This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/CSDE50874.2020.9411607, 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)

Biomarker CA125 Feature Engineering and Class Imbalance Learning Improves Ovarian Cancer Prediction

Xiaoyan Yang
School of Computer Science
The University of Sydney
Sydney, Australia
xyan2645@uni.sydney.edu.au

Matloob Khushi
School of Computer Science
The University of Sydney
Sydney, Australia
matloob.khushi@sydney.edu.au

Kamran Shaukat
School of Electrical Engineering and Computing
The University of Newcastle
Callaghan, NSW 2308, Australia
Kamran.shaukat@uon.edu.au

***Abstract*—Ovarian cancer is a fatal female reproductive cancer because it has no specific clinical manifestations and effective screening methods in the early stage. When it is discovered, it is already an advanced stage with a low cure rate. Therefore, it is of great significance to improve the diagnostic ability of early screening for ovarian cancer. In this study, we feature engineered CA125 by calculating the rate of change of CA125 in addition to selecting a few top-ranked important features from PLCO ovarian cancer dataset. The dataset was extremely imbalanced; the imbalance ratio (ratio of negative samples in the majority class to positive examples in the minority class) was 143.7. Twenty-three types of class-imbalanced learning methods were used in this study to improve the predictive ability of the model. We identified the decision tree method with the highest AUC value among nine classic classifiers (decision tree, AdaBoost, etc.) to build a model with the class imbalance method, and showed their comparison. We identified the decision tree using SVMSMOTE has the most robust predictive ability for ovarian cancer, with a PPV of 0.9041, AUC of 0.9532, the sensitivity of 0.7792, and specificity of 0.9982. The high PPV shown by the model selected in this study indicates that the true positive samples predicted by the model account for 90.41%. Compared with other studies, the PPV of this study increased by 81.3%. This study helps to improve the accuracy of early screening for ovarian cancer and makes the diagnosis of ovarian cancer more reliable.**

***Index Terms*—Machine Learning, Feature Engineering, Class Imbalance, Ovarian Cancer**

I. Introduction

Ovarian cancer is a female reproductive tumour that has rarely been paid attention to but has a very high fatality rate [1]. It is high mortality because it has no obvious symptoms in the early stage. When the patient is found to have cancer, it is usually a malignant tumour that is difficult to cure. The success rate of early treatment of ovarian cancer is high, the earlier the disease is detected, the better the danger can be avoided [2], [3]. Therefore, it is essential to carry out early ovarian cancer screening for women and improve the accuracy of the prediction method, which can save more women's lives.

Detection of serum CA125 (tumour marker) significantly elevated is currently one of the main methods of screening for ovarian cancer. However, the use of this method does not provide highly accurate screening results. Feature engineering has been widely used in the medical informatics, and other domains [4]–[8]. Therefore, this project uses the change rate of CA125 and combines feature engineering to select some factors related to the diagnosis of ovarian cancer. Then, apply them to machine learning to build predictive models to improve the ability to identify patients with ovarian cancer. In addition, the majority of negative samples in the majority of ovarian cancer data sets are far more than the minority positive samples. When using machine learning to predict it, the classification will be more skewed towards the majority class, resulting in inaccurate predictions by the model [9]. Many pieces of research use class-imbalanced learning to solve such problems [10]–[12]. This study used class imbalance learning and classifiers to build a machine learning model to evaluate cases, to improve the predictive ability of early ovarian cancer patients.

II. Literature Review

A. Research status of using CA125 to predict ovarian cancer

Cancer Antigen 125 (CA125) is an important marker of ovarian cancer [13]. Many studies use CA125 as a predictor of ovarian cancer [14]–[18]. The higher the level of CA125, the greater the risk of ovarian cancer. It is used in many stages of ovarian cancer, whether it is early diagnosis or later cancer treatment [19]. Jacobs et al. used age, ultrasound score, menopausal status, clinical impression score and serum CA125 level as features to distinguish benign and malignant ovarian tumours in 1990 [20]. The result of their experiment was 81% sensitivity and 75% specificity. In 1995, Skates et al. conducted a horizontal screening test of longitudinal markers on menopausal women without ovarian cancer in Stockholm [21]. They used linear regression to predict ovarian cancer, using the logarithm of (CA125II+4) as the main predictor. The intercept and slope of the model are used to distinguish different groups of people. In Skates' study, the slope of CA125 can achieve higher predictive power than using CA125 alone. The result of his experiment was 99.7% specificity and 16% positive predictive value (PPV).

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/CSDE50874.2020.9411607, 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)

In 2012, the longitudinal assessment of CA125 speed was applied to the field of early screening for ovarian cancer by Xu et al. [22]. They used part of the data in the PLCO dataset for ovarian cancer screening to training the model. They used the rate of change of CA125 (where the difference between the last two CA125 measurements is the numerator and the corresponding difference between the two detection times is the denominator) as the core feature to predict ovarian cancer positive. Their model predicts a sensitivity of 62.5%, a specificity of 98.4%, and a PPV of 9.1%. Sasamoto et al. also used five data sets related to ovarian cancer (PLCO, NEC, NHS, NHSII, EPIC) to predict cancer in 2019 [23]. They used a binary classification model to study all PLCO ovarian cancer data. Then a third of the PLCO data set was used to combine with other data sets to construct a linear CA125 model (this was used to show the variability of CA125). They also tried to screen some characteristics, such as women's ethnicity, smoking status, etc. In their experiment, the model AUC using the entire PLCO ovarian cancer data set was 0.64. The actual value of CA125 was linearly correlated with the predicted value of the model but not the same. Therefore, it is essential to explore the predictive characteristics of ovarian cancer. The above about using CA125 levels and calculating related values as features to predict ovarian cancer provides strong theoretical support for this study.

B. Class-imbalanced Learning

Data imbalance refers to the relative disparity in the number of different categories of data in the data set. The majority category accounts for a larger number, and the minority category accounts for a smaller percentage [24]. In the ovarian cancer data set, it is also an unbalanced data set, where most of the classes are negative samples, and the minority are positive samples. In the field of data mining, the classifier is prone to skew the classification to the majority class when predicting such a data set. In order to solve the problem of classification deviation, class imbalance learning is proposed to be applied to machine learning to improve the prediction ability of the classification model. Class imbalance learning can be divided into three categories: data-level methods, algorithm-level methods, and hybrid methods [25].

Data-level methods are relatively widespread among imbalanced learning methods, and they often occur in the data preprocessing. The data-level method is to resample the majority or minority classes in the data set to adjust the proportion of various samples in the data set. In order to reduce skew in the data set as much as possible, it processes the data from different classes. Data-level methods can be mainly divided into three methods: undersampling, oversampling, and hybrid sampling [26]. The algorithm-level method solves the class imbalance problem by assigning different weights to the minority class, and the majority class increases the preference for the minority class samples and reduces the attention to the majority class samples to improve the classification accuracy [27]. The hybrid methods combine data-level methods with the algorithm-level methods to form some ensemble methods to improve the predictive ability of the model [28].

III. Method

This research mainly has the following stages: data preprocessing (data extraction, calculation of CA125 rate of change, data cleaning), feature engineering, data scaling and split data set, selection of baseline classifiers, combine the baseline classifier with class imbalance learning to find a predictive model with a strong ability to identify patients. The following will elaborate on the work content.

A. Data Preprocessing

The data used in this study is the data in The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. PLCO is the National Cancer Institute (NCI) in the United States that conducts an annual screening test for women older than 55 years old (n=39103) [29]. The women participating in the test were divided into two groups: the intervention group and the routine care group. Women in the intervention group received an annual CA125 review for six years, while the routine group did not undergo CA125 testing. NCI also asked these women about their backgrounds, such as age and family genetic history. This data set belongs to the imbalanced data set, in which the confirmed samples of ovarian cancer are far fewer than the normal samples.

In studies on the prediction of ovarian cancer, the medication history, family medical history, and surgical history of the samples were also related to the diagnosis of ovarian cancer, so they were included in this study for subsequent feature screening [22], [23]. However, there are many samples in this data set that are not valuable for research or have missing values, so the data will be initially extracted to make the samples in the data set meaningful.

We also calculated the rate of change of CA125 level by taking the difference between the last two measurements of CA125 divided by the corresponding measurement time difference. The final calculated rate of change in the CA125 level of the sample represents the change in their daily CA125 level.

B. Feature Engineering

The feature engineering of this study is mainly based on finding important features using F test in the analysis of variance to perform univariate screening of 27 features (only the relationship between a single variable and the label is studied). We ranked the F scores of 27 features and features with F score larger than 2 are selected as the predictive features in training the model. This study uses the 10-Fold cross-validation. Too few positive samples per fold may cause the model's PPV value to fluctuate wildly. Therefore, in order to ensure the rigour of the research, this research ignores the feature of 'whether the hysterectomy' contains a large number of null values to ensure that the positive samples in the data set are not extremely scarce.

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/CSDE50874.2020.9411607, 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)

Finally, two numerical features used in this study are the latest value of CA125 level and the rate of change of CA125. In addition, five categorical features are used including i) age at Menopause ii) the number of ovaries removed iii) whether they have had benign ovarian tumours, iv) ever take female hormones and v) ever take birth control pills. The factor 'whether ovarian cancer is confirmed' is used as the predictive label of this study. There are 26449 normal samples and 184 confirmed samples in the data set used in subsequent studies. This data set is hugely class imbalanced, and its imbalance ratio is 143.7.

C. Data Scaling and Model Verification Method

Scaling the data to the same range helps reduce the model's preference for individual features. We used MinMaxScaler to scale numerical features and OridinalEncoder to convert categorical features. This study uses 10-Fold cross-validation to train the model. We averaged the verification results of 10 training sessions to obtain a model evaluation result of 10-Fold cross-validation. Then we repeated the above process five times and averaged result of 5 times 10-Fold cross-validation, making the result more robust. It is worth to note that our samples in the test dataset were real data samples without applying any class imbalance processing. Such treatment makes the research more reliable.

D. Baseline Classifier Selection and Imbalance Learning

This study uses nine classical classifiers in machine learning: Logistic Regression, K-Nearest Neighbor (KNN) Classifier, Decision Tree, Random Forest, Gradient Boosting Decision Tree (GBDT), Adaptive Boosting (AdaBoost), eXtreme Gradient Boosting (XGBoost), Multilayer Perceptron (MLP), Bagging (SVC). These nine classifiers, without the participation of any imbalance learning method, use the inherent classifier to classify and predict the data set. The different results obtained from the nine classifier training models are used for comparison. The model with the best final effect will be used as the baseline model. It will be combined with the imbalance learning method to improve the predictive ability of the model.

This research mainly discusses imbalance learning in data level method and hybrid methods. In the data level, the under-sampling methods used in this study are Random Under-Sampling (RUS), ALL-KNN, Cluster Centroids, EditedNN, RepeatENN, Instance Hardness Threshold (IHT), NearMiss, Neighbourhood Cleaning Rule (NCR), One-Sided Selection (OSS), Tomek Link (TL). Oversampling methods: random oversampling (ROS), adaptive synthetic sampling (ADASYN), SMOTE, SVMSMOTE, SMOTENC, BorderlineSMOTE, KmeansSMOTE. Hybrid sampling methods: SMOTEENN, SMOTETomek. The hybrid-level methods used in this study are BalancedBagging, RUSBoost, Easyensemble, BalancedRandomForest. Besides, because the hybrid methods already have classifiers involved, it is not combined with the baseline classifier selected in this study, but directly with other model comparisons.

E. Evaluation Method

1) Evaluate the Classification Model: In machine learning, the classification ability of a model is often measured by a confusion matrix [30].

TABLE I
CONFUSION MATRIX

		Predicted	
Data Class		Negative	Positive
Actual	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

It compares the predicted value obtained after each training model with the true value in training set to determine how many samples are correctly classified and which are incorrectly judged. The specific evaluation form is shown table I.

2) Imbalance Ratio: This study uses the imbalance ratio to measure the proportion of the majority samples and the minority samples in the training set [31].

$$IR = \frac{Samples_{Majority}}{Samples_{Minority}} \tag{1}$$

The data-level method in imbalance learning will adjust the distribution of each class samples in the data set. We can understand the impact of imbalance learning on the predictive ability of the model by observing the changes in the number and distribution of samples in the training set. The calculation method of the imbalance ratio is shown in the Eq 1.

IV. RESULTS

A. Baseline Classifier

In this study, nine different classic classifiers were selected, and 10-FOLD cross-validation was performed five times to calculate the average AUC of the model. When evaluating the classifier, the higher the AUC, the stronger the model's ability to predict ovarian cancer. The higher the value of PPV, the stronger the model's ability to diagnose positive cases. In Fig.1, the AUC and PPV values of the nine classifiers are shown.

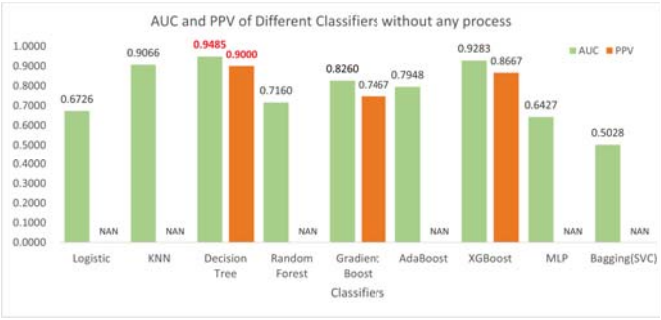

Fig. 1. Baseline Classifier

From the comparison results of nine classifiers, the decision tree obtained the highest AUC value of 0.9485, and it also obtained the highest PPV value of 0.9000. LinearSVC using

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/CSDE50874.2020.9411607, 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)

bagging performed the worst, with an AUC value of only 0.5028. Based on the above situation, this study chooses a decision tree as the final baseline classifier.

B. Classifier Combine with Imbalance Learning

In this study, the AUC values of the model obtained by applying the class imbalance method to the decision tree and the combined imbalance model are shown in Fig. 2. In the figure, green represents the ensemble methods, blue represents the hybrid sampling method, yellow represents the under-sampling methods, and purple represents the over-samplings. The orange dashed line represents the AUC of the baseline classifier.

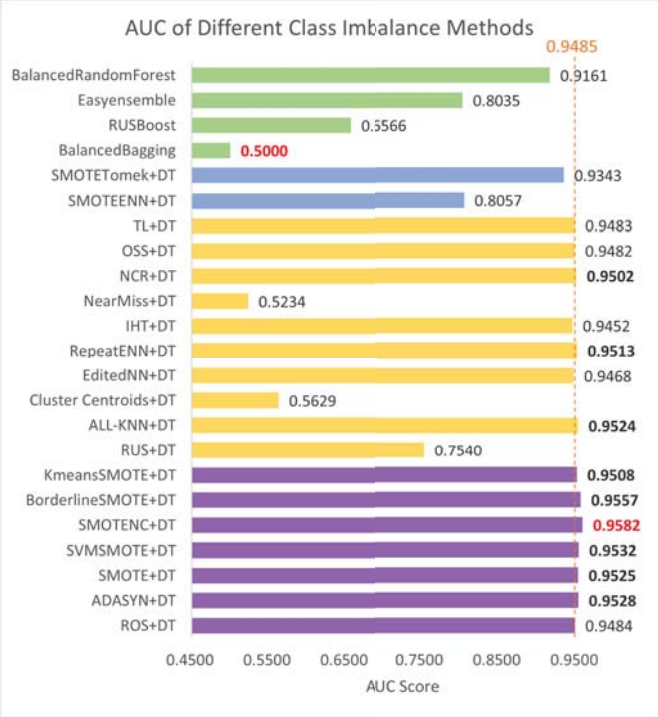

Fig. 2. AUC of Imbalance Learning Methods

In the comparative imbalance learning model results, three undersampling methods (NCR, Repeat ENN, ALL-KNN) and six oversampling methods (KmeansSMOTE, BorderlineSMOTE, SMOTENC, SVMSMOTE, SMOTE, ADASYN) are higher than the baseline. It shows that the oversampling method is more suitable to be used to construct the classification model of ovarian cancer. On the other hand, we also calculated the standard deviation (SD) of various methods. The SD of the hybrid method is 0.1806, the SD of hybrid sampling methods is 0.0910, the SD of the under-sampling method is 0.1792, and the SD of the over-sampling method is 0.0032. The comparison shows that the hybrid method has the largest fluctuation when predicting ovarian cancer, and the performance of the oversampling method is relatively stable.

C. Imbalance Ratio in Imbalance Learning

The data-level method in imbalance learning will resample the samples in the training set. We counted the number of samples adjusted by each resampling method and calculated the imbalance rate (IR) in the training set. The results are shown in table II.

TABLE II
CLASS DISTRIBUTION FOR DATA-LEVEL METHODS

Method	IR	Negative	Positive
Baseline	143.92	23804.00	165.40
ROS	1.00	23804.00	23804.00
ADASYN	1.00	23804.00	23774.60
SMOTE	1.00	23804.00	23804.00
SVMSMOTE	1.81	23804.00	13165.40
SMOTENC	1.0 0	23804.00	23804.00
BSMOTE	1.00	23804.00	23804.00
KMeansSMOTE	1.00	23804.00	23804.00
RUS	1.00	165.40	165.40
ALLKNN	140.78	23285.40	165.40
Cluster Centroids	1.00	165.40	165.40
EditedNN	141.21	23356.80	165.40
RepeatENN	139.95	23147.80	165.40
IHT	126.14	20864.20	165.40
NearMiss	1.00	165.40	165.40
NCR	141.22	23357.40	165.40
OSS	142.23	23525.20	165.40
Tomek Linkes	143.50	23735.00	165.40
SMOTEENN	0.96	21909.08	22936.04
SMOTETomek	1.00	23692.60	23692.60

IR= Imbalance Ratio, ROS = Random Over-Sampling, BSMOTE = Borderline SMOTE, RUS = Random Under-Sampling, IHT = Instance Hardness Threshold, NCR = Neighbourhood Cleaning Rule, OSS = One Sided Selection

In the performance of IR value, the IR of most oversampling and hybrid methods is around 1. However, only a few under-sampling methods have IR close to 1, and other under-sampling IRs are still greater than 100.

D. Further Comparison of Models

When predicting cancer, the PPV value can be regarded as the probability that the model predicts a positive case of ovarian cancer. The larger the PPV of the model, the more it can be used to diagnose whether ovarian cancer is positive. In order to better diagnose the samples for ovarian cancer, we compared the PPV values of the models that passed the baseline and combined their AUC and sensitivity. The results are shown in table III.

TABLE III
FURTHER COMPARE THE SELECTED MODELS

Model	PPV	AUC	Sensitivity
ADASYN+DT	0.7725	0.9528	0.7898
SMOTE+DT	0.7797	0.9525	0.7898
SVMSMOTE+DT	0.9041	0.9532	0.7792
SMOTENC+DT	0.7548	0.9582	0.8069
BorderlineSMOTE+DT	0.8802	0.9557	0.7845
KmeansSMOTE+DT	0.7849	0.9508	0.7845
ALL-KNN+DT	0.6276	0.9524	0.8287
RepeatENN++DT	0.5109	0.9513	0.8398
NCR+DT	0.7008	0.9502	0.8015

DT= Decision Tree,NCR = Neighbourhood Cleaning Rule

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/CSDE50874.2020.9411607, 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)

In the figure, the PPV of SVMSMOTE is 0.9041, which is much higher than that of other methods. Although its AUC is not the highest, it ranks second, only 0.005 less than SMOTENC. Moreover, in 'sensitivity', all methods higher than SVMSMOTE have lower PPV and AUC, far inferior to SVMSMOTE. This study not only considers AUC when diagnosing cancer patients, but also considers the contribution of PPV to positive samples, and then considers sensitivity. In summary, the cancer prediction model finally selected in this study is a decision tree using SVMSMOTE.

E. Compare with Other Related Research

Stakes et al. used Stockholm dataset to predict ovarian cancer cases [21]. Xu et al. also used the part of PLCO related to ovarian cancer for research [22]. The researchers compared the SVMSMOTE decision tree using class imbalance learning with other studies in Fig.3.

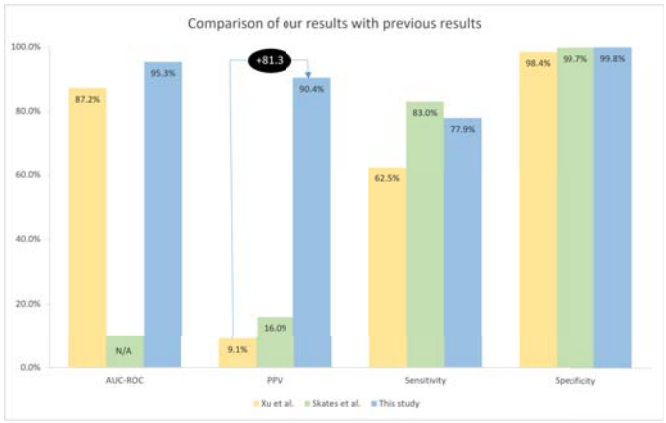

Fig. 3. Compare Our Model with Other Studies

It can be found from the results in the figure that the results of this study surpass other people's studies in AUC, PPV and specificity, and sensitivity is similar to skates. It is worth noting that the PPV obtained by the model selected in this study is much higher than that of the other two studies, indicating that the SVMSMOTE decision tree helps predict patients with positive ovarian cancer.

V. Discussion

In this study, the decision tree with the highest AUC value was selected as the baseline model from nine classic classifiers, because the better the classification ability of the baseline model, the better the prediction ability of the model that applies imbalance learning. Besides, the data set used in the research is extremely imbalanced. This paper explored the principle of data-level imbalance method to improve the model's predictive ability by exploring the sample distribution generated by the imbalance learning method in the training set. The study found that the oversampling method achieves the balance of classes in the data set by increasing the number of positive samples (minority class) to make it close to the number of negative samples (majority class). The under-sampling method reduces the gap between the two categories as much as possible by discarding some minority samples. The comprehensive sampling method not only adjusted the negative samples but also increased the number of positive samples. After applying class imbalance learning, the oversampling method shows its high stability and obtains the lowest standard deviation. Moreover, 6 model methods use over-sampling methods that have better predictive ability than baseline. On the contrary, the standard deviation of ensemble methods is the highest, and no method is better than the baseline classifier. Moreover, the same re-sampling method, the under-sampling method is not as stable as the over-sampling method, and the performance is not as good as the over-sampling method. Only three methods have AUC higher than the baseline. It shows that when dealing with extremely imbalance ovarian cancer data sets, the oversampling method is more worthy of being applied than other imbalance methods.

Based on the research purpose of this study, "In order to more accurately diagnose patients with ovarian cancer", the researchers integrated the three parameters of AUC, PPV, and Sensitivity to screen further the models that exceeded the baseline. Finally, the SVMSMOTE model was selected. In the case of ensuring high AUC, this study pays more attention to the model's PPV, because the higher the model's PPV, the more reliable the model is to diagnose ovarian cancer. Compared with the study of Xu et al., which uses the PLCO data set, the AUC of the model combining SVMSMOTE and decision tree selected in this study increased by 8.1%, PPV increased by 81.3%, specificity increased by 0.1%, and sensitivity increased by 14.8%. It shows that the model proposed in this study is meaningful for the diagnosis of ovarian cancer.

VI. Conclusion

In order to improve the predictive ability of ovarian cancer-positive patients, this study shows that the rate of change of CA125 and 6 other features selected by feature engineering improves the prediction of ovarian cancer. Since the number of malignant tumour samples and benign samples in this data set is very different, therefore class-imbalanced methods needed to be investigated. We show that class imbalance learning applied to the best-performing decision tree among the nine classic classifiers improves the model's predictive ability for positive cases.

Experiments show that among all the models that apply imbalance learning, the decision tree combined with SVMSMOTE has the best comprehensive performance, and it shows the highest PPV of 0.9041. Moreover, its AUC (0.9532), sensitivity (0.7792) and specificity (0.9982) are among the top. The most important contribution of this research is that the diagnostic ability of the proposed model for positive cases has been greatly improved. Compared with other scholars doing similar research, this research has achieved surpassing in AUC, PPV, and specificity. Especially the PPV increased by 81.3%. PPV represents the probability that samples diagnosed as positive by the model have ovarian cancer. The improved PPV indicates that the model is more reliable in diagnosing cancer. This discovery has important significance in the early

screening stage of ovarian cancer, and it improves the accuracy of diagnosis. In previous screenings for ovarian cancer, misdiagnosing women as positive and treating them may be counterproductive. Among the characteristics selected in this study, female reproductive surgery is related to the diagnosis of ovarian cancer, which also shows that misdiagnosis can cause significant harm to women. This study improves the model's ability to predict actual positive patients and enhances the accuracy of early screening for ovarian cancer, hoping to save more lives.

In future studies, more forms of CA125 levels could be considered in addition to finding more features related to the early prediction of ovarian cancer. Furthermore, new imbalance learning methods can also be explored to improve the problem of data skewness caused by class imbalance.

ACKNOWLEDGMENT

The authors thank the National Cancer Institute for access to NCI's data collected by the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

REFERENCES

[1] N. Howlader, A. Noone, M. Krapcho, D. Miller, A. Brest, M. Yu, J. Ruhl, Z. Tatalovich, A. Mariotto, D. Lewis *et al.*, "Seer cancer statistics review, 1975-2017, national cancer institute," *Bethesda, MD, April*, 2020.

[2] E. J. Pavlik *et al.*, *Ovarian Cancer Screening*. MDPI AG-Multidisciplinary Digital Publishing Institute, 2018.

[3] A. Gentry-Maharaj and U. Menon, "Screening for ovarian cancer in the general population," *Best Practice & Research Clinical Obstetrics & Gynaecology*, vol. 26, no. 2, pp. 243–256, 2012.

[4] H. Barlow, S. Mao, and M. Khushi, "Predicting high-risk prostate cancer using machine learning methods," *Data*, vol. 4, no. 3, p. 129, 2019.

[5] T. M. Alam, K. Shaukat, I. A. Hameed, S. Luo, M. U. Sarwar, S. Shabir, J. Li, and M. Khushi, "An investigation of credit card default prediction in the imbalanced datasets," *IEEE Access*, 2020.

[6] S. K. P. J. K. M. M. Philip Mehrgardt, Seid Miad Zandavi and M. Khushi, "U-net segmented adjacent angle detection (usaad) for automatic analysis of corneal nerve structures," *Data*, vol. 5, no. 2, 2020.

[7] A. Panta, M. Khushi, U. Naseem, P. Kennedy, and D. Catchpoole, "Classification of neuroblastoma histopathological images using machine learning." *Neural Information Processing. ICONIP 2020. Lecture Notes in Computer Science*, vol. 12534, 2020.

[8] M. Khushi, J. Naseem, U amd Du, A. Khan, and S. K. Poon, "Data mining encode data predicts a significant role of sina3 in human liver cancer." *Neural Information Processing. ICONIP 2020. Lecture Notes in Computer Science*, vol. 12534, 2020.

[9] J. Attenberg and S. Ertekin, "Class imbalance and active learning," *Imbalanced Learning: Foundations, Algorithms, and Applications*, pp. 101–149, 2013.

[10] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou, "On the class imbalance problem," in *2008 Fourth international conference on natural computation*, vol. 4. IEEE, 2008, pp. 192–201.

[11] D.-C. Li, C.-W. Liu, and S. C. Hu, "A learning method for the class imbalance problem with medical data sets," *Computers in biology and medicine*, vol. 40, no. 5, pp. 509–518, 2010.

[12] N. F. Hordri, S. S. Yuhaniz, N. F. M. Azmi, and S. M. Shamsuddin, "Handling class imbalance in credit card fraud using resampling methods," *Int. J. Adv. Comput. Sci. Appl*, vol. 9, no. 11, pp. 390–396, 2018.

[13] R. Bast, M. Feeney, H. Lazarus, L. Nadler, R. Colvin, R. Knapp *et al.*, "Reactivity of a monoclonal antibody with human ovarian carcinoma." *The Journal of clinical investigation*, vol. 68, no. 5, pp. 1331–1337, 1981.

[14] R. C. Bast Jr, T. L. Klug, E. S. John, E. Jenison, J. M. Niloff, H. Lazarus, R. S. Berkowitz, T. Leavitt, C. T. Griffiths, L. Parker *et al.*, "A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer," *New England journal of medicine*, vol. 309, no. 15, pp. 883–887, 1983.

[15] V. Dochez, H. Caillon, E. Vaucel, J. Dimet, N. Winer, and G. Ducarme, "Biomarkers and algorithms for diagnosis of ovarian cancer: Ca125, he4, rmi and roma, a review," *Journal of ovarian research*, vol. 12, no. 1, p. 28, 2019.

[16] B. Y. Karlan, J. Thorpe, K. Watabayashi, C. W. Drescher, M. Palomares, M. B. Daly, P. Paley, P. Hillard, M. R. Andersen, G. Anderson *et al.*, "Use of ca125 and he4 serum markers to predict ovarian cancer in elevated-risk women," *Cancer Epidemiology and Prevention Biomarkers*, vol. 23, no. 7, pp. 1383–1393, 2014.

[17] G. Rustin, "Use of ca-125 to assess response to new agents in ovarian cancer trials," *Journal of Clinical Oncology*, vol. 21, no. 10S, 2003.

[18] J. Tcherkassova, C. Abramovich, R. Moro, C. Chen, R. Schmit, A. Gerber, and R. Moro, "Combination of ca125 and recaf biomarkers for early detection of ovarian cancer," *Tumor Biology*, vol. 32, no. 4, pp. 831–838, 2011.

[19] N. Scholler and N. Urban, "Ca125 in ovarian cancer. biomark med 1 (4): 513–523," 2007.

[20] I. Jacobs, D. Oram, J. Fairbanks, J. Turner, C. Frost, and J. Grudzinskas, "A risk of malignancy index incorporating ca 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer," *International Journal of Gynecology & Obstetrics*, vol. 35, no. 3, pp. 292–292, 1991.

[21] S. J. Skates, F.-J. Xu, Y.-H. Yu, K. Sjövall, N. Einhorn, Y. Chang, R. C. Bast, and R. C. Knapp, "Toward an optimal algorithm for ovarian cancer screening with longitudinal tumor markers," *Cancer*, vol. 76, no. S10, pp. 2004–2010, 1995.

[22] J.-L. Xu, J. Commins, E. Partridge, T. L. Riley, P. C. Prorok, C. C. Johnson, and S. S. Buys, "Longitudinal evaluation of ca-125 velocity and prediction of ovarian cancer," *Gynecologic oncology*, vol. 125, no. 1, pp. 70–74, 2012.

[23] N. Sasamoto, A. Babic, B. A. Rosner, R. T. Fortner, A. F. Vitonis, H. Yamamoto, R. N. Fichorova, L. J. Titus, A. Tjønneland, L. Hansen *et al.*, "Development and validation of circulating ca125 prediction models in postmenopausal women," *Journal of ovarian research*, vol. 12, no. 1, p. 116, 2019.

[24] C. C. Johnson, B. Kessel, T. L. Riley, L. R. Ragard, C. R. Williams, J.-L. Xu, S. S. Buys *et al.*, "The epidemiology of ca-125 in women without evidence of ovarian cancer in the prostate, lung, colorectal and ovarian cancer (plco) screening trial," *Gynecologic oncology*, vol. 110, no. 3, pp. 383–389, 2008.

[25] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, p. 27, 2019.

[26] R. F. Lima and A. C. M. Pereira, "A fraud detection model based on feature selection and undersampling applied to web payment systems," in *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, vol. 3. IEEE, 2015, pp. 219–222.

[27] C. Elkan, "The foundations of cost-sensitive learning," in *International joint conference on artificial intelligence*, vol. 17, no. 1. Lawrence Erlbaum Associates Ltd, 2001, pp. 973–978.

[28] S. Ahmed, A. Mahbub, F. Rayhan, R. Jani, S. Shatabda, and D. M. Farid, "Hybrid methods for class imbalance learning employing bagging with sampling techniques," in *2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*. IEEE, 2017, pp. 1–5.

[29] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "Smoteboost: Improving prediction of the minority class in boosting," in *European conference on principles of data mining and knowledge discovery*. Springer, 2003, pp. 107–119.

[30] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," *Information Sciences*, vol. 513, pp. 429–441, 2020.

[31] N. Noorhalim, A. Ali, and S. M. Shamsuddin, "Handling imbalanced ratio for class imbalance problem using smote," in *Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017)*. Springer, 2019, pp. 19–30.