Advances in Computational Design, *Vol. 6, No. 3 (2021) 191-205*
https://doi.org/10.12989/acd.2021.6.3.191

Improved performance of machine learning algorithms for prognosis of cervical cancer

Mamta Arora*[1,2], Sanjeev Dhawan[1b] and Kulvinder Singh[1c]

[1]*Department of CSE, U.I.E.T., Kurukshetra University, (Kurukshetra), INDIA*
[2]*Department of CST, Manav Rachna University, (Haryana), INDIA*

(Received June 11, 2020, Revised May 9, 2021, Accepted May 12, 2021)

Abstract. With the progression of artificial intelligence in medical services, the world has achieved many benefits. The constant improvement of existing artificial intelligence techniques becomes a boon in the medical field for assisting healthcare providers. In current years, the diagnosis of cancers using machine learning techniques for timely decisions has gained popularity. Cancer is preventable and can be cured with early and timely diagnosis. Cervical cancer is one of the foremost cancers in other female cancers which ranked at the fourth position. The objective of this study to develop a model that provides a timely and cost-effective cervical cancer risk prediction score by using supervised machine learning techniques in amalgamation with dimensionality reduction techniques. The dimensionality reduction techniques help in providing the prediction with a minimum number of features. The experimental investigation on cervical cancer risk factor reveals that Random Forest classifier using recursive feature elimination with cross-validation technique gives 93%.

Keywords: artificial intelligence; machine learning; classification; support vector machine; k-nearest neighbor; random forest; decision tree; naive bayes; cancer; cervical cancer

1. Introduction

In the field of the medical era, the great cause of mortality and morbidity is determined by the disease of cancer. According to Eaton L. (2020), "A scientific paper published in the American Cancer Society journal, CA: A Cancer Journal for Clinicians", which reports by 2020, there will be an estimated 1.8 million new cancer cases diagnose and 606, 520 cancer deaths in the United States. Cervical cancer is female cancer that develops in the cervix. The cervix is the lower part of the uterus which is connected to the vagina. The main cause of cervical cancer is human papillomavirus (HPV) (Shruti *et al*. 2020). HPV virus is present in every human body. Many times immune system of the body prevents this virus to do any harm to the body. However, in a few cases, the virus stays in the body for many years which starts affecting the cells of the cervix

*Corresponding author, Ph.D. Research Scholar, E-mail: imamta.arora@gmail.com
[a]Assistant Professor, E-mail: mamta@mru.edu.in
[b]Assistant Professor, E-mail: sdhawan2015@kuk.ac.in
[b]Assistant Professor, E-mail: ksingh2015@kuk.ac.in

 ISSN: 2383-8477 (Print), 2466-0523 (Online)

Table 1 Related work in the prognosis of cervical cancer

Author	Dataset	Technique	Results
K. Fernandes *et al*. (2018)	Clinical Data	Hybrid methods of deep learning	0.6875 (auc)
R. Vidya *et al*. (2016)	NCBI	CART Algorithm	(acc) 83.87%
Y. E. Kurniawati, *et al*. (2016)	Tumor Registry	SVM	(acc) 80.18 %
J. Kahng *et al*. (2015)	Clinical data from Bucheon St Mary's Hospital	SVM	(acc) 74.41%
C.-C. Chang *et al*. (2013)	Tumor Registry	MARS	(acc) 86.00%
M. Kusy *et al*. (2013)	Clinical Data	GEP	(aucroc) 0.72
K. Fernandes *et al*. (2017)	Clinical data	Transfer Learning with Partial observability	(rmse) 35.11
B. Obrzut *et al*. (2017)	Clinical data	PNN	(aucroc) 0.818
M. Arora *et al*. (2020)	Clinical data	Supervised Machine Learning	(acc) 92%
M. Arora *et al*. (2019)	Intel Mobile ODT cervix images	Deep neural network	(acc) 56%

which causes cancer. Thus it takes many years in converting from a pre-cancerous to a cancerous state. So with the timely screening test, the risk of developing cervical cancer can be abridged. With the facility of computer-aided tools and programs screening tests of the mass population can be done in a timely and cost-effective manner. These days the machine learning techniques are playing a big role in the diagnosis of cancer and other illness. This gives the motivation in developing a computer-based algorithm that can assist the health care providers in providing a timely diagnosis so that it can decrease the rate of mortality mentioned by Arora M. (2020). Several machine learning algorithms are used in this paper for conducting experimental work. The dataset for the experiment is taken from the "UCI Machine learning repository" which is contributed by Dua, D. and Graff, C. (2019). Machine learning is now becoming a state-of-art in the medical domain. Kourous *et al*. (2015), gave a comprehensive review on machine learning techniques in cancer prognosis and prediction. They discussed the ML techniques commonly used in the medical domain for disease prediction. Most of the studies are using a supervised ML approach in predicting the disease outcome. They grouped the usage of the ML approach into three domains. Firstly, they summarized the use of ML methods for cancer susceptibility prediction, secondly for cancer recurrence prediction, and thirdly for the cancer patients survival prediction. The author highlighted that usage of multimodal heterogeneous data with classification and dimensionality reduction techniques can be proved a good tool in cancer prediction. Author Lakshmi and Krishnaveni (2016) demonstrated the use of various data mining techniques such as Multilayer Perceptron, Bayes Classifier, SVM classifier, and Bayesnet for classifying the Pap smear images into 7 types. The dataset used in this experiment is obtained from the Herlev university database. They also incorporated a feature selection approach by dimensionality reduction based on a correlation matrix. The highest accuracy reported by authors is 90% for seven class classifiers. Author Ceylan and Pekel (2017) have implemented a multi-label classification for pre-diagnosis of cervical cancer using four machine learning algorithms namely Naïve Bayes, J48 Decision Tree, Sequential Minimal Optimization, and Random Forest. The performance of all four algorithms was evaluated on accuracy (AC), Exact Match (EM), Hamming Loss (HL), and Rank

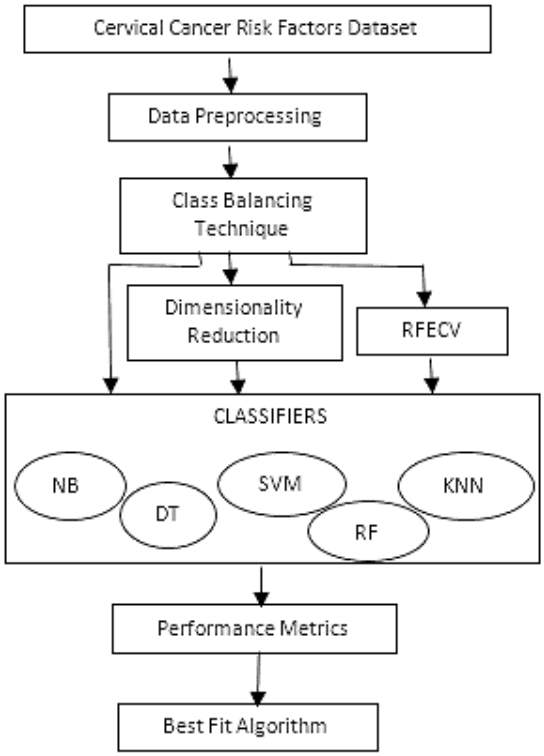

Fig. 1 Proposed Architecture

Loss (RL). The highest accuracy is reported as 86% by Sequential Minimal Optimization (SMO) technique. Author Rajvir Kaur *et al*. (2018) presented a Comparative Evaluation of Accuracy Machine Learning Classification Techniques for Diagnosis of breast cancer and cervical cancer. The breast cancer data is acquired from Wisconsin (original) breast cancer dataset which is contributed by O. L. Mangasarian (1995) whereas the cervical cancer dataset is obtained from Kaggle. They implemented ensemble bagged-based classifiers on both breast and cervical cancer datasets. In the study (Mortazavi 2021, Mortazavi and Togan 2021) author suggests the techniques for improving the interactive search by applying the Bayesian module with metaheuristic algorithms that can help in various engineering problems. The author (Mortazavi 2019) demonstrated a novel hybrid metaheuristic optimization technique, named as Interactive Fuzzy Search algorithm (IFSA). This technique combined features of interactive Particle Swarm Optimizer (iPSO) and Teaching-learning Based Optimizer (TLBO) techniques with a fuzzy module. The proposed technique validated various problems and gives the optimum solution in choosing hyperparameter variables.

2. Materials and methods

The layout of the experimental work is shown in Fig. 1. The work started with dataset collection. The data is suffering from a missing value problem which is resolved in the data pre-

Table 2 Attributes of cervical cancer dataset (Dua and Graff 2019)

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

processing step. The output of the data pre-processing step is clean data which becomes the input to the class balancing process. The SMOTE class balancing technique is used in the class

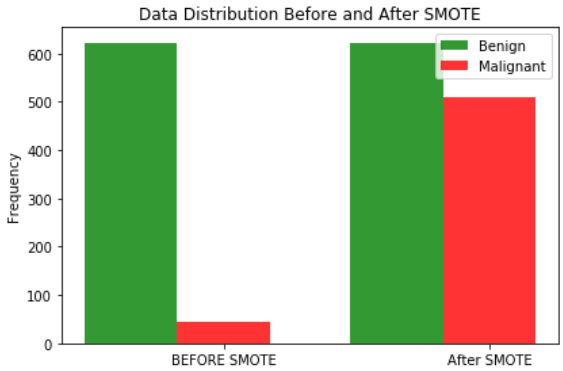

Fig. 2 Data distribution in benign and malignant before and after class balancing

balancing step. The data is then passed to various classifiers for prediction. The classifiers used in this work are Naïve Bayes, Decision Tree, and K nearest neighbor, Support vector machine, and random forest. These models are then evaluated using performance metrics. The model with the highest accuracy is selected and reported as a result of this experiment. A detailed explanation of every step is further given as sub-sections.

2.1 Data collection

The data used in the experimental work is cervical cancer (Risk factors) dataset, collected from "Hospital Universitario de Caracas" in Caracas, Venezuela. It is publicly available on the UCI repository. The dataset contains clinical information includes demographic information, habits, and historic medical records. The data consist of 858 instances and 36 attributes. The attribute information is shown in Table 2. In Table 2, the last 4 attributes represent the label. In this study, the biopsy is used as a target, and attributes from 1-30 used as features.

2.2 Data pre-processing

The Data consist of missing values. The missing values in the dataset are represented with '?'. These missing values are handled by the following python script.

data=data.drop (["STDs: Time since first diagnosis", "STDs: Time since last diagnosis"], axis=1)
data=data.replace ('?', np.NaN)
data=data.dropna (axis=0)

Out of 30 attributes listed in Table 2, two attributes namely: "STDs: Time since first diagnosis" and "STDs: Time since last diagnosis" having missing values for 787 instances. These two columns are not contributing much information, thus dropped from the dataset. After dropping these two columns, rows containing the missing values are also dropped off. It results in 668 instances.

2.3 Class balancing technique

After the data pre-processing step, the exploratory data analysis is performed which reflects the data distribution is imbalanced. It contains 623 benign and 45 malignant instances. This issue is

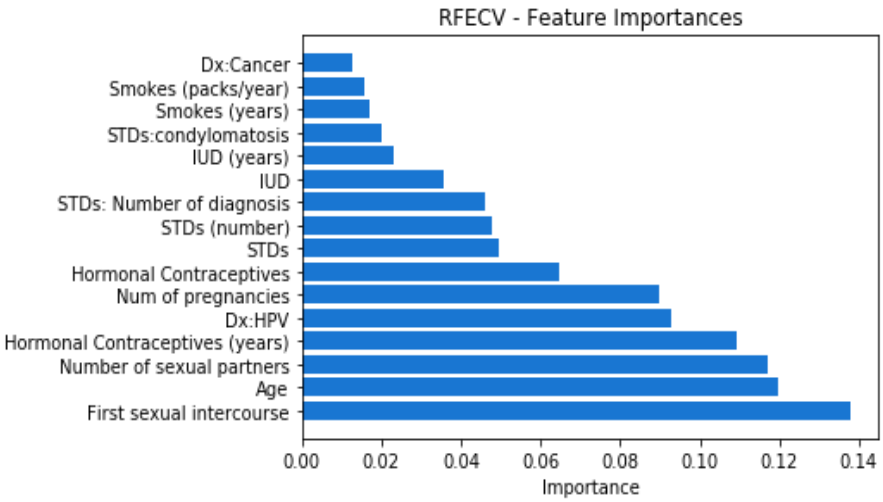

Fig. 3 Recursive feature elimination with cross-validation

resolved by applying the class balancing technique namely SMOTE. SMOTE results in 623 and 509 benign and malignant instances respectively. The pictorial representation of original data and data after applying SMOTE class balancing technique is shown in Fig. 2.

2.4 Dimensionality reduction

The dimensions of the dataset are reduced to two components using PCA which is available in Python's Scikit-Learn library. PCA is an unsupervised machine learning technique as it works only by considering the attributes of the dataset, not the labels.

2.5 Recursive Feature Elimination with Cross-Validation (RFECV)

This is another technique that uses feature importance to reduce the dimensions recursively. After running the RFECV on the cervical cancer dataset Fig. 3 is obtained that depicts the features on X-axis and their importance value on Y-axis. A total of 16 features are selected out of 28 features.

2.6 Classifiers

A total of five classifiers were used for prediction of the cervical cancer namely Naive Bayes, Random Forest, Support vector machine, K-nearest neighbor, and Decision tree. A detailed explanation of each one of them is given in sub-sections.

2.6.1 Naive Bayes

Naive Bayes is a classification algorithm that is based on the Bayes theorem. The Naive Bayes algorithm can predict better only when all the variables are categorical. It belongs to the family of probabilistic models. It is known as naive because it takes the assumption those features that are used to build the model are independent of each other. The formula used for calculating the

probability is given below:

$$P(c|x) = \frac{p(x \vee c).p(c)}{p(x)} \quad (1)$$

$$P(c \vee X) = p(c \vee x_1)p(c \vee x_2)....p(c \vee x_n).p(c)$$

P is termed as posterior probability. P(c|x) is the posterior probability of class (c, target) given predictor (x, attributes). P(c) is the prior probability of class. P(x|c) is the likelihood which is the probability of the predictor given class. P(x) is the prior probability of the predictor.

2.6.2 K-Nearest Neighbour

K Nearest Neighbor is one of the machine learning algorithms that groups the data into k classes. The value of the k parameter depends on the available training instances. The grouping is done by using one of the distance functions. These distance functions are classified into two groups. One group of functions are used for calculating the distance for continuous values whereas the other group of functions is used for calculating the distance of categorical data. The most commonly used distance functions are Hamming distance, Euclidean distance, Minkowski distance, and Manhattan distance. Equations of these functions are mentioned below:

$$\text{Euclidean} = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \quad (2)$$

$$\text{Manhattan} = \sum_{i=1}^{k}|x_i - y_i| \quad (3)$$

$$\text{Minkowski} = \left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q} \quad (4)$$

$$\text{Hamming} = \sum_{i=1}^{k}|x_i - y_i| \quad (5)$$

$$x = y => D = 0$$
$$x \neq y => D = 1$$

2.6.3 Random forest tree

Random forest tree is one of the supervised machine learning algorithms which is first introduced by Leo Breiman (Leo 2001). This algorithm is most commonly used in cancer prediction and gives higher prediction accuracy. In this approach, multiple trees are generated randomly thus named as random forest. Each Tree votes for one of the target classes. In the case of cervical prediction cancer, tree vote either for benign or malignant class. Forest having the maximum vote will be reported as a prediction result. Random forest techniques can be used for solving both regression and classification problems. For the regression problem, the average of all the trees is taken as the prediction value.

2.6.4 Decision tree

A decision tree is a predictive tree-based machine learning algorithm that was first introduced by J. Ross Quinlan in the late 1970s. Initially, it was named as ID3 algorithm as it works based on inference system and concept learning system. The working of the decision tree starts with the splitting of the entire dataset into the number of subsets. The tree consists of the decision node and the leaf node. The decision node represents the predictor and the leaf node represents the target

Table 3 Evaluation metrics

Accuracy	$(TP + TN)/(TP + TN + FP + FN)$
Sensitivity/Recall	$TP/(TP + FN)$
Specificity	$TN/(TN + FP)$
Precision	$TP/(TP + FP)$
F1-Score	$(2 * precision * recall)/precision + recall$

Table 4 Performance of various classifiers

Clf	Acc	Sen	Spec	Prec	F1	Time
NB	14.18	8.87	80	6.61	12.21	0.0087
SVM	80.6	86.29	10	5.56	7.14	1.3385
DT	91.79	99.19	0	0	0	0.0082
KNN	72.39	77.42	10	3.45	5.13	0.0149
RF	92.54	100	0	0	0	2.3659

class. It works by using the concept of entropy and information gain. The formula for calculating the entropy and information gain is given below.

$$\text{Entropy} = \sum -p_i \log_2 p_i \tag{6}$$

$$\text{Information Gain} = \text{Entropy}(y) - \text{Entropy}\left(\frac{y}{x}\right) \tag{7}$$

Where y is the dependent variable and x is the independent variable.

2.6.5 Support vector machine

Support vector machine introduced by Vapnik in 1995. It is most widely used for solving pattern recognition problems. It is a supervised machine learning algorithm and can be used to solve both linear and nonlinear problems. For solving the problems it works behind the principle of finding the best hyperplane that classifies the classes more appropriately. The core principle of the Support vector machine is to minimize the cost function. The formula used for the minimization of the cost function is given below as Eq. (8).

$$\min_{\theta} C \sum_{i=1}^{m} \left[y^{(i)} \text{cost}_1(\theta^T x^i) + \left(1 - y^{(i)}\right) \text{cost}_1(\theta^T x^i) \right] + \frac{1}{2} \sum_{i=1}^{n} \theta_j^2 \tag{8}$$

2.7 Evaluation metrics

The performance on the prediction of cervical cancer through risk factors of the different classifiers is evaluated using accuracy, sensitivity, specificity, precision, and F1- Measures. The formula used for calculating performance metrics is listed in Table 3.

3. Results and discussions

Table 5 Confusion Matrix of NB Classifier

	BPre(0)	MPre(1)
BAct(0)	11	113
MAct(1)	2	8

Table 6 Confusion Matrix of SVM Classifier

	BPre(0)	MPre(1)
BAct(0)	107	17
MAct(1)	9	1

Table 7 Confusion Matrix of DT Classifier

	BPre(0)	MPre(1)
BAct(0)	123	1
MAct(1)	10	0

Table 8 Confusion Matrix of KNN Classifier

	BPre(0)	MPre(1)
BAct(0)	96	28
MAct(1)	9	1

Table 9 Confusion Matrix of RFT Classifier

	BPre(0)	MPre(1)
BAct(0)	11	113
MAct(1)	2	8

The classification performed on cervical cancer (Risk Factors) dataset using various classifiers viz. Naïve Bayes, Support Vector Machine, K-Nearest Neighbor, Decision Tree, and Random Forest Tree. The performance of all the classifiers is given in Table 4. Each row in the table is corresponding to the performance of one classifier. The header row of Table 4 is abbreviated as 'clf' for the classifier, 'Acc' for accuracy, 'Sen' for sensitivity, 'Prec' for precision, 'F1' for F1 Score, 'Time' for computation time in seconds.

Table 5 represents the confusion matrix of the Naïve Bayes Classifier after performing the classification on original features of the cervical cancer dataset. In Table 6 BAct and MAct represent the actual Benign and Malignant target which is represented by 0 and 1. Whereas, BPre and MPred represent predicted Benign and Malignant values which are also represented by 0 and 1 respectively. The NB classifier is the least performer among all the other classifiers with an accuracy of 14.18%.

Table 6 represents the confusion matrix of the Support Vector Machine after performing the classification on original features of the cervical cancer dataset. The accuracy reported by the SVM classifier is 80.6% and the F1 score is 7.14%. The time taken by the classifier is the second-highest time 1.3385 seconds.

Table 7 represents the confusion matrix of the Decision Tree classifier after performing the

Table 10 Performance of various classifiers after applying PCA

Clf	Acc	Sen	Spec	Prec	F1	Time
NB	81.34	86.29	20	10.53	13.79	0.0037
SVM	79.1	83.06	30	12.5	17.65	0.0638
DT	81.34	87.1	10	5.88	7.41	0.004
KNN	70.15	72.58	40	10.53	16.67	0.0058
RF	79.1	83.06	30	12.5	17.65	3.1207

Table 11 Performance of various classifiers after applying RFECV

Clf	NB	SVM	DT	KNN	RF
Acc	-	76.12	84.33	-	91.79
CM	-	[101 239 1]	[112 129 1]	-	[122 29 1]

Table 12 Classification Accuracy with special dimension reduction cases on cervical cancer Dataset

Clf	CASE1	CASE2	CASE3	CASE4	CASE5
NB	14.18	12.69	11.19	11.94	62.69
SVM	79.85	79.1	78.36	81.34	76.12
DT	88.81	88.81	84.33	63.43	51.49
KNN	74.63	69.4	71.64	70.9	73.13
RF	92.54	92.54	91.79	91.79	92.54

classification on original features of the cervical cancer dataset. The accuracy of the classifier is calculated as 91.79%. This classifier takes 0.0082 seconds which is the least computation time among all the other classifiers.

Table 8 represents the confusion matrix of K nearest neighbor classifier after performing the classification on original features of the cervical cancer dataset. The accuracy calculated for KNN is 72.39 with an F1- score of 5.13.

Table 9 represents the confusion matrix of the Random Forest Tree classifier after performing the classification on original features of the cervical cancer dataset. This classifier reported the highest accuracy with 92.54 among all the other implemented classifiers. The time taken by RFT is the highest.

Several dimensionality reduction techniques like PCA, LDA, KDA, etc. can be used for reducing the number of features which in turn reduces the computation time and other resources required while performing the classification task. Fig. 4 represents the PCA technique on the cervical cancer dataset that wraps all the features into two principal components viz. pca1 and pca2.

All the 5 classifiers viz. NB, SVM, DT, KNN, and RFT are used for performing the prediction task using two principal components on the original feature set. These classifiers were also implemented with the original features. The accuracy and time taken by both the implementation are then compared to obtained better results in fewer time frames. Table 10 represents the performance of various classifiers after applying the PCA on the cervical cancer dataset. The

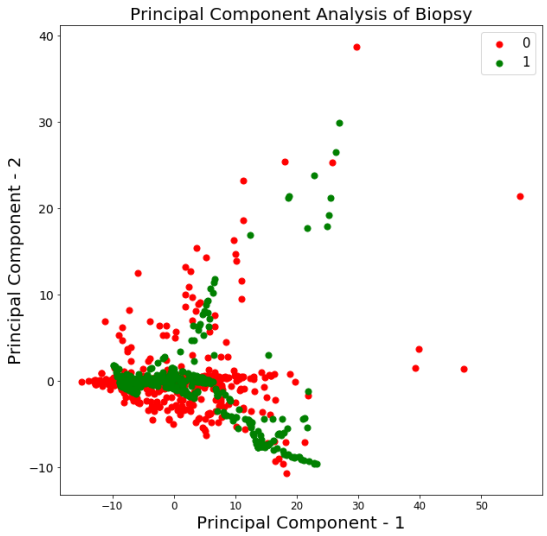

Fig. 4 PCA on Cervical Cancer Data Set

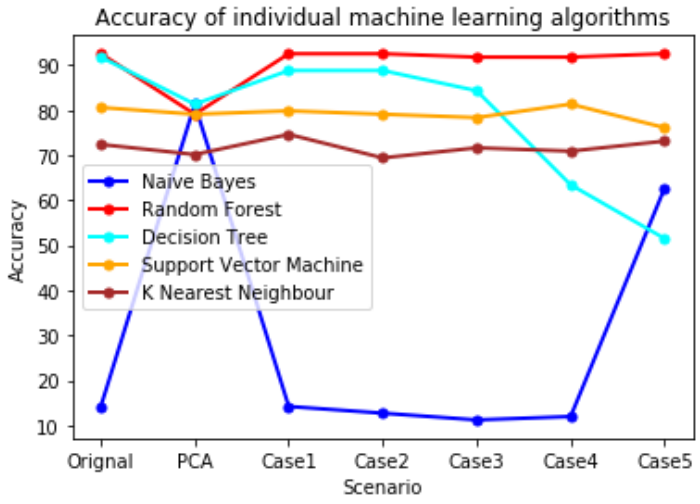

Fig. 5 Performance comparisons of different classifiers in terms of accuracy

analysis shows the accuracy of all the four classifiers except NB is reduced with PCA. The decrease was calculated as 1.5 for SVM, 10.45 for DT, 2.24 for KNN, and 13.44 for RF. Hence it is concluded that PCA with two principal components is not giving promising results when used on cervical cancer dataset.

Now the other technique used for experimenting with a fewer number of features is Recursive feature elimination with cross-validation. The principle behind the RFECV is to find the best results by using an optimum number of features. It eliminates the features by calculation feature importance value recursively and eliminates the feature which is least significant or contributing very little to the entire dataset. As mentioned earlier RFECV, use only 16 features out of 28. Once the features are selected then the classifiers are implemented with selected features. Table 11

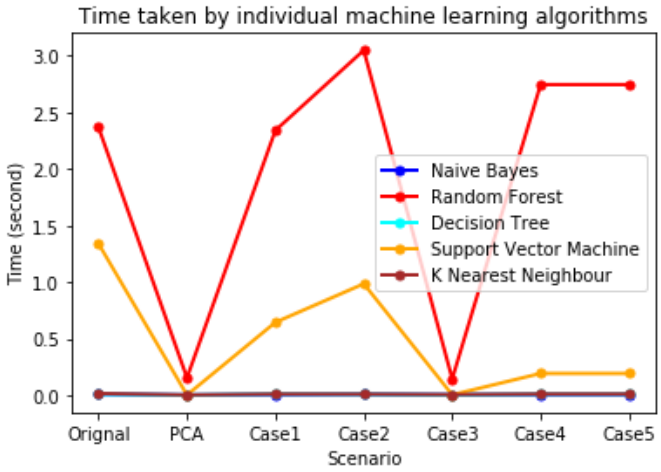

Fig. 6 Comparisons of different classifiers in terms of computation time taken

shows the accuracy and confusion matrix obtained by SVM, DT, and RF after applying RFECV. RFECV doesn't work with NB and KNN classifiers. The accuracy is better than PCA but slightly lesser than the accuracy gained using original features.

Further, the performance of all the five classifiers was measured in special five cases denoted by case 1, 2, 3, 4, and 5; where case 1 represents dataset with 15 features, case 2 represents the dataset with 13 features, case 3 represents the dataset with 10 features, case 4 represents the dataset with 7 features and case 5 represents the dataset with 5 features. Table 12 shows the performance in terms of the accuracy of all the five classifiers.

The comparison of all the five special cases along with the original feature set and PCA technique in terms of accuracy and computation time can be better analyzed through line charts shown in Figs. 5 and 6 respectively. Fig. 5 depicts the accuracy achieved in all 5 cases along with the original feature set and PCA technique for all the five classifiers. The highest accuracy is achieved by the RF classifier in all situations. The accuracy of RF almost remains fixed in the original dataset and from cases 1-5. The accuracy of the DT classifier is significantly decreased from case3 to case4. KNN and SVM perform almost the same in all situations. The NB classifier is the least performer among all the other classifiers.

Fig. 6 represents the computation time taken by all the classifiers. All the algorithms take the least computation time when working either with PCA or case3.

4. Conclusions

Cervical cancer is one of the foremost cancers in other female cancers. The mortality rate of this cancer can be decreased with the early diagnosis. With this intent, this study applies five machine learning techniques to the cervical cancer risk factor dataset that is obtained from the UCI repository. The five machine learning techniques namely naïve Bayes, k nearest neighbor, decision tree, random forest classifier, and support vector machine are implemented for obtaining the predictions. These algorithms are evaluated based on accuracy, sensitivity, specificity, and F1-support. In addition to this principal component analysis and random feature elimination with

cross-validation techniques are used for dimensionality reduction. In conclusion, the results obtained in this study show that a Random forest tree using recursive feature elimination with cross-validation technique gives more promising results than other machine learning techniques. The limitation of the proposed model is that it is validated on a small dataset that suffers from the imbalance data issue. The imbalance data problem is taken care of by using SMOTE class balancing technique. In the future, we will try to use a model for breast and lung cancer prediction. The proposed model has significant implications in medical science, especially in the prognosis and diagnosis of cervical cancers.

5. Future Scope

This research has thrown up many questions in need of further investigation. Further work needs to be done to establish whether the deep learning technique can be used in improving the predictive accuracy of cervical cancer.

Acknowledgments

The authors would like to acknowledge that they have used cervical cancer (Risk Factors) Data Set. It is obtained from the UCI machine learning repository. The link of the dataset used for experimenting is https://archive.ics.uci.edu/ml/datasets/Cervical+cancer+%28Risk+Factors%29.

References

Arora, M., Dhawan, S. and Singh, K. (2019), *Advances in Intelligent Systems and Computing*, Springer Nature, Singapore.

Arora, M., Dhawan, S. and Singh, K., (2018), "Deep learning: overview, architecture, framework & applications", *Int. J. Latest Trends Eng. Technol.*, **10**(1), 379-384. http://dx.doi.org/10.21172/1.101.69.

Arora, M., Dhawan, S. and Singh. K. (2020), "Data driven prognosis of cervical cancer using class balancing and machine learning techniques", *EAI Endorsed Transactions Energy Web*, 1-9. https://doi: 10.4108/eai.13-7-2018.164264.

Arora, M., Dhawan, S., Singh, K. (2018), "Deep neural network for transformation zone classification", *In First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, Jalandhar, India, 213-216, https://doi.org/10.1109/ICSCCC.2018.8703327.

Breiman, L. (2001), "Random forests", *Machine Learning*, **45**(1), 5-32. https://doi.org/10.1023/A:1010933404324.

Ceylan, Z. and Pekel, E. (2017), "Comparison of multi-label classification methods for prediagnosis of cervical cancer", *Int. J. Intell. Syst. Appl. Eng.*, **5**(4), 232-236. https://doi.org/10.18201/ijisae.2017533896.

Chang, C.C., Cheng, S.L., Lu, C.J. and Liao, K.H. (2013), "Prediction of recurrence in patients with cervical cancer using MARS and classification", *Int. J. Machine Learning Comput.*, **3**(1), 75-78. https://doi.org/10.7763/IJMLC.2013.V3.276.

Dua, D. and Graff, C. (2019), *UCI Machine Learning Repository Irvine*, University of California, School of Information and Computer Science. http://archive.ics.uci.edu/ml.

E.F. Krause (1987), *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*. California: Addison-Wesley.

Eaton, L. (2003), "World cancer rates set to double by 2020", *BMJ* **326**(7392).

https://doi:10.1136/bmj.326.7392.728/a.

Fernandes, K., Cardoso, J.S. and Fernandes, J. (2017), "Transfer learning with partial observability applied to cervical cancer screening", *In Iberian conference on pattern recognition and image analysis*, https://doi.org/10.1007/978-3-319-58838-4_27.

Fernandes, K., Chicco, D., Cardoso, J.S. and Fernandes, J. (2018), "Supervised deep learning embeddings for the prediction of cervical cancer diagnosis", *Peer J. Comput. Sci.*, **4**(e154). https://doi.org/10.7717/peerj-cs.154.

Gurusamy, K. and Krishnaveni, K. (2016), "Feature extraction and feature set selection for cervical cancer diagnosis", *Indian J. Sci. Technol.*, **9**(19), 1-7. https://doi.org/10.17485/ijst/2016/v9i19/93881.

Jain, S., Sood, M. and Pal, S. (2020), *Advances in Computational Intelligence Techniques*, Springer, Singapore.

Kahng, J., Kim, E.H., Kim, H.G. and Lee, W. (2015), "Development of a cervical cancer progress prediction tool for human papillomavirus-positive Koreans: A support vector machine-based approach", *J. Int. Medic. Res.*, **43**(4), 518-525. https://doi.org/10.1177/0300060515577846.

Kaur, R. and Ginige, J.A. (2018), "Comparative evaluation of accuracy of selected machine learning classification techniques for diagnosis of cancer: a data mining approach", *World Academy of Sci. Eng. Technol.: Int. J. Biomedic. Biolo. Eng.*, **12**(2), 19-25. http://scholar.waset.org/1307-6892/10008834.

Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V., and Fotiadis, D.I. (2014), "Machine learning applications in cancer prognosis and prediction", *Comput. Struct. Biotechnol. J.*, **13**, 8-17. https://doi.org/10.1016/j.csbj.2014.11.005.

Kurniawati, Y.E., Permanasari, A.E. and Fauziati, S. (2016), "Comparative study on data mining classification methods for cervical cancer prediction using pap smear results", *1st International Conference on Biomedical Engineering (IBIOMED)*, 1-5. https://doi.org/10.1109/ibiomed.2016.7869827.

Kusy, M., Obrzut, B. and Kluska, J. (2013), "Application of gene expression programming and neural networks to predict adverse events of radical hysterectomy in cervical cancer patients", *Medic. Biologic. Eng. Comput.*, **51**, 1357-1365. https://doi.org/10.1007/s11517-013-1108-8.

Mangasarian Olvi, L., Nick Street, W. and Wolberg, H. (1995), "Breast cancer diagnosis and prognosis via linear programming", *Operat. Res.*, **43**(4), 570-577. https://doi.org/10.1287/opre.43.4.570.

Mortazavi, A. (2019), "Interactive fuzzy search algorithm: A new self-adaptive hybrid optimization algorithm", *Eng. Appl. Artificial Intell.*, **81,** 270-282. https://doi.org/10.1016/j.engappai.2019.03.005.

Mortazavi, A. (2021), "Bayesian interactive search algorithm: A new probabilistic swarm intelligence tested on mathematical and structural optimization problems", *Advan. Eng. Softw.*, **155**, 102994. https://doi.org/10.1016/j.advengsoft.2021.102994.

Mortazavi, A. and Vedat T. (2021), "Solving of distinct engineering optimization problems using metaheuristic algorithms", *In Nature-Inspired Metaheuristic Algorithms for Engineering Optimization Applications*, 111-136.

Obrzut, B., Kusy, M., Semczuk, A., Obrzut, M. and Kluska, J. (2017), "Prediction of 5-year overall survival in cervical cancer patients treated with radical hysterectomy using computational intelligence methods", *BMC Cancer*, **17**(1), 840. https://doi.org/10.1186/s12885-017-3806-3.

Parikh, D. and Menon, V. (2019), "Machine learning applied to cervical cancer data", *Int. J. Math. Sci. Comput.*, **5**(1), 53-64. https://doi.org10.5815/ijmsc.2019.01.05.

Paul E. Black (2020), "Manhattan distance", *In Dictionary of Algorithms and Data Structures*, https://www.nist.gov/dads/HTML/manhattanDistance.html

Paul E. Black, (2020), "Euclidean distance", *In Dictionary of Algorithms and Data Structures*, https://www.nist.gov/dads/HTML/euclidndstnc.html.

Sarwar, A., Ali, M., Suri, J. and Sharma, V. (2015), "Performance evaluation of machine learning techniques for screening of cervical cancer", *2nd International Conference on "Computing for Sustainable Global Development*, New Delhi (INDIA).

Suykens, J. and Vandewalle, J. (1999), "Least squares support vector machine classifiers", *Neur. Proc. Lett.*, **9**(1), 293-300. https://doi.org/10.1023/A:1018628609742.

Vapnik, V. (1998), *Statistical Learning Theory*, John Wiley, New York.

Vapnik, V. (2013), *The Nature of Statistical Learning Theory*, Springer science & business media.

Vidya, R. and Nasira, G.M. (2016), "Prediction of cervical cancer using hybrid induction technique: A solution for human hereditary disease patterns", *Indian J. Sci. Technol.*, **9**(30), 1-10. https://doi.org/10.17485/ijst/2016/v9i30/82085.

Wenying Y., Xin, G., Tongqing, X., Xiping, Y. and Maohong, J. (2019), "Cervical cancer risk prediction model and analysis of risk factors based on machine learning", *In Proceedings of the 2019 11th International Conference on Bioinformatics and Biomedical Technology (ICBBT'19)*, Association for Computing Machinery, New York, USA, 50-54. https://doi.org/10.1145/3340074.3340078.

Wu, W.M. and Zhou, H. (2017), "Data-driven diagnosis of cervical cancer with support vector machine-based approaches", IEEE Access, **5**(1), 25189-25195. https://doi.org/ 10.1109/ACCESS.2017.2763984.

CC