IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 11, 2017, accepted October 6, 2017, date of publication October 17, 2017, date of current version December 5, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2763984

Data-Driven Diagnosis of Cervical Cancer With Support Vector Machine-Based Approaches

WEN WU[1] AND HAO ZHOU[2]

[1]Department of Blood Transfusion, Jinan Military General Hospital, Jinan 250031, China
[2]Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Harbin 150001, China

Corresponding author: Hao Zhou (15754603750@163.com)

ABSTRACT Cervical cancer, as the fourth most common cause of death from cancer among women, has no symptoms in the early stage. There are few methods to diagnose cervical cancer precisely at present. Support vector machine (SVM) approach is introduced in this paper for cervical cancer diagnosis. Two improved SVM methods, support vector machine-recursive feature elimination and support vector machine-principal component analysis (SVM-PCA), are further proposed to diagnose the malignant cancer samples. The cervical cancer data are represented by 32 risk factors and 4 target variables: Hinselmann, Schiller, Cytology, and Biopsy. All four targets have been diagnosed and classified by the three SVM-based approaches, respectively. Subsequently, we make the comparison among these three methods and compare our ranking result of risk factors with the ground truth. It is shown that SVM-PCA method is superior to the others.

INDEX TERMS Cervical cancer, data-driven, SVM classification, SVM-RFE, SVM-PCA.

I. INTRODUCTION

As the current trend of automation, cyber-physical system (CPS) receives considerably increasing attention in research and application domains [1]. Monitored by computer-based algorithms, CPS associates physical with software components tightly and has been widely applied in health care field. By using network communication and complex physical systems, medical cyber-physical systems (MCPS) monitor and control patients' bodies based on modern medical technology. In the medical field, cancer is an inevitable topic. Cancer is one of the primary causes of morbidity and mortality in the world. As the fourth most common inducement of cancer and the fourth most common cause of death from cancer [2], cervical cancer is one of the most dangerous diseases among women.

Cervical cancer, arising from the cervix, is caused by the changes of genes that control the growth and division function of the cells. This cancer has the ability to spread from the cervix to other parts in the body. In the early stage, no signs can be observed. Regular check-up is the only way to detect cervical cancer in time. The symptoms appear in the late stage, including vaginal bleeding and pelvic pain. In addition, cervical cancer may spread to other organs, like abdomen and lungs. In this stage, namely advanced cervical cancer, the symptoms may be fatigue, leg pain, bone fractures, weight loss and back pain. Magnetic resonance imaging (MRI) and diffusion-weighted imaging (DWI) techniques can detect cervical cancer to some extent [3], [4]. However, people in developing countries have low problem awareness to routine screening. Furthermore, the lack of physician expertise and limited medical equipment make cervical cancer become a major cause of mortality in low-income countries.

Human papillomavirus (HPV) is the leading cause of cervical cancer [5]. Cigarette smoking, contraceptives using, multiple pregnancies and some other factors may cause cervical cancer as well. For instance, the risk of cervical cancer will be increased by two to three times if the HPV-infected patient smokes [6]. Meanwhile, it is shown that the incidence of cervical cancer for the women using contraceptives is three times higher than those without using contraceptives. Furthermore, the occurrence will rise to four times if contraceptives are used for over 10 years. As for multiple pregnancies, female HPV-infected patients with no pregnancies enjoy lower incidence of cervical cancer compared with those with more than one fullterm pregnancies [7].

In recent years, many detection methods were proposed and applied in the field to provide timely diagnosis, including data-driven approaches. Some commonly used data-driven methods include principal component analysis (PCA), particle swarm optimization (PSO), fuzzy positivistic C-means clustering, linear regression (LR), artificial neural network (ANN), support vector machine (SVM) and so on [8]–[13].

SVM method, as one of the most popular classification approaches, was proposed by Vapnic *et al.* [14] in 1997. Based on the structural risk minimization (SRM) principle, SVM is able to realize the maximization of the margin between different classes. In recent years, SVM algorithm experiences some developments [15], [16] and has been applied to many spheres, such as airborne metal prediction [17], image co-segmentation [18], photovoltaic power forecast [19] and stock price prediction [20]. In some cases, standard SVM method is applied in combination with other algorithms, including SVM-RFE [21], least squares support vector machines (LS-SVM) [22], histograms of oriented gradients descriptor support vector machine (HOG-SVM) [23], SVM-PCA [24] and genetic algorithm support vector machine (GA-SVM) [25].

However, when it comes to the biomedical data, the shortcomings of standard SVM are revealed. Doctors usually need to know not only whether patients get a disease, but also which factor influences the disease most. Cervical cancer is hard to be detected in the early stage due to the absence of symptoms. In the diagnosis of cervical cancer, the extraction of relevant features is very important.

In this paper, we apply standard SVM, SVM-RFE and SVM-PCA to analyse the cervical cancer dataset [26] from the repository of University of California at Irvine (UCI). To the best of our knowledge, this is the first time that these approaches are applied to this dataset. Our work shows that SVM method can realize the classification of the cervical cancer. Meanwhile, combinations with RFE algorithm and PCA algorithm are able to reduce the computation burden and extract highly correlated risk factors, respectively.

The paper is organized as follows. The related approaches are reviewed in Section II. Section III focuses on the cervical cancer dataset and risk factors of cervical cancer. Then in Section IV, we discuss the performance of these three methods when applied to cervical cancer dataset. Finally, the conclusion is shown in Section V.

II. SVM AND RELEVANT APPROACHES

A. SVM

SVM approach was invented to solve data classification and regression problems by Vapnic [14]. In terms of classification, SVM can classify the coming data into different categories after training. It is in the training process that the learning model was built by dividing original data into different groups via their labels. Between the groups is a hyperplane constructed by SVM, which helps to predict the label of new data. The essence of SVM algorithm lies in the minimization:

$$\min C\sum_{i=1}^{m}[y^i cost_1(\theta^T x^i)+(1-y^i)cost_0(\theta^T x^i)]+\frac{1}{2}\sum_{i=1}^{n}\theta_j^2$$

where C means the error penalty factor, $\frac{1}{2}\sum_{i=1}^{n}\theta_j^2$ refers to the regularization term, $cost_1(\theta^T x^i)$ and $cost_0(\theta^T x^i)$ are cost function when y equals to one and zero respectively. However, the original dataset is in low-dimensional space and it is hard to realize the non-linear separation. In order to deal with this problem, the initial space need to be mapped into a higher dimension feature space. In this process, the concept of kernel function $k(x, y)$ is introduced. The new hyperplane is dependent on the linear combination of the image of feature vectors x_i with α_i. So the map relation of point x is:

$$\sum_{i=1}\alpha_i k(x_i, x) = constant$$

The sum plays an important role in measuring the relative nearness between test point and data points.

B. SVM-RFE

Although SVM algorithm can handle classification problems, it has some drawbacks. The model built by SVM takes advantage of a large amount of features, which leads to the expensive computation cost. Meanwhile, using all the features sometimes harms the final results because of the existence of noise and redundance. The SVM-RFE approach was proposed by Guyon [27] to solve the problem. SVM-RFE method is applied widely in electricity price prediction, infrared data analysis and classification of digitized mammograms [28]–[30]. In the first step, SVM-RFE applies SVM algorithm to each feature vector and accomplishes the training process. Then, the features are ranked through the training process based on weight and those features with little relevance to the prediction are removed. The elimination order depends on the relevance, in other words, the ranking. The removal will continue until only one feature, namely the most relevant one, remains.

$$w=\sum_{i=1}^{m}\alpha_i y_i x_i$$

Nonetheless, it does not mean that only one feature can accomplish the prediction perfectly. The combination of features is still essential. SVM-RFE provides a criterion for the choice of features, and we can choose the features through the ranking table. This not only reduces the risk of choosing poor features but also reduces the computation cost.

C. SVM-PCA

PCA, as one of the most effective multi-variate statistical analysis methods, has been applied to many fields successfully [31]–[33]. It decomposes the feature space into principal component subspace and redundant subspace through orthogonal linear transformation. The correlated variables are in the principal component subspace and the first principal component is the one with the highest variance. In order to realize the mapping from row vectors $x_{(i)}$ to principal component scores $t_{(i)} = (t_1, \ldots, t_m)_{(i)}$, PCA transformation can be described as a set of vectors of weights $w_{(k)} = (w_1, \ldots, w_p)_{(k)}$ mathematically.

$$t_{k(i)} = x_{(i)} w_{(k)}$$

where $i = 1, \ldots, n$, $k = 1, \ldots, m$, n and m mean the number of initial and projected dimension respectively. The weights vectors are in the following forms:

$$w_{(1)} = \underset{\|w\|=1}{\text{argmax}} \{w^T X^T X w\} = \text{argmax} \{\frac{w^T X^T X w}{w^T w}\}$$

$$w_{(k)} = \underset{\|w\|=1}{\text{argmax}} \{\|\hat{X}_k w\|^2\} = \text{argmax} \{\frac{w^T \hat{X}_k^T \hat{X}_k w}{w^T w}\}$$

where $\hat{X}_k = X - \sum_{s=1}^{k-1} X w_{(s)} w_{(s)}^T$. $\hat{X}_k$ is the remains of space X after $k-1$ subtraction. Similarly, SVM-PCA manages to eliminate the irrelevant features and features with little relativity, which increases the speed of prediction process.

III. CERVICAL CANCER DATA

The cervical cancer dataset was collected at Hospital 'Universitario de Caracas' in Caracas [26]. The data is represented by 32 risk factors, including demographic information, patient's habits and historic medical records. These features are shown in TABLE 1. Meanwhile, there are 4 target variables or labels: Hinselmann, Schiller, Cytology and Biopsy. Hinselmanns test refers to colposcopy using acetic acid. Meanwhile, colposcopy using Lugol iodine includes Schillers test, cytology and biopsy. Some patients did not answer all questions for individual privacy reason and accordingly the dataset needs to be pre-treated to deal with the missing values. Considering that the dataset belongs to imbalanced data, oversampling is applied in the pre-treatment process. After pre-treatment, risk factor 27 and 28 were removed as a consequence of lack of available values. Hence we need to analyze cervical cancer data of 668 patients with 30 features.

TABLE 1. Attributes of cervical cancer.

Number	Attributes Name
1	Age
2	Number of sexual partners
3	First sexual intercourse (age)
4	Num of pregnancies
5	Smokes
6	Smokes (years)
7	Smokes (packs/year)
8	Hormonal Contraceptives
9	Hormonal Contraceptives (years)
10	IUD
11	IUD (years)
12	STDs
13	STDs (number)
14	STDs: condylomatosis
15	STDs: cervical condylomatosis
16	STDs: vaginal condylomatosis
17	STDs: vulvo-perineal condylomatosis
18	STDs: syphilis
19	STDs: pelvic inflammatory disease
20	STDs: genital herpes
21	STDs: molluscum contagiosum
22	STDs: AIDS
23	STDs: HIV
24	STDs: Hepatitis B
25	STDs: HPV
26	STDs: Number of diagnosis
27	STDs: Time since first diagnosis
28	STDs: Time since last diagnosis
29	Dx: Cancer
30	Dx: CIN
31	Dx: HPV
32	Dx

IV. SIMULATION EXPERIMENTS AND ANALYSIS

When it comes to biomedical data, total accuracy is not the only criterion to evaluate an algorithm. It is the correct diagnosis that gets more attention. As for a dataset containing few data of malignant condition, an algorithm that predicts all the cancer samples belonging to benign cancer will sometimes enjoy a higher total accuracy than those predicting malignant samples correctly. Obviously, the former one is not what we pursued. As a consequence, sensitivity, specificity, positive predictive accuracy (PPA) and negative predictive accuracy (NPA) are applied so that the diagnosis condition can get ample explanations.

$$\begin{cases} \textit{Total Accuracy} = \dfrac{TP + TN}{TP + TN + FP + FN} \\ \textit{Sensitivity} = \dfrac{TP}{TP + FN} \\ \textit{Specificity} = \dfrac{TN}{TN + FP} \\ \textit{Positive Predictive Accuracy} = \dfrac{TP}{TP + FP} \\ \textit{Negative Predictive Accuracy} = \dfrac{TN}{TN + FN} \end{cases}$$

TP is true positive, refering to those malignant cancer samples which have been diagnosed correctly. TN means true negative, equaling to the number of uninfected people who get negative predictions. FP, false positive, is the number of samples without cervical cancer but have been classified into the positive group. Contrary to FP, FN is the number of undetected malignant cancer samples. What's more, total accuracy refers to the precision of the algorithm, which equals to the proportion of the correctly detected samples in all samples. Sensitivity, or recall, means the proportion of those malignant cancer samples which have been diagnosed correctly in all malignant samples. Similarly, specificity is the share of correctly diagnosed benign cancer samples in all benign samples. As for positive predictive accuracy, it reflects the precision of the model and refers to the ratio between samples which are malignant and are classified precisely and those malignant samples. Negative predictive accuracy is the ratio between benign samples which have been classified precisely and all the benign samples.

In this section, three SVM-based approaches are used to realize the classification of cervical cancer data. In order to evaluate the performance of these SVM-based approaches, 5-fold cross-validation is applied in the process. Four target variables, Hinselmann, Schiller, Cytology and Biopsy, will be diagnosed respectively. The total accuracy of these three methods are plotted in Figure 1 to 8. Due to the fact that SVM algorithm obtains the same result as those of SVM-RFE and SVM-PCA method with full features, the accuracy of SVM will not be plotted separately.

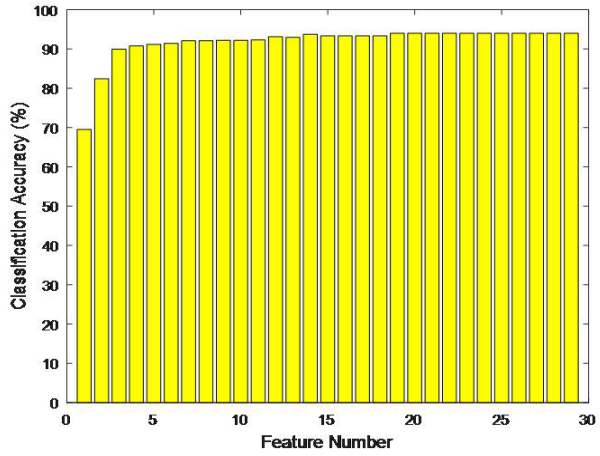


FIGURE 1. SVM-RFE on Hinselmann.

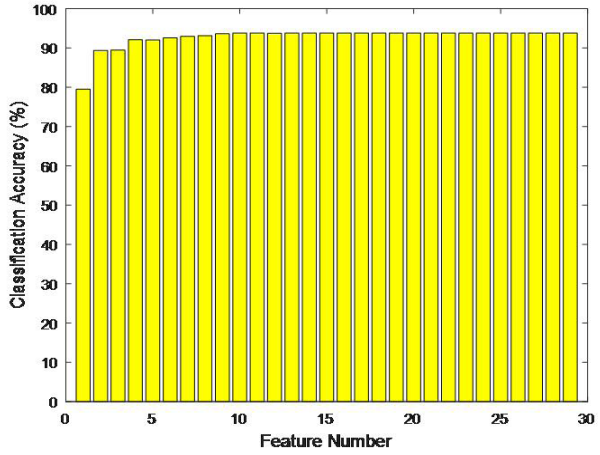


FIGURE 2. SVM-PCA on Hinselmann.

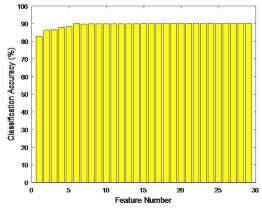


FIGURE 3. SVM-RFE on Schiller.

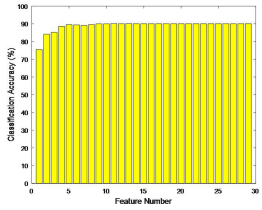


FIGURE 4. SVM-PCA on Schiller.

A. TARGET VARIABLE: HINSELMANN

Under Hinselmann's test, there are 638 benign and 30 malignant samples. The total accuracy of SVM algorithm with 30 features is 93.97% as shown in TABLE 2. In addition,

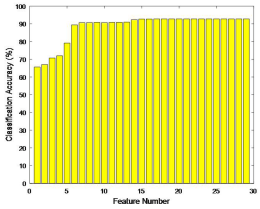


FIGURE 5. SVM-RFE on Citology.

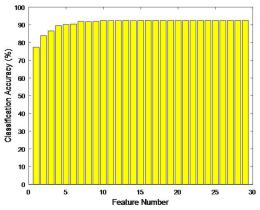


FIGURE 6. SVM-PCA on Citology.

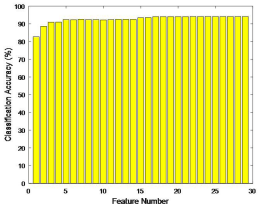


FIGURE 7. SVM-RFE on Biopsy.

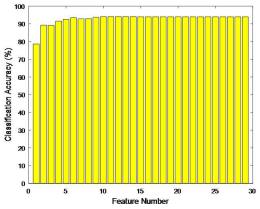


FIGURE 8. SVM-PCA on Biopsy.

TABLE 2. Target Variable (1): Hinselmann.

	SVM	SVM-RFE		SVM-PCA	
Feature Number	30	5	15	5	11
Total Accuracy (%)	93.97	90.77	93.69	92.09	93.79
Sensitivity(%)	100	100	100	100	100
Specificity(%)	89.96	84.63	89.49	86.83	89.65
PPA(%)	84.97	78.69	84.38	81.16	84.57
NPA(%)	100	100	100	100	100

its sensitivity and negative predictive accuracy reach one hundred percent. The specificity and positive predictive accuracy are up to 89.96% and 84.97% respectively. In SVM-RFE, the ranking is the first step. The sequence of risk factors according to relevance is shown in TABLE 3.

SVM-RFE works well when there are only 5 features selected. In this condition, we find that the sensitivity and negative predictive accuracy already reach 100 percent. Also, the specificity of SVM-RFE is nearly 85 percent. Similarly, its positive predictive accuracy is relatively close to the

TABLE 3. Attributes of cervical cancer on RFE sequence.

Number	Attributes Name
9	Hormonal Contraceptives (years)
4	Num of pregnancies
2	Number of sexual partners
3	First sexual intercourse (age)
13	STDs (number)
7	Smokes (packs/year)
26	STDs: Number of diagnosis
27	Dx: Cancer
29	Dx: HPV
17	STDs: vulvo-perineal condylomatosis
14	STDs: condylomatosis
12	STDs
6	Smokes (years)
30	Dx
11	IUD (years)
8	Hormonal Contraceptives
10	IUD
23	STDs: HIV
18	STDs: syphilis
5	Smokes
1	Age
16	STDs: vaginal condylomatosis
28	Dx: CIN
25	STDs: HPV
24	STDs: Hepatitis B
20	STDs: genital herpes
21	STDs: molluscum contagiosum
19	STDs: pelvic inflammatory disease
22	STDs: AIDS
15	STDs: cervical condylomatosis

standard one. At the same time, its total accuracy reaches 90 percent. When the number of features reaches 15, the performance of SVM-RFE algorithm can accomplish what is realized by SVM algorithm perfectly. Two of the indexes are the same as those of SVM and the distinction among the other three indexes is slight. After that, the performance of diagnosis nearly remains unchanged.

In comparison with SVM-RFE, SVM-PCA can manage to classify the data similarly. When 5 principal components are chosen, SVM-PCA method can basically actualize the function of SVM. The sensitivity and negative predictive accuracy are equal to those of SVM method, which are 100 percent both. As for specificities and positive predictive accuracy, although they are lower than those belonging to standard SVM, the difference is only 3 percent. The general accuracy is 92.09%. Different from SVM-RFE, SVM-PCA can work well when 11, not 15, components are trained. It performs exactly the same as SVM. On the other hand, its computation cost is lower than that of SVM definitely.

B. TARGET VARIABLE: SCHILLER

Concerning Schiller's test, the number of malignant samples reaches 63. SVM method gets the 90.18% general accuracy with Schiller variable. The sensitivity and NPA reach 99 percent. Meanwhile, the PPV and specificity are close to 80% and 85% respectively. Nevertheless, SVM-RFE can achieve the similar results with 7 features. When 18 risk factors are taken into account, the diagnosis performance of SVM-RFE is same as that of SVM. In terms of SVM-PCA, after 6 principal components extraction, the model gets the basic classification function of SVM. Taking advantage of 12 principal components, the model can accomplish what SVM obtains with 30 features. These results are shown in TABLE 4, Figure 3 and Figure 4.

TABLE 4. Target variable (2): schiller.

	SVM	SVM-RFE		SVM-PCA	
Feature Number	30	7	18	6	12
Total Accuracy (%)	90.18	90.08	90.18	89.49	90.18
Sensitivity(%)	98.73	98.73	98.73	98.99	98.99
Specificity(%)	84.63	84.46	84.63	83.14	84.30
PPA(%)	80.75	80.58	80.75	79.31	80.45
NPA(%)	99.03	99.03	99.03	99.21	99.22

C. TARGET VARIABLE: CITOLOGY

As for the Citology variable, the test shows that there are 38 malignant samples and the comparison results are in the TABLE 5. Besides two perfect diagnosis indexes, the accuracy, specificity and positive predictive accuracy of SVM are 92.75%, 87.92% and 83.00% respectively. Both SVM-RFE and SVM-PCA need 8 elements to realize the basic function. However, we find that the capability of SVM-PCA with 8 principal components is better than that of SVM-RFE with 8 relevant risk factors. In addition, at least 15 risk factors are needed in the modeling process of SVM-RFE to outperform the SVM algorithm. In spite of that, the model through SVM-PCA with 11 principal components works as well as SVM model.

TABLE 5. Target variable (3): citology.

	SVM	SVM-RFE		SVM-PCA	
Feature Number	30	8	15	8	11
Total Accuracy (%)	92.75	90.65	92.37	91.98	92.46
Sensitivity(%)	100	100	100	100	100
Specificity(%)	87.92	84.42	87.28	86.65	87.44
PPA(%)	83.00	79.10	82.26	81.54	82.44
NPA(%)	100	100	100	100	100

D. TARGET VARIABLE: BIOPSY

Compared with the other three variables, Biopsy test gives rise to different detection results. Based on Biopsy test, there are 623 benign and 45 malignant samples. The performance of these methods are shown in TABLE 6. In terms of the selection on necessary elements, SVM-RFE uses less elements to accomplish the modeling process and the model enjoys better

TABLE 6. Target variable (4): biopsy.

	SVM	SVM-RFE		SVM-PCA	
Feature Number	30	6	18	8	11
Total Accuracy (%)	94.13	92.39	94.03	93.45	94.03
Sensitivity(%)	100	100	100	100	100
Specificity(%)	90.21	87.32	90.05	89.09	90.05
PPA(%)	86.07	82.68	85.88	84.72	85.88
NPA(%)	100	100	100	100	100

interpretability than SVM-PCA. Six attributes are enough for the former algorithm. Although SVM-PCA needs 8 elements, we find that the accuracy of SVM-PCA outperforms SVM-RFE's a little. It is shown in Figure 7 and Figure 8 that the performance of SVM-RFE and SVM-PCA are similar. In order to build the model whose capability is as well as that of SVM, SVM-PCA needs only 11 components but SVM-RFE demands 18 factors. So we can conclude that even for Biospy variable, SVM-PCA excels SVM-RFE to some extent.

E. COMPARISON AND ANALYSIS

Based on the implementation on four target variables, we find that SVM-based methods can detect malignant samples and achieve the classification well. Taking advantage of 30 features, SVM method can obtain the following performance on average: the total accuracy of up to 90% and the sensitivity and negative predictive accuracy of nearly 100%. At the same time, the specificity and positive predictive accuracy are approximately 88% and 83% respectively. Both SVM-RFE and SVM-PCA methods can reduce the computation cost and make the diagnosis with less factors or components. These two methods only need at most 8 variables to reconstruct the SVM model and the fundamental classification can be done for these four targets. When the number of features comes to 18, SVM-RFE approach will fulfill the task for each of these four diagnosis. In comparison with SVM-RFE, when at most 12 principal components are provided, SVM-PCA is able to fulfill the task for the four targets. Furthermore, the significance test of these three SVM-based methods shows that are there are no significant difference among the total accuracy of these approaches and the value of significant level is 0.05.

Meanwhile, Original PCA method is applied to classify cervical cancer dataset. PCA T^2 and SPE statistics can detect faulty samples effectively. For instance, the classification results of cervical cancer data on Schiller variable are shown in Figure 9. According to the classification results, the total accuracy is 90.48%.

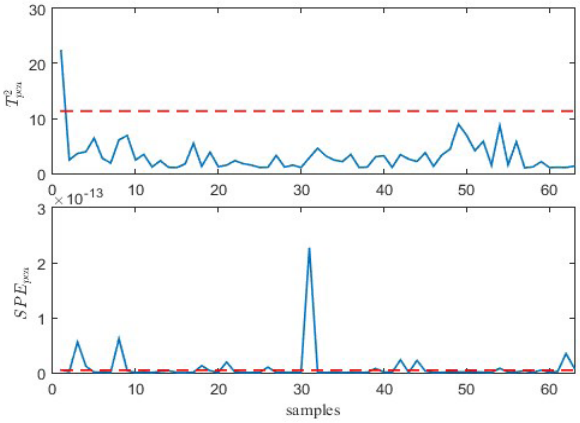


FIGURE 9. PCA for fault detection on malignant samples.

TABLE 7. Top ten relevant attributes for four variables on SVM-RFE sequence.

Hinselmann	Schiller	Citology	Biopsy
9	9	9	9
4	3	11	3
2	2	7	2
3	11	13	13
13	6	6	7
7	13	2	27
26	4	3	29
27	7	29	30
29	26	27	12
17	17	30	17

TABLE 7 focuses on the top ten relevant attributes for these four variables when SVM-RFE is applied to cervical cancer dataset. It is seen that Attribute 2, 3, 7, 9, 13, 27 and 29 appear in almost all these four columns. According to Table 1, we know the risk factors correspond to the numbers. Attribute 2 is the number of sexual partners and attribute 3 is the age when first sexual intercourse happens. Attribute 7 refers to the packs of smokes per years. Attribute 9 means the years that the patient starts to use hormonal contraceptives. Attribute 13 is the number of sexually transmitted diseases. Attribute 27 and 29 refer to the digital radiography about cancer and HIV respectively. According to modern medical research, these attributes do harm human body and increase the risk of cervical cancer. For instance, cigarette smoking, contraceptives using, multiple sexual partners and some other factors will lead to cervical cancer [6], [7]. This is consistent with our results.

V. CONCLUSION

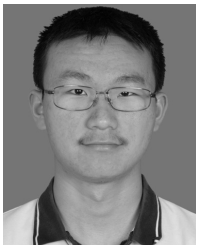
In this paper, some cervical cancer risk factors are reviewed and three SVM-based approaches are applied to the classification of cervical cancer dataset. The standard SVM method can classify the malignant cancer and benign cancer well. Both SVM-RFE and SVM-PCA are able to actualize the similar function with less features than SVM. More specifically, SVM-RFE and SVM-PCA enjoy the capability to reduce the feature numbers from 30 to 8 to accomplish the classification. Meanwhile, the classification speed can be improved prominently. Furthermore, although SVM method can classify the cervical cancer data precisely, its high computation cost shows as a limitation. SVM-RFE and SVM-PCA are able to solve the problem effectively. Compared with SVM-RFE, SVM-PCA holds better capability with same number of features.

REFERENCES

[1] S. K. Khaitan and J. D. McCalley, "Cyber physical system approach for design of power grids: A survey," in *Proc. IEEE Power Energy Soc. Gen. Meet.*, Vancouver, BC, Canada, Jul. 2013, pp. 1–5.
[2] *World Cancer Report*, World Health Org., Geneva, Switzerland, 2014.
[3] M. Exner *et al.*, "Value of diffusion-weighted MRI in diagnosis of uterine cervical cancer: A prospective study evaluating the benefits of DWI compared to conventional MR sequences in a 3T environment," *Acta Radiologica.*, vol. 57, no. 7, pp. 869–877, 2016.

[4] P. Z. Mcveigh, A. M. Syed, M. Milosevic, A. Fyles, and M. A. Haider, "Diffusion-weighted MRI in cervical cancer," *Eur. Radiol.*, vol. 18, no. 5, pp. 1058–1064, 2008.
[5] A. Gadducci, C. Barsotti, S. Cosio, L. Domenici, and A. G. Riccardo, "Smoking habit, immune suppression, oral contraceptive use, and hormone replacement therapy use and cervical carcinogenesis: A review of the literature," *Gynecol. Endocrinol.*, vol. 27, no. 8, pp. 597–604, 2011.
[6] P. Luhn, J. Walker, M. Schiffman, and R. E. Zuna, "The role of co-factors in the progression from human papillomavirus infection to cervical cancer," *Gynecol. Oncol.*, vol. 128, no. 2, pp. 265–270, 2013.
[7] (2015). *Cervical Cancer Prevention*. [Online]. Available: https://www.cancer.gov/types/cervical/hp/cervical-prevention-pdq
[8] J. V. Kresta, J. F. MacGregor, and T. E. Marlin, "Multivariate statistical monitoring of process operating performance," *Can. J. Chem. Eng.*, vol. 69, no. 1, pp. 35–47, 1991.
[9] J. L. Salmeron, S. A. Rahimi, A. M. Navali, and A. Sadeghpour, "Medical diagnosis of Rheumatoid Arthritis using data driven PSO-FCM with scarce datasets," *Neurocomputing*, vol. 232, pp. 104–112, Apr. 2017.
[10] S. Yin and Z. Huang, "Performance monitoring for vehicle suspension system via fuzzy positivistic C-means clustering based on accelerometer measurements," *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 5, pp. 2613–2620, Oct. 2015.
[11] P. L. Rodrigues, N. F. Rodrigues, J. C. Fonseca, J. Correia-Pinto, and J. L. Vilaca, "Automatic modeling of *pectus excavatum* corrective prosthesis using artificial neural networks," *Med. Eng. Phys.*, vol. 36, no. 10, pp. 1338–1345, 2014.
[12] S. Yin, H. Gao, J. Qiu, and O. Kaynak, "Descriptor reduced-order sliding mode observers design for switched systems with sensor and actuator faults," *Automatica*, vol. 76, pp. 282–292, Feb. 2017.
[13] S. Yin, H. Yang, and O. Kaynak, "Sliding mode observer-based FTC for Markovian jump systems with actuator and sensor faults," *IEEE Trans. Autom. Control*, vol. 62, no. 7, pp. 3551–3558, Jul. 2017.
[14] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.
[15] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi, "Prediction interval estimation of electricity prices using PSO-tuned support vector machines," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 322–331, Apr. 2015.
[16] F. Gruson, P. Le Moigne, P. Delarue, A. Videt, X. Cimetiere, and M. Arpilliere, "A simple carrier-based modulation for the SVM of the matrix converter," *IEEE Trans. Ind. Informat.*, vol. 9, no. 2, pp. 947–956, May 2013.
[17] X. Leng *et al.*, "Prediction of size-fractionated airborne particle-bound metals using MLR, BP-ANN and SVM analyses," *Chemosphere*, vol. 180, pp. 513–522, Aug. 2017.
[18] X. Liang, L. Zhu, and D.-S. Huang, "Multi-task ranking SVM for image cosegmentation," *Neurocomputing*, vol. 247, pp. 126–136, Jul. 2017.
[19] M. Malvoni, M. G. De Giorgi, and P. M. Congedo, "Data on support vector machines (SVM) model to forecast photovoltaic power," *Data Brief*, vol. 9, pp. 13–16, Dec. 2016.
[20] F. Wen, J. Xiao, Z. He, and X. Gong, "Stock price prediction based on SSA and SVM," *Proc. Comput. Sci.*, vol. 31, pp. 625–631, May 2014.
[21] X. Li, T. Liu, P. Tao, C. Wang, and L. Chen, "A highly accurate protein structural class prediction approach using auto cross covariance transformation and recursive feature elimination," *Comput. Biol. Chem.*, vol. 59, pp. 95–100, Dec. 2015.
[22] X. Song, L. Jian, and Y. Song, "A chunk updating LS-SVMs based on block Gaussian elimination method," *Appl. Soft Comput.*, vol. 51, pp. 96–104, Feb. 2017.
[23] A. Radman, N. Zainal, and S. A. Suandi, "Automated segmentation of iris images acquired in an unconstrained environment using HOG-SVM and GrowCut," *Digit. Signal Process.*, vol. 64, pp. 60–70, May 2017.
[24] Y. Huang, D. Wu, Z. Zhang, H. Chen, and S. Chen, "EMD-based pulsed TIG welding process porosity defect detection and defect diagnosis using GA-SVM," *J. Mater. Process. Technol.*, vol. 239, pp. 92–102, Jan. 2017.
[25] Z. Yin, Z. Fei, C. Yang, and A. Chen, "A novel SVM-RFE based biomedical data processing approach: Basic and beyond," in *Proc. 42nd Ann. Conf. IEEE Ind. Electron. Soc. (IECON)*, Florence, Italy, Oct. 2016, pp. 7143–7148.
[26] K. Fernandes, J. S. Cardoso, and J. Fernandes, "Transfer learning with partial observability applied to cervical cancer screening," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.*, 2017, pp. 243–250.
[27] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support processing systems," *Mach. Learn.*, vol. 30, nos. 2–3, pp. 271–274, 1998.
[28] Z. Shao, S. Yang, F. Gao, K. Zhou, and P. Lin, "A new electricity price prediction strategy using mutual information-based SVM-RFE classification," *Renew. Sustain. Energy Rev.*, vol. 70, pp. 330–341, Apr. 2017.
[29] F. E. Spetale, P. Bulacio, S. Guillaume, J. Murillo, and E. Tapia, "A spectral envelope approach towards effective SVM-RFE on infrared data," *Pattern Recognit. Lett.*, vol. 71, pp. 59–65, Feb. 2016.
[30] S. Yoon and S. Kim, "Mutual information-based SVM-RFE for diagnostic classification of digitized mammograms," *Pattern Recognit. Lett.*, vol. 30, no. 16, pp. 1489–1495, 2009.
[31] S. J. Qin, "Statistical process monitoring: Basics and beyond," *J. Chemometrics*, vol. 17, nos. 8–9, pp. 480–502, 2003.
[32] S. Ding *et al.*, "On the application of PCA technique to fault diagnosis," *Tsinghua Sci. Technol.*, vol. 15, no. 2, pp. 138–144, 2010.
[33] Y. Chen, "Reference-related component analysis: A new method inheriting the advantages of PLS and PCA for separating interesting information and reducing data dimension," *Chemometrics Intell. Lab. Syst.*, vol. 156, pp. 196–202, Aug. 2016.

WEN WU is a M.D. and an attending Physician with Jinan Military General Hospital. Her research interest is centered on the diagnosis of cancer. She is a Standing Member of the Clinical Transfusion Branch, Shandong Medical Doctor Association.



HAO ZHOU received the B.E. degree in automation from the Harbin Institute of Technology. He is currently pursuing the Ph.D. degree with Tsinghua University. His research interests are in data-driven fault diagnosis and deep learning.