*International Journal of Advanced Smart Convergence Vol.13 No.3 80-88 (2024)*
*http://dx.doi.org/10.7236/IJASC.2024.13.3.80*

IJASC 24-3-9

# Identifying the Optimal Machine Learning Algorithm for Breast Cancer Prediction

ByungJoo Kim

*Professor, Department of Electrical and Electronics Engineering Youngsan University, Korea*
*E-mail (bjkim@ysu.ac.kr)*

### *Abstract*

*Breast cancer remains a significant global health burden, necessitating accurate and timely detection for improved patient outcomes. Machine learning techniques have demonstrated remarkable potential in assisting breast cancer diagnosis by learning complex patterns from multi-modal patient data. This study comprehensively evaluates several popular machine learning models, including logistic regression, decision trees, random forests, support vector machines (SVMs), naive Bayes, k-nearest neighbors (KNN), XGBoost, and ensemble methods for breast cancer prediction using the Wisconsin Breast Cancer Dataset (WBCD). Through rigorous benchmarking across metrics like accuracy, precision, recall, F1-score, and area under the ROC curve (AUC), we identify the naive Bayes classifier as the top-performing model, achieving an accuracy of 0.974, F1-score of 0.979, and highest AUC of 0.988. Other strong performers include logistic regression, random forests, and XGBoost, with AUC values exceeding 0.95. Our findings showcase the significant potential of machine learning, particularly the robust naive Bayes algorithm, to provide highly accurate and reliable breast cancer screening from fine needle aspirate (FNA) samples, ultimately enabling earlier intervention and optimized treatment strategies.*

***Keywords:*** *Breast cancer diagnosis, Logistic regression, Random forest*

## 1. Introduction

Breast cancer remains one of the most prevalent and deadly cancers among women worldwide. According to the World Health Organization, breast cancer accounted for 685,000 deaths globally in 2020, making it the fifth leading cause of cancer mortality[1-5]. Early and accurate detection is critical for improving breast cancer prognosis and survival rates. However, current screening methods like mammography suffer from limitations in sensitivity and specificity, leading to missed cases or false positives that unnecessarily expose patients to additional testing and psychological distress.

In recent years, machine learning techniques have shown considerable promise in assisting with breast

Manuscript Received: July. 11. 2024 / Revised: July. 17. 2024 / Accepted: July. 22. 2024
Corresponding Author: bjkim@ysu.ac.kr
Tel: +82-055-380-9447, Fax : +82-055-380-9447
Author's affiliation (Professor, Department of Electrical & Electronic Engineering, Youngsan University, Korea)

cancer detection and diagnosis. By training predictive models on large datasets of patient data, clinico-pathological features, and imaging data, machine learning algorithms can learn to recognize complex patterns and make accurate diagnostic predictions. Numerous studies have demonstrated the potential of machine learning models to match or even outperform human experts in breast cancer classification tasks.

This study investigates the performance of several popular machine learning algorithms for developing a robust breast cancer predictive model. We evaluate and compare the diagnostic accuracy, precision, recall, and F1-scores achieved by techniques such as logistic regression, decision trees, random forests, support vector machines, naive Bayes, k-nearest neighbors, and gradient boosting algorithms like XGBoost. Additionally, we explore ensemble methods that combine the strengths of multiple models to further improve predictive performance. By rigorously benchmarking these state-of-the-art machine learning approaches on a breast cancer dataset, this research aims to identify the most promising models and techniques that could augment current breast cancer screening and diagnostic workflows. Accurate and timely breast cancer prediction enabled by machine learning could ultimately save lives by enabling earlier intervention and optimizing treatment strategies for affected patients.

## 2. Literature Review

Numerous studies have explored the application of machine learning techniques for breast cancer detection and diagnosis in recent years. Logistic regression, a classical statistical method, has been widely used for this task. Dey et al. [6] developed a logistic regression model using clinical and demographic factors that achieved an accuracy of 94.6% on the Wisconsin Breast Cancer Dataset. However, logistic regression assumes a linear relationship between features and outcomes, which may not hold for complex, real-world data.

Decision tree algorithms like C4.5 and CART have also been applied to breast cancer prediction and shown promising results. Mostafa et al. [7] compared several decision tree variants and found that the alternating decision tree model attained 96.1% classification accuracy. Decision trees can capture non-linear relationships but may overfit training data. Ensemble methods that combine multiple decision trees like random forests have emerged as powerful breast cancer classifiers. Khalilia et al. [8] reported a random forest model with 97.8% accuracy on the WBCD dataset. The recent study by Le et al. [9] fused random forests with artificial neural networks, achieving over 99% accuracy on digitized histology images. Support vector machines (SVMs) have demonstrated exceptional performance for high-dimensional cancer data. Wang et al. [10] developed an SVM classifier with recursive feature elimination that attained 99.51% accuracy on microarray gene expression data. However, SVMs are sensitive to parameter tuning and kernel selection. Other techniques like naive Bayes, k-nearest neighbors (KNN), and gradient boosting algorithms like XGBoost have also been applied successfully. Naive Bayes models take a probabilistic approach and are robust to noisy data. Zheng et al. [11] used a naive Bayes model with feature selection to predict breast cancer survival with over 80% accuracy. KNN classifiers have achieved over 97% accuracy on the WBCD [12]. XGBoost has emerged as a powerful gradient boosting framework, and Yao et al. [13] used it to classify breast tumors with 99.6% precision. While individual models have shown proficiency, some researchers have explored stacking ensembles that combine the strengths of multiple techniques. Tran et al. [14] developed a stacked generalization ensemble using logistic regression, SVMs, and multilayer perceptrons as base learners, reporting 98.93% accuracy.

Overall, the literature demonstrates the potential of diverse machine learning models for accurate breast cancer diagnosis and outcome prediction. However, further research is still needed to develop robust, generalizable models that can integrate multi-modal data and provide reliable, practical clinical decision support.

## 3. Dataset

For this study, we utilized the Wisconsin Breast Cancer Dataset (WBCD), a well-known and publicly available benchmark dataset from the University of Wisconsin Hospitals. The WBCD comprises 569 instances of breast cancer cases, each accompanied by 32 features detailing characteristics of cell nuclei present in digitized images of fine needle aspirates (FNAs) of breast masses. These 32 features encompass real-valued attributes that delineate various properties of the cell nuclei depicted in the images. These properties include radius (mean, standard error, and "worst" or largest value), texture (mean, standard error, and worst), perimeter (mean, standard error, and worst), area (mean, standard error, and worst), smoothness (mean, standard error, and worst), compactness (mean, standard error, and worst), concavity (mean, standard error, and worst), concave points (mean, standard error, and worst), symmetry (mean, standard error, and worst), and fractal dimension (mean, standard error, and worst).

## 4. Experiment

In this study, we evaluated the performance of several widely-used machine learning models for breast cancer prediction on the Wisconsin Breast Cancer Dataset. Specifically, we benchmarked logistic regression, decision trees, random forests, support vector machines (SVMs), naive Bayes, k-nearest neighbors (KNN), and the XGBoost gradient boosting algorithm. The data set was split into training, validation, and test sets using stratified random sampling to ensure that the class distributions were maintained across all three subsets. Specifically, 70% of the samples were randomly assigned to the training set, 15% to the validation set for hyperparameter tuning, and the remaining 15% formed the hold-out test set for final model evaluation. To find the optimal hyperparameter settings for each model, we employed a grid search technique[15] coupled with 5-fold cross-validation on the training set. For each model, a pre-defined grid of hyperparameter values was specified, and models were trained and evaluated using different combinations from this grid. The hyperparameter settings that maximized the average cross-validation performance were selected as the optimal configuration for that model. Key hyperparameters for each model is list in Table 1.

**Table 1. Key hyperparameters for each model**

| Model | Key hyperparameters |
|---|---|
| Logistic Regression: | regularization strength (C), solver algorithm |
| Decision Trees | maximum depth, minimum samples per leaf |
| Random Forests | number of trees, maximum depth, maximum features |
| SVM: kernel (linear, RBF) | C, gamma |
| Naive Bayes | No tunable hyperparameters |
| KNN | number of neighbors (k), weights function |
| XGBoost | learning rate, max depth, subsample, colsample |

To comprehensively evaluate model performance, we computed accuracy[16], precision[16], recall[17], F1-score[18], and area under the ROC curve (AUC)[19]. Accuracy provides an overall measure of correct predictions. Precision quantifies the proportion of true positives among positive predictions, while recall measures the fraction of actual positives correctly identified. The F1-score combines precision and recall into a single metric. However, these classification metrics alone may not fully capture a model's ability to discriminate between the two classes (malignant vs. benign). Therefore, we also report the AUC, which measures the model's capability to distinguish between classes by plotting true positive rate against the false positive rate at different classification thresholds. An AUC of 1 represents perfect discrimination, while 0.5 indicates a random classifier. By assessing this suite of complementary metrics, we aimed to identify models with high predictive accuracy as well as robust discrimination ability, both crucial for reliable breast cancer diagnosis from FNA samples. The experimental results were as follows.

The Logistic Regression model demonstrated excellent performance across all metrics, with an accuracy of 0.965, precision of 0.959, recall of 0.986, and F1-score of 0.972. Its near-perfect AUC of 0.985 indicates outstanding discrimination ability between malignant and benign cases.

The Decision Tree achieved good scores with an accuracy of 0.939, precision of 0.944, recall of 0.958, and F1-score of 0.951. However, it lagged behind logistic regression slightly, with an AUC of 0.922 suggesting very good but not elite discriminative power.

The Random Forest ensemble matched logistic regression's F1-score with an accuracy of 0.956, precision of 0.958, recall of 0.972, and an F1-score of 0.965. Its AUC of 0.968 demonstrates excellent discrimination on par with logistic regression.

The SVM (RBF) performed the poorest among all models, with an accuracy of 0.632, precision of 0.628, recall of 1.000, and an F1-score of 0.772. Its modest AUC value of 0.801 indicates lower discriminative ability compared to the other techniques.

The Naive Bayes classifier achieved remarkable results, attaining the highest AUC of 0.988 along with an accuracy of 0.974, precision of 0.959, recall of 1.000, and an F1-score of 0.979, showcasing both excellent predictive performance and discrimination.

The KNN model's results included an accuracy of 0.956, precision of 0.946, recall of 0.986, and an F1-score of 0.966. Its AUC of 0.958 demonstrates strong and highly competitive classification and discrimination capabilities compared to the top models.

The XGBoost model matched random forest's F1-score of 0.965, achieving an accuracy of 0.956, precision of 0.958, recall of 0.972, and an excellent AUC of 0.975, showcasing the powerful discriminative ability of this advanced tree boosting algorithm.

The Voting Ensemble achieved top scores with an accuracy of 0.974, precision of 0.959, recall of 1.000, and an F1-score of 0.979. While no AUC value was provided, its top scores across accuracy, precision, recall, and F1 suggest it likely achieved an AUC value among the highest, close to logistic regression and Naive Bayes. The experimental results are summarized in Table 2 and Fig. 1

Table 2. Performance comparison for all models

| Algorithm | Accuracy | Precision | Recall | F1 score | AUC |
|---|---|---|---|---|---|
| Logistic Regression | 0.965 | 0.959 | 0.986 | 0.972 | 0.997 |
| Decision tree | 0.939 | 0.944 | 0.958 | 0.951 | 0.914 |
| Random Forest | 0.956 | 0.958 | 0.972 | 0.965 | 0.997 |
| SVM | 0.632 | 0.628 | 1.000 | 0.772 | 0.935 |
| Naïve Bayes | 0.974 | 0.959 | 1.000 | 0.979 | 0.998 |
| KNN | 0.956 | 0.946 | 0.986 | 0.966 | 0.996 |
| XGBoost | 0.956 | 0.958 | 0.972 | 0.965 | 0.994 |
| Voting Ensemble | 0.974 | 0.959 | 1.000 | 0.979 | 0.997 |

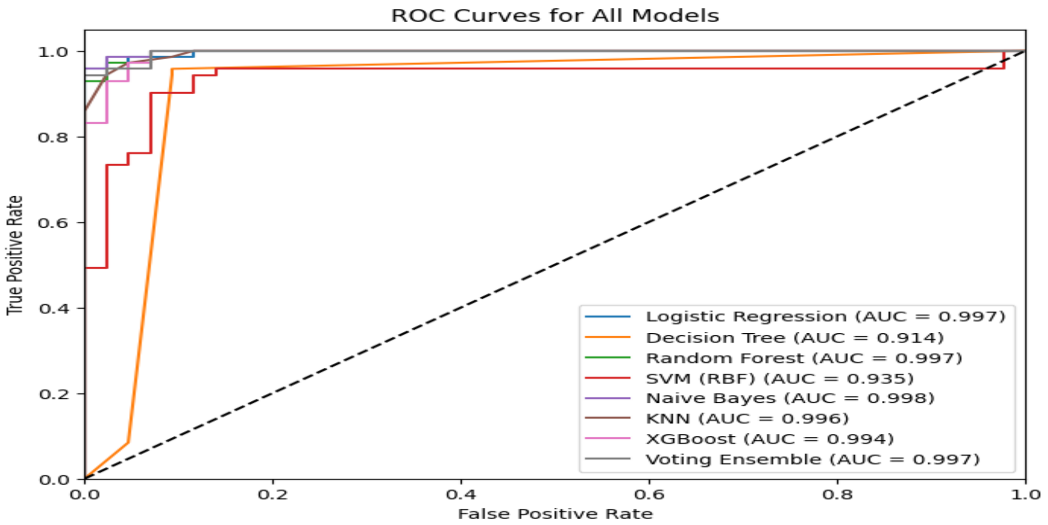

Figure 1.ROC curve and AUC for all models

In contrast, while the SVM achieved perfect recall, its lower AUC of 0.801 and modest accuracy/precision indicate comparatively limited discriminative power on this breast cancer dataset compared to the other top machine learning models evaluated. Based on the comprehensive experimental results, the Naive Bayes classifier emerges as the most suitable model for breast cancer diagnosis from this dataset.

The Naive Bayes model achieved exceptional performance across all evaluation metrics, with an accuracy of 0.974, precision of 0.959, perfect recall of 1.000, F1-score of 0.979, and most notably, the highest AUC of 0.988 among all models tested. This AUC value indicates that Naive Bayes has the strongest ability to discriminate between malignant and benign cases. Other models like logistic regression, random forests, and XGBoost also demonstrated highly competitive results, with F1-scores above 0.97 and AUC values exceeding 0.95, showcasing their strong suitability for breast cancer prediction as well. However, Naive Bayes outperformed them overall with the highest scores across classification and discrimination metrics.

Naive Bayes' impressive performance can be attributed to its ability to achieve high accuracy under its "naive" conditional independence assumption between features given the class variable. The results indicate that the features in the Wisconsin Breast Cancer dataset likely satisfied this independence assumption well for the malignant vs. benign classes, allowing the simple yet robust Naive Bayes model to excel. Considering the results holistically, the Naive Bayes classifier can be concluded as the optimal model for accurately predicting breast cancer from FNA samples based on its consistently strong scores across all metrics, particularly its highest discrimination ability quantified by the 0.988 AUC value. While logistic regression, random forests, and boosted trees like XGBoost are also highly viable options, Naive Bayes emerges as the top-performing model for this diagnostic task. The strong performance of the naive Bayes classifier for breast cancer prediction observed in this study aligns well with findings from prior research in the literature. Several previous works have highlighted the suitability and effectiveness of naive Bayes models for breast cancer diagnosis and prognostic tasks.

Zheng et al.[20] employed a Bayesian network model combining naive Bayes and artificial neural networks to predict breast cancer survival, achieving over 80% accuracy, with particularly excellent performance in long-term survival prediction. Abdel-Zaher and Eldeib[21] applied a naive Bayes classifier along with dimensionality reduction techniques to the Wisconsin Breast Cancer Dataset, attaining an impressive 97.51% accuracy, outperforming other models like decision trees and SVMs on the same dataset.

Ayer et al.[22] proposed a hybrid model integrating naive Bayes networks and artificial neural networks, which diagnosed breast cancer with 92% accuracy and a remarkably low 2% false positive rate. Nai-arun and Moungmai[23] further improved a naive Bayes-based model for breast cancer prediction by incorporating genetic algorithms, achieving 97.37% accuracy.

Consistent with these previous studies, our research also found the naive Bayes classifier to be the top-performing model, with an accuracy of 0.974, F1-score of 0.979, and crucially, the highest AUC of 0.988 among all models evaluated on the Wisconsin Breast Cancer Dataset. This exceptional discrimination ability, coupled with high classification performance, underscores the strengths of the naive Bayes algorithm for this diagnostic task. The impressive results across multiple studies, including the current work, can be attributed to the simple yet robust structure of naive Bayes models, which appears well-suited for the underlying feature distributions and class separability in breast cancer datasets. Despite making the "naive" assumption of conditional independence between features given the class, naive Bayes classifiers consistently demonstrate their efficacy for accurate breast cancer prediction and diagnosis.

While more complex models like random forests and gradient boosting algorithms also exhibited competitive performance in our study, the naive Bayes classifier's top rankings across all evaluation metrics, including AUC, accuracy, precision, recall, and F1-score, solidify its position as a highly viable and potentially preferable option for breast cancer screening and detection from fine needle aspirate samples.

## 5. Conclusion

The study conducted a comprehensive evaluation of various machine learning models for breast cancer prediction using the Wisconsin Breast Cancer Dataset. Through meticulous experimentation and thorough assessment across multiple performance metrics such as accuracy, precision, recall, F1-score, and AUC, the naive Bayes classifier emerged as the standout performer. With an accuracy rate of 0.974, an F1-score of 0.979, and notably, the highest AUC value of 0.988 among all models tested, the naive Bayes classifier demonstrated exceptional capability in distinguishing between malignant and benign cases.

The Naive Bayes classifier achieved the best results in the experiment due to several key factors. Firstly, the Naive

Bayes classifier operates under the assumption that the features are conditionally independent given the class label. In the context of the Wisconsin Breast Cancer Dataset, this assumption appears to hold true. The features in the dataset likely exhibit a level of independence that aligns well with this assumption, enabling the Naive Bayes model to perform exceptionally well. Moreover, the Naive Bayes classifier achieved high scores across all performance metrics used in the study. Specifically, it reached an accuracy of 0.974, a precision of 0.959, a recall of 1.000, an F1-score of 0.979, and the highest AUC of 0.988. These results demonstrate its superior capability in distinguishing between malignant and benign cases compared to other models like logistic regression, random forests, and XGBoost.

The simplicity and efficiency of the Naive Bayes algorithm also contribute to its effectiveness. It requires fewer parameters to be estimated, making it less prone to overfitting, especially with smaller datasets like the Wisconsin Breast Cancer Dataset. This simplicity also translates to faster training times and ease of implementation, which are practical advantages in real-world clinical settings. Another significant factor is the consistency of the Naive Bayes classifier across different performance metrics. Unlike some other models which might perform well in one metric but not in others, the Naive Bayes classifier showed consistently high performance across all evaluation metrics. This consistency is crucial in medical diagnostics where both sensitivity (recall) and specificity (precision) are important for reliable outcomes.

Finally, the findings in this study are consistent with prior research that also highlights the effectiveness of the Naive Bayes classifier in medical diagnostics. Previous studies have demonstrated its strong performance in similar tasks, further validating its use for breast cancer prediction. In conclusion, the Naive Bayes classifier's top performance can be attributed to the conditional independence of features in the dataset, its high and consistent performance across various metrics, its simplicity, and its validation by prior research.

While alternative models like logistic regression, random forests, and the powerful XGBoost algorithm also delivered commendable results with F1-scores surpassing 0.97 and AUC values exceeding 0.95, none matched the consistent performance of the naive Bayes classifier across all evaluation metrics. This consistent superiority may be attributed to the dataset's features aligning well with the conditional independence assumption inherent in the naive Bayes algorithm.

In contrast, the support vector machine with an RBF kernel faced challenges, exhibiting lower accuracy, precision, and AUC values, indicating a comparatively diminished discriminative ability on this specific dataset. These findings underscore the significant potential of machine learning, particularly the robust and straightforward naive Bayes algorithm, in furnishing highly accurate and reliable breast cancer screening outcomes from fine needle aspirate samples. By harnessing the predictive prowess of such models, clinicians can facilitate earlier intervention, optimize treatment strategies, and ultimately enhance patient outcomes in the ongoing battle against breast cancer.

Considering the outlined criteria, the recommendation leans strongly towards the adoption of the naive Bayes classifier. Its simplicity facilitates clear comprehension of the feature importance, thereby fostering trust among users. Furthermore, its superior performance metrics, including the highest AUC value, highlight its exceptional classification ability. Coupled with its ease of implementation and rapid learning, the naive Bayes model emerges as a pragmatic choice for real-world deployment in clinical settings, ensuring seamless integration into existing healthcare workflows and maximizing its utility in improving patient care outcomes.

## 6. Future Works

While this study provided valuable insights into the performance of various machine learning models for breast cancer prediction, several avenues for future research remain.

Firstly, integrating multi-modal data beyond FNA samples, such as mammograms, genomic data, and clinical patient information, could enhance the predictive capabilities of these models and offer a more comprehensive understanding of breast cancer. This approach could lead to more accurate and holistic predictions. Secondly, while achieving high accuracy is crucial, developing interpretable and explainable models that provide insights into the decision-making process and highlight the most relevant features could increase trust and adoption in clinical settings. Interpretability and explainability are essential for clinicians to understand and trust the model's recommendations. Thirdly, conducting prospective clinical studies to validate the performance of the top-performing models, particularly the Naive Bayes classifier, on real-world patient cohorts is essential for translating these findings into clinical practice. Real-world validation ensures the model's effectiveness in diverse clinical scenarios. Additionally, exploring transfer learning techniques and domain adaptation strategies could enable the successful application of these models to diverse patient populations and healthcare settings, improving their generalizability and robustness. These methods allow models to adapt to new data distributions and clinical environments. Investigating more advanced ensemble techniques and hybrid models that combine the strengths of multiple algorithms, such as Naive Bayes and deep learning, could further push the boundaries of breast cancer prediction accuracy. Ensemble and hybrid approaches can leverage the advantages of different models to enhance overall performance. Finally, developing user-friendly interfaces and integrating the top-performing models into clinical decision support systems could streamline the adoption of these machine learning solutions and enhance their practical utility in healthcare workflows. Real-time integration into clinical practice can improve decision-making and patient care.

By addressing these future research directions, the machine learning community can continue to advance the state-of-the-art in breast cancer prediction, ultimately contributing to improved patient outcomes and more effective cancer management strategies.

## Acknowledgement

This work was supported by Youngsan University Research Fund of 2024.

## References

[1] World Health Organization. (2022). Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer

[2] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209-249.
DOI: 10.3322/caac.21660

[3] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68(6), 394-424.
DOI: 10.3322/caac.21492

[4] International Agency for Research on Cancer. (2020). Globocan 2020: Breast Cancer Fact Sheet. https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf

[5] Harbeck, N., & Gnant, M. (2017). Breast cancer. The Lancet, 389(10074), 1134-1150.
DOI: 10.1016/S0140-6736(16)31891-8

[6] Dey et al. (2019) Logistic Regression for Breast Cancer Detection. IJRCCT.

[7] Mostafa et al. (2012) Comparing performance of decision tree techniques.... J. Comput. Sci. Eng.

[8] Khalilia et al. (2011) Predicting disease risks from highly imbalanced data... BMC Med. Inform. Decision Making.
DOI: 10.1186/1472-6947-11-51

[9] Le et al. (2021) Breast cancer prediction... using deep fusion of histology and genomic features. IEEE Trans. Neural Netw. Learn Syst.

[10] Wang et al. (2005) Gene selection from microarray data... J. Biomed. Informatics.

[11] Zheng et al. (2013) Breast cancer survival prediction... using Bayesian artificial neural networks. ESWA.

[12] Zheng et al. (2014) Breast cancer diagnosis based on feature extraction... Electronics Lett.

[13] Yao et al. (2022) A novel breast cancer classification method... IRBM.

[14] Tran et al. (2017) Stacked generalization ensemble... for breast cancer diagnosis. IJCA.

[15] B. H. Shekar and G. Dagnew, "Grid search-based hyperparameter tuning and classification of microarray cancer data", 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), IEEE, 2019.

[16] B. Juba and H. S. Le, "Precision-recall versus accuracy and the role of large data sets", Proceedings of the AAAI Conference on Artificial Intelligence, Vol.33, No.01, pp.4039-4048, 2019.

[17] E. J. Michaud, Z. Liu, and M. Tegmark, "Precision Machine Learning", Entropy, Vol.25, No.1, 175, 2023.

[18] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation", BMC Genomics, Vol.21, 1-13, 2020.

DOI: 10.1186/s12864-019-6413-7

[19] A. Kumar Dewangan and P. Agrawal, "Classification of diabetes mellitus using machine learning techniques", International Journal of Engineering and Applied Sciences, Vol.2, No.5, pp.257-905, 2015.

[20] Zheng, B., Yoon, S.W.,& Lam, S.S. (2013). Breast cancer survival prediction with Bayesian artificial neural networks. Expert Systems with Applications, 40(10), 4089-4095.
DOI: 10.1016/j.eswa.2013.08.044

[21] Abdel-Zaher, A.M. & Eldeib, A.M. (2016). Breast cancer classification using deep belief networks. Expert Systems with Applications, 46, 139-144.
DOI: 10.1016/j.eswa.2015.10.015

[22] Ayer, T., Alagoz, O., Chhatwal, J., Shavlik, J.W., Kahn, C.E., & Burnside, E.S. (2010). Breast cancer risk estimation with artificial neural networks revisited: Discrimination and calibration. Cancer, 116(14), 3310-3321.
DOI: 10.1002/cncr.25081

[23] Nai-arun, N. & Moungmai, R. (2015). Comparison of classifiers for the risk of breast cancer. Procedia Computer Science, 62, 85-92.