Breast Cancer Prediction using Feature Selection and Ensemble Voting

Quang H. Nguyen
School of Information & Communication Technology
Hanoi University of Science and Technology
Hanoi, Vietnam
quangnh@soict.hust.edu.vn

Trang T. T. Do
Institute for Infocomm Research
*A*STAR*
Singapore, Singapore
dotrang@alumni.nus.edu.sg

Yijing Wang
Institute of Systems Science
National University of Singapore
Singapore, Singapore
e0267665@u.nus.edu

Sin Swee Heng
Institute of Systems Science
National University of Singapore
Singapore, Singapore
e0267555@u.nus.edu

Kelly Chen
Institute of Systems Science
National University of Singapore
Singapore, Singapore
kelly.chen@u.nus.edu

Wei Hao Max Ang
Institute of Systems Science
National University of Singapore
Singapore, Singapore
e0267589@u.nus.edu

Conceicao Edwin Philip
Institute of Systems Science
National University of Singapore
Singapore, Singapore
e0269012@u.nus.edu

Misha Singh
Institute of Systems Science
National University of Singapore
Singapore, Singapore
misha.singh@u.nus.edu

Hung N. Pham
School of Information & Communication Technology
Hanoi University of Science and Technology
Hanoi, Vietnam
hungpn@soict.hust.edu.vn

Binh P. Nguyen
School of Mathematics and Statistics
Victoria University of Wellington
Wellington, New Zealand
binh.p.nguyen@vuw.ac.nz

Matthew C. H. Chua
Institute of Systems Science
National University of Singapore
Singapore, Singapore
mattchua@nus.edu.sg

***Abstract*—Breast cancer is the most common cause of cancer among women worldwide. This paper analyses the performance of supervised and unsupervised models for breast cancer classification. Data from Wisconsin Breast Cancer Dataset is used in this paper. Feature selection is processed through scaling and principal component analysis. Final results indicate that Ensemble Voting approach is ideal as a predictive model for breast cancer. The raw data has 569 cases of breast cancer. The data is split into training and testing sets in the ration 70:30, respectively. The benchmark model is then created using Random Forest method. Various models are trained and tested on the data after Feature Scaling and Principle Component Analysis. Cross-validation is performed which showed that our model is stable. Among all the evaluated models, only four models, i.e., Ensemble - Voting Classifier, Logistics Regression, SVM Tuning and AdaBoost returned with accuracy of at least 98%. Based on results of the precision and recall, ROC-AUC, F1-measure and computational time of the models, the Ensemble showed the most potential in breast cancer classification of the given dataset.**

Keywords — Breast Cancer, Ensemble Voting Classification, SVM, Random Forest, Perception, Logistics Regression, KNN, Stochastic Gradient Descent, XGBoost, Extremely Randomised Trees, AdaBoost

I. Introduction

Breast cancer is the most common cancer among women worldwide, claiming the lives of hundreds of thousands of women each year and affecting countries at all levels of modernization according to the World Health Organization [1]. If it is diagnosed in early stages, the chances of survival are higher [1]. The detection of the pattern of symptoms using data mining is a very important technique to correctly understand hidden patterns [2]. Data mining is a collection of techniques for efficient and automated discovery of previously unknown patterns in large databases [3]. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends [3]. Recent statistics of breast cancer in the United States of America is 12.4% in 2017 [4].

In this paper, we analyse the performance of supervised or unsupervised learning models such as Random Forest, KNN, SVM, Logistic Regression, AdaBoost, and Perceptron. These models are used for classifying the Wisconsin Breast Cancer Dataset (WBCD) from UCI Machine learning depository. They have been tested by many researchers and were proven efficient in many aspects. This paper seeks to evaluate the

978-1-7281-0525-3/19/$31.00 ©2019 IEEE

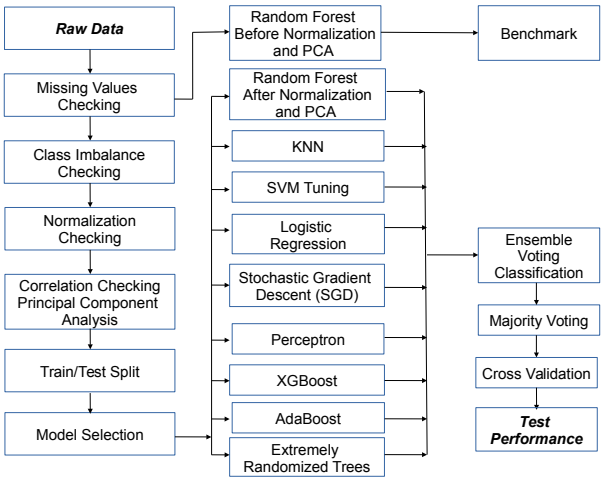

Fig. 1. Flow diagram

breast cancer prediction through feature selections using the ensemble voting classification method from various models.

II. Materials and Methods

A. Data preparation

The data used is the Wisconsin Breast Cancer Dataset (WBCD) taken from the UCI machine learning repository. The dataset contained 569 instances taken from needle aspirates from patients' breasts, of which 357 cases were identified as "benign" and the remaining 212 cases were classified as "malignant".

The dataset was analysed and pre-processed with the following steps below before building the models (Figure 1).

1) Missing Value Checking

This check ensures that the conclusion of the model is not affected by missing values within the dataset.

2) Class Imbalance Checking

In the dataset, the ratio between the two classes, Benign (B) = 0 and Malignant (M) = 1, is 63:37, respectively. This reveal a close gap of 0.26, which does not warrant imbalanced data treatment.

3) Normalization Checking

All variables should have the same scale for fair comparison between them. Based on Figure 2, there is evident difference in scale for each variable. Therefore, feature scaling is needed to ensure fair treatment.

4) Correlation Checking

A correlation matrix was done to check features that showed highly-correlation that will inadvertently affect the model in similar ways. For our dataset, based on the correlation matrix (Figure 3), values above 0.7 indicate a high correlation between each two features which will bias the result. In this situation, principal component analysis needs to be performed.

5) Train/Test Split

Train the baseline models and the ensemble model on the training and test sets, which are in the ratio of 7:3, respectively.

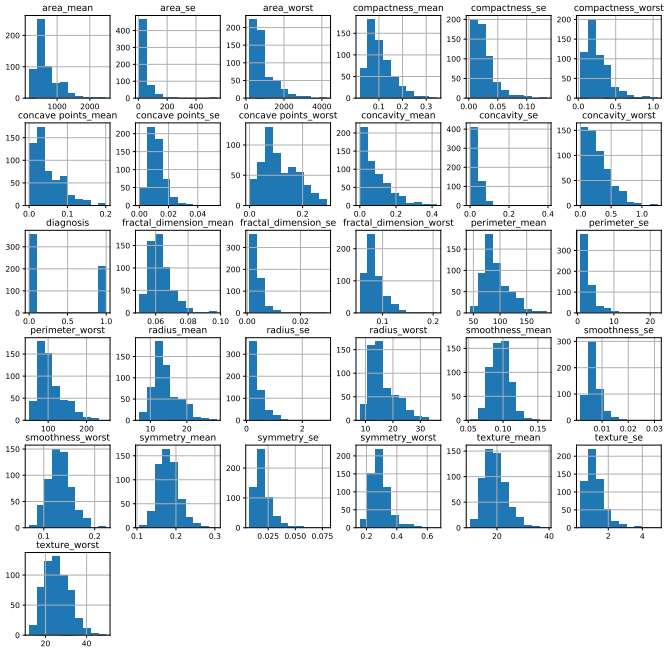

Fig. 2. Normalization charts

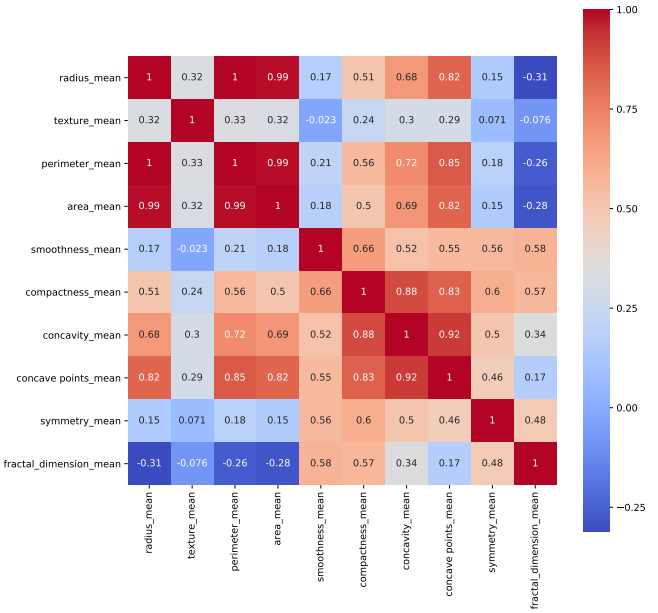

Fig. 3. Correlation table on multi-features

B. Methods

To ensure proper rigour, objectivity and generalizability of the solutions, due importance was placed on the key modelling aspects as detailed below.

1) Benchmark Model – Random Forest (RF): Deterministic approach of generating multiple variations of the decision tree, through optimal splitting. Decision tree is a hierarchical model with rules that split variables into homogeneous zones, this means it generates an internal unbiased estimate of the

	Before Scaling				After Scaling			
Feature	Min	Mean	Max	Std.Dev	Min	Mean	Max	Std.Dev
Radius_mean	6.9810	14.1273	28.1100	3.5240	-2.0296	-1.3700E-16	3.9713	1.0009
Texture_mean	9.7100	19.2896	39.2800	4.3010	-2.2292	6.8700E-17	4.6519	1.0009
Perimeter_mean	43.7900	91.9690	188.5000	24.2990	-1.9845	-1.2500E-16	3.9761	1.0009
Area_mean	143.5000	654.8891	2501.0000	351.9141	-1.4544	-2.1900E-16	5.2505	1.0009
Smoothness_mean	0.0526	0.0964	0.1634	0.0141	-3.1121	1.7500E-16	4.7709	1.0009
Compactness_mean	0.0194	0.1043	0.3454	0.0528	-1.6101	2.0000E-16	4.5684	1.0009
Concavity_mean	0.0000	0.0888	0.4268	0.0797	-1.1149	3.7500E-17	4.2436	1.0009
Concave points_mean	0.0000	0.0489	0.2012	0.0388	-1.2618	-3.7500E-17	3.9279	1.0009
Symmetry_mean	0.1060	0.1812	0.3040	0.0274	-2.7441	1.8700E-16	4.4848	1.0009
Fractal_dimension_mean	0.0500	0.0628	0.0974	0.0071	-1.8199	4.5300E-16	4.9109	1.0009
Radius_se	0.1115	0.4052	2.8730	0.2773	-1.0599	2.5000E-16	8.9069	1.0009
Texture_se	0.3602	1.2169	4.8850	0.5516	-1.5543	-1.0300E-16	6.6553	1.0009
Perimeter_se	0.7570	2.8661	21.9800	2.0219	-1.0440	-3.5000E-16	9.4620	1.0009
Area_se	6.8020	40.3371	542.2000	45.4910	-0.7378	-1.3100E-16	11.0418	1.0009
Smoothness_se	0.0017	0.0070	0.0311	0.0030	-1.7761	4.4000E-16	8.0300	1.0009
Compactness_se	0.0023	0.0255	0.1354	0.0179	-1.2981	1.8100E-16	6.1435	1.0009
Concavity_se	0.0000	0.0319	0.3960	0.0302	-1.0575	1.6200E-16	12.0727	1.0009
Concave points_se	0.0000	0.0118	0.0528	0.0062	-1.9134	1.2500E-17	6.6496	1.0009
Symmetry_se	0.0079	0.0205	0.0790	0.0083	-1.5329	8.1200E-17	7.0719	1.0009
Fractal_dimension_se	0.0009	0.0038	0.0298	0.0026	-1.0970	6.2400E-18	9.8516	1.0009
Radius_worst	7.9300	16.2692	36.0400	4.8332	-1.7269	-8.2400E-16	4.0942	1.0009
Texture_worst	12.0200	25.6772	49.5400	6.1463	-2.2240	1.2500E-17	3.8859	1.0009
Perimeter_worst	50.4100	107.2612	251.2000	33.6025	-1.6934	-3.7500E-16	4.2873	1.0009
Area_worst	185.2000	880.5831	4254.0000	569.3570	-1.2224	0.0000E+00	5.9302	1.0009
Smoothness_worst	0.0712	0.1324	0.2226	0.0228	-2.6827	-2.3700E-16	3.9554	1.0009
Compactness_worst	0.0273	0.2543	1.0580	0.1573	-1.4439	-3.3700E-16	5.1129	1.0009
Concavity_worst	0.0000	0.2722	1.2520	0.2086	-1.3058	7.4900E-17	4.7007	1.0009
Concave points_worst	0.0000	0.1146	0.2910	0.0657	-1.7451	2.2500E-16	2.6859	1.0009
Symmetry_worst	0.1565	0.2901	0.6638	0.0619	-2.1610	2.7500E-16	6.0460	1.0009
Fractal_dimension_worst	0.0550	0.0839	0.2075	0.0181	-1.6018	2.1200E-16	6.8469	1.0009
Diagnosis(Target Variable)	0.0000	0.3726	1.0000	0.4839	0.0000	0.3726	1.0000	0.4839

Fig. 4. Descriptive statistics of original data provided

TABLE I
PCA TABLE. NC: NUMBER OF COMPONENTS, SP: SCREEN PLOT, EV: EXPLAINED VARIANCE, CEV: CUMULATIVE EXPLAINED VARIANCE

NC	SP	EV	CEV
1	12.70	0.44	43.71
2	5.37	0.18	62.18
3	2.82	0.10	71.9
4	1.98	0.07	78.72
5	1.65	0.06	84.4
6	1.21	0.04	88.56
7	0.67	0.02	90.85
8	0.48	0.02	92.49
9	0.40	0.01	93.85
10	0.35	0.01	95.04

generalization error as the forest building progresses and does not require cross-validation [5]. A random forest is a type of meta estimator which fits several decision tree classifiers on different dataset samples and average it to raise the predictive accuracy and reduce over-fitting. The sub-sample trees are similar to original input sample size but are trained with bootstrap of the training set.

2) Feature Scaling: Wide range of values of raw data will affect the result of PCA and final model. For example, if one of the features has a broad range of values, the distance will be governed by this feature. Hence, scaling was performed so that each feature contributes approximately proportionately to the final distance (Figure 4).

3) Principal Component Analysis: Principal Component Analysis (PCA) is a feature extraction technique which takes an orthogonal transformation to convert a set of observations of possibly correlated parameters into a set of values of linearly uncorrelated parameters called principal components [6].

4) Modeling: Based on the Eigenvalue Criteria and Cumulative Explained Variance Criteria, eigenvalue components greater than 1 (Table I) are retained. Hence, we extract 6 components account for 88.56% variance of the raw variables which is reasonable and acceptable. Subsequently, we use these 6 components as our new features to build the model for analysis.

1. K-nearest neighbours algorithm

Classification of predictors according to cluster of similar behaviour. This is a form of optimization that seek to find the nearest point to a target variable point [7].

2. Support Vector Machines

Support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. SVMs have been successfully applied to several real-world problems, such as handwritten character and digit recognition, face recognition, text categorization and object detection in machine vision. SVMs has several advanced properties, including the ability to handle large feature space, effective avoidance of over fitting, and information condensing for the given data set.

Each kind of classifier needs a metric to measure the similarity or distance between patterns. SVMs classifier uses inner product as metric. If there are dependent relationships among pattern's attributes, such information will be accommodated through additional dimensions, and this can be realized by a mapping. In SVMs, the above is realized through kernel function [3].

$$K(x, y) =< \phi(x), \phi(y) > \tag{1}$$

Among different kernel functions, linear, polynomial, and radial basis function kernels are the most widely used and compared in various domain problems [8].

3. Logistic Regression

Probabilistic odds of an event is a linear combination of independent variables against a dichotomous (binary) dependent variable. Goodness of fit is assessed through Hosmer-Lemeshow test [9] [10].

4. Stochastic Gradient Descent

Gradient descent is a method of optimization and stochastic gradient descent (SGD) is an incremental gradient descent for finding the minimum of function. Stochastic is an approximation of gradient descent optimization. SGD is not used for Gated Recurrent Unit-Support Vector Machine (GRU-SVM), Nearest-Neighbour and SVM [7].

5. Perceptron

Perceptron is a neural network that decides if an input belongs to a specific class, based on weighted feature vector. Multilayer perceptron consists of layers of nodes that enable approximation of functions [7], [11].

$$\sum_{m}^{i=1} \omega_i x_i \quad f(x) = \begin{cases} 1 & if \quad w.x + b > 0 \\ 0 & otherwise \end{cases} \tag{2}$$

6. AdaBoost

Boosting is an ensemble method that start on a base classifier from the training dataset. AdaBoost is a boosting ensemble method which is building on up weighted classification [12].

7. XGBoost

XGBoost is a type of gradient tree boosting which allows for regularization, in order to avoid overfitting. Tianqi Chen, one of the developers for XGBoost, explained that AdaBoost

and XGBoost are two different way of boosting. XGBoost works for generic loss function, while AdaBoost is derived for classification with exponential loss. Loss or cost function mapped real number to an event, thereby representing "cost" associated with the event. Loss represents "price paid" for inaccurate prediction in classification problems.

Differences are in the regularization (introduction of additional information): AdaBoost is a meta-estimator while XGBoost has more customizable parameters. Both can be used in ensemble to be cross-validated. XGBoost is faster and can handle large amount of data.

8. Extremely Randomised Trees (ERT)

Non-deterministic approach of splitting tree, i.e. variable splitting value are done at random. It can be considered as a more randomized version of random forest. ERT is trained with complete training set.

9. Ensemble Voting Classification

Voting Classification is a selection method used to vote the models according to accuracy score, with highest ranking accorded to model with the highest accuracy score [13]. The majority voting approach allows the selection of models that can provide the best accuracy score [14]. The formula for the ensemble voting classification is written as $class(x) = argmax(\sum k(f_k(x), c_i))$.

Majority voting-based ensemble classifier is a simple approach. In this approach, each classifier assigns equal weight, which represents all classifiers equally.

C. Model Engineering Cross Validation

To minimise the out-of sample error and improve the precision of the model, Cross-Validation is performed on the training and testing data by using multiple tests sets and average the out-of-sample errors [15]. The formula is:

$$CV(\lambda) = \frac{1}{K}\sum_{K}^{k=1} E_k(\lambda) \tag{3}$$

This is a common approach whereby the training and testing data are further randomly separated into folds that evenly stratified the train/test sets, either 10 folds. Every point in the sets of data will occur exactly once, results in identical training and test sets. One important feature of cross-validation is to ensure that each test set is randomly assigned to avoid systemic biases in the data; this in turn allows estimation for the out of sample error in the predictions.

Majority Voting/Hard Voting

Hard voting is the simplest case of majority voting, which is widely used in classification tasks. Here, we predict the class label $\widehat{y}$ via majority (plurality) voting of each classifier C_j:

$$\widehat{y} = mode(C_1(x), C_2(x), ..., C_m(x)) \tag{4}$$

Assuming that we combine three classifiers that classify a training sample as follows: classifier 1 $\rightarrow$ class 0, classifier 2 $\rightarrow$ class 0, classifier 3 $\rightarrow$ class 1, $\widehat{y} = mode(0, 0, 1) = 0$. Via majority vote, we would classify the sample as "class 0".

TABLE II
PREDICTION RESULTS: ACCURACY (ACC), PRECISION (PREC), RECALL, F1 SCORE, ROC, RUN TIME IN SECONDS OF MODELS (1. SVM TUNING, 2. LOGISTIC REGRESSION, 3. ADABOOST, 4. STOCHASTIC GRADIENT DESCENT, 5. PERCEPTRON, 6. XGBOOST, 7. RANDOM FOREST, (AFTER SCALING AND PCA), 8. BENCHMARK (RF BEFORE SCALING AND PCA, 9. KNN (K-NEAREST NEIGHBOR), 10. EXTREMELY RANDOMIZED TREES, 11. ENSEMBLE VOTING CLASSIFIER)

Model	Acc	Prec	Recall	F1	ROC	Run time
1	98.83	0.99	0.99	0.99	0.9844	0.002389
2	98.83	0.99	0.99	0.99	0.9844	0.001517
3	98.25	0.98	0.98	0.98	0.9797	0.145240
4	97.66	0.98	0.98	0.98	0.9750	0.001662
5	97.08	0.97	0.97	0.97	0.9673	0.001244
6	96.49	0.96	0.96	0.96	0.9625	0.020180
7	95.91	0.96	0.96	0.96	0.9516	0.019644
8	95.32	0.95	0.95	0.95	0.9266	0.022151
9	95.32	0.96	0.95	0.95	0.9375	0.000758
10	94.74	0.95	0.95	0.95	0.9340	0.022392
11	98.83	0.99	0.99	0.99	0.9844	0.202470

III. RESULTS AND DISCUSSION

The results are in Table II. We select the top 4 models with above 98% prediction accuracy for comparison.

- Ensemble - Voting Classifier
- Logistics Regression
- SVM Tuning
- AdaBoost

Accuracy: Ensemble - voting classifier, SVM tuning and logistics regression have the same prediction accuracy of 98.83% while AdaBoost has a predicted score of 98.25%. This study shows ensemble and standalone model can produce similar prediction results. From the results of Precision and Recall, it is difficult to distinguish which model is the best classifier except AdaBoost.

ROC-AUC: Ensemble - voting classifier, SVM tuning and logistics regression have the same ROC-AUC of 0.9844 while AdaBoost has ROC-AUC score of 0.9797. From the results of ROC-AUC, it is difficult to distinguish which model is the best classifier except AdaBoost.

F1-Measure: In this F-measure, we are only interested in the correctly predicted true positives, whereas the true negatives are not considered, the F-score is calculated using weighted average of precision and recall, given by this formula.

$$F1 = \frac{2}{\frac{1}{recall} + \frac{1}{precision}} = \frac{2 \times precision \times recall}{precision + recall} \tag{5}$$

Ensemble - voting classifier, SVM tuning and logistics regression have the same F1 score of 0.99 while AdaBoost has F1 score of 0.98.

Computational Time: When compare with the computational time, the ensemble - voting classifier has the longest training time of 2.02×10^{-1} seconds as it requires to predict the majority votes among the 7 models, while the logistics regression is 1.51×10^{-3} seconds, SVM tuning is 2.38×10^{-3}

seconds and AdaBoost is 1.45×10^{-1} seconds. In this paper, our dataset is small, therefore the effect of the training time is insignificant. More significant impact on training time will be observed when the size of the dataset is increased for future studies. From the observations, it can be seen that logistics regression is the most efficient machine learning algorithm. However, we may need to further evaluate the results with a larger dataset. It is highly recommended for tuning model such as SVM to optimize the hyper parameters setting before building into the baseline of the ensemble. This is more computationally efficient than training it directly from the baseline model as this will increase the computational time during the majority voting.

IV. Conclusions

Final model: Prior to the training of models, the dataset has to be pre-processed and scaling has to be done so that each feature contributes approximately proportionately to the final distance. This is followed by PCA feature extraction to transform the set of values to linearly uncorrelated principal components. To minimize the overfitting problem, we used 10-fold cross validation technique which splits the data into 90% for training and 10% for validation test for each baseline model. The process is repeated a few times until we get a stable accuracy score from the average 10-fold. The number of k-fold is subjective to debate and the most widely use numbers are k = 5 or 10.

Limitation: The number of baseline models has to be in odd numbers and greater than one. XGBoost, Adaboost, SGD and SVM are considered black box models subject to acceptance among industrial practices and regulations.

Future work: We look forward to seeing how this ensemble model can be deployed in other fields of cancer prediction studies. This can be extremely beneficial to the clinical and healthcare sector in terms of reducing human-errors and false-positives in judgement and assessment, which can result in very costly diagnosis and invasive procedures. Additionally, we would like to determine if this model can be extended a wider class of patients in determining the relative risk of patients developing breast cancer, as well as include other features that would help improve the specificity and accuracy of the model.

In this study, multi-layer perceptron model is not included, therefore future works will include neural network models (MLFF BP) or deeper learning model (100+ layers) or any state-of-the-art models, such as the enhanced k-nearest neighbours [16], linear iterative clustering [17] and kernel dictionary learning [18]–[20], as the baseline model to improve the overall accuracy of the ensemble model.

References

[1] Centers For Disease Control and Prevention. Breast cancer facts. http://www.nationalbreastcancer.org/breast-cancer-facts, 2016.

[2] Ashutosh Kumar Dubey, Umesh Gupta, and Sonal Jain. Analysis of k-means clustering approach on the breast cancer Wisconsin dataset. *(International Journal of Computer Assisted Radiology and Surgery*, 11(11):2033–2047, 2016.

[3] Aruna S, Rajagopalan SP, and Nandakishore LV. Knowledge based analysis of various statistical tools in detecting breast cancer. *Computer Science and Information Technology*, pages 37–45, 2011.

[4] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432, 2017.

[5] Cheng-Min Chao, Ya-Wen Yu, Bor-Wen Cheng, and Yao-Lung Kuo. Construction the model on the breast cancer survival analysis use support vector machine, logistic regression and decision tree. *Journal of Medical Systems*, 38(10):106, 2014.

[6] Animesh Hazra, Subrata Kumar Mandal, and Amit Gupta. Study and analysis of breast cancer cell detection using naïve bayes, svm and ensemble algorithms. *International Journal of Computer Applications*, 145(2):0975–8887, 2016.

[7] Agarap AFM. On breast cancer detection: An application of machine learning algorithms on the Wisconsin diagnostic dataset. In *ICMLSC 2018*, Feb 2018.

[8] Min-Wei Huang, Chih-Wen Chen, et al. SVM and SVM ensembles in breast cancer prediction. *PloS One*, 12(1):e0161501, 2017.

[9] I Barco, M García Font, et al. A logistic regression model predicting high axillary tumour burden in early breast cancer patients. *Clinical and Translational Oncology*, 19(11):1393–1399, 2017.

[10] Rimi Murakami, Chiehfeng Chen, et al. Lovastatin lowers the risk of breast cancer: a population-based study using logistic regression with a random effects model. *SpringerPlus*, 5(1):1932, 2016.

[11] Siddhant Gada, Het Sheth, Meet Chheda, Ashwini Swain, and Kriti Srivastava. Triple-technique diagnosis using machine learned classifiers. *International Journal of Science and Research*, 6(10), 2017.

[12] Bala MS and Lakshmi GVR. Efficient ensemble classifiers for prediction of breast cancer. *International Journal of Advanced Research in Computer Science and Software Engineering*, 6(3), 2016.

[13] Sandra L Taylor and Kyoungmi Kim. A jackknife and voting classifier approach to feature selection and classification. *Cancer Informatics*, 10:CIN–S7111, 2011.

[14] Md Rafiul Hassan, M Maruf Hossain, James Bailey, Geoff Macintyre, Joshua WK Ho, and Kotagiri Ramamohanarao. A voting approach to identify a small number of highly predictive genes using multiple classifiers. *BMC Bioinformatics*, 10(1):S19, 2009.

[15] Nilam Fitriah, Sastra Kusuma Wijaya, et al. EEG channels reduction using PCA to increase XGBoost's accuracy for stroke detection. In *AIP Conference Proceedings*, volume 1862, page 030128, 2017.

[16] Binh P. Nguyen, Wei-Liang Tay, and Chee-Kong Chui. Robust biometric recognition from palm depth images for gloved hands. *IEEE Transactions on Human-Machine Systems*, 45(6):799–804, Dec 2015.

[17] Binh P. Nguyen, Hans Heemskerk, Peter T. C. So, and Lisa Tucker-Kellogg. Superpixel-based segmentation of muscle fibers in multi-channel microscopy. *BMC Systems Biology*, 10(Suppl 5:124):39–50, Dec 2016.

[18] Xuan Chen, Binh P. Nguyen, Chee-Kong Chui, and Sim-Heng Ong. Automated brain tumor segmentation using kernel dictionary learning and superpixel-level features. In *Proceedings of the International Conference on Systems, Man, and Cybernetics (SMC 2016)*, pages 2547–2552, Budapesh, Hungary, 9–12 Oct 2016. IEEE.

[19] Xuan Chen, Binh P. Nguyen, Chee-Kong Chui, and Sim-Heng Ong. Reworking multilabel brain tumor segmentation – an automated framework using structured kernel sparse representation. *IEEE Systems, Man, and Cybernetics Magazine*, 3(2):18–22, Apr 2017.

[20] Xuan Chen, Binh P. Nguyen, Chee-Kong Chui, and Sim-Heng Ong. An automatic framework for multi-label brain tumor segmentation based on kernel sparse representation. *Acta Polytechnica Hungarica*, 14(1):25–43, Apr 2017.