Submitted: 26 October, 2022 Accepted: 11 May, 2023 Published: 15 August, 2023 DOI:10.22514/ejgo.2023.050

ORIGINAL RESEARCH

European Journal of Gynaecological Oncology

Effectiveness of artificial intelligence algorithms in identification of patients with high-grade histopathology after conisation

Marko Mlinarič[1,*,†], Maša Mlinarič[2,†], Miljenko Križmarić[3,†], Iztok Takač[4,5,†], Alenka Repše Fokter[6,†]

[1]Outpatient clinic for gynaecology and obstetrics, 1410 Zagorje ob Savi, Slovenia
[2]General Hospital Trbovlje, 1420 Trbovlje, Slovenia
[3]University of Maribor Faculty of Medicine, 2000 Maribor, Slovenia
[4]University Clinic of Gynaecology and Perinatology, University Medical Centre Maribor, 2000 Maribor, Slovenia
[5]Department of Gynaecology and Perinatology, Faculty of Medicine, University of Maribor, 2000 Maribor, Slovenia
[6]Department of Pathology and Cytology, General Hospital Celje, 3000 Celje, Slovenia

***Correspondence**
info@ginekoloska-ambulanta.si
(Marko Mlinarič)

† These authors contributed equally.

Abstract

The aim of this study was to compare effectiveness of various artificial intelligence classification algorithms in identifying patients with high-grade final histopathology of conisation based on last PAP smear result and risk factors for development of uterine cervical dysplasia and cancer. The data of 1475 patients who underwent conisation surgery at University Clinical Centre Maribor between 1993–2005 were analysed. Synthetic Minority Oversampling Technique (SMOTE) algorithm was employed for the imbalanced data correction. Various classification algorithms were tested with Weka open-source software. The 10-fold cross validation was used to define testing and hold-out set for analysis. Random Forest (RF) classification algorithm was better than the other tested algorithms and achieved 89.42% correct classifications (baseline ZeroR classification 63.4%, sensitivity 96.80%, specificity 76.60%, kappa 0.7632, Area under Receiver Operation Characteristic curve (AUC ROC) 0.911, Precision Recall curve (PRC) Area 0.916, and Matthews Correlation Coefficient (MCC) 0.771. Random Forest (RF) algorithm correctly identified majority of patients with final high-grade histopathology of conisation from patients dataset based on last PAP smear result and risk factors of developing high-grade dysplasia and carcinoma. Such algorithms can help clinicians to identify high-risk patients in future. An invitation could be sent to patients who did not participate in organized screening program, thus preventing the serious disease. Further studies are required in this regard.

Keywords

Uterine cervical dysplasia; Uterine cervical cancer; Conisation; Artificial intelligence

1. Introduction

Cervical cancer is preventable. Measures include cervical cancer screening programmes, treatment of early lesions, and immunisation against Human Papilloma Virus (HPV) [1, 2]. In 2020, cervical cancer was the fourth most frequently diagnosed cancer with estimated 604,000 new cases and fourth leading death cause in women, claiming 342,000 lives worldwide [3].

In Slovenia, 104 new cases of cervical cancer were diagnosed (Crude Incidence Rate = 10/100,000, Age Standardised Incidence Rate = 9), and 1056 cases of High Grade Squamous Intraepithelial Lesion (HSIL) in 2019. In this period, 220,301 PAP smears from 206,323 women were analysed [4, 5].

Conisation, and Large Loop Excision of Transformation Zone (LLETZ) are surgical procedures. They are preferred as the first treatment of dysplastic changes on uterine cervix [6]. In Slovenia, 2017 conisation procedures were performed in year 2019. 283 patients (14%) had no dysplasia from final histopathology results, 400 (20%) had conisation due to low-grade squamous intraepithelial lesions (LSIL) and 1334 (66%) due to HSIL (cervical intraepithelial neoplasm (CIN)) [4]. In Slovenia, there has been a decline in number of conisations compared to LLETZ [7].

There are numerous risk factors reported in literature that contribute to the development of dysplastic changes on uterine cervix. Dysplasia can progress to cervical carcinoma. The risk factors include early coitarche (first sexual intercourse), socioeconomic and marital status, long term usage of hormonal contraception, numerous sex partners, parity, sexually transmitted diseases (HPV, Human Immunodeficiency Virus (HIV), Herpes simplex virus (HSV), Chlamydia), factors affecting long-term infections such as genetics, immunological impairment (HIV infection), and sex hormones. Other risk factors are related to HPV (genotype, numbers of viral copies), obesity, and smoking [8–19].

HPV is an important risk factor necessary for the development of cervical dysplasia and cancer [20, 21]. Nearly all women acquire HPV infection after the initiation of sexual activity. Infection can be transitory, meaning that it clears spontaneously and does not progress to dysplasia [22]. DNA

This is an open access article under the CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/).

HPV test is positive in 8.9% of the patients younger than 35 compared to 3.3% in above 35 years' age [23].

Deep machine learning—artificial neural networks (ANN)—are a part of machine learning which stems from artificial intelligence (AI). AI algorithms are utilized for the classification and regression problem solving. Algorithms are either supervised or unsupervised. A medical problem or solution is predicted or identified with better accuracy using appropriate algorithm than only with random guessing. AI algorithms assist in finding data connections that cannot be seen with naked eye. AI algorithms are employed in various medicine fields [24, 25].

In our previous research, it was evaluated if ANN can identify patients for the high-risk final histopathology of conisation based on last PAP smear result and risk factors for developing cervical dysplasia and carcinoma. Artificial Neural Networks—Multi Layer Perceptron (MLP) worked better than majority algorithm in the dataset prepared with SMOTE. Baseline ZeroR prediction was 63.4%. 77.87% classifications were correct with sensitivity 80.0%, specificity 74.2%, positive predictive value (PPV) 84.3%, negative predictive value (NPV) 74.2%, F-Measure 0.780, MCC 0.533, AUC ROC (ROC) 0.814, AUC PRC (PRC) 0.802, and kappa 0.532. However, MLP performance was not sufficient for every-day clinical practice [26]. In this study, various classification algorithms were chosen to evaluate their performance on our dataset of patients.

2. Materials and methods

Data of patients who had undergone conisation surgery in the University Clinical Centre Maribor between 1993-2005 were used. In the database information was stored regarding age at the time of surgery, age at coitarche, number of intimate partners, number of times woman was pregnant (regardless of outcome), socio-economic status, smoking behaviour, marital status, contraception type, dysmenorrhea, vaginal discharge, impaired coagulation, colposcopy result, additional cervical smears, HPV 16, 18, 31, 33 and other potential pathogens, PAP smear result before procedure, histopathology result of cervical biopsy prior to conisation, conisation reasons, vaginal therapy before conisation, conisation type, post conisation complications (if any), histopathological findings after conisation, and information regarding whether the margins were free of disease or not. The used data were anonymised. Only the data of potential risk factors for HSIL were used: age at conisation time, age at coitarche, age at first period, number of intimate partners, number of pregnancies (births and abortions; both legal and spontaneous), contraception, socioeconomic and marital status, smoking behaviour and the last PAP smear result. The final histopathological result of cone was also included. Only the patients having complete data were used for analysis.

The database contained 1475 patients with complete data. 26 patients (1.8%) were without dysplasia in the final histological result of conisation, 160 (10.8%) had LSIL and 1289 (87.4%) had HSIL. In patients without dysplastic changes, 16 patients (61.5%) had high-risk PAP smear (III, IV, V). Last PAP smear was high-risk in 127 patients (79.4%) having LSIL and 1199 (93.0%) with HSIL. The goal was to classify patients into high- or low-risk groups for the final histological result of conisation. These groups were defined as:

a. NO-HSIL: Patients having non-dysplastic changes, CIN1 and CIN1–2.

b. HSIL: Patients with CIN 2, 2–3, 3, CIS (carcinoma *in situ*) and CA (carcinoma).

There were 1289 (87.4%) patients in HSIL group and 186 (12.6%) in NO-HSIL. Patients median age at the time of procedure for HSIL group was 33 (28–40) and 37 (30–46) for NO-HSIL ($p < 0.01$). Median age at menarche was 13 (12–14) for HSIL group and 14 (12–15) for NO-HSIL ($p = 0.436$). Median age at 1st intercourse was 18 (17–18) for HSIL group and 18 (17–19) for NO-HSIL ($p < 0.035$). The number of sex partners was 2 (1–4) for HSIL group and 2 (1–3) for No-HSIL ($p < 0.004$). The number of births in HSIL and NO-HSIL groups was 0 (0–1) ($p = 0.940$). No statistical differences were found between the groups regarding spontaneous abortions ($p = 1.0$), legal abortions ($p = 0.139$), socioeconomic status ($p = 0.823$), marital status ($p = 0.725$) and smoking habits (0.163). HSIL and NO-HSIL groups were statistically different upon comparing the age at the time of procedure ($p < 0.01$), age at first intercourse ($p < 0.035$), number of sex partners ($p < 0.004$) and last PAP smear result ($p < 0.01$).

In patients' group without HSIL, 57% tested HPV 16 negative and 27% tested positive (16% were not tested). In patients' group with HSIL, 54% tested negative and 33% tested positive (14% were not tested). In NO-HSIL group, 65% tested HPV 18 negative while 21% tested positive (15% were not tested). In HSIL group, 60% tested negative and 27% tested positive (13% were not tested). HPV was not routinely tested in Slovenia during this period. Initially, the patients without HPV tests (HPV 16, 18, 31 or 33) were removed. However, the analysis of removed patients revealed that many patients with HSIL would be omitted. Chi-square test ($\chi^2 = 0.631$, $p = 0.202$) found no significant differences in HPV 16, 18 statuses and HSIL presence in our group of patients. Therefore, patients without HPV testing were retained and HPV status from the analysis was removed. Patients' statuses of HPV 16 and HPV 18 in HSIL and NO-HSIL groups are presented in Table 1.

2.1 Dealing with imbalanced data

Imbalanced data (meaning that one of the attributes represents a low proportion of the dataset) hinders accurate classification. Baseline prediction for the majority class is high and low for the minority class. In our dataset, NO-HSIL was the minority group having 12.6% patients. Patients without HSIL represent majority group in real life. 87.4% correct classification would be achieved on our dataset if majority class classifier was used.

Methods to deal with the imbalanced data include the following:

a. Under-sampling: Randomized reduction of majority class to match the minority class.

b. Over-sampling: The n-fold replication of minority class to match the majority class.

c. SMOTE: Synthetic Minority Oversampling Technique; it creates new synthetic instances having similar characteristics as the original ones of minority class [27, 28].

TABLE 1. Number of patients with HPV 16 and 18 statuses in HSIL and NO-HSIL group.

	HPV 16		HPV 18	
	NO-HSIL group Frequency	HSIL group Frequency	NO-HSIL group Frequency	HSIL group Frequency
Positive	51	419	39	342
Negative	106	693	120	775
Not performed	29	177	27	172
Total	186	1289	186	1289

HSIL: high-grade squamous intraepithelial lesion; HPV: Human Papilloma Virus.

SMOTE method was chosen to deal with the imbalanced data because of two reasons. Instances with important properties could be removed with under-sampling method. The classification system could become less effective because of the reduced number of instances available for training and evaluation. Instances are duplicated with the over-sampling method, wherein many instances with exact same properties could be found in training and evaluation sets. This might lead to the unrealistic better performance of classification system. SMOTE algorithm create new instances with similar characteristics to those already present in minority class. No important instances from majority class were lost with this algorithm, and no existing instances from minority class were multiplied.

The minority class was enlarged by adding synthetic instances through SMOTE algorithm. The new minority class had 744 patients (36.60%). Number of patients in majority class had not changed having 1289 patients (63.40%).

2.2 Experiment with Weka

Weka (1999–2022, version 3.8.6, The University of Waikato, Hamilton, New Zealand) is an open-source application employed for data mining. Besides ANN, it has features such as Classification trees, Logistic regression, K-nearest neighbours, Bayesian networks and many others [29]. The classification algorithms testing can be made using various approaches. Testing can be conducted on the whole dataset or on a part using multiple techniques. The database can be split into the training and testing set by percentage, or tested by comparing the original database against a separate database—a training database that is imported in Weka. Additionally, n-fold cross validation can also be employed. When dataset is split into training and testing parts, the majority of important instances can end up in the same part. This is likely when instances containing important characteristic represent low proportion of all cases. N-fold cross validation can minimise the chances of this scenario. N-fold cross validation splits the dataset into n parts (usually 10-fold cross validation is employed). One part is used for testing and n-1 parts as training set. All combinations of n-1 and n/n are then tested against each other. A new algorithm is created by combining the outcomes from all tests. This algorithm is then tested on whole dataset. The 10-fold cross validation was used in this study [30].

WEKA expressed the classification algorithm efficiency with:

1. TP Rate: True positive rate.
2. FP Rate: False positive rate.
3. Precision.
4. Recall.
5. F-Measure.
6. MCC: Matthews Correlation Coefficient.
7. ROC Area: Area under Receiver Operation Characteristic curve.
8. PRC Area: Area under the Precision/Recall Curve.
9. Correct: Percentage of correct classified instances.
10. Kappa: Kappa statistic.
11. ZeroR: Percentage of correctly classified instances where instances are classified as members of majority class.

2.3 Selection of algorithms

It was decided to test Multi-Layer Perceptron (MLP), Naïve Bayes algorithm, Logistic Regression, Random Tree, Tree J48, Random Forest, One Rule, and Voted Perceptron algorithm. Classification algorithms performance could be improved with ensemble methods, bagging and boosting [31]. The algorithms used were: the MLP algorithm and the algorithm with the best results. These were used in conjunction with AdaBoost and Bagging ensemble method to enhance algorithm performances and to later evaluate if performance improves. Bagging and boosting ensemble methods are also the part of Weka software.

3. Results

Baseline ZeroR prediction for all the classification algorithms was 63.40% which also represents patients' percentage in majority class. Voted perceptron among all the tested algorithms had the lowest performance with 65.81% correct classifications (sensitivity 95.10%, specificity 15.10%, PPV 66.00%, NPV 64.00%, F-measure 0.583, MCC 0.175, ROC area 0.600, PRC area 0.612, and Kappa 0.121).

Naive bayes provideed 70.24% correct classifications (sensitivity 69.50%, specificity 71.50%, PPV 80.90%, NPV 57.50%, F-measure 0.707, MCC 0.397, ROC area 0.769, PRC area 0.774 and Kappa 0.390).

Logistic algorithm produced 73.73% correct classifications (sensitivity 82.40%, specificity 58.70%, PPV 77.60%, NPV 65.80%, F-measure 0.7340, MCC 0.422, ROC area 0.799, PRC area 0.800, and Kappa 0.4208).

MLP being an artificial neural network had 77.87% correct classifications (sensitivity 80.00%, specificity 74.20%, PPV 84.30%, NPV 68.10%, F-measure 0.7800, MCC 0.5330, ROC

area 0.814, PRC area 0.802 and Kappa 0.5318). MLP with bagging ensemble method performed better having 81.80% correct classifications (sensitivity 86.20%, specificity 74.20%, PPV 85.30%, NPV 75.60%, F-measure 0.8180, MCC 0.6060, ROC area 0.8530, PRC area 0.8460 and Kappa 0.6063). Parameters for MLP model were batch size 100, number of hidden layers 1, learning rate for weight updates 0.3, momentum applied to weight updates 0.2, training time 500 epochs, and validation threshold 20.

RulesOne algorithm generated 81.36% correct classifications (sensitivity 96.90%, specificity 54.40%, PPV 78.70%, NPV 91.00%, F-measure 0.8000, MCC 0.5980, ROC area 0.7570, PRC area 0.7380 and Kappa 0.5610).

Random Tree algorithm produced 81.65% correct classifications (sensitivity 85.60%, specificity 74.90%, PPV 85.50%, NPV 75.00%, F-measure 0.8170, MCC 0.6050, ROC area 0.8020, PRC area 0.7630 and Kappa 0.6045).

Tree J48 had 82.73% correct classifications (sensitivity 89.70%, specificity 70.70%, PPV 84.10%, NPV 79.80%, F-measure 0.8250, MCC 0.6210, ROC area 0.8510, PRC area 0.8400 and Kappa 0.6188).

Random Forest gave the best results of 89.42% correct classifications (sensitivity 96.80%, specificity 76.60%, PPV 87.80%, NPV 93.30%, F-measure 0.8920, MCC 0.7710, ROC area 0.9110, PRC area 0.9160 and Kappa 0.7632). Attribute importance based on the average impurity decrease (and number of nodes using that attribute) for age was 0.45 (6632), menarche 0.39 (4569), age at 1st intercourse 0.35 (3494), number of sex partners 0.34 (3214), number of births 0.34 (2836), legal abortions 0.33 (2132), spontaneous abortions 0.29 (960), contraception type 0.27 (1105), last PAP smear result 0.26 (1037), smoking habits 0.26 (947), marital status 0.25 (775), and for socio-economic status 0.24 (536). Model parameters were as follows: Bag size percent 100, Batch size 100, Numlterations 100, and number of execution slots 1. Performance was slightly lower when using AdaBoost ensemble method as compared to the original Random Forest algorithm. Improvements were minimal with the Bagging method.

The results are presented in Table 2. Comparisons of TP Rate, FP Rate, Precision, Recall, F-Measure, MCC, ROC Area, and PRC Area regarding the selected classification algorithms are depicted in Fig. 1. Sensitivity, Specificity, PPV and NPV are given in Fig. 2. F-Measure, MCC, ROC Area and PRC Area are exhibited in Fig. 3. ROC curves of the best performing Random Forest algorithm and the worst performing Voted Perceptron algorithm are provided in Fig. 4, PRC curves of the best performing Random Forest algorithm and the worst performing Voted Perceptron algorithm are provided in Fig. 5.

4. Discussion

Cervical cancer is preventable [1]. There are numerous known factors increasing the likelihood of the development of dysplastic changes which later progress to cancer [8–19]. It is known from every-day practice that some patients having such risk factors may never get the disease. It is also possible that patients develop disease without having risk factors. The examples include smoking and lung cancer [32].

Artificial intelligence is being employed in cervical cancer programmes (screening, diagnosis, and treatment).

Cervical cytology is the vital part of screening programmes. Many studies evaluated the AI usage for analysing cytology. Pictures of PAP smears—conventional or liquid based cytology (LBC)—were digitalised and analysed by using AI. In 1993, Mango and Laurie employed computer assistance for the cervical cancer screening through artificial neural network (ANN). They used automated camera, automated microscope, and the robotic arm designed for loading and unloading slides containing PAP smears, which were stored in container. ANN achieved 96% sensitivity that was higher than the sensitivity of cytologists (81%) [33].

Sompawong *et al*. [34] used ANN on LBC PAP smears. They reached 91.7% accuracy, specificity and sensitivity and 57.8% mean average precision.

ANN utility in Rural Kenya was described by Holmström *et al*. [35]. They achieved 95.7% sensitivity, and 84.7% specificity. Human examinator had reached 78.4% specificity, 100% sensitivity, ROC area 0.94, and NPV 99–100%. It was concluded that such a model could be helpful in countries where health sector resources were scarce and lacked trained professionals [35].

Bao *et al*. [36] and Turic *et al*. [37] employed AI assisted cytology for screening cervical cancer in China. AI assisted cytology exhibited 5.8% improvement in the detection sensitivity of CIN2+ lesions compared to human reading. Specificity was slightly declined. Concordance level between human and AI assisted reading was 94.7%. Kappa of 0.92 represented nearly flawless consensus [36, 37].

Colposcopy is an important diagnostic step. Karakitsos *et al*. [38] studied whether AI usage could identify patients requiring colposcopy and colposcopy directed biopsies. They employed LBC and several biomarkers. The sensitivity (comprising of training and evaluation sensitivity) was 85.16%, specificity 98.01%, PPV 85.71%, NPV 97.92%, and overall accuracy 96.42% [38]. Similar results were obtained by Pouliakis *et al*. [39] with sensitivity 83.28%, specificity 94.26%, PPV 79.04%, and NPV 95.06%.

Besides the triage of patients for colposcopy referral, it is important to correctly evaluate colposcopic images. AI algorithms were also tested in this field. Chandran *et al*. [40] studied AI usage in colposcopic pictures analysis. AI algorithms were successful in the analysis with the sensitivity 92.4%, and specificity 96.2%. Kappa of 0.88 indicated a significant link between real and predicted changes in colposcopic image analysis [40].

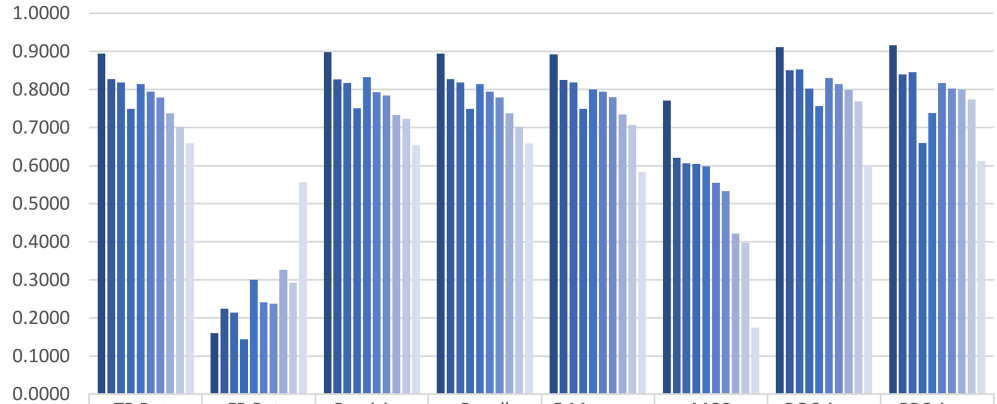
From previous studies, it is apparent that AI can be helpful in all the fields of cervical cancer screening: analysis of cytology (conventional and LBC), triage of patients for colposcopy, and colposcopic images analysis. AI can even help in countries with low resources for health care professionals. Pictures could be uploaded to cloud and analysed at remote location. Screening programmes might get compromised in the times of crisis, such as COVID-19 pandemic. AI could help in such scenarios [41].

Weegar *et al*. [42] analysed AI algorithms in predicting cervical cancer through electronic health records. They used clinical codes, lab results, and clinical events in text format.

TABLE 2. Results of tested classification algorithms.

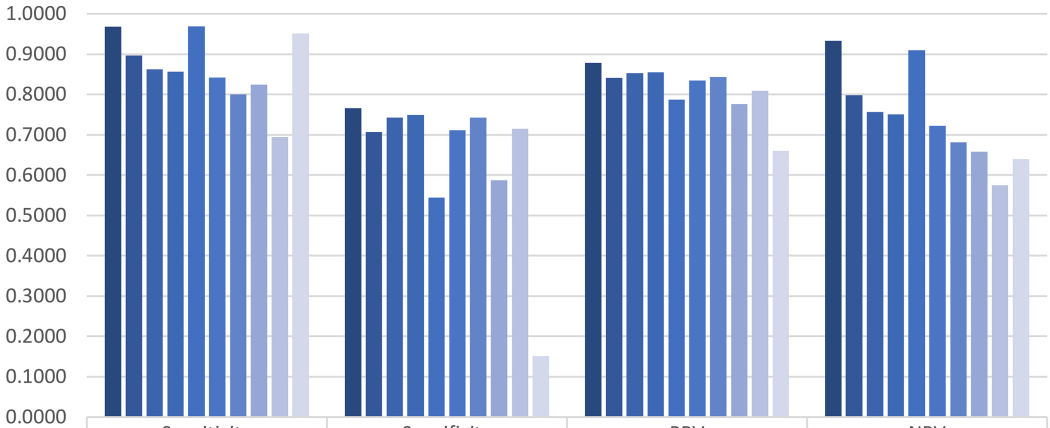
RAW SMOTE	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	Correct	Kappa	ZeroR
Random Forest Bagging	0.9730	0.2390	0.8760	0.9730	0.9220	0.7740	0.9130	0.9240	Yes	0.8952	0.7647	0.6340
	0.7610	0.0270	0.9420	0.7610	0.8420	0.7740	0.9130	0.9110	No			
	0.8950	0.1620	0.9000	0.8950	0.8920	0.7740	0.9130	0.9200	weigh_Avg			
Random Forest	0.9680	0.2340	0.8780	0.9680	0.9210	0.7710	0.9110	0.9190	Yes	0.8942	0.7632	0.6340
	0.7660	0.0320	0.9330	0.7660	0.8410	0.7710	0.9110	0.9100	No			
	0.8940	0.1600	0.8980	0.8940	0.8920	0.7710	0.9110	0.9160	weigh_Avg			
RandomForest Adaboost	0.9590	0.2350	0.8760	0.9590	0.9160	0.7560	0.8910	0.8930	Yes	0.8879	0.7497	0.6340
	0.7650	0.0410	0.9150	0.7650	0.8330	0.7560	0.8910	0.8850	No			
	0.8880	0.1640	0.8900	0.8880	0.8850	0.7560	0.8910	0.8900	weigh_Avg			
Tree J48	0.8970	0.2930	0.8410	0.8970	0.8680	0.6210	0.8510	0.8570	Yes	0.8273	0.6188	0.6340
	0.7070	0.1030	0.7980	0.7070	0.7500	0.6210	0.8510	0.8110	No			
	0.8270	0.2240	0.8260	0.8270	0.8250	0.6210	0.8510	0.8400	weigh_Avg			
MLP-Bagging	0.8620	0.2580	0.8530	0.8620	0.8570	0.6060	0.8530	0.8880	Yes	0.8180	0.6063	0.6340
	0.7420	0.1380	0.7560	0.7420	0.7490	0.6060	0.8530	0.7740	No			
	0.8180	0.2140	0.8170	0.8180	0.8180	0.6060	0.8530	0.8460	weigh_Avg			
Random Tree	0.8560	0.2510	0.8550	0.8560	0.8550	0.6050	0.8020	0.8230	Yes	0.8165	0.6045	0.6340
	0.7490	0.1440	0.7500	0.7490	0.7490	0.6050	0.8020	0.6590	No			
	0.8170	0.2120	0.8160	0.8170	0.8170	0.6050	0.8020	0.7630	weigh_Avg			
RulesOne	0.9690	0.4560	0.7870	0.9690	0.8680	0.5980	0.7570	0.7820	Yes	0.8136	0.5610	0.6340
	0.5440	0.0310	0.9100	0.5440	0.6810	0.5980	0.7570	0.6620	No			
	0.8140	0.3000	0.8320	0.8140	0.8000	0.5980	0.7570	0.7380	weigh_Avg			
MLP Adaboost	0.8420	0.2890	0.8350	0.8420	0.8380	0.5550	0.8300	0.8720	Yes	0.7939	0.5545	0.6340
	0.7110	0.1580	0.7220	0.7110	0.7160	0.5550	0.8300	0.7220	No			
	0.7940	0.2410	0.7930	0.7940	0.7940	0.5550	0.8300	0.8170	weigh_Avg			
MLP	0.8000	0.2580	0.8430	0.8000	0.8210	0.5330	0.8140	0.8670	Yes	0.7787	0.5318	0.6340
	0.7420	0.2000	0.6810	0.7420	0.7100	0.5330	0.8140	0.6910	No			
	0.7790	0.2370	0.7840	0.7790	0.7800	0.5330	0.8140	0.8020	weigh_Avg			
Logistic	0.8240	0.4130	0.7760	0.8240	0.7990	0.4220	0.7990	0.8630	Yes	0.7373	0.4208	0.6340
	0.5870	0.1760	0.6580	0.5870	0.6210	0.4220	0.7990	0.6910	No			
	0.7370	0.3260	0.7330	0.7370	0.7340	0.4220	0.7990	0.8000	weigh_Avg			
Naive Bayers	0.6950	0.2850	0.8090	0.6950	0.7480	0.3970	0.7690	0.8290	Yes	0.7024	0.3901	0.6340
	0.7150	0.3050	0.5750	0.7150	0.6380	0.3970	0.7690	0.6780	No			
	0.7020	0.2920	0.7230	0.7020	0.7070	0.3970	0.7690	0.7740	weigh_Avg			
VotedPerception	0.9510	0.8490	0.6600	0.9510	0.7790	0.1750	0.5630	0.6660	Yes	0.6581	0.1213	0.6340
	0.1510	0.0490	0.6400	0.1510	0.2440	0.1750	0.6650	0.5180	No			
	0.6580	0.5560	0.6530	0.6580	0.5830	0.1750	0.6000	0.6120	weigh_Avg			

TP Rate: true positive rate; FP Rate: false positive rate; MCC: Matthews Correlation Coefficient; ROC Area: area under the Receiver Operator Characteristic curve; PRC Area: area under Precision Recall curve; Class: YES—classification for class HSIL, No—classification for class NO-HSIL, weigh_Avg—weighted average for classes Yes and No; Correct: correctly classified instances (sum of true positives and true negatives); Kappa: kappa statistic; ZeroR: percentage of correctly classified instances in case where all instances are classified as members of the majority class HSIL; SMOTE: Synthetic Minority Oversampling Technique; MLP: Multi Layer Perceptron.



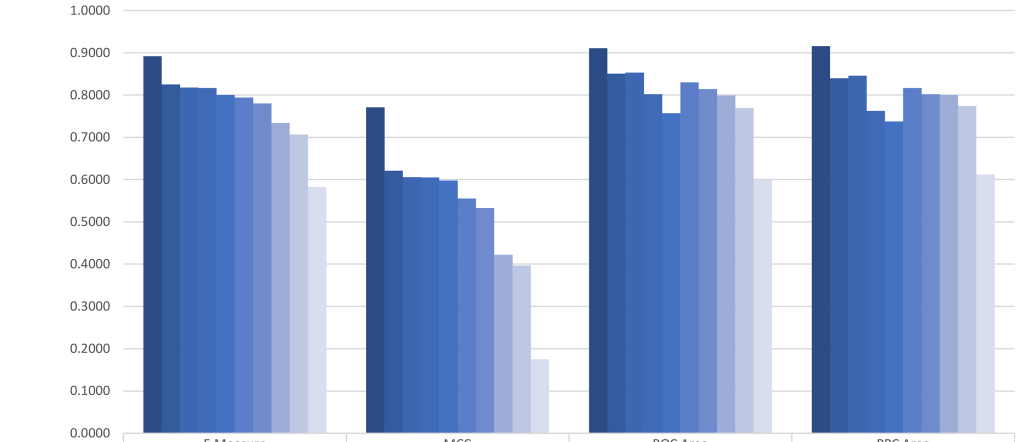
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Random Forest	0.8940	0.1600	0.8980	0.8940	0.8920	0.7710	0.9110	0.9160
Tree J48	0.8270	0.2240	0.8260	0.8270	0.8250	0.6210	0.8510	0.8400
MLP-Bagging	0.8180	0.2140	0.8170	0.8180	0.8180	0.6060	0.8530	0.8460
Random Tree	0.7490	0.1440	0.7500	0.7490	0.7490	0.6050	0.8020	0.6590
RulesOne	0.8140	0.3000	0.8320	0.8140	0.8000	0.5980	0.7570	0.7380
MLPAdaboost	0.7940	0.2410	0.7930	0.7940	0.7940	0.5550	0.8300	0.8170
MLP	0.7790	0.2370	0.7840	0.7790	0.7800	0.5330	0.8140	0.8020
Logistic	0.7370	0.3260	0.7330	0.7370	0.7340	0.4220	0.7990	0.8000
Naive Bayers	0.7020	0.2920	0.7230	0.7020	0.7070	0.3970	0.7690	0.7740
VotedPerceptron	0.6580	0.5560	0.6530	0.6580	0.5830	0.1750	0.6000	0.6120

FIGURE 1. Comparison of TP Rate, FP Rate, Precision, Recall, F-Measure, MCC, ROC Area and PRC Area of the selected classification algorithms. TP: true positive; FP: false positive; MCC: Matthews Correlation Coefficient; ROC Area: area under the Receiver Operator Characteristic curve; PRC Area: area under Precision Recall curve; MLP: Multi Layer Perceptron.



	Sensitivity	Specificity	PPV	NPV
Random Forest	0.9680	0.7660	0.8780	0.9330
Tree J48	0.8970	0.7070	0.8410	0.7980
MLP-Bagging	0.8620	0.7420	0.8530	0.7560
Random Tree	0.8560	0.7490	0.8550	0.7500
RulesOne	0.9690	0.5440	0.7870	0.9100
MLPAdaboost	0.8420	0.7110	0.8350	0.7220
MLP	0.8000	0.7420	0.8430	0.6810
Logistic	0.8240	0.5870	0.7760	0.6580
Naive Bayers	0.6950	0.7150	0.8090	0.5750
VotedPerceptron	0.9510	0.1510	0.6600	0.6400

FIGURE 2. Sensitivity, specificity, PPV and NPV of selected algorithms. PPV: positive predictive value; NPV: negative predictive value; MLP: Multi Layer Perceptron.



	F-Measure	MCC	ROC Area	PRC Area
Random Forest	0.8920	0.7710	0.9110	0.9160
Tree J48	0.8250	0.6210	0.8510	0.8400
MLP-Bagging	0.8180	0.6060	0.8530	0.8460
Random Tree	0.8170	0.6050	0.8020	0.7630
RulesOne	0.8000	0.5980	0.7570	0.7380
MLPAdaboost	0.7940	0.5550	0.8300	0.8170
MLP	0.7800	0.5330	0.8140	0.8020
Logistic	0.7340	0.4220	0.7990	0.8000
Naive Bayers	0.7070	0.3970	0.7690	0.7740
VotedPerceptron	0.5830	0.1750	0.6000	0.6120

FIGURE 3. Comparison of F-Measure, MCC, ROC Area and PRC Area of selected classification algorithms. MCC: Matthews Correlation Coefficient; ROC Area: area under the Receiver Operator Characteristic curve; PRC Area: area under Precision Recall curve; MLP: Multi Layer Perceptron.

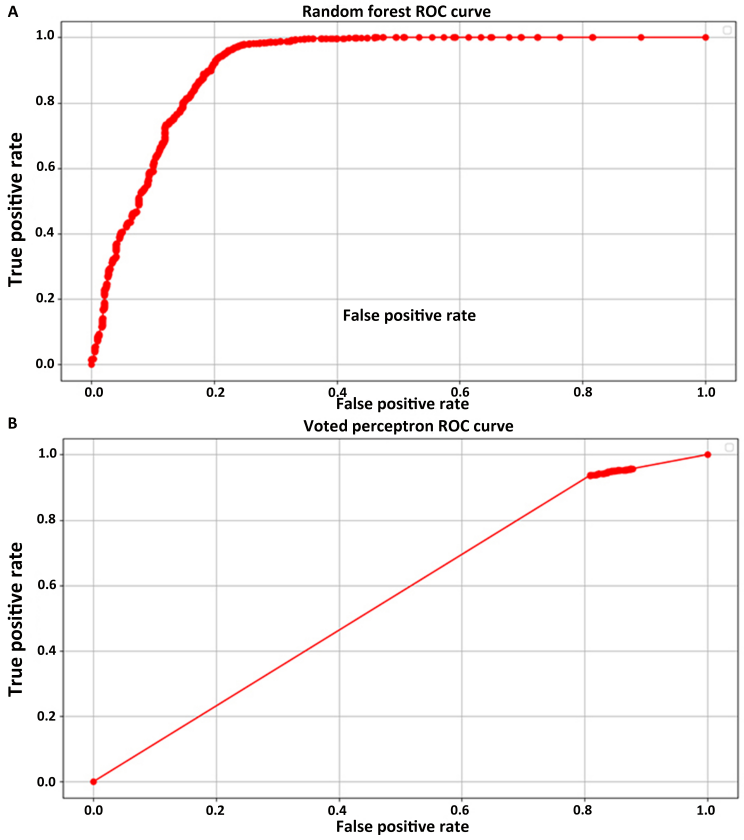


FIGURE 4. ROC area. A: The best performing Random Forest algorithm with ROC area = 0.911 (ROC area = 0.5 means the performance of random guessing, and ROC area = 1 means the ideal performance). B: The worst performing Voted Perceptron algorithm with ROC area = 0.600 (ROC area = 0.5 means the performance of random guessing, and ROC area = 1 means the ideal performance). ROC: Receiver Operator Characteristic.

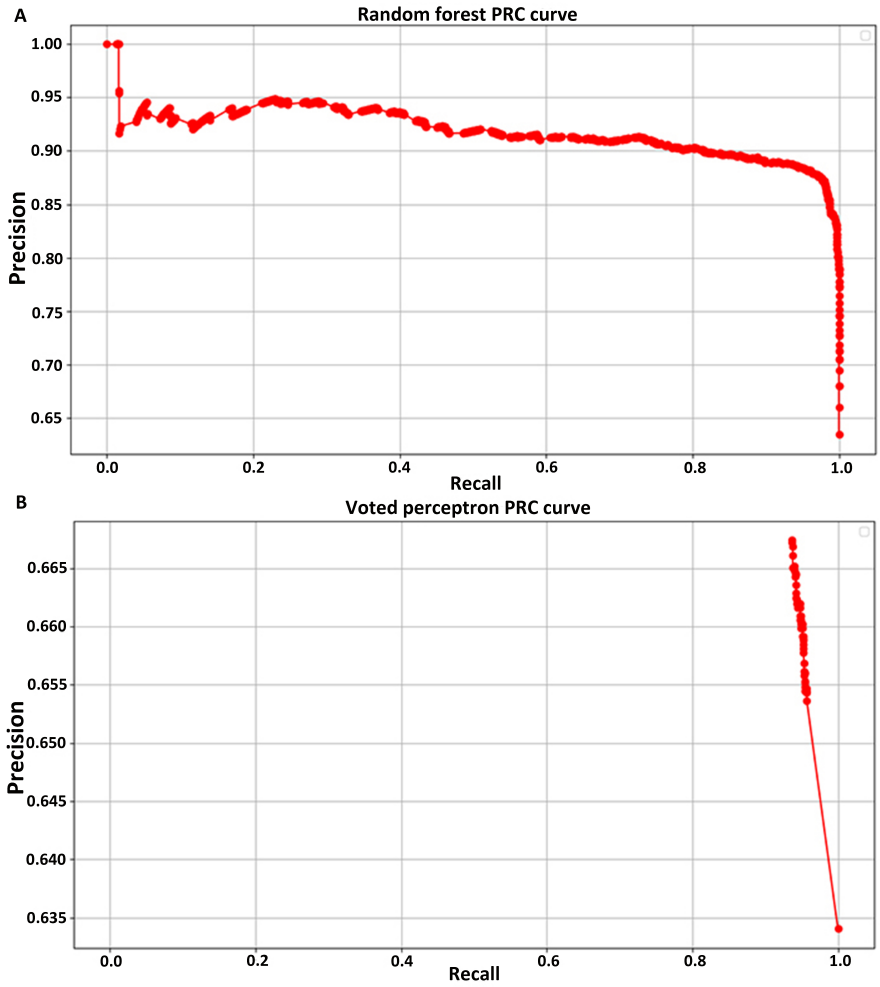


FIGURE 5. PRC area. A: The best performing Random Forest algorithm with PRC area = 0.9160. B: The worst performing Voted Perceptron algorithm with PRC area = 0.6120. PRC: Precision Recall.

The performances of four classifiers were studied. RF performed the best with AUC 0.7, one year before the diagnosis and 0.97 up to one day before diagnosis [42].

Majority of studies used AI algorithms in conjunction with image databases to predict risk of cervical cancer development. Conversely, our study tackled the predicting risk of HSIL and cervical cancer development by using a database containing patients' risk factors. From literature, it was found that a study published by Al-Wesabi *et al*. [43] used similar settings as the ones in our study. They analysed a cervical cancer dataset. Age, age at first sexual intercourse, number of pregnancies, smoking status, hormonal contraceptives, and sexually transmitted diseases (STDs) (genital herpes) were the main features for predicting cervical cancer development with high accuracy (97.5%). Decision tree with SMOTE had 91.77% accuracy, 91.46% sensitivity, 92.10% specificity, and precision 92.59% [43].

In our previous work, it was proved that with the use of ANN, based only on the factors that contributed to the development of dysplastic changes and cancer in uterine cervix, we identified more high-risk patients than with random guessing [26]. This study was an extension of our previous work and based on same dataset. The goal was to compare different AI algorithms and evaluate, which of the chosen AI algorithms has better performance on our database.

RF was the best choice for this task among the tested algorithms. Because of the imbalanced data, SMOTE method was employed to match the minority and majority class. In real life, patients without disease represent majority which was opposite from the situation in our database.

In Slovenia, efforts were made to reduce cervical cancer incidence [44]. "Classic methods" are unlikely to be sufficient in eliminating the cervical cancer. Patients who do not attend or drop out from regular screening present one of the major problems. Clinicians could recognise high-risk patients with such algorithms and took more active approaches, thus preventing patients from serious illness.

A large anonymous database with patients classified as "ill" or "healthy" would contribute to the improvement in AI algorithms regarding sensitivity, specificity, PPV, NPV, and kappa in identifying high-risk patients. However, as research ethics committees have limited experience with the review of the newly trending big data research in the field of healthcare, caution is advised when dealing with the ethics implications of such databases [45].

5. Conclusions

AI algorithms are a powerful tool in medicine, uterine cervical pathology, and cervical cancer screening. This work proves that AI can be a useful tool in this field of medicine. With AI assistance, we can identify patients at the risk of developing HSIL or cervical cancer based on last PAP smear result and the risk-factors for developing cervical dysplasia and cancer. Methods for equalizing imbalanced data are required before starting the classification. Identified patients could receive special care which may prevent them from acquiring the disease. This would be especially important for the patients that dropped out from cervical cancer screening programme.

Further studies are however needed. It is important to standardize the settings (risk factors of developing dysplastic cervical changes and cervical cancer, methods of balancing minority and majority classes, and the ratio between majority and minority classes). A general agreement on these settings can make all future studies comparable and compatible.

ABBREVIATIONS

AUC: Area under the Curve; AI: Artificial intelligence; ANN: artificial neural networks; CIN: Cervical intraepithelial neoplasm; HIV: human immunodeficiency virus; HPV: human papilloma virus; HSIL: high-grade squamous intraepithelial lesion; LBC: Liquid based cytology; LLETZ: Large Loop Excision of Transformation Zone; LSIL: low-grade squamous intraepithelial lesion; MCC: Matthews Correlations Coefficient; MLP: multi-layer perceptron; NPV: negative predictive value; PPV: Positive predictive value; PRC: precision/recall curve; RF: Random Forest; ROC: Receiver Operator (characteristic) curve; SMOTE: Synthetic Minority Oversampling Method.

AVAILABILITY OF DATA AND MATERIALS

The datasets generated and analyzed during the current study are not publicly available due to the sensitive nature of the information it contains.

AUTHOR CONTRIBUTIONS

IT—composed the dataset and created data collection methodology. IT, MMar and ARF—designed the study. MMar—performed analysis of the data with Weka. MK—performed statistical analysis of the data. MMar and MMaš—wrote the article. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Our study has been approved by Medical Ethics Committee of the Republic of Slovenia on 11 October 2015, No.: 0120-553/2015-2 KME63/11/15. All patients gave informed consent to participate in the study.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Cooper DB, McCathran CE. Cervical dysplasia. 1st edn. StatPearls Publishing: Treasure Island (FL). 2022.

[2] World Health Organization. WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention. 2nd edn. World Health Organization: Geneva. 2021.

[3] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2021; 71: 209–249.

[4] Zadnik V, Zagar T. SLORA: Slovenia and cancer. Epidemiology and cancer registry. Istitute of Oncology Ljubljana. Available at: www.slora.si (Accessed: 21 October 2022).

[5] Ivanuš U, Jerman T, Gašper Oblak U, Meglič L, Florjančič M, Strojan Fležar M, *et al*. The impact of the COVID-19 pandemic on organised cervical cancer screening: the first results of the Slovenian cervical screening programme and registry. The Lancet Regional Health—Europe. 2021; 5: 100101.

[6] Hecken JM, Rezniczek GA, Tempfer CB. Innovative diagnostic and therapeutic interventions in cervical dysplasia: a systematic review of controlled trials. Cancers. 2022; 14: 2670.

[7] Lasič A, Ivanuš U, Jerman T, Smrkolj Š, Cvjetičanin B, Lukanovič D, *et al*. Analysis of conizations in Slovenia 2009–2018: diagnosis, treatment and outcomes of cervical precancerous lesions in Slovenia. 2019. Available at: http://dirros.openscience.si/IzpisGradiva.php?lang=slv&id=11590 (Accessed: 21 October 2022).

[8] Cervical Cancer Screening Program and Registry ZORA, Institute of Oncology Ljubljana. ZORA: Slovenian national cervical cancer screening programme and registry. 2022. Available at: https://zora.onko-i.si/en (Accessed: 21 October 2022).

[9] Agarossi A, Delli Carpini G, Sopracordevole F, Serri M, Giannella L, Gardella B, *et al*. High-risk HPV positivity is a long-term risk factor for recurrence after cervical excision procedure in women living with HIV. International Journal of Gynecology & Obstetrics. 2021; 155: 442–449.

[10] Sulistyawati D, Faizah Z, Kurniawati EM. An association study of cervical cancer correlated with the age of coitarche in Dr. Soetomo hospital Surabaya. Indonesian Journal of Cancer. 2020; 14: 3–7.

[11] Nagelhout G, Ebisch RM, Van Der Hel O, Meerkerk G, Magnée T, De Bruijn T, *et al*. Is smoking an independent risk factor for developing cervical intra-epithelial neoplasia and cervical cancer? A systematic review and meta-analysis. Expert Review of Anticancer Therapy. 2021; 21: 781–794.

[12] Dahlman D, Li X, Magnusson H, Sundquist J, Sundquist K. Cervical cancer among Swedish women with drug use disorders: a nationwide epidemiological study. Gynecologic Oncology. 2021; 160: 742–747.

[13] Alimena S, Davis J, Fichorova RN, Feldman S. The vaginal microbiome: a complex milieu affecting risk of human papillomavirus persistence and cervical cancer. Current Problems in Cancer. 2022; 46: 100877.

[14] Xiao T, Ou CQ, Yang J, Wang C, Yang M, Yu T, *et al*. The risk factors for cervical cytological abnormalities among women infected with non-16/18 high-risk human papillomavirus: cross-sectional study. JMIR Public Health and Surveillance. 2022; 8: e38628.

[15] Wencel-Wawrzeńczyk A, Lewitowicz P, Lewandowska A, Saługa A. Sexual behavior and the awareness level of common risk factors for the

development of cervical, anogenital and oropharyngeal cancer among women subjected to HR HPV DNA-testing. International Journal of Environmental Research and Public Health. 2022; 19: 9580.

[16] Kim JY, Lee DW, Kim MJ, Shin JE, Shin YJ, Lee HN. Secondhand smoke exposure, diabetes, and high BMI are risk factors for uterine cervical cancer: a cross-sectional study from the Korea national health and nutrition examination survey (2010–2018). BMC Cancer. 2021; 21: 880.

[17] Yang Z, Zhang Y, Stubbe-Espejel A, Zhao Y, Liu M, Li J, *et al*. Vaginal microbiota and personal risk factors associated with HPV status conversion—a new approach to reduce the risk of cervical cancer? PLOS ONE. 2022; 17: e0270521.

[18] Gajsek US, Dovnik A, Takac I, Ivanus U, Jerman T, Zatler SS, *et al*. Diagnostic performance of p16/Ki-67 dual immunostaining at different number of positive cells in cervical smears in women referred for colposcopy. Radiology and Oncology. 2021; 55: 426–432.

[19] Saraiya M, Cheung LC, Soman A, Mix J, Kenney K, Chen X, *et al*. Risk of cervical precancer and cancer among uninsured and underserved women from 2009 to 2017. American Journal of Obstetrics and Gynecology. 2021; 224: 366.e1–366.e32.

[20] zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. Nature Reviews Cancer. 2002; 2: 342–350.

[21] Araldi RP, Sant'Ana TA, Módolo DG, de Melo TC, Spadacci-Morena DD, de Cassia Stocco R, *et al*. The human papillomavirus (HPV)-related cancer biology: an overview. Biomedicine & Pharmacotherapy. 2018; 106: 1537–1556.

[22] Bosch FX, Burchell AN, Schiffman M, Giuliano AR, de Sanjose S, Bruni L, *et al*. Epidemiology and natural history of human papillomavirus infections and type-specific implications in cervical neoplasia. Vaccine. 2008; 26: K1–K16.

[23] Zorzi M, Del Mistro A, Giorgi Rossi P, Laurino L, Battagello J, Lorio M, *et al*. Risk of CIN2 or more severe lesions after negative HPV-mRNA E6/E7 overexpression assay and after negative HPV-DNA test: concurrent cohorts with a 5-year follow-up. International Journal of Cancer. 2020; 146: 3114–3123.

[24] Kononenko I, Kukar M. Machine learning and data mining: introduction to principles and algorithms. 1st edn. Horwood Publishing: Chichester, UK. 2008.

[25] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in Medicine. 2001; 23: 89–109.

[26] Mlinaric M, Krizmaric M, Takac I, Repse Fokter A. Identification of women with high grade histopathology results after conisation by artificial neural networks. Radiology and Oncology. 2022; 56: 355–364.

[27] Mohammed R, Rawashdeh J, Abdullah M. Machine learning with oversampling and undersampling techniques: overview study and experimental results. 2020 11th International Conference on Information and Communication Systems (ICICS). Irbid, Jordan and 07–09 April 2020. IEEE. 2020.

[28] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. 2002; 16: 321–357.

[29] Witten IH, Frank E, Hall MA. Data mining: practical machine learning tools and techniques. 3rd edn. Morgan Kaufmann: Burlington, MA. 2011.

[30] Airola A, Pahikkala T, Waegeman W, De Baets B, Salakoski T. An experimental comparison of cross-validation techniques for estimating the area under the ROC curve. Computational Statistics & Data Analysis. 2011; 55: 1828–1844.

[31] Maclin R, Opitz D. An empirical evaluation of bagging and boosting. American Association for Artificial Intelligence. 1997; 546–551.

[32] Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. Clinics in Chest Medicine. 2011; 32: 605–644.

[33] Mango LJ. Computer-assisted cervical cancer screening using neural networks. Cancer Letters. 1994; 77: 155–162.

[34] Sompawong N, Mopan J, Pooprasert P, Himakhun W, Suwannarurk K, Ngamvirojcharoen J, *et al*. Automated pap smear cervical cancer screening using deep learning. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Berlin, Germany and 23–27 July 2019. IEEE. 2019.

[35] Holmström O, Linder N, Kaingu H, Mbuuko N, Mbete J, Kinyua F, *et al*. Point-of-care digital cytology with artificial intelligence for cervical cancer screening in a resource-limited setting. JAMA Network Open. 2021; 4: e211740.

[36] Bao H, Sun X, Zhang Y, Pang B, Li H, Zhou L, *et al*. The artificial intelligence-assisted cytology diagnostic system in large-scale cervical cancer screening: a population-based cohort study of 0.7 million women. Cancer Medicine. 2020; 9: 6896–6906.

[37] Turic B, Sun X, Wang J, Pang B. The role of AI in cervical cancer screening. 1st edn. IntechOpen: London. 2021.

[38] Karakitsos P, Chrelias C, Pouliakis A, Koliopoulos G, Spathis A, Kyrgiou M, *et al*. Identification of women for referral to colposcopy by neural networks: a preliminary study based on LBC and molecular biomarkers. Journal of Biomedicine and Biotechnology. 2012; 2012: 303192.

[39] Pouliakis A, Karakitsou E, Chrelias C, Pappas A, Panayiotides I, Valasoulis G, *et al*. The application of classification and regression trees for the triage of women for referral to colposcopy and the estimation of risk for cervical intraepithelial neoplasia: a study based on 1625 cases with incomplete data from molecular tests. BioMed Research International. 2015; 2015: 914740.

[40] Chandran V, Sumithra MG, Karthick A, George T, Deivakani M, Elakkiya B, *et al*. Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images. BioMed Research International. 2021; 2021: 5584004.

[41] Mendez MJG, Xue P, Quiao Y. Cervical cancer elimination in the era of COVID-19: the potential role of artificial intelligence (AI)-guided digital colposcope cloud platform. European Journal of Gynaecological Oncology. 2022; 43: 160–162.

[42] Weegar R, Sundström K. Using machine learning for predicting cervical cancer from Swedish electronic health records by mining hierarchical representations. PLOS ONE. 2020; 15: e0237911.

[43] Choudhury A, Al-Wesabi YMS, Won D. Classification of cervical cancer dataset. In: Proceedings of the 2018 IISE annual conference. Orlando. 2018; 1456–1461.

[44] World Health Organization. Turning the tide: Slovenia's success story of fighting cervical cancer. 2020. Available at: https://www.who.int/europe/news/item/17-12-2020-turning-the-tide-slovenia-s-success-story-of-fighting-cervical-cancer (Accessed: 20 October 2022).

[45] Ferretti A, Ienca M, Velarde MR, Hurst S, Vayena E. The challenges of big data for research ethics committees: a qualitative Swiss study. Journal of Empirical Research on Human Research Ethics. 2022; 17: 129–143.

How to cite this article: Marko Mlinarič, Maša Mlinarič, Miljenko Križmarić, Iztok Takač, Alenka Repše Fokter. Effectiveness of artificial intelligence algorithms in identification of patients with high-grade histopathology after conisation. European Journal of Gynaecological Oncology. 2023; 44(4): 66-75. doi: 10.22514/ejgo.2023.050.