J Immunol Clin Microbiol
Available online at: https://dergipark.org.tr/tr/pub/jicm
DOI: 10.58854/jicm.1609786

Journal of Immunology and Clinical Microbiology

ORIGINAL ARTICLE / ÖZGÜN MAKALE

Prediction of high-risk human papillomavirus after conization by machine learning methods

Konizasyon operasyonu sonrasında yüksek riskli HPV kalıcılığının makine öğrenme yöntemi ile ön görülmesi

Erol Ozcekic[1], Akin Usta[2], Duygu Lafci[2], Orkun Cetin[2],
Yener Ozel[3] Gökberk Kozak[1]

[1] Department of Computer Technology, Balıkesir Vocational School, Balıkesir University, Balıkesir, Turkiye,
[2] Department of Obstetrics and Gynaecology, Faculty of Medicine, Balıkesir University, Balıkesir, Turkiye,
[3] Department of Medical Microbiology, Faculty of Medicine, Balıkesir University, Balıkesir, Turkiye

Geliş: 13.01.2025, Kabul: 12.05.2025

Abstract

Objective: This study aimed to use machine learning, a branch of artificial intelligence, to predict the persistence of high-risk HPV in women who have undergone conization surgery.
Materials and Methods: This retrospective study was conducted between 2018 and 2023 in the Gynecology and Obstetrics Clinic of Balıkesir University Health Practice and Research Hospital. A dataset of 69 female patients between the ages of 23-67 years; for the prediction of HPV status 1 year after the conization operation, the patients' data were recorded according to the criteria we determined, and these data were analyzed and classified using machine learning methods. Various Machine Learning methods such as Gradient Boosting, Support Vector Machine (SVM), Catboost, Random Forest (RF), and Naive Bayes (NB) are used here.
Results: We found the highest accuracy rate in Random Forest, and Catboost with 76 %. Gradient Boosting followed with a score of 67%, and Naive Bayes and Support Vector Machine (SVM) performed considerably lower, with scores of 48% and 43%, respectively.
Conclusions: Our results show that machine learning, a novel use of artificial intelligence, is effective in predicting the persistence of high-risk HPV. Further studies with more data will be a promising and useful tool for HPV and cervical cancer screening in the future.
Keywords: Cervical intraepithelial neoplasia, Human papillomavirus, Artificial intelligence, Machine learning method

Corresponding Author: Duygu Lafci Department of Obstetrics and Gynaecology, Faculty of Medicine, Balıkesir University, Balıkesir, Turkiye, E mail: duygulafci94@gmail.com

How to Cite: Ozcekic E, Usta A, Lafci D, Cetin O, Ozel Y, Kozak G. Prediction of high-risk human papillomavirus after conization by machine learning methods. *Journal of Immunology and Clinical Microbiology* 2025;10(1):11-22

Öz

Amaç: Bu çalışma, konizasyon ameliyatı geçiren kadınlarda yüksek riskli HPV'nin kalıcılığını tahmin etmek için yapay zekanın bir dalı olan makine öğrenimini kullanmayı amaçladı.
Gereç ve Yöntem: Bu retrospektif çalışma 2018-2023 yılları arasında Balıkesir Üniversitesi Sağlık Uygulama ve Araştırma Hastanesi Kadın Hastalıkları ve Doğum Kliniğinde gerçekleştirildi. 23-67 yaş arası 69 kadın hastadan oluşan veri seti; Konizasyon operasyonundan 1 yıl sonra HPV durumunun tahmini için hastaların verileri belirlediğimiz kriterlere göre kayıt altına alındı ve bu veriler makine öğrenmesi yöntemleri kullanılarak analiz edildi ve sınıflandırıldı. Burada Gradient Boosting, Support Vector Machine (SVM), Catboost, Random Forest (RF) ve Naive Bayes (NB) gibi çeşitli Makine Öğrenimi yöntemleri kullanılmaktadır.
Bulgular: En yüksek doğruluk oranını %76 ile Random Forest ve Catboost'ta bulduk. Bunu %67 puanla Gradient Boosting takip ederken, Naive Bayes ve Support Vector Machine (SVM) sırasıyla %48 ve %43 puanlarla oldukça düşük performans gösterdi.
Sonuçlar: Bu çalışma, yapay zekanın yeni bir kullanımı olan makine öğreniminin, yüksek riskli HPV'nin kalıcılığını tahmin etmede etkili olduğunu göstermektedir. Daha fazla veri içeren ileri çalışmalar gelecekte HPV ve rahim ağzı kanseri taraması için umut verici ve faydalı bir araç olacaktır.
Anahtar Kelimeler: Servikal intraepitelyal neoplazi, İnsan papilloma virüsü, Yapay zeka, Makine öğrenme yöntemi

INTRODUCTION

Papillomaviruses are small, non-enveloped viruses with an approximately 8 kb circular double-stranded DNA genome containing a non-coding regulatory long control region and eight open reading frames (1). Human papillomavirus (HPV) represents the most common sexually transmitted infectious agent globally. The virus, which has approximately 220 different subtypes, can infect the skin and mucous membranes (2,3). Of these, 40 subtypes have the potential to infect the anogenital tract and especially subtypes 16 and 18 have been identified as critical high-risk types for the development of cervical cancer (CC) (4).

In 2020, 604,000 new cases of CC were diagnosed, 342,000 of which resulted in death. It is recognized as the fourth leading cause of cancer deaths in women (5). Almost all CC cases (99.7%) can be attributed to the presence of HPV(6). The development of CC is significantly influenced by persistent infections from high-risk human papillomavirus (HR-HPV) strains (7,8). Fourteen HR-HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) are associated with CC (9,10). The most common types of HR-HPV are HPV-16 and HPV-18. In our country, where HPV-based cervical cancer screening has been implemented since 2014, women between the ages of 30-65 are screened. Colposcopy is recommended for women who are HPV 16 and/or 18 positive, regardless of cytology. In other high-risk HPV-positive women, co-testing is recommended after 1 year in the presence of normal cytology. At the same time, colposcopy is recommended in the presence of abnormal cytology (11). Conization is the process of cone-shaped removal of the cervix by covering the transformation zone according to the colposcopy result.

While cervical cancer screening protocols are increasingly moving towards HPV-based tests, the accuracy of abnormal cytological findings in the detection of dysplastic lesions is still of great importance. Conventional cervical cytology and colposcopy show higher accuracy rates as the grade of lesions increases. Detection of HPV may help to avoid unnecessary surgical procedures, especially in the case of ASCUS (12).

Our aim in this study is to develop a new strategy to minimize the unnecessary administration of treatments, such as colposcopy and conization. The new prediction strategy will reduce high healthcare costs and maximize the benefits of CC screening worldwide. Based on this knowledge, we aimed to predict the recurrence and persistence of HPV in HPV-positive patients using machine learning methods, a branch of artificial intelligence (AI).

MATERIAL AND METHODS

Study design

This study included 69 female patients between the ages of 23 and 67 years (mean 26) who applied to the Obstetrics and Gynecology Outpatient Clinic of xxx University Health Practice and Research Hospital for routine gynecological examination or gynecological complaints between 2018 and 2023.

Demographic data such as age, BMI, blood group, gravida, parity, abortion history, smoking, history of use of oral contraceptives or intrauterine device, and HPV vaccination [GARDASIL (types 6,11,16,18) and GARDASIL-9 (types 6,11,16,18,31,33,45,52,58)] were obtained from the file archive.

Table 1 Demographic data of the patients

Age	43* (23-67)
BMI	26,6* (17,2-37,2)
Gravide	3* (0-8)
Parity	3* (0-5)
Abortion	1* (0-5)

*(*median)*

Age, BMI, gravida, parity, and abortion were entered into the artificial intelligence model with the patients' exact values (Table 1).

Patients' ABO-Rh values were trained as group 1 (A Rh+), group 2 (0 Rh+), group 3 (B Rh+), group 4 (A Rh-), group 5 (0 Rh-), group 6 (AB Rh+), group 7(no data) in the artificial intelligence system. Smoking was classified as minimum 5 pack/year smokers (group 1), non-smokers (group 2), and no data (group 3). Patients with Gardasil-4 vaccines were classified as Group 1, patients

Table 2 Demographic data of patients (each section is grouped separately in the artificial intelligence)

Group	AB0-Rh	Smokers	HPV vaccination	OC or IUD
1	20(ARh+)	26(+)	5 (Gardasil4)	13 (+)
2	17(0Rh+)	23(-)	4 (Gardasil9)	36(-)
3	13(BRh-)	20(no data)	40 (no vaccination)	20(no data)
4	10(Arh-)		20(no data)	
5	4(0Rh-)			
6	3(ABRh+)			
7	2(no data)			

(HPV: Human papillomavirus, OC: Oral Contraceptive, IUD: Intrauterine device) with Gardasil-9 vaccines as Group 2, those with no HPV vaccination as Group 3, and those with no data as Group 4. Patients with any history of oral contraceptives or intrauterine device use or with use for less than 5 years were classified as group 1, those with a history of use for more than 5 years were classified as group 2 and no data was group 3 (table 2).

Additionally, values included in the artificial intelligence evaluation were the HPV results before conization and the histopathological examination results following conization.

HPV results before conization were classified as HPV 16-18 positive group 1, HR-HPV other types (HPV 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59) positive group 2, and HR-HPV negative group 3. Patients with no data were classified as group 4 and patients who were positive for HR-HPV 16-18 and others together were classified as group 5 (Table 3).

The colposcopic biopsy results were categorized by the artificial intelligence system as follows: HPV 16-18 positive group 1, HR-HPV (other types, including HPV 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) positive group 2, and HR-HPV negative group 3 (Table 4).

The conization pathology results were classified as CIN 1 (Group 1), CIN 2 and/or CIN 3 (Group 2), and negative Group 3 (Table 5).

Demographic data for some patients were not available; however, their age and BMI values were included in the artificial intelligence models, calculated using the median value, and the remaining demographic data were coded as unavailable.

Random Forest, Gradient Boosting, SVM, CatBoost, and Naive Bayes were used as artificial intelligence models for the study. All models aimed to predict the HPV results in 1 year after conization. HPV values one year later were coded as HPV 16-18 class 1, HPV other class 2, and HPV negative class 3 in the artificial intelligence system. After 1 year, there were no patients with HPV 16-18 and the other group together, so no additional class was created.

Table 3 Distribution of HPV types before conisation operation

	Number of Observations	Observation Percentage	Group entry to AI
HPV-16-18	22	31,8%	1
HPV other	24	34,7%	2
Negative	11	15,9%	3
No data	5	7,2%	4
HPV 16-18 and others	7	10,1%	5

Table 4 Distribution of colposcopic biopsy results before conisation operation

CIN Types	Number of Observations	Observation Percentage	Group entry to AI
CIN-I	31	44,9%	1
CIN-II-III	23	47,8%	2
Negative	5	7,2%	3

Table 5. Distribution of conisation histopathological results results

CIN Types	Number of Observations	Observation Percentage	Group entry to AI
CIN-I	13	18,8%	1
CIN-II-III	23	23%	2
Negative	24	34,7%	3

In conclusion, HPV results before conization, conization pathology results, and the demographic data listed above were used to train the artificial intelligence to assess the predictability of HPV results 1 year after conization.

2.2. DNA Isolation

The presence and genotype distribution of HPV-DNA in cervical swabs were investigated by real-time PCR (RT-PCR). Swab samples obtained from the patients during gynecological examination were transferred to a Digene HC2 DNA collection device (Qiagen, Germany), a liquid-based transport medium, and sent to the microbiology laboratory. DNA isolation from these cervical swab samples was performed using the EZ1 Advanced XL isolation robot (Qiagen, Germany) with the EZ1 virus mini kit, following the manufacturer›s instructions.

2.3. Detection of HR-HPV Positivity by Real-Time PCR

The presence of HPV-DNA was first detected with the QiaScreen commercial kit (Qiagen, Germany) on the Qiagen Rotor-Gene Q device (Qiagen, Germany). HPV 16, 18, and other HPV types (31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) were detected in DNA samples isolated with this kit.

2.4. Patient Selection in Colposcopic **Examination**

CC screening involved colposcopic examination in patients with HPV16 and 18, who were grouped as high-risk HPV (HPV subgroups associated with CC) and other HR-HPV-positive patients with abnormal cytological findings in cytological examinations. In colposcopy, firstly, the mucus in the cervix was cleared, and then 3% acetic acid was applied for at least 30 seconds, followed by treatment with Lugol solution. The cervix was initially examined classically using low magnification. Colposcopy, columnar epithelium, squamous epithelium, and transformation zone were evaluated. BX specimens were taken from the areas with leukoplakia, hyperkeratosis, aceto-white epithelium, fine or coarse punctuation, atypical vascularity, and mosaicism and subjected to histological and immunohistochemical examination.

2.5. Conization

Treatment of CIN lesions involves ablation or excision of the cervix to include the site of transformation, followed by histologic evaluation and assessment of lesion margins. In the conization procedure, the bladder is emptied to prevent bladder injury in anesthetized patients. Following appropriate wiping and draping, the cervix was held with a tenaculum to facilitate cervical movement. With the help of a scalpel and cautery, the cervical lesion was excised in a cone shape at least 5-7 mm deep from the surface of the CIN, including the cervical lesion and the transformation site. The material was subjected to histologic and immunohistochemical examination (13).

2.6. Statistical Analysis

Machine learning is a set of statistical techniques that imitate the human mind, analyze data, and develop correlations to facilitate decision-making when it is difficult or impractical to perform the necessary procedures (14). In this study, Gradient Boosting (GB), Support Vector Machine (SVM), Catboost, Random Forest (RF), and Naive Bayes (NB) algorithms were used for various methods, including artificial intelligence, machine learning, and deep learning.

2.6. Processing of data with machine learning methods

Our computer system that we utilize has the following specifications: Intel® Core™ i5-9400F CPU (2.9 GHz, 9 MB cache, 6 cores), 16 GB RAM, and Windows 10 Pro 64-bit operating system. The data was transferred to a computer program in MATLAB (The MathWorks Inc., MATLAB version: 9.13.0 R2022b) to perform several machine learning methods.

GB is an algorithm for finding the minimum of the function concerning the gradient model parameter by taking the derivative of the loss function with respect to the predicted value (15).

CatBoost is an algorithm that is a member of the Gradient Boosted Decision Trees family, which enables the use of categorical, heterogeneous data by using big data content in machine learning studies (16). SVM is a method developed to perform both linear and non-linear multivariate classification tasks. SVM transforms the input vectors into a high-dimensional feature space and generates a linear decision boundary to distinguish between different classes of data points. SVM is highly effective in processing small and limited datasets (17). The essence of the random forest algorithm is that it creates many decision trees and tries to aggregate the predictions on the trees. It is used in the field of machine learning for classification, regression, and other tasks. Naive Bayes, which works with Bayesian theory, is a probability-based algorithm that calculates the probability of belonging to a particular class. It is used for classification and other tasks in the field of machine learning (18).

TP (true positive) signifies the count of samples accurately predicting the positive category, TN (true negative) represents the count of samples accurately predicting the negative category, FP (false positive) denotes the count of samples incorrectly predicting the positive category, and FN (false negative) indicates the count of samples incorrectly predicting the negative category (19).

The classification status was analyzed by using the findings such as accuracy recall, precision, and F1 score related to the classification prediction. Accuracy is the proportion of accurate predictions out of the total number of predictions. Precision is the ratio of positive predictions that truly belong to the positive class to all positive predictions. Recall is the portion of positive examples in the dataset predicted as positive. The F1 Score is a machine learning evaluation metric used to evaluate the performance of a classification model. The F1 Score is defined as the harmonic mean of precision and recall (19,20). The formulas used to calculate these metrics are listed in the range of Equation 1 - Equation 3 (22,23).

Equations;

1-Accuracy=(TP+TN)/(TP+TN+FP+FN)

2-Precision=TP/(TP+FP)

3-F1Score=2×((Precision×Sensitivity)/(Precision+Sensitivity))

RESULTS

69 patients were included in the study, 22 (31,8%) of them were positive for HPV types 16-18, 24 (34,7%) of them for other HR-HPV types, and 7 (%10,1) patients for HR-HPV types 16-18 and others. 11 (15,9%) patients were negative for HR-HPV types. In other words, HR-HPV was positive in 62 (89.85%) and negative in 7 (10.15%) of 69 patients before the conization operation. In 5 patients, the HPV result could not be obtained before the conization operation. The distribution of HPV types is shown in Table 3.

The histopathological colposcopic biopsy results before the conization operation were evaluated, 31 (44.9%) of 69 patients had CIN-I (LSIL- Cervical intraepithelial neoplasia), 33 (47.8%) had CIN II and/or III (HSIL- High grade cervical intraepithelial lesion), 5 (7.2%) of the patients had no cytology. The distribution of histopathology biopsy values is summarised in Table 4.

For predicting the HPV result, the histopathological evaluation result of conization was also taken into consideration. Of the 69 patients who underwent conization operation, 13 (18.8%) had CIN I, 32 (46.3%) had CIN II and/or III, and 34 (24,7%) had negative results. This distribution is shown in Table 5.

The confusion matrices obtained as a result of the classification of the dataset by Random Forest, GradientBoosting, SVM, Catboost, and Naive Bayes methods are shown in Figure 1-5.

Regarding model performance, both Random Forest and CatBoost demonstrated superior predictive capabilities, achieving scores of 0.76 (76%) each. Gradient Boosting followed with a score of 0.67 (67%), indicating solid but slightly less robust performance. On the other hand, Naive Bayes and Support Vector Machine (SVM) performed considerably lower, with scores of 0.48 (48%) and 0.43 (43%), respectively. These results suggest that tree-based ensemble methods, particularly Random Forest and CatBoost, are more effective for this classification task, while linear models, such as SVM, and probabilistic models, like Naive Bayes, may not capture the underlying patterns in the dataset as effectively (Figure 6).

As a result of the analyses performed, the precision, recall, and F1 score performance criteria values of each class are shown in Table 6, Table 7, and Table 8, respectively.

Table 6 Class 1 HPV16-18

Model	Precision	Recall	F1 Sc.
SVM	0.00	0.00	0.00
GB	0.00	0.00	0.00
RF	0.00	0.00	0.00
NB	0.00	0.00	0.00
CB	1.00	0.25	0.40

Table 7 Class 2 HR-HPV other

Model	Precision	Recall	F1 Sc.
SVM	0.69	0.56	0.62
GB	0.88	0.88	0.88
RF	0.80	1.00	0.89
NB	0.75	0.56	0.64
CB	0.83	0.94	0.88

Table 8 Class 3 HPV negative

Model	Precision	Recall	F1 Sc.
SVM	0.00	0.00	0.00
GB	0.00	0.00	0.00
RF	0.00	0.00	0.00
NB	0.11	1.00	0.20
CB	0.00	0.00	0.00

Figure 1 Support Vector Machine (SVM)

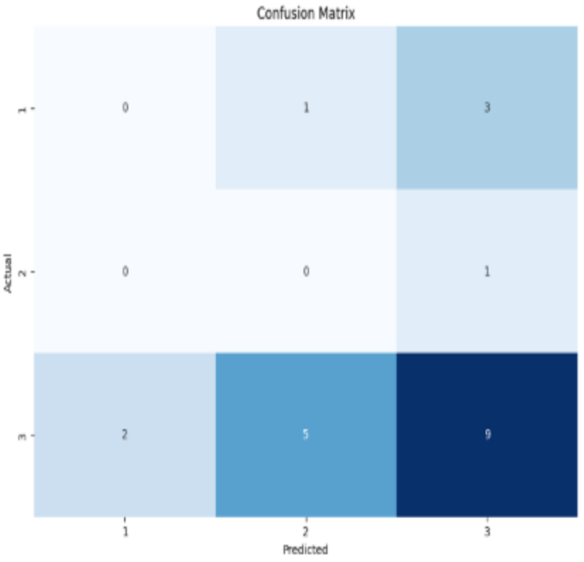


Figure 2 Naive Bayes

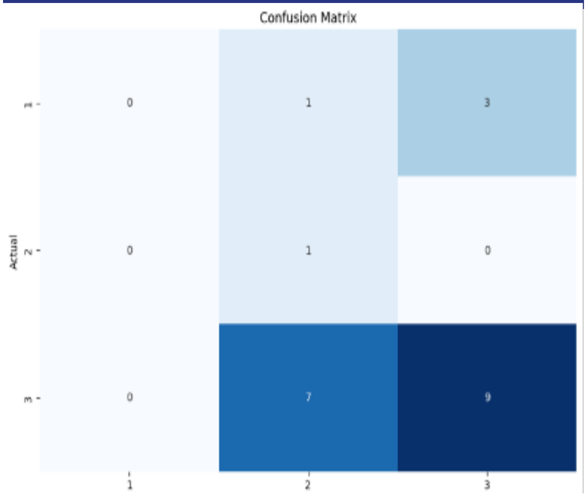


Figure 3 Random Forest

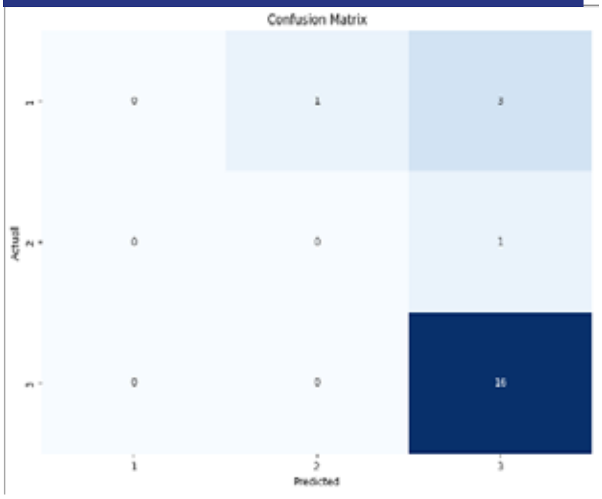


Figure 4 Gradient Boosting

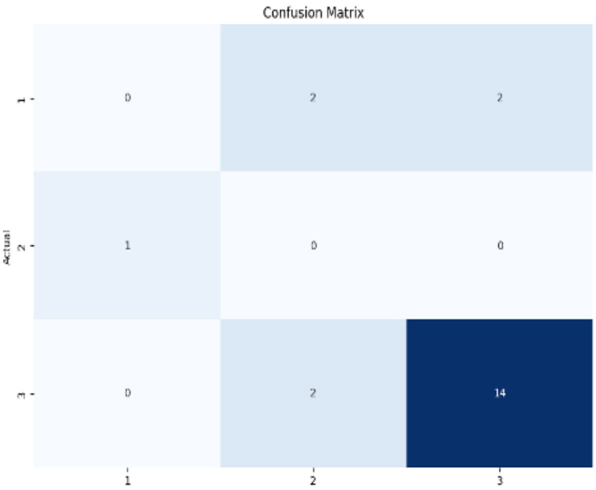


Figure 5 Catboost

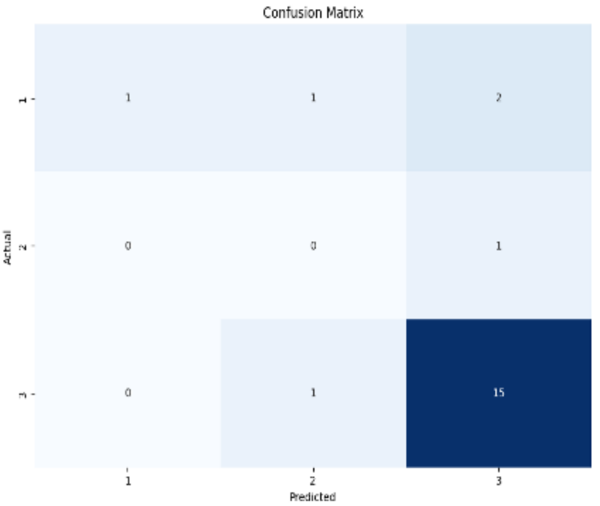
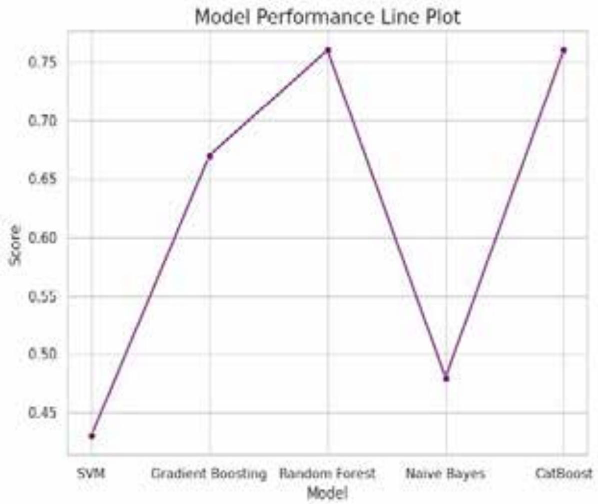


Figure 6 Predictive of Model Performance



DISCUSSION

CC is a severe health problem that affects approximately 500,000 women worldwide every year. Among the risk factors for the development of CC, the leading risk factor is exposure to HPV. Nearly all cases of CC are attributed to HPV, a virus that can be transmitted from one person to another through sexual activity. Several variations of HPV exist. Certain strains of HPV can induce alterations in a woman›s cervix, potentially progressing to CC, whereas other types can result in the development of genital or skin warts (24). In the majority of women, HPV typically resolves on its own; however, if it persists, there is a potential risk of it progressing to CC over time. The virus is mainly acquired through sexual activity. Moreover, an elevated risk of developing CC is associated with long-term use of birth control pills (5 years or more), having given birth to 3 or more children, having multiple sexual partners, smoking, and immune system disorders (25).

The PAP Smear Test, the most commonly used screening method, is a cytological screening test based on the collection and examination of cervical cells. After this scan, the doctor will visualize the cervix with a colposcopy device. This cytological screening test detects precursor cervical lesions. With cancer treatment at this stage, cancer development can be prevented before the cells turn into cancer. The PAP test, employed for early diagnosis, is a readily applicable, cost-effective, non-invasive, highly sensitive screening method that minimizes mortality, morbidity, and the burden of treatment (26,27). Another important method for early diagnosis and screening of CC is the HPV DNA test. Of all available cervical screening tests, the HPV test is the most objective and reproducible. The association between the HPV DNA test and CC has been established, with the presence of HPV DNA demonstrated in 99.9% of CC patients (27).

Cytological evaluation of cervical smears and simultaneous HPV DNA testing of the same sample is called «co-test». Co-testing is currently the most widely accepted screening method for women over 30 years of age (28). American Society of Colposcopy and Cervical Pathology and American College of Obstetricians and Gynecologists recommend that women aged 30-65 years should be screened every five years with a co-test. Since HPV infections in people under the age of thirty are quite transient, a standard test is not recommended for this age group. Therefore, screening with cytology only is recommended every three years between 21 and 30 years of age (27).

Previous studies have shown that HPV persistence persists for a long time in 10-20% of patients and results in cervical squamous intraepithelial lesions (SIL). HSIL has a high carcinogenic risk and should be treated on time for viral DNA integration, while unnecessary screening and treatment should be avoided as LSIL is mostly associated with transient HPV infection, and 60% of these are known to resolve within 1 year (29,30).

Artificial intelligence is a field of computer science that focuses on analyzing complex medical data and can be used to predict diagnosis, treatment, and outcomes in a variety of clinical situations (28) [35]. Artificial intelligence applications such as machine learning methods have been widely used in medicine and significant gains have been achieved. Research on using artificial intelligence applications in HPV screening methods has started to appear in the literature (31). Xue et al. stated that the diagnostic performance of the artificial intelligence-assisted liquid-based cytology (AI-LBC) triage approach was compared with cytologist evaluations and HPV16/18 genotyping, and the sensitivity of AI-LBC was found to be similar to that of cytologists, but significantly higher than HPV16/18 typing in CIN II+ detection (23). It is reported that AI-LBC reduced the colposcopy rate by 10 % compared to cytologists. Colposcopy diagnosis and referral to biopsy are essential components of CC screening programs (22). However, the performance of these programs is limited by the need for experienced colposcopists. Xue et al. developed an AI-assisted colposcopic companion diagnostic system to grade colposcopic impressions and guide biopsies by processing colposcopic images, clinical information, and pathology results of 19,435 patients to train the AI algorithm, evaluate its performance, and validate the results. The concordance between colposcopic impressions graded by their AI algorithm and pathology findings was higher than that of colposcopies interpreted by colposcopists. It showed higher sensitivity than colposcopists for detecting pathologic high-grade squamous intraepithelial lesions or worse (HSIL+) (22).

Another use of artificial intelligence applications in cervical cancer screening is to analyze cervical cytology on histopathological images. Tan et al. and Bumrungthai et al. stated that by collecting cytological test images diagnosed by pathologists from several partner hospitals in different regions, a deep convolutional neural network model that can assist pathologists in cervical cancer screening is designed (24,32). The images were divided into a training dataset (13,775 images), a validation dataset (2301 images), and a test dataset (408,030 images from 290 scanned replicates) for training and impact evaluation of the faster region-based convolutional neural network system. It has been reported that the system has a sensitivity of 99.4 % and specificity of 34.8%, can distinguish between negative and positive cells, and has high sensitivity in the evaluation of ASCUS, LSIL, and HSIL (24).

Ali et al. stated that various machine learning algorithms such as Instance-Based K-nearest neighbor (IBK), KStar, Split-Point and Attribute Reduced Classifier (SPAARC), Random Tree (RT), and Random Forest (RF) are used to detect early-stage cervical cancer (25). The input features of the study include parameters such as the patient›s age, number of sexual partners, age at first sexual intercourse, and gestational age. The output features of the study are positive or negative biopsy, cytology, Hinselmann's and Schiller's results (26). As a result of the analyses performed in Ali et al.'s study, the performance of the methods was measured with the help of accuracy, precision, sensitivity, F1 score, and specificity metrics (25). Among all the applied classifiers, the best classifiers for the accuracy metric are RT for biopsy and cytology, RF for Hinselmann, and IBK for Schiller. The best sensitivity, precision, and F1 score values were obtained with RT for biopsy and cytology, RF for Hinselmann, and IBk for Schiller. The best values for specificity were obtained with RF for biopsy, Hinselmann and Schiller, and SPAARC for cytology(25).

Using machine learning methods to detect cervical cancer, prediction models are developed in Lilhore et al.'s study (33). In these models, SVM, RF, DT, and Boruta Analysis methods were used to search for the model that provides the most effective feature selection and classification service for cervical cancer datasets. As a result of the study, Boruta analysis and random forest methods performed better than other methods in terms of accuracy, precision, and other parameters in determining the risk and type of cervical cancer.

Currently, there has been extensive research and advances in machine learning algorithms for diagnosing, predicting, and forecasting the onset of numerous diseases.

In this study, both Random Forest and CatBoost showed superior prediction abilities with 76% prediction rate in predicting HPV persistence using machine learning methods of AI. Machine learning, a branch of AI, can predict HPV persistence 1 year later based on age, smoking, pregnancy history, BMI, baseline HPV status, and colposcopic and conization histopathology results. These results suggest that the machine learning method of AI can be used to develop an effective prediction strategy for the incidence of persistence in high-risk HPV-positive women.

Our study included a relatively small sample size of 69 patients, and both the missing data and the small number of patients result in less reliable and less generalizable findings. Therefore, prospective and more comprehensive studies on this topic are needed to evaluate the performance of AI strategies. For the first time in the literature, we evaluated the predictive performance of machine learning methods in screening for human papillomavirus (HPV) persistence after a conization operation. If we expect high-risk CIN after the following HPV result, planned conization can be performed more thoroughly by gynecologists to reduce recurrences. Adopting this method can not only increase the success of the final cervical conization but also reduce overtreatment procedures and unnecessary deep conizations. This AI-linked strategy maximizes the benefits of CC screening while minimizing potential unnecessary healthcare costs globally. CIN is a disease that patients are concerned about, and AI is creating a method that enables more objective counseling about patients› prognosis.

CONCLUSION

Our results demonstrate that machine learning, a novel application of artificial intelligence, is effective in predicting the persistence of high-risk human papillomavirus (HPV). Further studies with more data will be a promising and useful tool for HPV and cervical cancer screening in the future.

ACKNOWLEDGMENTS

Conflict of interest statement: The authors report no conflict of interest.

Funding: This study received no extramural funding

Author Contribution Concept: EÖ, YÖ., Design: AU, DL, OC, Data Collection or Processing: AU, DL, OC, Analysis or Interpretation: GK., YÖ, EÖ, Literature Search: DL, GK, YÖ, Writing: EÖ, AU, DL.

REFERENCES

1. Villiers E.M.; Fauquet C.; Broker T.R.; Bernard H.U.; zur Hausen H. Classification of papillomaviruses. Virology. 2004, 324(1), 17-27.

2. Coser J.; Boeira T.R.; Wolf J.M.; Cerbaro K.; Simon D.; Lunge V.R. Cervical human papillomavirus infection and persistence: a clinic-based study in the countryside from South Brazil. Braz J Infect Dis. 2016, 20(1), 61-68.

3. Tumban E. A current update on human papillomavirus-associated head and neck cancers. Viruses. 2019, 11(10), 922.

4. Cuzick J.; Cuschieri K.; Denton K.; Hopkins M.; Thorat M.A.; Wright C.; Cubie H.; Moore C.; Kleeman M.; Austin J.; et al. Performance of the Xpert HPV assay in women attending for cervical screening. Papillomavirus Res. 2015, 1, 32-37.

5. Sung H.; Ferlay J.; Siegel R.L.; Laversanne M.; Soerjomataram I.; Jemal A.; Bray F. Global Cancer Statistics 2020: GLO-BOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021, 71(3), 209-249.

6. Okunade K.S. Human papillomavirus and cervical cancer. J Obstet Gynaecol. 2020, 40(5), 602-608.

7. Parkin D.M.; Bray F. The burden of HPV-related cancers. In HPV Vaccines and Screening in the Prevention of Cervical Cancer; Bosch F.X., Cuzick J., Schiller J.T., Garnett G.P., Meheus A., Franco E.L., Wright T.C., Eds.; Vaccine: Amsterdam, Holland, 2006; Volume 24S3, pp. S3/11-S3/25.

8. Rossi P.G.; Ricciardi A., Cohet C.; Palazzo F; Furnari1 G; Valle S; Largeron N; Federici A. Epidemiology and costs of cervical cancer screening and cervical dysplasia in Italy. BMC Public Health. 2009, 9(1), 71.

9. Nyári T.A; Kalmár L; Deák J; Szõllõsi J, Farkas I, Kovács L. Prevalence and risk factors of human papilloma virus infection in asymptomatic women in southeastern Hungary. Eur J Obstet Gynecol Reprod Biol. 2004, 115(1), 99-100.

10. Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. Clin Sci (Lond). 2006, 110(5), 525-541.

11. Ulusal Kanser Danışma Kurulu ETvTAKR, T.

12. Aydogan Kirmizi D, Baser E, Demir Caltekin M, Onat T, Sahin S, Yalvac ES. Concordance of HPV , conventional smear, colposcopy, and conization results in cervical dysplasia. Diagn Cytopathol. Ocak 2021;49(1):132-9.

13. Basu P, Taghavi K, Hu SY, Mogri S, Joshi S. Management of cervical premalignant lesions. Curr Probl Cancer. 2018 Mar-Apr;42(2):129-136. doi: 10.1016/j.currproblcancer.2018.01.010. Epub 2018 Jan 11. PMID: 29428790

14. Shah A.; Shah M.; Pandya A.; Sushra R.; Sushra R.; Mehta M.; Patel K.; Patel K. A comprehensive study on skin cancer detection using artificial neural network (ANN) and convolutional neural network (CNN). Clinical eHealth. 2023, 6, 76-84.

15. Zhang Z, Zhao Y, Canes A, Steinberg D, Lyashevska O; written on behalf of AME Big-Data Clinical Trial Collaborative Group. Predictive analytics with gradient boosting in clinical medicine. Ann Transl Med. 2019 Apr;7(7):152. doi: 10.21037/atm.2019.03.29. PMID: 31157273; PMCID: PMC6511546.)

16. (Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data. 2020;7(1):94. doi: 10.1186/s40537-020-00369-8. Epub 2020 Nov 4. PMID: 33169094; PMCID: PMC7610170.)

17. Cui L.; Zhang Q.; Shi Y.; Yang L.; Wang Y.; Wang J.; Bai C. A method for satellite time series anomaly detection based on fast-DTW and improved-KNN. Chin J Aeronaut. 2023, 36, 149-159.

18. Maheswari S, Pitchai R. Heart Disease Prediction System Using Decision Tree and Naive Bayes Algorithm. Curr Med Imaging Rev. 2019;15(8):712-7.

19. Rashed A.E.E.; Elmorsy A.M.; Atwa A.E.M. Comparative evaluation of automated machine learning techniques for breast cancer diagnosis. Biomed Signal Process Control. 2023, 86, 105016.

20. Qian L.; Bai J.; Huang Y.; Zeebaree D.Q.; Saffari A.; Zebari D.A. Breast cancer diagnosis using evolving deep convolutional neural network based on hybrid extreme learning machine technique and improved chimp optimization algorithm. Biomed Signal Process Control. 2024, 87, 105492.

21. Navazi F.; Yuan Y.; Archer N. An examination of the hybrid meta-heuristic machine learning algorithms for early diagnosis of type II diabetes using big data feature selection. Healthcare Anal. 2023, 4, 100227.

22. Xue P.; Tang C.; Li Q.; Li Y.; Shen Y.; Zhao Y.; Chen J.; Wu J.; Li L.; Wang W.; et al. Development and validation of an artificial intelligence system for grading colposcopic impressions and guiding biopsies. BMC Med. 2020, 18(1), 406.

23. Xue P.; Xu H.M.; Tang H.P.; Wu W.Q.; Seery S.; Han X.; Ye H.; Jiang Y.; Qiao Y.L. Assessing artificial intelligence enabled liq-uid-based cytology for triaging HPV-positive women: a population-based cross-sectional study. Acta Obstet Gynecol Scand. 2023, 102(8), 1026-1033.

24. Tan X.; Li K.; Zhang J.; Wang W.; Wu B.; Wu J.; Li X., Huang X. Automatic model for cervical

cancer screening based on con-volutional neural network: a retrospective, multicohort, multicenter study. Cancer Cell International. 2021, 21, 35.

25. Ali M.M.; Ahmed K; Bui F.M.; Paul B.K.; Ibrahim S.M.; Quinn J.M.W.; Moni M.A. Machine learning-based statistical analysis for early stage detection of cervical cancer. Comput Biol Med. 2021, 139, 104985.

26. Fusco E.; Padula F.; Mancini E.; Cavaliere A.; Grubisic G. History of colposcopy: a brief biography of Hinselmann. J Prenat Med. 2008, 2, 19-23.

27. Gecer M.; High-risk Human Papillomavirus (hrHPV) Prevalence and Genotype Distribution among Turkish Women. J Cytol. 2023, 40, 42-48

28. Stuebs, F.A.; Dietl, A.K.; Behrens, A.; Adler, W.; Geppert, C.; Hartmann, A.; Knöll, A.; Beckmann, M.W.; Mehlhorn, G.; Schulmeyer, C.E.; et al. Concordance Rate of Colposcopy in Detecting Cervical Intraepithelial Lesions. Diagnostics. 2022, 12, 2436

29. Guan P, Howell-Jones R, Li N, Bruni L, de Sanjosé S, Franceschi S, vd. Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. Int J Cancer. 15 Kasım 2012;131(10):2349-59.

30. Zang L, Hu Y. Risk factors associated with HPV persistence after conization in high-grade squamous intraepithelial lesion. Arch Gynecol Obstet. Aralık 2021;304(6):1409-16.

31. Hou X.; Shen G.; Zhou L.; Li Y.; Wang T.; Ma X. Artificial Intelligence in Cervical Cancer Screening and Diagnosis. Front On-col. 2022, 12 ,851367.

32. Bumrungthai, S.; Ekalaksananan, T.; Kleebkaow, P.; Pongsawatkul, K.; Phatnithikul, P.; Jaikan, J.; Raumsuk, P.; Duangjit, S.; Chuenchai, D.; Pientong, C. Mathematical Modelling of Cervical Precancerous Lesion Grade Risk Scores: Linear Regression Analysis of Cellular Protein Biomarkers and Human Papillomavirus E6/E7 RNA Staining Patterns. Diagnostics. 2023, 13, 1084.

33. Lilhore U.K.; Poongodi M.; Kaur A.; Simaiya S.; Algarni A.D.; Elmannai H.; Vijayakumar V.; Tunze G.B.; Hamdi M. Hybrid model for detection of cervical cancer using causal analysis and machine learning techniques. Comput Math Methods Med. 2022, 2022, 4688327.