International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

A Novel SVM-KNN Classifier for Cervical Cancer Diagnosis using Feature Reduction and Imbalanced Learning Techniques

[1]K.Lavanya, [2]Devireddy Syamala, [3]Kotha Vineetha Vani, [4]Choragudi Gipsy

ABSTRACT-- *Cervical cancer is one sort of prenatal tumors and a large portion of the complexities of cancer threatening causes to deaths which were identified in most of the countries. There are different risk factors related to cancer threatening development. The number of methodologies developed to predict this cancer such as Decision Tree (DT), K-nearest neighbors (KNN), Support vector machine (SVM), Random Forest (RF), Logistic Regression (LR), Principal Component Analysis (PCA) and Logistic Regression (LR). However, it is observed that most of the medical data suffer from class imbalance issues. The work in this paper proposed an ensemble classifier using SVM and KNN with an oversampling technique called Synthetic Minority Oversampling Technique (SMOTE) for Cervical Cancer. Also, work extended to applied set of feature reduction techniques to reduce computation tasks and to improve model accuracy. However, in this cancer data total 4 target variables: Hinselmann, Schiller, Cytology, and Biopsy are considered associated with 32 risk factors. Moreover, the study used the number of benchmarks like Accuracy, Sensitivity, Specificity, Positive Prediction Accuracy (PPA) and Negative Prediction Accuracy (NPA) for the performance analysis. The results showed that the proposed ensemble classifier method to be proven efficient for cervical cancer analysis compared to standard methods.*

Key Words*-- Classification, Cervical Cancer, Feature Selection, and Regularization Method.*

I. INTRODUCTION

Cervical cancer malignancy is one kind of gynecological disease due to irregular menstruation, spotting, weight losses and etc. There are distinctive hazard areas for each kind of gynecological malignant growths. The cervical malignant growth has capacity to spread to different organs, similar to guts and lungs. This is the most frequently occurred disease in the developing countries especially in women [1]. More than 80% of people in world are affected by the disease. Cervix harm is poisonous tumor that occurs right when the cervix tissue cells begin to create and reproduce bizarre without controlled cell division and cell end. Cervical malignancy is the fourth most

[1] *Lakireddy Bali Reddy College of Engineering), L. B. Reddy Nagar, Mylavaram, Krishna District, Andhra Pradesh, INDIA*

[2] *Lakireddy Bali Reddy College of Engineering), L. B. Reddy Nagar, Mylavaram, Krishna District, Andhra Pradesh, INDIA*

[3] *Lakireddy Bali Reddy College of Engineering), L. B. Reddy Nagar, Mylavaram, Krishna District, Andhra Pradesh, INDIA*

[4] *Lakireddy Bali Reddy College of Engineering), L. B. Reddy Nagar, Mylavaram, Krishna District, Andhra Pradesh, INDIA*

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

sickness happened among ladies. It is caused because of changes of qualities that controls the development and division of cells. The significant manifestations identified with the malignancy can't be recognized in beginning time.

The main problem is to identify the symptoms in early stage so that we can reduce the amount cancer victims. This can be done using machine learning algorithms. The learning algorithms, through which we can provide the ability to the system to learn automatically and improve from experience, by using machine learning. However, these algorithms are categorized into four types they are Supervised, unsupervised, semi supervised and reinforcement. Number of classification techniques developed to predict this cancer such as Decision Tree (DT), K-nearest neighbor (KNN), Support vector machine (SVM), Random Forest (RF), Logistic Regression (LR), Principal Component Analysis (PCA) and Logistic Regression (LR) [1-7].A simple machine learning based classification method is KNN, it can gather the information of malignant growth patients and arrange them into distinguish classes based on its nearest neighbors. The malignant assigned to class that is in KNN, where K is positive number. Moreover, support vector machine (SVM) [8], is a machine learning technique applied to various applications. But for some of its application, retained low classifying accuracy especially in medical applications due to the kernel function calculation difficulty. To resolve such kind of problems, SVM classifier is combined with the K-nearest neighbor (KNN) classifier [9] and produced excellent results to many of applications. However, a number of variations to SVM includes, SVM+RFE, SVM+PCA [10] and SVM+GA [11]. However, in medical data like cervical cancer it is very difficult to diagnose patient without knowing the influence factors. So, there is a need of identify suitable predictors using feature selection techniques prior to model prediction. In case of cervical cancer data is suffers with class imbalance and hence there is a need of applying resampling strategies along with feature selection. To do this proposed method introduced number of oversampling techniques, includes SMOTE [12], RSMOTE [13] and improved SMOTE [14] to retained class imbalance property. Later applied SVM, SVM with RF, SVM with Bayesian and SVM with MLP classifier to analyze the data collected from repository called University of California at Irvine (UCI).The work in this paper results that ensemble KNN-SVM with SMOTE algorithm reduce computational burden to the model compared standard selection methods includes MLP, RFE and PCA.

Basic classifier methods along with feature selection techniques is provided in Section 2. Later, in Section 3, work is focused on SMOTE applied to Cancer data to recover miss-classification. In Section 4, Results and Discussion of proposed method mentioned. At end work is concluded in Section 5.

II. Methods

2.1. Classification Techniques:

2.1.1. SVM:

The aim of the SVM algorithms is to minimizing an error with the help of raising the margin over two components one is hyperplane and another is data .Consider the training data set corresponds two separate

classes represented as in binary classification and in multi-classification , with this there is maximal function to coefficient to be defined as:

$$W(\beta) = \sum_{i=1}^{k} \beta_i - \frac{1}{2} \sum_{i,j=1}^{k} \beta_i \beta_j Y_i Y_j l(X_i . X_j), \quad (1)$$

$$0 \le \beta_i, i = 1, \cdots, n \text{ and } \sum_{i=1}^{k} \beta_i Y_i = 0 \quad . \quad (2)$$

Where set of samples having coefficients > 0, to be considered as support vector very near to hyperplane. Moreover, there is a discriminant function which is used to represent the rule of separating to be mentioned as:

$$f(X) = \text{sign} \sum_{i=1}^{k} \beta_i Y_i l(X_i . X) - b) \quad (3)$$

2.1.2. KNN:

Among many machine learning algorithms one of the easy and simple is K-nearest neighbor (KNN) [17, 18] .The algorithm in which sample points classified depends on K nearest neighbors. Consider the data set S= and is used for training. From this is called predictors set and is known to be class label data. The case of binary classification to be represented as and in multi-classification it to be .From the training data set S, KNN algorithm generate regions to the input space which is to be represented as:

$$R(X) = \{X^{'} \mid D_m \le \hat{d}_k\} \quad (4)$$

Where is known to be distance matrix and is to be order statics to distance matrix.

Now, from the KNN estimated posterior probability for the sample point of observation X to be represented as:

$$f(Y \mid X) = \frac{f(X \mid Y) f(Y)}{f(X)} \cong \frac{k[Y]}{k} \quad (5)$$

Next, evaluated the decision function for the predictor set 'X' using the value k[y] .Further it helps to find associated class with maximum value i.e., max(k[y]).

$$G^{*}(X) = \begin{cases} 1, k[Y=1] \ge k[Y=-1], \\ -1, k[Y=-1] \ge k[Y=1], \end{cases} \quad (6)$$

Modified to multi-classification equation (6) and is represented as:

$$G^{*}(X) = \begin{cases} j, k[Y=m] \ge k[Y=-m], m \in k \\ -j, k[Y=-m] \ge k[Y=m], m \in k \end{cases} \quad (7)$$

From (7) , the decision function which maximizes the posterior probability and the decision rule corresponds to binary classification is generated using equation from (4-7) to be represented as:

$$G^{*}(X) = sign(ave X_{i \in R(X)} Y_i) \quad (8)$$

2.1.3. SVM-KNN Classifier Algorithm:

It is observed that, classification using SVM results wrong distribution to samples near to hyperplane and also degrades classification accuracy. Such kind of cases KNN algorithms is used instead SVM. The case where samples far from the hyperplane as usual SVM Classification is used.

Algorithm: SVM-KNN Classifier

1. *if* $T_{test} \neq \Theta$, *get* $x \in T_{test}$, *if* $T_{test} = \Theta$, *stop*; *where* T_{test} is test set
2. *calculate* $G^*(X) = \sum_{i=1}^{k} \beta_i Y_i l(X_i . X) - b)$;
3. *if* $|G^*(X)| > \in$, *calculate* $G^*(X) = f(X) = \text{sign} \sum_{i=1}^{k} \beta_i Y_i l(X_i . X) - b)$ //SVM Classifier

   if $|G^*(X)| < \in$, *calculate* $G^*(X) = sign(ave X_{i \in R(X)} Y_i)$. //KNN Classifier

   Where $\in$ is distance threshold $0 < \in < 1$
4. $T = T - X$, *go to step* 1.

2.2. Features Selection Techniques: FEE and PCA

The work in this paper focused on two popular feature selection methods one among is Principle Component Analysis (PCA) and other one is Recursive Feature Elimination (RFE). The principal mechanism of both methods is dimensional reduction and are used to optimize number of features into small set. The resultant features to be considered as final set of features are helps to improve the model performance. However, in RFE model features are correlated with individual ranks and weights based on importance and is calculated using (9).

$$w = \sum_{i=1}^{k} X_i Y_i \beta_i \quad (9)$$

But in the case of PCA, feature space is equivalently converted into principal components space using orthogonal linear transformation [19, 20].It is observed that all the time highest ranking feature to be considered as first principal component.

For every sample in equivalent principal component is with associated weights and PCA transformation of input to be represented as:

$$P_i = X_i * w, \text{ where i=1,…,k} \quad (10)$$

III. Synthetic Minority Oversampling Technique (SMOTE)

In multi-classification class imbalance is one of the most important problem and it is retained using resampling and are over sampling and under-sampling. However, under-sampling will reduce some important information related to dataset .The work in this paper focus on over-sampling techniques to reduce class imbalance. The number

of over-sampling techniques includes SMOTE [21], an ensemble SMOTE with boosting procedure [22]. The new approach which combined both over-sampling and under-sampling methods [23].It is a popular method used in many applications includes breast cancer [24], intruder detection [25] and speech recognition [26]. The technique is to estimate good training models using k-nearest neighbors (i.e., k=1 of k>2) in the minority class. Later, the resampled data employed for the classification.

The overall idea described as follows:

Algorithm: SMOTE

Input: Minority Class Data Set A

Output: Resampled Data Set $\overset{*}{\text{A}}$

1. Consider the set A, for every sample $X \in A$ obtain K-nearest neighbors
2. Total k samples in A represented as $(X_1, X_2, ..., X_k)$ and generated new samples from k-nearest neighbours and stored into new set called $\overset{*}{\text{A}}$
3. The new sample in $\overset{*}{\text{A}}$ is generated with the formula (11):

$$X_{resample} = X + rand(0,1) * || X - X_k || \quad (11)$$

IV. Results and Discussion

4.1.Dataset

Cervical cancer data set includes 858 samples with 32 instances from four classes: Hinselmann, Schiller, Cytology and Biopsy [2]. The complete description about data set shown in Table1.

4.2.Evaluation Metrics

The performance of proposed classification algorithm over standard methods evaluated with the following measurements includes: accuracy, sensitivity, specificity and RMSE shown in (12-16).

$$Accuracy = \frac{\text{TP}}{\text{TP} + \text{TN} + F\text{P} + F\text{N}} \quad (12)$$

$$Sensivity = \frac{\text{TP}}{\text{TP} + F\text{N}} \quad (13)$$

$$Specificity = \frac{\text{TN}}{\text{TN} + F\text{P}} \quad (14)$$

$$Positive\ Predictive\ Accuracy\ (PPA) = \frac{\text{TP}}{\text{TP} + F\text{P}} \quad (15)$$

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

$$Negative\ Predictive\ Accuracy\ (NPA) = \frac{TN}{TN + FN} \quad (16)$$

Table 1: Attributes and their types

S. no.	Attribute name	Type	S. no.	Attribute name	Type
1	Age	int	19	STDs: pelvic inflammatory disease	bool
2	Number of sexual partners	int	20	STDs: genital herpes	bool
3	First sexual intercourse (age)	int	21	STDs: molluscum contagiosum	bool
4	Num of pregnancies	int	22	STDs: AIDS	bool
5	Smokes	bool	23	STDs: HIV	bool
6	Smokes (years)	bool	24	STDs: Hepatitis B	bool
7	Smokes (packs/year)	bool	25	STDs: HPV	bool
8	Hormonal contraceptives	bool	26	STDs: Number of diagnosis	int
9	Hormonal contraceptives years	int	27	STDs: time since first daignosis	int
10	IUD	bool	28	STDs: time since last daignosis	int
11	IUD (years)	int	29	Dx: cancer	bool
12	STDs	bool	30	Dx: CIN	bool
13	STDs (number)	int	31	Dx: HPV	bool
14	STDs: condylomatosis	bool	32	Dx	bool
15	STDs: cervical condylomatosis	bool	33	Hinselmann: target variable	bool
16	STDs: vaginal condylomatosis	bool	34	Schiller: target variable	bool
17	STDs: vulvo-perinerl condylomatosis	bool	35	Cytolagy: target variable	bool
18	STDs: syphilis	bool	36	Biopsy: target variable	bool

4.3. Proposed Work Implementation:

The work in this paper proposed, an ensemble classifier using SVM and KNN approaches on cervical cancer using resampling technique SMOTE. To enhance model efficiency feature reduction techniques also applied to be mentioned RFE and PCA. The proposed method implementation is depicted in Figure 1, first method start with pre-processing include Imputation[15,16,27,28], Normalization, and Resampling using SMOTE. Second, applied set of feature selection techniques which was already mentioned to retrieve optimal features. Finally different classifiers is applied include: KNN, SVM and KNN-SVM.

The ensemble SVM-KNN classification techniques shows better results in terms of accuracy in cervical cancer data. However, while performing classification with ensemble method leads to issue of class imbalance and it affects the performance of model accuracy for all four target tests. To retain solution to such kind of misclassifications work introduced resampling algorithm (i.e., SMOTE). Hence, proposed SMOTE-SVM-KNN results shows overall improvement in different measures accuracy: 1% to 3%, specificity: 1% to 3%, and PPA: 1% to 4% in two target tests i.e., Biopsy and Cithology. The figure 2 and 3 shows the accuracy improvement results of proposed methods over standard with two target tests.

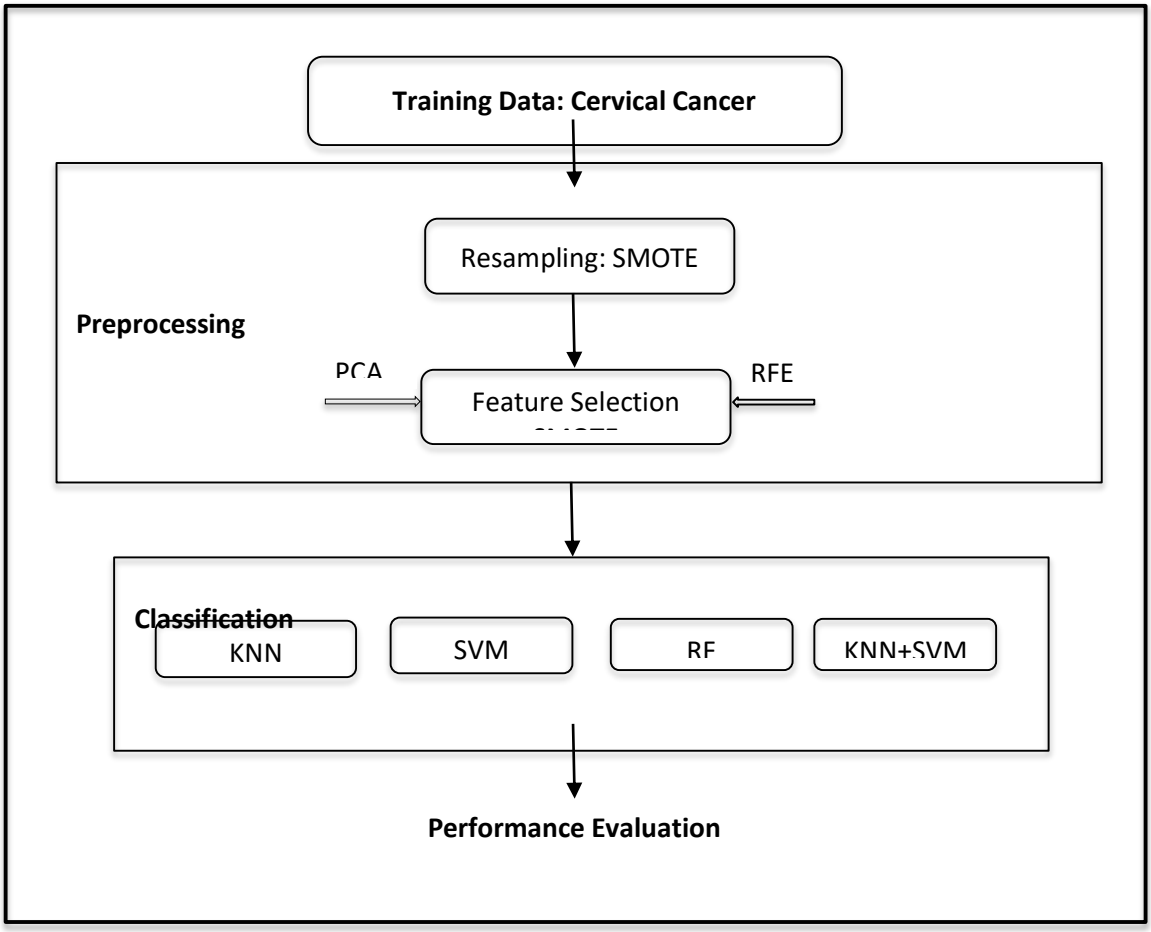

Fig 1: Proposed Classification Framework for cervical cancer diagnosis

However, RFE and PCA methods were chosen for feature selection and also retain low computational time. Also results showed that these techniques produces good classifications rates on cancer data. The relevant attributes of cervical cancer data set using proposed method SMOTE-SVM-KNN-RFE for the four target tests shown in Table2. The features that are 3,7and 9 available in all target tests .However, the feature 29 to be appeared in three test data sets among four. Similarly 4, 8, 10, 17 and 26 appear in two out of four tests.

Table 2: Top features for SMOTE-SVM-KNN-RFE

Hinselmann	Schiller	Cithology	Biopsy
3	8	2	3
10	7	3	7
26	12	9	12
7	4	29	30
2	29	5	9
17	9	7	11

9	26	18	28
4	3	27	29
8	17	10	15
14	6	13	27

The proposed algorithms, SMOTE-SVM-KNN-RFE and SMOTE-SVM-KNN-PCA produced promising better results compared to existing methods proposed on [10] in terms of accuracy, specificity and PPA.

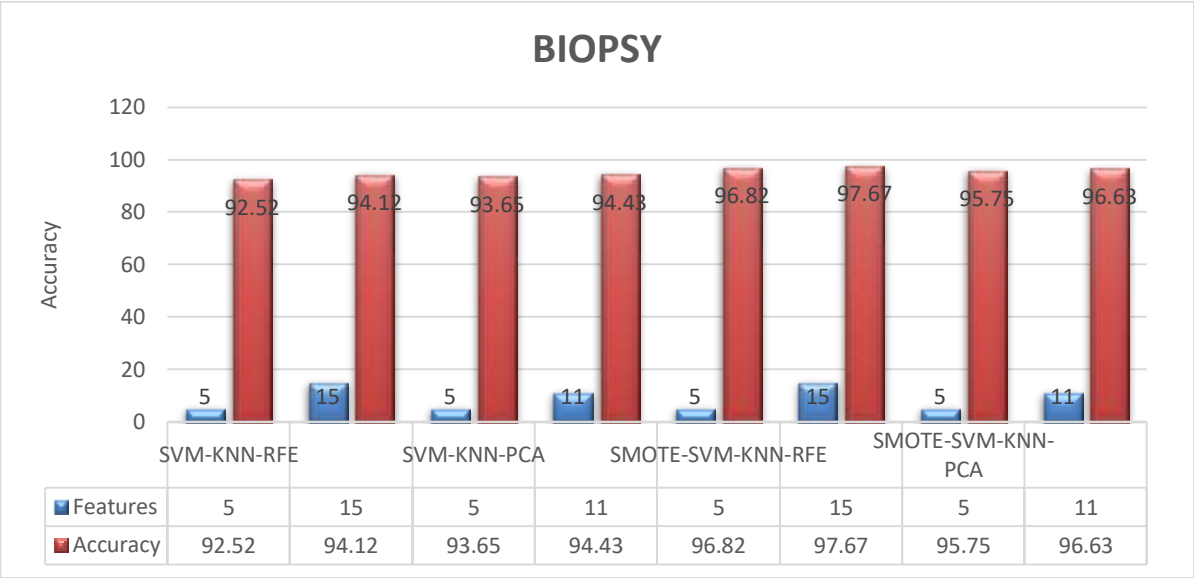

Fig 2. Accuracy: Results of proposed method over standard in BIOPSY test

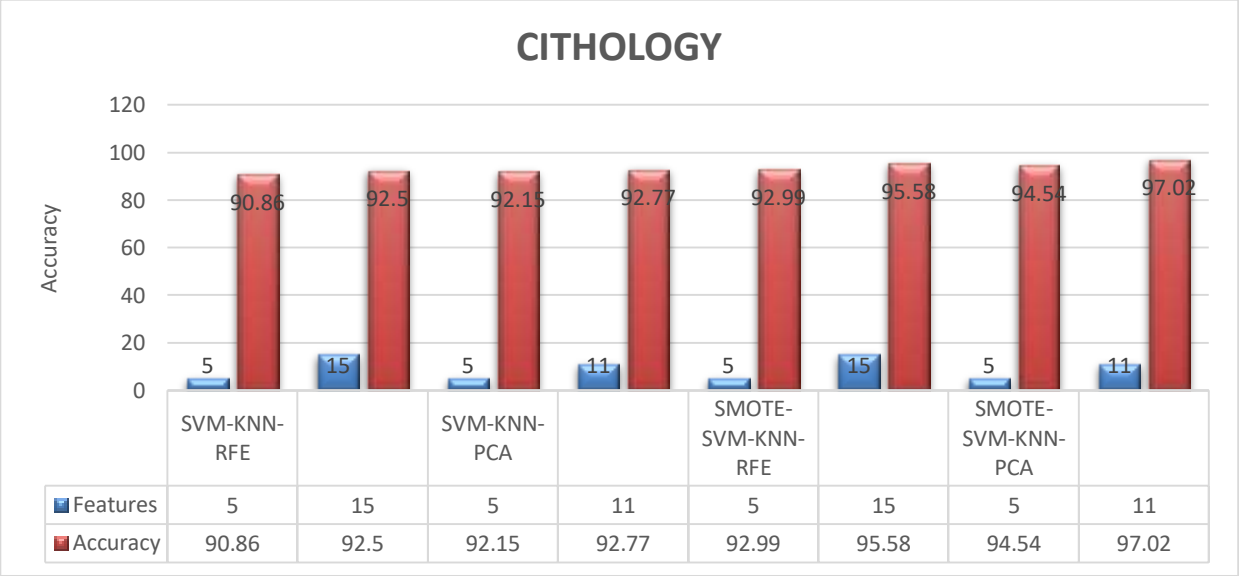

Fig 3. Accuracy: Results of proposed method over standard in COTHOLOGY test

The Table 3 and 4 provides results of proposed over standard methods where performance measures improved Accuracy: 94.03 % to 97.67% with feature 15, Specificity: 90.21 to 93.69 with feature 15, PPA: 86.07 to 89.52 with feature 15 in case of Biopsy test. Similarly in case of Citology test, results improved in terms of Accuracy: 92.75% to 97.02% with feature 11, Specificity: 87.92 to 93.49 with feature 15, PPA: 83to 87.78 with feature 15.

TABLE 3. Results performance comparison of BIOPSY test

	SVM	**SVM-RFE**		**SVM-PCA**		**SVM-KNN-RFE**		**SVM-KNN-PCA**		**SMOTE-SVM-KNN-RFE**		**SMOTE-SVM-KNN-PCA**	
Features	30	5	15	5	11	5	15	5	11	5	15	5	11
Accuracy (%)	94.13	92.39	94.03	93.45	**94.03**	92.52	94.12	93.65	94.43	96.82	**97.67**	95.75	96.63
Sensitivity (%)	100	100	100	100	100	100	100	100	100	100	100	102.1	100
Specificity (%)	**90.21**	87.32	90.05	89.09	90.05	87.45	90.08	89.29	90.45	91.75	**93.69**	91.39	92.65
PPA (%)	**86.07**	82.68	85.88	84.72	85.88	82.81	85.98	84.92	84.28	87.11	**89.52**	87.02	88.48
NPA (%)	100	100	100	100	100	100	100	100	100	100	100	100	100

TABLE 4. Results performance comparison of CITHOLOGY test

	SVM	**SVM-RFE**		**SVM-PCA**		**SVM-KNN-RFE**		**SVM-KNN-PCA**		**SMOTE-SVM-KNN-RFE**		**SMOTE-SVM-KNN-PCA**	
Features	30	5	15	5	11	5	15	5	11	5	15	5	11
Accuracy (%)	**92.75**	90.65	92.37	91.98	92.46	90.86	92.5	92.15	92.77	92.99	95.58	94.54	**97.02**
Sensitivity (%)	100	100	100	100	100	100	100	100	100	100	100	100	100
Specificity (%)	**87.92**	84.42	87.28	86.65	87.44	84.63	87.41	86.82	87.75	86.76	**93.49**	89.21	92
PPA (%)	**83**	79.1	82.26	81.54	82.44	79.31	82.39	81.71	82.75	81.44	85.47	84.15	**87.78**
NPA (%)	100	100	100	100	100	100	100	100	100	100	100	100	100

V. Conclusion:

At present cancer has become one of the main problems in the world. In the early stages, we cannot identify the symptoms of this disease and it has become a major problem. In this paper, our main theme is to identify the symptoms of cancer in early-stage such that it can be cured. For this, we applied some machine learning algorithms to predict the disease. As per now, machine learning has become the fastest technique for predicting the symptoms in the early stage, so we have used this technique in it. In this paper, the three methods named such as SVM, Recursive Feature elimination and K-nearest neighbor have been applied to the data and results have been calculated. Based on parameters like sensitivity, specificity, false negative, false positivity the values are predicted. By using this type of techniques we can reduce some amount of disease occurring in women all over the world. In the future, this paper helps to predict the values in early-stage and helps in maintaining good health care in society.

REFERENCES

1. 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.
2. 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.
3. 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.
4. S. Ding et al., ‘‘On the application of PCA technique to fault diagnosis,’’Tsinghua Sci. Technol., vol. 15, no. 2, pp. 138–144, 2010.
5. 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.
6. X. Liang, L. Zhu, and D.-S. Huang, ‘‘Multi-task ranking SVM for image segmentation,’’ Neurocomputing, vol. 247, pp. 126–136, Jul. 2017.
7. 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.
8. V. Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag, New York, 1995.
9. Li R., Ye S. W., Shi Z. Z., 2002, Chinese Journal of Electronics, 30(5), 745
10. Wu, W. and H. Zhou, 2017, “Data-Driven Diagnosis of Cervical Cancer With Support Vector MachineBased Approaches”, IEEE Access, 5:p. 25189-25195.
11. S. Di Martino, F. Ferrucci, C. Gravino, and F. Sarro, "A genetic algorithm to configure support vector machines for predicting faultprone components," in Product-Focused Software Process Improvement, ed: Springer, 2011, pp. 247-261.
12. Chawla, N.V., Lazarevic, A., Hall, L.O. and Bowyer, K.: SMOTEBoost: Improving prediction of the Minority Class in Boosting. 7th European Conference on Principles and Practice of Knowledge Discovery in Databases, Cavtat Dubrovnik, Croatia (2003) 107-119
13. Gustavo, E.A., Batista, P.A., Ronaldo, C., Prati, Maria Carolina Monard: A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. SIGKDD Explorations 6 (1) (2004) 20-29.
14. Andrew Estabrooks, Taeho Jo and Nathalie Japkowicz: A Multiple Resampling Method for Learning from Imbalanced Data Sets. Comprtational Intelligence 20 (1) (2004) 18-36
15. Lavanya K., Reddy, L., & Reddy, B. E. (2019). Distributed Based Serial Regression Multiple Imputation for High Dimensional Multivariate Data in Multicore Environment of Cloud. International Journal of Ambient Computing and Intelligence (IJACI), 10(2), 63-79. doi:10.4018/IJACI.2019040105. (SCOPUS)(Web of Science)(ESCI).

16. K. Lavanya, L. S. S. Reddy and B. Eswara Reddy, ”Modelling of Missing Data Imputation using Additive LASSO Regression Model in Microsoft Azure”, Journal of Engineering and Applied Sciences,2018,Vol 13,Special Issue 8,pp:6324-6334. (SCOPUS)
17. Liu Yang and Rong Jin, “Distance metric learning: a comprehensive survey,” Tech. Rep., Department of Computer Science and Engineering, Michigan State University, 2006.
18. C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer, New York, NY, USA, 2007
19. S. Ding et al., ‘‘On the application of PCA technique to fault diagnosis,’’ Tsinghua Sci. Technol., vol. 15, no. 2, pp. 138–144, 2010.
20. 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.
21. Chawla, N.V., Bowyer,K.W., Hall, L.O., Kegelmeyer W.P.: SMOTE: Synthetic Minority Over-Sampling Technique. Journal of Artificial Intelligence Research 16 (2002) 321-357
22. Chawla, N.V., Lazarevic, A., Hall, L.O. and Bowyer, K.: SMOTEBoost: Improving prediction of the Minority Class in Boosting. 7th European Conference on Principles and Practice of Knowledge Discovery in Databases, Cavtat Dubrovnik, Croatia (2003) 107-119
23. Gustavo, E.A., Batista, P.A., Ronaldo, C., Prati, Maria Carolina Monard: A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. SIGKDD Explorations 6 (1) (2004) 20-29.
24. D. A. Cieslak, N. V. Chawla, and A. Striegel, ‘‘Combating imbalance in network intrusion datasets,’’ in Proc. IEEE Int. Conf. Granular Comput., May 2006, pp. 732–737.
25. A. Fallahi and S. Jafari, ‘‘An expert system for detection of breast cancer using data preprocessing and Bayesian network,’’ Int. J. Adv. Sci. Technol., vol. 34, no. 9, pp. 65–70, 2011.
26. Y. Liu, N. V. Chawla, M. P. Harper, E. Shriberg, and A. Stolcke, ‘‘A study in machine learning from imbalanced data for sentence boundary detection in speech,’’ Comput. Speech Lang., vol. 20, pp. 468–494, Oct. 2006.
27. K. Lavanya, L. S. S. Reddy and B. Eswara Reddy ,”A Study of High-Dimensional Data Imputation Using Additive LASSO Regression Model”, Computational Intelligence in Data Mining,2019 Advances in Intelligent Systems and Computing 711. (SCOPUS).
28. Lavanya.K, L.S.S.Reddy, B. Eswara Reddy, “Multivariate Missing Data Handling with Iterative Bayesian Additive Lasso (IBAL) Multiple Imputation in Multicore Environment on Cloud”, Volume 5, Issue 5 , May 2019,International Journal on Future Revolution in Computer Science & Communication Engineering (IJFRSCE), PP: 54 – 58.