Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 6, No.4, October 2024, pp. 02404017-01 ~ 02404017-09
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v6i4.986

Optimized Stacking Ensemble Classifier for Early Cancer Detection Using Biomarker Data

K. Jegadeeswari[1], R. Rathipriya[2*]

[1]Department of Computer Science, Periyar University, Salem, Tamilnadu, India.

[2]Assistant Professor, Department of Computer Science, Periyar University, Salem, Tamilnadu, India.

* rathi_priyar@periyaruniversity.ac.in

Abstract. Ovarian cancer ranks sixth globally as a major cause of death among women, with a five-year survival rate below 50%, largely due to late detection. Early detection is crucial to lower mortality rates. This paper introduces an Optimized Stacking Ensemble Classifier (OSEC) for early ovarian cancer detection using biomarkers. The model comprises two layers: the first layer includes base classifiers optimized with Particle Swarm Optimization (PSO), while the second layer is a meta-classifier integrating Support Vector Machine (SVM), Logistic Regression (LR), and Random Forest(RF) models fine-tuned through grid search. Among the three datasets evaluated, the Blood Routine dataset showed the best performance with a stacked RF meta-classifier, achieving: 94.29% accuracy. The Stacked RF model also outperformed others, reaching 92.82% accuracy on the Serum dataset and 92.77% on the Malignant Ovarian Tumor (MOT) dataset, consistently excelling in precision, recall, and f1-score.

Keywords: Cancer Detection, Ensemble Methods, Biomarkers, Hyperparameter Optimization, Machine Learning Optimization, Particle Swam Optimization, Stacking Ensemble.

(Received 2024-08-22, Accepted 2024-09-09, Available Online by 2024-09-20)

1. Introduction

Ovarian cancer (OC) is a serious and prevalent condition among women, currently with no definitive cure. It accounts for 2.5% of all female cancers [1], with patients facing a lifetime risk of 2.7% and a fatality rate of 5%. This high fatality rate is largely due to late-stage diagnosis and the lack of early symptom detection. Early identification of OC can significantly enhance survival rates. Traditional diagnostic methods such as chemotherapy, ultrasonography, and helical CT scanning [2], though valuable, have limitations. Chemotherapy, while effective in treating advanced OC, has a high recurrence rate of 60% to 80% within five years. Ultrasonography and Helical CT scanning are essential for tumor detection and treatment planning but struggles to reliably distinguish between benign and malignant

pelvic masses, complicating early diagnosis [3]. Gynecologists [4] often face the challenge of distinguishing between benign and malignant pelvic masses, which may indicate tumors. Both ultrasonography and helical CT scanning are crucial for early-stage tumor detection. Several biomarkers [5], including CA125, CA72-4, and HE4, provide valuable information about the presence, progression, or response to treatment of OC. Higher levels of CA-125, CA-72-4 and HE4 in the blood indicate the presence of ovarian cancer. Normal blood values are typically less than 35 U/mL for CA-125, less than 140 pmol/L for HE4, and below 6.9 U/mL for CA 72-4. Levels within these ranges suggest the presence of a benign ovarian tumor. Bast and Mills [6] considered the properties of combined biomarkers with special emphasis on CA 125, HE 4, and CA 72-4 in the finding of ovarian cancer at an early stage. Concluding their review of the literature, the authors opined that the integration of these biomarkers improves the diagnostic performance compared to the use of individual biomarkers. Nonetheless, a pointed weakness is methodological fluctuations in biomarker levels; they remain influenced by the different characteristics of patients, and, therefore, research that includes numerous and highly diverse populations of patients is required. Medeiros [7] examined the diagnostic performance of CA-125 in ovarian cancer. Although CA-125 is useful in diagnosing ovarian cancer, it is not very sensitive or specific, especially in stage I disease. This stresses the need for more biomarkers and better diagnostics. Moore and Miller [8] reported on clinical trials regarding HE4 test efficiency in the case of ovarian cancer, stating that HE4 raises the overall test sensitivity when complemented with CA-125. However, they often include small numbers of patients and must be reproduced in greater numbers of racially diverse subjects. However, existing methods using these biomarkers often lack the precision needed for early detection [9-12], leading to a gap in the effective early diagnosis of OC. This research gap highlights the need for more accurate and reliable predictive models. To address these shortcomings, this study introduces an Optimized Stacking Ensemble Classifier (OSEC) developed to improve early detection of ovarian cancer by analyzing biomarkers within clinical datasets. The clinical dataset includes various biomarkers categorized into three types: blood routine (Red Blood Cell Distribution Width, Lymphocyte Cell Count, Basophil Percentage, Mean Corpuscular Hemoglobin, ect), serum (Anion Gap, Albumin, Sodium, Phosphorus, Alkaline Phosphatase, Aspartate Aminotransferase (AST), Potassium, Chlorine, ect), and Malignant Ovarian Tumor (MOT) (Carbohydrate Antigen 125, Alpha-Fetoprotein, Carbohydrate Antigen 19-9, Human Epididymic Protein 4, Carbohydrate Antigen 72-4, Carcinoembryonic Antigen) markers. Unlike existing methods, which often rely on single or non-optimized models [15], the proposed OSEC uses a two layered ensemble approach. Initially, it uses four distinct base ensemble classifiers such as Gradient Boosting (GBClassifier), Random Forest (RFClassifier), Extra Trees (ETClassifier), and XGBoost (XGBClassifier), along with non-ensemble classifiers such as Decision Trees, Logistic Regression, SVM, KNN, and Naive Bayes. These base models are individually tuned using Particle Swarm Optimization (PSO) for hyperparameter optimization (HPO). The prediction from these optimized base classifiers are then aggregated and used to train the meta-classifiers such as LR, SVM, and RF in the second layer. Here, Grid search-based hyperparameter optimization is applied to fine-tune the hyperparameter of the meta-classifiers. This two-stage HPO refines the overall predictive performance of the OSEC model.

1.1. Contributions

This sub-section outlines the key contributions of the proposed OSEC model. They are

1. **Two-layered Ensemble Approach**: Proposes a novel two layered ensemble method for early detection ovarian cancer, enhancing predictive accuracy and robustness
2. **Two-stage Hyperparameter Optimization:** Uses advanced hyperparameter tuning through PSO

The rest of this paper is organized as follows: Section 2 sums up the outcomes of ensemble and non-ensemble methods, as well as the techniques related to hyperparameters tuning. Section 3 further explains the conceptual framework and by extension the methodology to be used in the research study. The last section is devoted to the further discussion of the results of using the proposed methodology for analysis. Hence, Section 5 concludes with the following summary of the findings.

2. Methodology

The research study used a clinically approved raw dataset comprising samples from individuals with both benign and malignant ovarian tumors. The OSEC applied classification techniques to detect ovarian cancer at an early stage.

2.1. Data Processing

The raw dataset went through multiple preprocessing stages, such as data cleaning, data scaling, SMOTE and splitting. It contains information from 349 individual patients along with 49 biomarkers. For data scaling, the values were standardized using the equation 1 Here, x_i – each individual, μ - mean of the population.Following these preprocessing steps, SMOTE was applied to manage class imbalance.

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \tag{1}$$

2.2. Optimized Stacking Ensemble Classifier (OSEC)

The proposed OSEC methodology incorporated a comprehensive approach to developing a two-layer ensemble classifier model for early ovarian cancer detection. The process began with a cancer dataset specifically curated for training and testing purposes. A subset of this dataset was used as the training set to train individual base classifiers. These base models, included both ensemble methods (GBClassifier, RFClassifier, ETClassifier, XGBClassifier) and non ensemble methods (DT, LR, KNN, KVM, Naive Bayes). These base models underwent hyperparameter tuning using Particle Swarm Optimization (PSO). In PSO [13], each particle represents a set of hyperparameters and moves through the search space based on its velocity. The velocity of V_i of each particle updated using the following equation 2

$$V_i^{t+1} = w.V_i^t + c_1.r_1.\left(P_i^t - X_i^t\right) + c_2.r_2.\left(G^t - X_i^t\right) \tag{2}$$

Where w is inertia weight, c_1 and c_2 are coefficient controlling the influence of personal global best position, r_1 and r_2 are random factors. The particle position X_i updated by adding this velocity as mentioned in equation 3

$$X_i^{t+1} = X_i^t + V_i^{t+1} \tag{3}$$

The performance of these configurations was evaluated using a fitness function, and the optimal hyperparameters were identified through multiple iterations of this process, thus enhancing the models' performance and accuracy. The base models then generated predictions on the training set, which were aggregated into meta (P_1, P_2,… P_n). These meta features served as input for a meta model, a classifier trained to make final predictions. Integration techniques such as Stacked LR, Stacked SVM, Stacked RF were implemented, with grid search used to optimize the meta model. This ensured the accurate and robust predictions.

Grid search systematically work through multiple combinations of hyperparameter values. It is to find the optimal set of hyperparameters for a meta classifier that maximizes the model's accuracy. Define the hyperparameters space X_i to search space in equation 4. Each hyperparameter h_i, a set of possible values.

$$X_i = (h_1, h_2, \ldots h_m) \tag{4}$$

The total number of combinations of grid points) as follows in equation 5

$$Total\ Combination = \prod_{i=1}^{m} n_i \tag{5}$$

Select the set of hyperparameters h^* that provides the best average performance $M(h_m)$ in equation 6

$$h^* = \arg \max_{h_m \epsilon X_i} M(h_m) \tag{6}$$

Algorithm 1 outlines the hyperparameter optimization of the classifiers used in the two stage HPO approach. This optimization process enhances the performance of the base Models, allowing them to better capture complex patterns within the data.

Algorithm 1. Hyperparameter Optimization

Input : hyperparameters X_i of each models M
Output : fitness Score of X_i

- Initialize model with Hyperparameters X_i
- Train the model $M(X_i)$ using training data
- Evaluate the model $M(X_i)$ on validation data.
- Calculate the performance metric such as accuracy

$$fitness(X_i) = Accuracy(M(X_i))$$

- Return the fitness score

The model's performance was assessed using a designated performance metrics to assess its effectiveness in ovarian cancer detection. Finally, the trained model was tested on a separate test set to provide valuable insights into its performance and delivered the final prediction for ovarian cancer detection. Algortihm-2 illustrates the steps of the proposed OSEC.

Algorithm -2 : Optimized Stacking Ensemble Classifier

Input: { X: Input features (biomarkers) , y: True labels (0 for benign, 1 for malignant) , K: Number of folds for cross-validation, Base classifiers: {C_1, C_2, …, C_m}, Meta-classifier: M }
Output: Predicted labels for new data instances
Procedure:

1. **Data Preprocessing:**
   - Handle missing values and outliers in X.
   - Standardize X to have zero mean and unit variance.
2. **Class Imbalance Handling:** Apply SMOTE to balance class distribution in X and y.
3. **Split Data:** Split X and y into training and test sets.
4. **Initialize meta_features:**
   - Create an empty array meta_features to store predictions from base classifiers.
5. **First Layer - Hyperparameter Tuning of Base Classifiers using PSO:**
   - For each base classifier C_i in {C_1, C_2, …, C_m}:
     - **Initialize** pBest (best score using algorithm 1) and gBest (best parameters).
     - **For** each hyperparameter set:
       - Train and validate the base classifier C_i using cross-validation (calculate fitness using Algorithm 1).
       - If the current fitness is better, update pBest and gBest using equations (2) and (3).
     - Train the final classifier C_i with gBest parameters.
     - Store predictions of C_i on training data in meta_features.
6. **Second Layer - Hyperparameter Tuning of Meta-classifier using Grid Search:**
   - Train meta-classifier M on meta_features using grid search to optimize hyperparameters.
7. **Evaluate Performance:**
   - Predict the test data using the trained meta-classifier M.
   - Calculate performance metrics (e.g., Accuracy, ROC AUC) for the final predictions y_{final}

3. Result and Discussion

The study data involved data of 349 patients and the data was collected from the Soochow University' Affiliated Hospital [14]. This analysis is a retrospective study based on 171 patients with

ovarian cancer diagnosed between July 2011 and July 2018 as well as 178 patients with benign ovarian tumors. The dataset comprises a set of 49 attributes that were identified through pathological diagnosis.

The dataset is categorized into three subgroups: Blood Routine contains 19 biomarkers, Serum contains 22 biomarkers, and MOT contains 6 biomarkers. For the experiment, both for datasets, the data was split into two sets: 80% for training and 20% for testing. The performance of the classifiers was evaluated using the following metrics Accuracy, Precision, recall, F1-score and log-loss. Table 1 explains the best hyperparameters settings of each classifier model along with Two-stage HPO technique.

Table 1. Hyperparameters values of Ensemble and non ensemble Classifier models

ML Models	**HPO Techniques**	**Best values**
DT Classifier	PSO	criterion: entropy, max-depth: 5, max-features: None , max-leaf-nodes: 12 , min-samples_leaf: 2 , min-samples_split:10, splitter: best,
Logistic Regression	Grid Search	Regularization Strength NAÏVE : 10.0, Solver : liblinear, Multi-class : ovr, Class-weight : balanced
SVM	Grid Search	Kernel : poly, Gamma : auto, C : 1, degree : 2
KNN	PSO	n_neighbors-values : 5, weights-values : distance, algorithm-values:auto,
Naive Bayes	PSO	Smoothing Parameter (alpha): 1.5
GB Classifier	PSO	Learning-rate: 0.1, n-estimators: 50 , max-depth: 2, subsample: 0.8, min-samples_split:2 , min_samples-leaf: 1 ,
ET Classifier	PSO	n-estimators: 100, max-depth : 20, min-samples_split:5, min-samples-leaf: 1,
XGB Classifier	PSO	Max-depth: 2, n-estimators: 60, learning-rate:0.1,
RF Classifier	Grid Search	n-estimators: 50, max-depth: 5, min-samples_split:15, min-samples-leaf:6

Table 1 presents the results of applying Two-stage HPO like PSO and grid search across various classifiers. Listing the best hyperparameter values for each classifier. The performance analysis of experiments using these HPO to the proposed two-layered ensemble approach namely OSEC. Table 2 presents the performance analyses of the non-ensemble base models and the OSEC on different test sets namely Blood sentinels of routine, Serum, MOT.

Table 2. Comparative Performance analysis of non ensemble base models and OSEC

Data set	Metric	Benchmark non ensemble Models					Proposed OSEC Models		
		DT	LR	SVM	KNN	Naive Bayes	Stacked LR	Stacked SVM	Stacked Random Forest
Blood Routine	Accuracy(M)	0.914	0.944	0.885	0.914	0.914	0.914	0.911	0.919
	Precision(M)	0.846	0.889	0.846	0.860	0.874	0.874	0.860	0.904
	Recall(M)	0.794	0.841	0.811	0.815	0.838	0.838	0.815	0.651
	f1(M)	0.850	0.890	0.846	0.862	0.874	0.874	0.862	0.910
	Logloss(M)	4.661	3.116	4.661	4.146	3.631	3.631	4.146	3.087
Serum	Accuracy(M)	0.852	0.882	0.823	0.852	0.852	0.872	0.852	0.882
	Precision(M)	0.784	0.827	0.784	0.798	0.812	0.812	0.798	0.847
	Recall(M)	0.732	0.779	0.749	0.753	0.776	0.776	0.753	0.809

	f1(M)	0.788	0.828	0.784	0.800	0.812	0.812	0.800	0.848
	Logloss(M)	4.720	3.175	4.720	4.030	3.690	3.690	4.205	3.146
MOT	Accuracy(M)	0.872	0.902	0.843	0.872	0.872	0.892	0.872	0.902
	Precision(M)	0.804	0.847	0.804	0.818	0.832	0.832	0.818	0.867
	Recall(M)	0.752	0.799	0.769	0.773	0.796	0.796	0.773	0.829
	f1(M)	0.808	0.848	0.804	0.820	0.832	0.832	0.820	0.868
	Logloss(M)	4.703	3.158	4.703	4.188	3.673	3.673	4.188	3.129

Table 2 shows the comparative performance analysis of hyperparameter tuned non-Ensemble base classifier and OSEC for OC Classification. The stacked RF model achieved the highest accuracy of 91.86% on the Blood Routine dataset, with exceptional precision (90.38%) and a strong recall of 82.94% on the MOT dataset. The model also recorded a balanced F1 score of 90.97% and the lowest log loss of 3.0865 on the Blood Routine dataset. Among the three datasets, the Blood Routine dataset shows the best overall performance with the stacked random forest. Figure 1 shows the comparative analyses of the non ensemble base models and the stacked models.

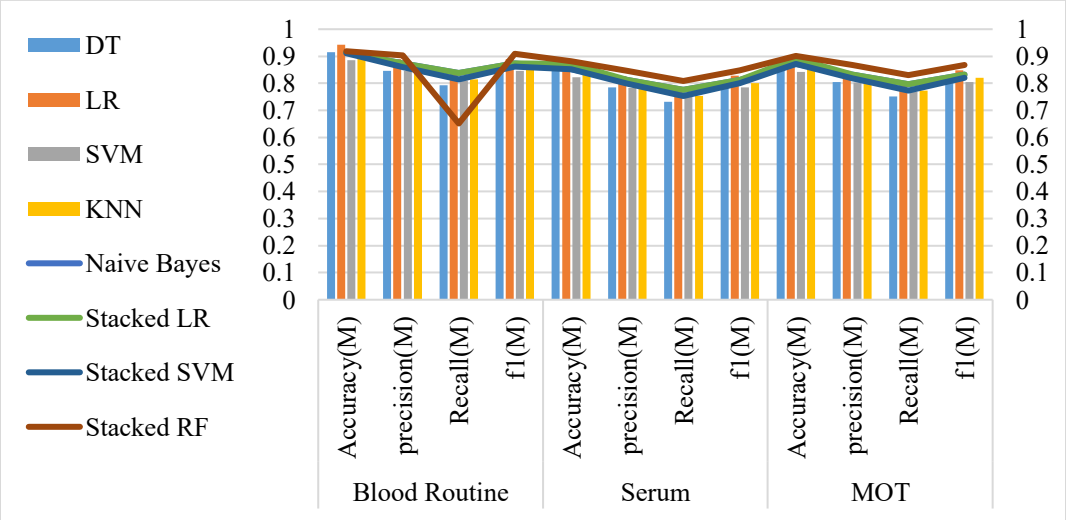

Figure 1. Comparative analyses of the non ensemble base models and the OSEC performances

The figure 1 presents performance metrics of various non-ensemble models across three different datasets. The x-axis denotes the metrics, while the left y axis scales performance scores (0 to 1) of base models. The right y axis highlights meta models, helping in comparing performance across datasets. Table 3, presents the comparative analyses of the ensemble base models and the OSEC on diverse test sets: Blood Routine, Serum, and MOT.

Table 3. Comparative Performance analysis of Ensemble base models and OSEC

Data set	Metric	Benchmark ensemble Models				Proposed OSEC Models		
		GB	RF	ET	XGB	Stacked LR	Stacked SVM	Stacked Random Forest
Blood Routine	Accuracy(M)	0.921	0.931	0.921	0.921	0.928	0. 9483	0.943
	Precision(M)	0.881	0.896	0.853	0.867	0.908	0.938	0.933
	Recall(M)	0.845	0.849	0.801	0.822	0.872	0.849	0.906

	f1(M)	0.882	0.897	0.857	0.869	0.909	0.896	0.944
	Logloss(M)	3.560	3.045	4.590	4.075	3.533	4.036	2.989
Serum	Accuracy(M)	0.859	0.889	0.859	0.859	0.912	0.898	0.928
	Precision(M)	0.828	0.876	0.845	0.849	0.896	0.849	0.906
	Recall(M)	0.783	0.787	0.739	0.760	0.872	0.872	0.849
	f1(M)	0.884	0.924	0.880	0.896	0.909	0.904	0.896
	Logloss(M)	3.649	3.134	4.698	3.989	3.563	4.088	3.019
MOT	Accuracy(M)	0.879	0.909	0.879	0.879	0.927	0.908	0.928
	Precision(M)	0.839	0.854	0.811	0.825	0.908	0.894	0.913
	Recall(M)	0.803	0.807	0.759	0.780	0.872	0.849	0.906
	f1(M)	0.840	0.855	0.815	0.827	0.909	0.896	0.924
	Logloss(M)	3.622	3.107	4.652	4.137	3.553	4.068	3.009

Table 3 shows the comparative performance analysis of hyperparameter tuned Ensemble base classifier and OSEC for OC Classification. The stacked RF attained the maximum Accuracy of 94.29% on the Blood Routine dataset, along with peak precision (93.27%) high recall (90.55%) and the f1 score of 94.38%, indicating the balanced performance. Figure 2 shows the comparative analyses of the ensemble base models and the OSEC.

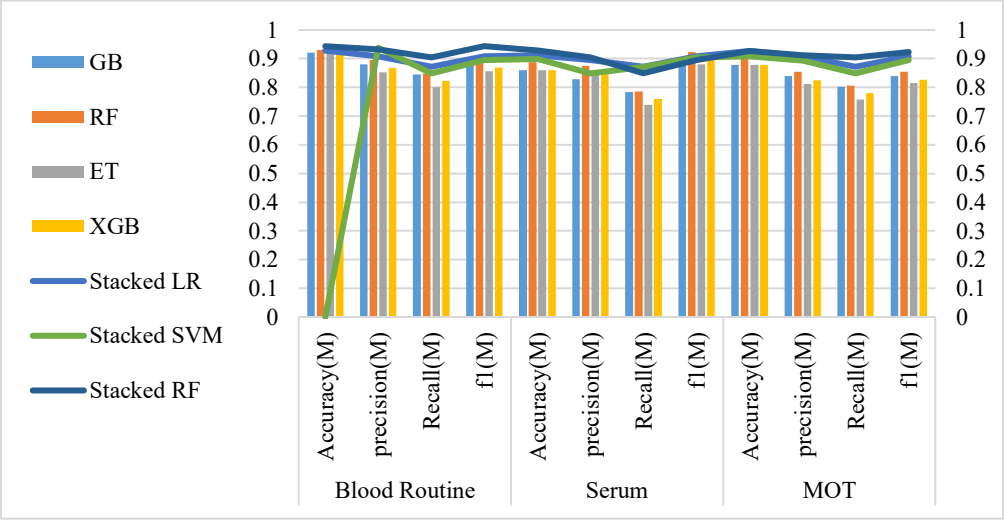

Figure 2. Comparative analyses of the ensemble base models and the OSEC performance

The figure3 presents performance metrics of various ensemble models across three different datasets. The x-axis denotes the metrics, while the left y axis scales performance scores (0 to 1) of base models. The right y axis highlights meta models, helping in comparing performance across datasets.

4. Conclusion

Ovarian cancer (OC) remains a critical global health challenge for women, often diagnosed at advanced stages due to lack of awareness and limited screening methods. Early detection is crucial for improving survival rates. This study introduced the Optimized Stacking Ensemble Classifier (OSEC), which utilized biomarkers CA-125, CA-72-4, and HE4 commonly assessed in blood, to differentiate between malignant and benign ovarian tumors. The proposed novel OSEC methodology employed a two-layered approach and use of two stage HPO for base classifier and meta classifier models. The Stacked Random Forest model outperformed all other models across the datasets, achieving the highest accuracy:

94.29% on the Blood Routine dataset, 92.82% on the Serum dataset, and 92.77% on the MOT dataset. These results highlight OSEC model's potential to enhance early detection and improve patient health. Future research should focus on integrating additional biomarkers and validating the model across a broader hyperparameters search space. Practical applications include its potential use in routine screening to guide treatment plans and reduce mortality.

Data Availability: The Ovarian cancer dataset is publicly available in the UCI repository.
Acknowledgments: No funding agency provided support for this work.
Declarations: others have no Conflicts of interests.

Reference

[1] Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018, 68, 394–424.

[2] Torre, L.A.; Trabert, B.; DeSantis, C.E.; Miller, K.D.; Samimi, G.; Runowicz, C.D.; Siegel, R.L. Ovarian cancer statistics, 2018. CA Cancer J. Clin. 2018, 68, 284–296.

[3] Marchetti, C.; Pisano, C.; Facchini, G.; Bruni, G.S.; Magazzino, F.P.; Losito, S.; Pignata, S. First-line treatment of advanced ovarian cancer: Current research and perspectives. Expert Rev. Anticancer Ther. 2010, 10, 47–60.

[4] Wang, J.; Gao, J.; Yao, H.; Wu, Z.; Wang, M.; Qi, J. Diagnostic Accuracy(L) of serum HE4, CA125 and ROMA in patients with ovarian cancer: A meta-analysis. Tumor Biol. 2014, 35, 6127–6138. [CrossRef]

[5] K. Jegadeeswari, R. Rathipriya and J. Renugadevi, "Fusion Learning of Regression Models for Missing Data Imputation in Breast Cancer Dataset," *2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)*, Raipur, India, 2023, pp. 1-14, doi: 10.1109/ICAIIHI57871.2023.10489656.

[6] Bast, R.C., Hennessy, B., & Mills, G.B. (2009). The biology of ovarian cancer: new opportunities for translation. Nature Reviews Cancer, 9(6), 415-428.

[7] Medeiros, L.R., Rosa, D.D., da Rosa, M.I., & Bozzetti, M.C. (2009). Accuracy(L) of CA 125 in the diagnosis of ovarian tumors: a quantitative systematic review. European Journal of Obstetrics & Gynecology and Reproductive Biology, 142(2), 99-105.

[8] Moore, R.G., Miller, M.C., Disilvestro, P., Landrum, L.M., Gajewski, W., Ball, J.J., & Skates, S.J. (2012). Evaluation of the diagnostic Accuracy(L) of the Risk of Ovarian Malignancy Algorithm in women with a pelvic mass. Obstetrics and Gynecology, 118(2), 280-288.

[9] Kinkel, K., Hricak, H., Lu, Y., Tsuda, K., & Filly, R.A. (2005). US characterization of ovarian masses: a meta-analysis. Radiology, 217(3), 803-811.

[10] Bookman, M.A. (2004). Standard treatment in advanced ovarian cancer in 2005: the state of the art. International Journal of Gynecological Cancer, 15(3), 212-220.

[11] Cruz, J.A., & Wishart, D.S. (2006). Applications of machine learning in cancer prediction and prognosis. Cancer Informatics, 2, 59-77.

[12] Chawla, N.V., Bowyer, K.W., Hall, L.O., & Kegelmeyer, W.P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357.

[13] Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. Proceedings of ICNN'95 - International Conference on Neural Networks, 4, 1942-1948.

[14] Wolpert, D.H. (1992). Stacked generalization. Neural Networks, 5(2), 241-259.

[15] Lu, M.; Fan, Z.; Xu, B.; Chen, L.; Zheng, X.; Li, J.; Znati, T.; Mi, Q.; Jiang, J. Using machine learning to predict ovarian cancer. Int. J. Med. Inform. 2020, 141, 104195.

[16] M. T. Anwar, Automatic Complaints Categorization Using Random Forest and Gradient Boosting, Advance Sustainable Science, Engineering and Technology, vol. 3, no. 1, p. 0210106, Apr. (2021), doi: 10.26877/asset.v3i1.8460.
[17] Darmawan, R. A., & Hidayat, E. Y. Comparison of gradient boosting and random forest models in the detection system of Rakaat during prayer. Advance Sustainable Science, Engineering and Technology, (2024),6(1), 02401015. https://doi.org/10.26877/asset.v6i1.17886.