Tak et al. Scr Med 2022 Sep;53(3):205-11. DOI:10.5937/scriptamed53-38848
ORIGINAL ARTICLE

Optimised Feature Selection and Cervical Cancer Prediction Using Machine Learning Classification

Amit Tak,[1] Puran Mal Parihar,[2] Dharmendra Singh Fatehpuriya,[3] Yogesh Singh[4]

1. RVRS Government Medical College, Bhilwara, Rajasthan, India.
2. Geetanjali Medical College, Udaipur, India.
3. JLN Medical College, Ajmer, Rajasthan, India.
4. Zoram Medical College, Falkawn, Mizoram, India.

Abstract

Background: Screening and early detection play a key role in cervical cancer prevention. The present study predicts the outcome of various diagnostic tests used to diagnose cervical cancer using machine learning algorithms.
Methods: The present study ran various cervical cancer risk factors on a machine learning (ML) classifier to predict outcomes of Hinselmann, Schiller, cytology and biopsy. The dataset is publicly available on the Machine Learning Repository website of the University of California Irvine. The imbalanced dataset was pre-processed using oversampling methods. The significantly varied features between the two levels of a response variable were used to train the machine learning classifiers on MATLAB. The classifiers used were Decision Trees, Support Vector Machine, K-Nearest Neighbours and Ensemble learning classifiers. The performance metrics of the classifiers were expressed as accuracy, the area under the receiver operator characteristic (AU-ROC) curve, sensitivity and specificity.
Results: The Fine Gaussian SVM classifier was the best to classify Hinselmann, cytology and biopsy with the accuracy of 97.5 %, 62.5 % and 98 %, respectively. However, Boosted trees performed best in the classification of Schiller with 81.3 % accuracy.
Conclusion: The present study selected optimised features among multiple risk factors to train various ML classifiers to predict cervical cancer.

Key words: Biopsy; Cervical cancer; Cytology; Hinselmann; Machine learning; Schiller.

Correspondence:
YOGESH SINGH
T: +8764429456
E: dryogeshsingh4u@gmail.com

ARTICLE INFO
Received: 26 June 2022
Revision received: 4 August 2022
Accepted: 4 August 2022

Introduction

Cervical cancer is the fourth most common cancer in women worldwide. An estimated 570,000 cervical cancer cases were diagnosed and 311,000 women died from cervical cancer worldwide in 2018.[1, 2] It is a disease in which healthy cells grow abnormally on the surface of the cervix, forming a mass of cells called a tumour and spreading to other parts of the body such as the bladder, rectum, lungs, vagina and liver. Women under 50 years of age are mostly affected by this disease. Cervical cancer control includes primary prevention (vaccination against human papillomavirus), secondary prevention (screening and treatment of pre-cancerous lesions), tertiary prevention (diagnosis and treatment of invasive cervical cancer) and palliative care.[3] The risk factors for cervical cancer include age, human papillomavirus infection, early sexual activity, long-term use of the hormonal contraceptive pill, sexually transmitted infections and genetics.[4] Screening has a significant role in the early diagnosis of cervical malignancy. Screening procedures for cervical cancer include cytology, Schiller, Hinselmann and the standard biopsy test to recognise cervical cancer.[5]

This article should be cited as follows: Tak A, Parihar PM, Singh Fatehpuriya D, Singh Y. Optimised feature selection and cervical cancer prediction using machine learning classification. Scr Med 2022 Sep;53(3):205-11.

Classification of diseases based on artificial intelligence methods helps in prediction of disease and survival rate.[6, 7] Recently, diseases have been classified using computer vision, machine learning (ML) and deep learning (DL) algorithms.[8-10] The various ML classifiers, including Logistic Regression (LR), Decision Tree (DT), Artificial Neural Networks (ANN), Support Vector Machine (SVM) and Naive Bayes (NB), along with feature optimisation methods such as Chicken Swarm optimisation can be used for prediction.[11-14] The present study used ML methods to predict the outcome of various methods used for diagnosing cervical cancer.

Material and Methods

The present study ran various cervical cancer risk factors on a machine learning (ML) classifier to predict the outcome of various diagnostic tests, including Hinselmann, Schiller, cytology and biopsy. The study dataset consisted of a random sample of medical records of 858 patients who attended gynaecology service at Hospital Universitario de Caracas in Caracas, Venezuela, between 2012 and 2013. The dataset was publicly available on the Machine Learning Repository website of the University of California Irvine (UCI ML).[15] Fifteen factors among 32 cervical cancer risk factors were used, including the age of the patient, sexual activity (number of sexual partners and age of first sexual intercourse), number of pregnancies, smoking behaviour, use of contraceptives (hormonal and intrauterine devices) and historical records of sexually transmitted diseases (STDs) to predict indications of response variables. The four response variables include Hinselmann, Schiller, cytology and biopsy (Table 1). Hinselmann's test refers to colposcopy using acetic acid. In Schiller's test, Lugol iodine was used to visualise the uterine cervix.[16]

The dataset contained many missing values as many patients did not answer all the questions for privacy reasons. The dataset was imbalanced and oversampling methods were employed during pre-processing. The significantly different features between the two levels of a response variable were used for ML classification. The ML classifier application on MATLAB 2019a was used for classification.[17]

The classifiers used in this application included Decision Trees, Support Vector Machine (SVM), K-Nearest Neighbours (KNN) and Ensemble learning classifiers. The decision trees included complex, medium and simple tree classifiers. Similarly, the SVMs included linear, quadratic, cubic, fine Gaussian, medium Gaussian and coarse Gaussian classifiers. The ensemble classifiers had boosted trees, bagged trees and RUS boosted tree classifiers.

Table 1: List of predictors and response variables used machine learning classification

SNo	Attribute name	Type	Predictor
1	Age (years)	Integer	Predictor
2	Number of sexual partners	Integer	Predictor
3	Age of first sexual intercourse (year)	Integer	Predictor
4	Number of pregnancies	Integer	Predictor
5	Smokes (yes/no)	Boolean	Predictor
6	Smokes (years)	Integer	Predictor
7	Smokes (packs/year)	Integer	Predictor
8	Hormonal contraceptives (yes/no)	Boolean	Predictor
9	Hormonal contraceptives (years)	Integer	Predictor
10	Intrauterine devices (yes/no)	Boolean	Predictor
11	Intrauterine devices (years)	Integer	Predictor
12	Sexually transmitted disease (STDs) (yes/no)	Boolean	Predictor
13	Number of STDs	Categorical	Predictor
14	STDs (years since the first diagnosis)	Integer	Predictor
15	STDs (years last diagnosis)	Integer	Predictor
16	Hinselmann	Boolean	Response
17	Schiller	Boolean	Response
18	Cytology	Boolean	Response
19	Biopsy	Boolean	Response

Statistical analysis

After assumption checked, the quantitative data were expressed in median (IQR) and compared using the non-parametric Mann-Whitney's U test. The categorical data were expressed in percentage and the relationship between discrete variables was found using a Chi-squared test. The performance metrics of the machine learning classifier were expressed as accuracy, the area under the receiver operator characteristic (AU-ROC) curve, sensitivity and specificity. The JASP version 0.16.2 and MATLAB 2019a were used for statistical analysis.[18] The significance level was considered at 5 %.

Results

Considering Hinselmann as a response variable, the features which were different between the two response levels were age [$W = 2348$; $p < 0.001$],

Table 2: Comparison of continuous predictors across binary outcomes of Hinselmann, Schiller, cytology and biopsy

	Responses							
	Hinselmann		Schiller		Cytology		Biopsy	
	W	p	W	p	W	p	W	p
Age	2348	< 0.001	1168.5	0.906	1135.5	0.158	1135.5	0.158
Number of sexual partners	2262.5	< 0.001	1341.5	0.152	1229.5	0.414	1229.5	0.414
First sexual intercourse	1559	0.780	962	0.160	1277.5	0.627	1277.5	0.627
Number of pregnancies	1768.5	0.106	1144	0.955	1527.5	0.240	1527.5	0.240
Smokes (years)	1565	0.694	675	< 0.001	1556	0.059	1556	0.059
Smokes (packs/year)	1605	0.487	739.5	< 0.001	1592	0.026	1592	0.026
Hormonal Contraceptives (years)	2237.5	< 0.001	1365	0.101	1188.5	0.277	1188.5	0.277
IUD* (years)	1414.5	0.385	1252	0.204	1142.5	0.057	1142.5	0.057
STDs*: Time since the first diagnosis	1781.5	0.103	1137	0.915	1174	0.244	1174	0.244
STDs*: Time since the last diagnosis	2060.5	< 0.001	1134.5	0.900	1141.5	0.168	1141.5	0.168

**IUD: Intrauterine devices; STDs: Sexually transmitted diseases*

Table 3: Relationship between categorial predictors and binary outcomes of Hinselmann, Schiller, cytology and biopsy

	Hinselmann		Schiller		Cytology		Biopsy	
	χ^2 value	p	χ^2 value	p	χ^2 value	p	χ^2 value	p
Smokes	0.045	0.832	20.308	< 0.001	4.833	0.028	0.437	0.509
Hormonal Contraceptives	18.084	< 0.001	1.510	0.219	2.056	0.152	0.679	0.410
IUD*	0.978	0.323	1.524	0.217	3.690	0.055	1.515	0.218
STDs* (number)	28.195	< 0.001	6.020	0.111	5.975	0.113	2.711	0.438

**IUD: Intrauterine devices; STDs: Sexually transmitted diseases*

Table 4: Performance metrics of the best classifier for Hinselmann, Schiller, cytology and biopsy classification

	Hinselmann	Schiller	Cytology	Biopsy
Features (Serial number based on Table 1)	1, 2, 8, 9, 13, 15	5, 6, 7	5, 7	2, 3, 4, 14, 15
Classifier	Fine Gaussian SVM#	Boosted Trees	Fine Gaussian SVM#	Fine Gaussian SVM#
Accuracy	97.5 %	81.3 %	62.5 %	98 %
AUC*	1	0.83	0.66	0.98
Sensitivity	100 %	69 %	100 %	100 %
Specificity	95 %	94 %	25 %	96 %

**AUC: Area under the receiver operating characteristic curve; #SVM: Support vector machine*

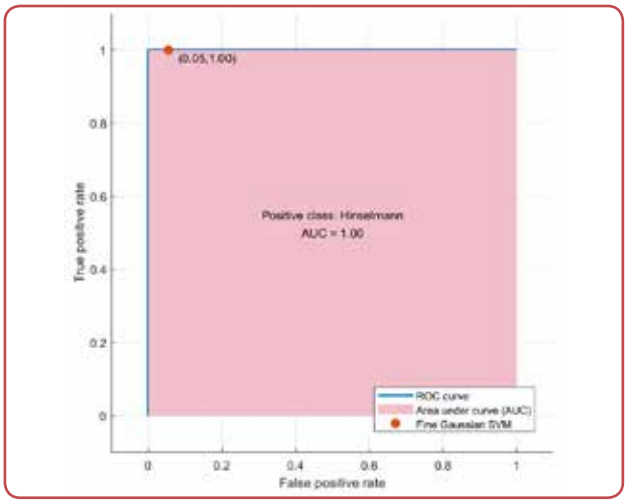

Figure 1: Receiver operating characteristic curve showing performance metrics of Fine Gaussian SVM classifier for Hinselmann classification

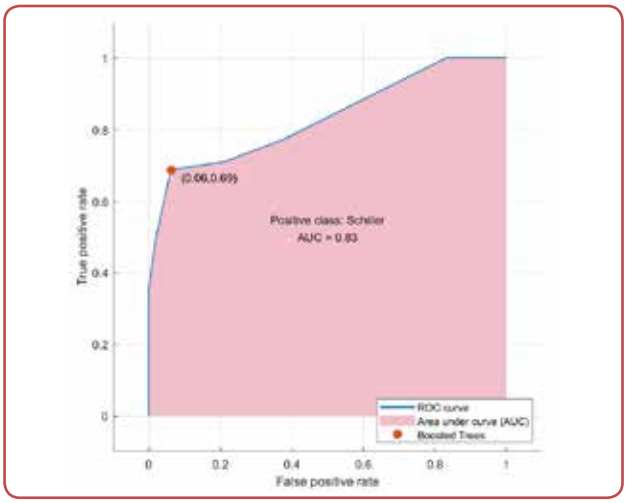

Figure 2: Receiver operating characteristic curve showing performance metrics of Boosted Trees classifier for Schiller classification

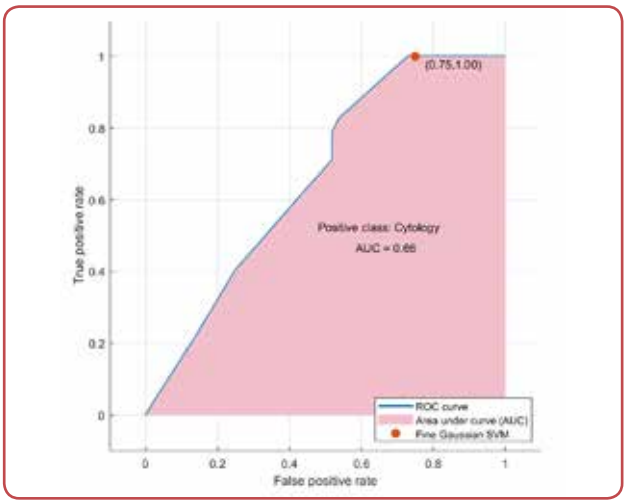

Figure 3: Receiver operating characteristic curve showing performance metrics of Fine Gaussian SVM classifier for Cytology classification

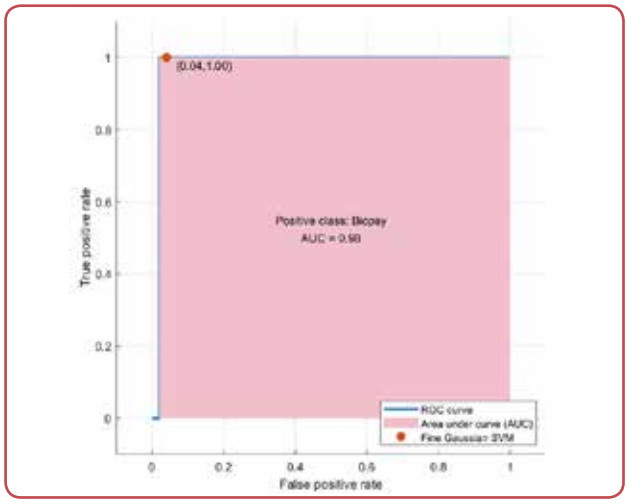

Figure 4: Receiver operating characteristic curve showing performance metrics of Fine Gaussian SVM classifier for Biopsy classification

number of sexual partners [W = 2262.5; p < 0.001], hormonal contraceptives (yes/no) [c^2 = 18.08; p < 0.001], hormonal contraceptives (years) [W = 2237.5; p < 0.001], STDs (number) [c^2 = 28.20; p < 0.001], and STDs: Time since the last diagnosis [W = 2060.5; p < 0.001]. Similarly, for Schiller as the response variable, the features significantly differed were: Smokers (yes/no) [c^2 = 20.31; p < 0.001], Smokers (years) [W = 675; p < 0.001], Smokers (packs/year) [W = 739.5; p < 0.001]. In case of cytology as a response variable, Smokers (yes/no) [c^2 = 4.83; p = 0.028], Smokers (packs/year) [W =1592; p = 0.026] were significantly differed. In case of biopsy, features differed were number of sexual partners [W = 1855.5; p < 0.001], age of the first sexual intercourse [W = 732.5; p < 0.001], number of pregnancies [W = 1647; p = 0.004], STDs (time since first diagnosis) [W = 1603.5; p = 0.013] and STDs (time since last diagnosis) [W = 1567; p = 0.026] (Table 2 and Table 3).

The Fine Gaussian SVM classifier was the best model to classify Hinselmann, cytology and biopsy. However, Boosted trees performed best in the classification of Schiller. Table 4 shows the performance metrics of various classifiers. (Table 4; Figures 1-4).

Discussion

Cervical cancer is a primary cause of premature mortality in women worldwide. Screening and early diagnosis are preventive strategies for better management of cervical cancer. Machine learning methods can be used to process vast amounts of cancer data and is readily accessible to the medical research community to upgrade the survival rate of patients.[19, 20] The present study used various ML algorithms to predict indications for various examinations to diagnose cervical cancer. The Fine Gaussian SVM classifier was the best model to classify Hinselmann, cytology and biopsy. However, Boosted trees performed best in the classification of Schiller (Table 4).

In a similar study, Nagadeepa et al used RF, SVM and Deep Learning (DL) models like Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN), for cervical cancer prediction. The SVM showed the highest accuracy (97 %), followed by CNN (95/3 %), RF (94 %) and ANN (95.2 %), respectively.[21] Ali et al used ML classifier models to predict cervical cancer from various examinations using clinical data. The random tree classifier showed better results for cytology (98.65 %) and biopsy (98.33 %), whereas the Instance-Based K-nearest neighbour (IBK) with random forest classifier provided higher accuracy for Hinselmann (99.16 %) and Schiller (98.58 %).[22] Nithya et al predict cervical cancer using random forest, rpart, $C_{5.0,}$ KNN and SVM algorithms after optimised feature selection. Contrary to the present study, the random forest and $C_{5.0}$ classifier models showed higher accuracy in predicting cervical cancer.[23] Zahras et al used a deep convolutional neural network to predict the outcome of Hinselmann, cytology, Schiller and biopsy for diagnosing cervical cancer. The results of deep convolutional neural network classification were comparable to the present study, with an accuracy of about 90 % for each target.[24] Asaduzzaman et al developed a system to predict the risk of cervical cancer using machine learning models including AdaBoost, Logistics Regression,

SVM, Neural Network, kNN, Naïve Bayes, Decision Tree, CN2 rule Inducer, Random Forest and Quadratic Classifier. The most significant factors that contributed to cervical cancer were the number of children, age at first intercourse, age of husband, Pap test and age.[25] Chaudhuri et al developed a 3-Stage Hybrid feature selection approach and a Stacked Classification model to evaluate the cervical cancer dataset obtained from the UCI Machine Learning Repository with 35 features and one outcome variable. In Stage 1, researchers used a Genetic Algorithm (GA) and Logistic Regression Architecture (LRA) for Feature Selection and selected twelve features well correlated with the class but not among themselves. Stage-2 utilises the same GA and LRA for Feature Selection to select five features. In Stage 3, Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), Extra Trees (ET), Random Forest (RF) and Gradient Boosting (GDB) were used with the five features to identify patients with or without cancer. The classifiers showed improved performance metrics with reduced features. In the 66-34 split, all five machine learning methods except NB recorded 97 % accuracy with five features. Also, the Stacked model produced higher than 96 % accuracy with five features in 66-34 and 80-20 splits and 10-fold cross-validation.[26] Sobar et al in a study, predicted cervical cancer using machine learning classifiers based on behaviour and its determinants. The Naïve Bayes and Logistic Regression showed 91.67 % and 87.5 % accuracy, respectively.[27] Ceylan et al predicted cervical cancer early on using a multi-label classification technique. For multi-label classification, problem transformation methods such as Binary Relevance (BR), Classifier Chains (CC), Conditional Dependency Networks (CDN) and Label Combination were used. Sequential Minimal Optimisation, Naïve Bayes, Random Forest and J48 Decision Tree machine learning classifiers were compared for their exact match, accuracy, hamming loss and ranking loss. Except for J48-BR and J48-CDN algorithms, the accuracy percentage and exact match were over 80 %. All algorithms with CC and LC methods had nearly equal accuracy, exact match, hamming loss and ranking loss. RF algorithms based on CC and LC methods showed better performance, followed by J48-CC and J48-LC methods.[28] Gupta et al tried the random forest regression technique for the early detection of cervical cancer. Researchers used recall-based scores to check performance. The aim was to achieve higher recall scores and reduce false-positive values. The recall scores for Hinselmann, Schiller, cytology and biopsy were 0.920, 0.972, 0.912 and 0.996, respectively.[29] High performance can be achieved by reducing variance and bias in ML models. To achieve this, Ahishakiye et al used an ensemble ML classifier including a decision tree, Classification and Regression Trees, Naïve Bayes Classifier, K-Nearest Neighbour and Support Vector Machine. The method showed an accuracy of 87.91 % in cervical cancer classification.[30] Sagala et al applied different data mining algorithms (SVM, Naïve Bayes and KNN) on four different medical tests (biopsy, cytology, Hinselmann and Schiller) as target variables. The Naïve Bayes classifier outperforms other classifiers after evaluation using the 10-fold cross-validation method.[5]

Many datasets have been characterised by low sample size, outliers and multiple risk factors. The dataset issues such as outliers and data imbalance were addressed by Ijaz et al in a random forest classification model. Researchers used density-based spatial clustering of applications with noise (DBSCAN) and isolation forest (iForest) for outlier detection. The synthetic minority over-sampling technique (SMOTE) and SMOTE with Tomek link (SMOTETomek) were used for data imbalance. The four protocols were compared: (1) DBSCAN + SMOTETomek + RF, (2) DBSCAN + SMOTE+ RF, (3) iForest + SMOTETomek + RF and (4) iForest + SMOTE + RF. The iForest with SMOTE and iForest with SMOTETomek had better performance than DBSCAN with SMOTE and DBSCAN with SMOTETomek.[31] Similarly, Ali et al used three feature transformation methods, including log, sine function and Z-score, before performing supervised classification training. Random Tree showed the best accuracy for the biopsy (98.33 %) and cytology (98.65 %) classification, whereas Random Forest and Instance-Based K-nearest neighbour (IBC) was the best for Hinselmann (99.16 %) and Schiller (98.58 %) respectively. The logarithmic method performed best for biopsy datasets, whereas the sine function showed superior performance for cytology. Both logarithmic and sine functions were superior for the Hinselmann dataset, while Z-score performed best for the Schiller dataset.[22] Similarly, Fernandes et al proposed a computationally automated strategy to predict biopsy results from cervical risk factors. The strategy consists of joint and fully supervised optimisation of dimensionality reduction. Further, the approach was instantiated with deep learning architectures, which showed results (AUC = 0.6875) that outperformed previously developed methods, such as denoising autoencoders.[15] Chauhan et al compared various ML classifiers for predicting cervical cancer, including Logistic Regression, Naive Bayes, Support Vector Machine, K-Nearest Neighbour, Linear Discriminant Analysis, Multi-Layer

Perceptron, Decision Tree and Random Forest. The authors used Synthetic Minority Oversampling Technique for the data imbalance issue. Fivefold cross-validation was used on scaled data and unscaled data obtained by Min-Max scaling, standard scaling and normalisation. RF, SVM and DT showed higher performance in cervical cancer diagnosis. The optimised features were selected using univariate feature selection and Recursive feature elimination (RFE). Overall performance of Random Forest predictor with RFE (RF-RFE) is superior to all others being implemented. The outcome of Random Forest with Recursive feature elimination was greater than other machine learning classifiers.[32] With an appropriate dataset and optimised feature selection, machine learning methods and classification can detect cervical cancer in its early stages.

Conclusion

The present study selected optimised features among multiple risk factors to train various ML classifiers to predict cervical cancer. The results showed the Fine Gaussian SVM classifier is the best model for Hinselmann, cytology and biopsy, whereas Boosted trees performed best in the classification of Schiller.

Acknowledgements

None.

Conflict of interest

None.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394–424.
2. Zhang S, Xu H, Zhang L, Qiao Y. Cervical cancer: Epidemiology, risk factors and screening. Chinese J Cancer Res 2020 Dec 31;32(6):720-8.
3. Kisling LA, M Das J. Prevention strategies. [Updated 2022 May 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. [cited 2022 Jun 6]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK537222/
4. Momenimovahed Z, Momenimovahed Z, Salehiniya H. Incidence, mortality and risk factors of cervical cancer in the world. Biomed Res Ther 2017 Dec;4(12):1795–811.
5. Sagala NTM. A Comparative study of data mining methods to diagnose cervical cancer. J Phys Conf Ser 2019 Aug;1255(1):012022. doi:10.1088/1742-6596/1255/1/012022.
6. Parikh RB, Manz C, Chivers C, Regli SH, Braun J, Draugelis ME, et al. Machine learning approaches to predict 6-month mortality among patients with cancer. JAMA Netw Open 2019 Oct 2;2(10):e1915997. doi: 10.1001/jamanetworkopen.2019.15997.
7. Arezzo F, La Forgia D, Venerito V, Moschetta M, Tagliafico AS, Lombardi C, et al. A machine learning tool to predict the response to neoadjuvant chemotherapy in patients with locally advanced cervical cancer. Appl Sci 2021;11(2):823. doi:10.3390/app11020823
8. Kuruvilla A, Jayanthi B. Analysis and review on feature selection and classification methods on cervical cancer. Ictact J Soft Comput 2022;12(2):2551-8.
9. Jahan S, Islam MDS, Islam L, Rashme TY, Prova AA, Paul BK, et al. Automated invasive cervical cancer disease detection at early stage through suitable machine learning model. SN Appl Sci 2021 Oct;3(10):1–17.
10. R KK, Kanimozhi N. Prediction of cervical cancer using machine learning and deep learning algorithms the Creative Commons Attribution License (CC BY 4.0). Int J Trend Sci Res Dev 2020 Oct;4(6):426–30.
11. Mehmood M, Rizwan M, Gregus M, Abbas S. Machine learning assisted cervical cancer detection. Front Public Heal 2021 Dec 23;9:788376. doi: 10.3389/fpubh.2021.788376.
12. Alsmariy R, Healy G, Abdelhafez H. Predicting cervical cancer using machine learning methods. Int J Adv Comput Sci Appl 2020;11(7):173–84.
13. Asadi F, Salehnasab C AL. Supervised algorithms of machine learning for the prediction of cervical cancer. J Biomed Phys Eng 2020;10(4):513–22.
14. Moldovan D. Cervical cancer diagnosis using a chicken swarm optimization based machine learning method. Int Conf EHB 2020 Dec;10;1–4.
15. Fernandes K, Chicco D, Cardoso JS, Fernandes J. Supervised deep learning embeddings for the prediction of cervical cancer diagnosis. Peer J Comput Sci 2018 May 14;4:e154. doi: 10.7717/peerj-cs.154.
16. Wu W, Zhou H. Data-driven diagnosis of cervical cancer with support vector machine-based approaches. IEEE Access 2017;5:25189-95.
17. Team MATLAB. Statistics and Machine Learning Toolbox. Natick, Massachusetts: The Mathworks Inc; 2019.
18. JASP Team. JASP (Version 0.16.2.0)[Computer software] [Internet]. 2022. [cited 2022 Jun 6]. Available from: https://jasp-stats.org/.
19. Razali N, Mostafa SA, Mustapha A, Wahab MHA, Ibrahim NA. Risk factors of cervical cancer using classification in data mining. J Phys Conf Ser 2020 Apr;1529(2):022102. doi:10.1088/1742-6596/1529/2/022102.
20. Tanimu JJ, Hamada M, Hassan M, Kakudi HA, Abiodun JO. A machine learning method for classification of cervical cancer. Electronics 2022;11(3):463. doi:10.3390/electronics11030463.
21. CH N, Sai PP, Madhuri G, Reddy KS, Simha B, Reddy DV. Artificial Intelligence based Cervical Cancer Risk Prediction Using M1 Algorithms. 2022 Int Conf Emerg Smart Comput Informatics 2022 Mar;1–6. doi: 10.1109/ESCI53509.2022.9758241.

22. Ali MM, Ahmed K, Bui FM, Paul BK, Ibrahim SM, Quinn JMW, et al. Machine learning-based statistical analysis for early stage detection of cervical cancer. Comput Biol Med 2021 Dec;139:104985. doi: 10.1016/j.compbiomed.2021.104985.
23. Nithya B, Ilango V. Evaluation of machine learning based optimized feature selection approaches and classification methods for cervical cancer prediction. SN Appl Sci 2019 Jun;1(6):1–16.
24. Zahras D, Rustam Z. Cervical cancer risk classification based on deep convolutional neural network. Proc ICAITI 2018 - 1st Int Conf Appl Inf Technol Innov Towar A New Paradig Des Assist Technol Smart Home Care 2018 Jul 2;149–53.
25. Asaduzzaman S, Ahmed MR, Rehana H, Chakraborty S, Islam MS, Bhuiyan T. Machine learning to reveal an astute risk predictive framework for Gynecologic Cancer and its impact on women psychology: Bangladeshi perspective. BMC Bioinformatics 2021 Apr 24;22(1):213. doi: 10.1186/s12859-021-04131-6.
26. Chaudhuri AK, Ray A, Banerjee DK, Das A. A multi-stage approach combining feature selection with machine learning techniques for higher prediction reliability and accuracy in cervical cancer diagnosis. Int J Intell Syst Appl 2021 Oct 8;13(5):46–63.
27. Sobar, Machmud R, Wijaya A. Behavior determinant based cervical cancer early detection with machine learning algorithm. Adv Sci Lett 2016 Oct 1;22(10):3120–3.
28. Ceylan Z, Pekel E. Comparison of multi-label classification methods for prediagnosis of cervical cancer. Int J Intell Syst Appl Eng 2017 Dec;5(4):232–6.
29. Gupta A, Anand A, Hasija Y. Recall-based Machine Learning approach for early detection of Cervical Cancer. 2021 6th International Conference for Convergence in Technology (I2CT) 2021;1-5. doi:10.1109/I2CT51068.2021.9418099.
30. Ahishakiye E, Wario R, Mwangi W TD. Prediction of cervical cancer basing on risk factors using ensemble learning. 2020 IST-Africa Conference (IST-Africa) 2020;1-12.
31. Ijaz MF, Attique M, Son Y. Data-driven cervical cancer prediction model with outlier detection and over-sampling methods. Sensors (Basel) 2020 May 15;20(10):2809. doi: 10.3390/s20102809.
32. Chauhan NK, Singh K. Performance assessment of machine learning classifiers using selective feature approaches for cervical cancer detection. Wirel Pers Commun 2022 Jan;124(3):2335–66.