, & Dr.R.Manikandan. (2021). Detection of Selfish Nodes Through Reputation Model In Mobile Adhoc Network - MANET. Turkish Journal of Computer and Mathematics Education, 12(9), 2404–2410. https://turcomat.org/index.php/turkbilmat/article/view/3720