Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 04, December 2022.

EARLY PREDICTION OF CERVICAL CANCER USING MACHINE LEARNING TECHNIQUES

Mohammad Subhi Al-Batah[1], Mazen Alzyoud[2], Raed Alazaidah[3], Malek Toubat[4], Haneen Alzoubi[2] and Areej Olaiyat[5]

(Received: 28-Aug.-2022, Revised: 16-Oct.-2022, Accepted: 30-Oct.-2022)

ABSTRACT

According to recent studies and statistics, Cervical Cancer (CC) is one of the most common causes of death worldwide and mainly in the developing countries. CC has a mortality rate of around 60%, in poor developing countries and the percentages could go even higher, due to poor screening processes, lack of sensitization and several other reasons. Therefore, this paper aims to utilize the high capabilities of machine-learning techniques in the early prediction of CC. In specific, three well-known feature selection and ranking methods have been used to identify the most significant features that help in the diagnosis process. Also, eighteen different classifiers that belong to six learning strategies have been trained and extensively evaluated against primary data consisting of five hundred images. Moreover, an investigation regarding the problem of imbalance class distribution which is common in medical datasets is conducted. The results revealed that LWNB and RandomForest classifiers showed the best performance in general and considering four different evaluation metrics. Also, LWNB and logistic classifiers were the best choices to handle the problem of imbalance class distribution which is common in medical diagnosis tasks. The final conclusion which could be made is that using an ensemble model which consists of several classifiers such as LWNB, RandomForest and logistic classifiers is the best solution to handle this type of problems.

KEYWORDS

Cervical cancer, Classification, Feature selection, Machine learning, Medical diagnosis.

1. INTRODUCTION

According to the Jordanian Ministry of Health (MoH) statistics, cancerous diseases are the second cause of death in Jordan. Globally, huge efforts from all nations have been implicated in the last century into building a strong understanding of pathophysiology, genetic changes and clinical presentation of different cancers and recruiting this knowledge in developing new methods of treatment, new screening methods and improving prognosis among cancer patients [1].

CC is a gynaecological malignancy that occurs mainly in middle-aged women, due to unregulated division of cells in cervical mucosa of females' reproductive system. Usually, females come to the clinic with chief complaints of vaginal bleeding and abnormal vaginal discharge [2].

CC almost exclusively develops in cervical cells with pre-existing human papilloma which induces dysplasia (abnormal cell growth that is premalignant) that remains latent with no symptoms for decades before developing into absolute CC [3].

Human Papilloma Virus (HPV) is a sexually transmitted infection that occurs mainly in individuals with multiple sex partners and who have not been vaccinated against carcinogenic HPVs [3]. Early detection of this dysplasia before developing into cancer is the cornerstone in fighting against CC [4].

Although CC-related incidences and deaths have dramatically decreased in developed countries- thanks to huge improvements in screening procedures [4], CC is still a huge challenge, especially to developing countries. It is the most deadly type of cancer in women in developing countries that cannot overcome the problem of lacking sufficient number of health-care professionals who are well

1. M. S. Al-Batah is with Department of Computer Science, Faculty of Science and Information Technology, Jadara University, Irbid, Jordan. Email: albatah@jadara.edu.jo

2. M. Alzyoud and H. Alzoubi are with Faculty of Information Technology, Al-al-Bayt University, Jordan. Emails: malzyoud@aabu.edu.jo and haneen123shada@gmail.com

3. R. Alazaidah is with Computer Science Department, Faculty of Information Technology, Zarqa University, Jordan. Email: razaidah@zu.edu.jo

4. M. Toubat is with Faculty of Medicine, JUST, Irbid, Jordan. Email: mhtoubat19@med.just.edu.jo

5. A. Olaiyat is with Faculty of Pharmacy, Yarmouk University, Irbid, Jordan. Email: 2017507139@sec.yu.edu.jo

trained in implementing this procedure for high-risk populations [4]. This signifies the importance of developing a computerized screening test using artificial intelligence and machine learning strategies [5].

Therefore, this paper aims to achieve the main following objectives:

1. To identify the most relevant and significant features that highly facilitate early prediction of CC.
2. To determine the best classifier that could be used to classify and predict the existence of CC among the large number of classifiers that belong to different learning strategies and use several evaluation metrics.
3. To determine the best classifier in handling the problem of imbalance class distribution which is common and familiar in the medical diagnostic field.

The main motivation for this research is to determine the best classification algorithms to use when attempting to predict CC; hence utilizing these algorithms in designing and programming a tool to automate the prediction of CC.

The rest of this paper is organized as follows: Section 2 surveys the related work dedicated to the prediction of CC. Section 3 describes the main steps of the conducted research and discusses the results obtained. Section 4 concludes the paper and lists out some future research-work horizons.

2. RELATED WORK

Classification is one of the main supervised learning tasks in machine learning. This task aims to accurately predict the class label for unseen instance [6]. In general, classification is divided into two main types: Single Label Classification (SLC) and Multi Label Classification (MLC) [7]. The former enforces each instance or example in the dataset to be linked to only one class label. Therefore, class labels in SLC are always mutually exclusive [7].

The latter allows instances in the dataset to be linked or associated with one class label or more. Hence, class labels in MLC are not mutually exclusive and have some kind of correlation among them, since they share the same values of features [8].

Moreover, SLC is divided into two sub-types: Binary Classification (BC) and Multi Class Classification (MCC). The former considers datasets with two class labels only, while the latter considers datasets with more than two class labels [9]-[10].

Classification as a machine-learning task has been utilized in several research papers related to CC. In [11], an attempt to combine the conventional diagnosis procedures and tests with machine learning to early predict abnormal cells, which highly increases the parentage of the complete cure of CC. This paper considered a large number of pap-smear test images which have been trained using deep learning techniques. The final proposed model was capable of predicting abnormal cells related to CC with accuracy of 74.04% only.

Ilyas and Ahmad (2021) [12] attempted to increase the accuracy of predicting CC by depending on an ensemble model. Therefore, eight different classifiers from different learning approaches have been utilized in predicting CC. Their study showed the significance of depending on several classifiers compared to depending on only one classifier when attempting to predict CC. This study could be improved by considering more classifiers and more learning strategies.

In [13], an ant-colony optimization algorithm has been proposed. The proposed algorithm has been trained on a dataset collected by the University of California. Support Vector Machine (SVM) has been used as the base classifier and showed a good performance (accuracy = 95.45%) compared with other algorithms which have been trained on the same dataset. The proposed algorithm has been evaluated using only one evaluation metric (accuracy). Also, the proposed algorithm should be evaluated against a larger number of algorithms.

A recent research that aimed to predict CC using MRI images has been conducted in [14]. Two main objectives have been achieved in this research. The first objective considered proposing an automatic system for early prediction of CC using image-processing techniques. The second objective aimed to enhance the performance of pre-trained Deep Convlutional Neural Networks (DCNNs) using Transfer

Learning (TL). In this paper, five classifiers were used to classify the input-image dataset into two class labels: benign or malign. Also, five evaluation metrics have been used in the evaluation phase of the five considered classifiers. Finally, according to the evaluation results, RandomForest (RF) classifier showed a better performance than the other four classifiers.

Another research that utilized machine-learning techniques in the early prediction of CC can be found in [15]. This research utilized the high capabilities of machine-learning techniques in the feature-selection step and the classification step. Unfortunately, the best evaluation result of the proposed algorithm was very low (best result for Area Under the Curve (AUC) metric was less that 0.69%) compared with other state-of- the-art algorithms.

A data-driven CC prediction model has been proposed in [16]. The proposed model not only aimed to predict CC, but also considered the problems of outliers and over-sampling. The prediction model only considered RF as a classifier. The model has been deployed through a mobile application that collects significant features related to CC and uses them in the prediction step of CC. The evaluation phase of the proposed model considered several evaluation metrics such as accuracy, precision, recall and F1-score. One of the main shortcomings of this model according to the authors themselves is the slow performance and the need to high memory during running the mobile-application software.

An ensemble model which combined the results of three different machine-learning algorithms to predict CC using Pap-smear test was proposed in [17]. The proposed model managed to predict CC using K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Multi-layer Perceptron (MLP) with a high accuracy rate (accuracy = 97.83%). The research concluded with a great potential of machine learning to highly and accurately predict CC. One of the main limitations for this research was depending on only the accuracy metric in the evaluation step while ignoring other significant evaluation metrics, such as precision, recall and F1-score.

In [18], an empirical analysis to determine the best classification algorithm among three classification algorithm has been performed. The paper considered Naïve Bayes (NB), Iterative Dichotomiser3 (ID3) and C4.5 classifiers. The analysis has been carried out using only one dataset and considering accuracy only as an evaluation metric. The paper concluded that NB outperformed the two other classifiers with accuracy being equal to (81%).

In [19], a research model that consisted of four main phases has been proposed. This research model consists of data pre-processing step, predictive model selection and pseudo-code. Also, several classifiers, such as KNN, Random Forest, SVM, Logistic Regression (LR), have been evaluated using three evaluation metrics. The research concluded the significance of using Random Forest, Decision Tree and several other classifiers in the prediction phase of CC.

3. Methodology, Results and Analysis

In this section, a comprehensive description regarding the methodology, results and analysis is presented. Firstly, in Section A, the research methodology is presented. Secondly, in Section B, the dataset is described. Thirdly, in Section C, the steps of feature selection and ranking are introduced. Finally, in Section D, the classifiers and evaluation metrics considered are introduced with the results obtained and their analysis.

A. Research Methodology

The methodology of this research is illustrated in Figure 1. As can be seen from Figure 1, the methodology consists of seven main steps. The first main step considers the collecting of data from hospitals and several specialized medical centres. Then, the segmentation process is performed as

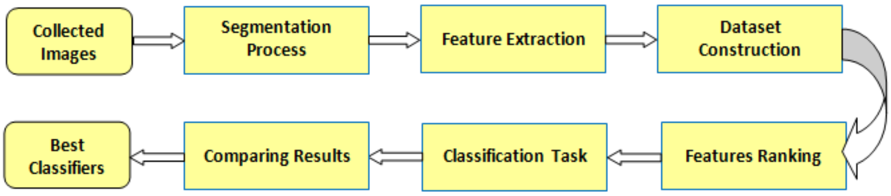

Figure 1. Research-methodology main steps.

explained in Section B. After that, several related features are extracted from the collected images as illustrated in Section C. The next step aims to construct a single-label dataset based on the data collected from the previous step. Then, three different feature-ranking techniques are applied on the dataset. The final steps aim to classify the data, obtain the results and identify the best classifiers among eighteen different classifiers based on several evaluation metrics, as extensively discussed in Section D. More information regarding these main steps can be found in the following sub-sections.

B. Dataset Description

One dataset has been considered in this research. This dataset has been constructed after performing several steps. Firstly, 500 images have been collected from different hospitals and specialized medical centres in Jordan. All images in this research have been captured using an automatic glass capturing system which has been designed specifically for this purpose. This system consists mainly of three main components: a high-resolution digital camera, a high-quality digital microscope and a personal computer. All images have been captured using 100X and 400X magnification, as recommended by both pathologists and cytologists. Each image has been labelled as Normal, Low-grade Squamous Intra-epithelial Lesion (LSIL) or High-grade Squamous Intra-epithelial Lesion (HSIL) by three domain experts and the final class of the image is determined by considering the majority. Figure 2 depicts a sample of the captured images. Figure 2.a represents a “Normal” class, Figure 2.b represents an “HSIL” class and Figure 2.c represents an “LSIL” class.

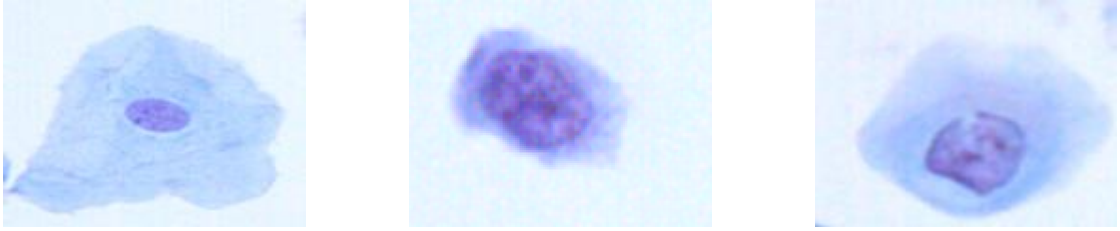

Figure 2.a Figure 2.b Figure 2.c

Figure 2. Sample of the captured images.

Secondly, a segmentation process has been applied on the collected images using Adaptive Fuzzy Moving K-means (AFMKM) clustering algorithm [20]. The main goal for applying AFMKM on the collected images is to differentiate the main three parts of the CC cell image: nucleus, cytoplasm and background. Thirdly, nine features are extracted from each CC cell image using both the nucleus and the cytoplasm parts. These features are: size, grey level, perimeter, red, green, blue, intensity1, intensity2 and saturation. Intensity1and saturation were computed using Equations (1) and (3), respectively [21]. Intensity 2 was computed using Equation (2) [22].

$$Intensity1 = \frac{1}{3}(Red + Green + Blue) \tag{1}$$

$$Intensity2 = (0.299\ Red) + (0.587\ Green) + (0.114\ Blue) \tag{2}$$

$$Saturation = \sqrt{c_1^2 + c_2^2} \tag{3}$$

where;

$$c_1 = \text{Red} – 0.5 \text{ Green} - 0.5 \text{ Blue} \tag{4}$$

$$c_2 = \sqrt{3}/_2\ Green + \sqrt{3}/_2\ Blue \tag{5}$$

Therefore, in total, the constructed dataset consists of eighteen features and five hundred instances. Each instance has been assigned to one class label only from three different class labels. These class labels are: Normal, LSIL and HSIL.

It is worth mentioning that the frequency of the three class labels: Normal, LSIL and HSIL was: 376, 79 and 45, respectively. Hence, as in most medical-diagnostic datasets, the considered dataset in this research suffers from the problem of imbalance class distribution. Therefore, this fact should be highly considered when attempting to identify the best classifier to deal with such kind of data.

Table 1 depicts a description of the features used in the considered dataset, such as data type, minimum and maximum values, average and standard deviation. It is worth mentioning that the original dataset consists of eighteen different features.

Table 1. Characteristics of the features of the CC dataset.

No.	Name	Data Type	Minimum Value	Maximum Value	Average	Slandered Deviation
1	Nucleus Area	Integer	33.00	5845.00	607.47	567.82
2	Cytoplasm Area	Integer	151.00	33222.00	13104.05	7958.00
3	Nucleus Grey Level	Real	83.14	195.14	141.87	17.59
4	Cytoplasm Grey Level	Real	130.90	220.53	192.49	15.26
5	Nucleus Perimeter	Integer	22.00	1272.00	140.40	112.44
6	Cytoplasm Perimeter	Integer	81.00	13816.00	2957.13	2482.46
7	Nucleus Red	Real	83.14	195.14	141.87	17.59
8	Cytoplasm Red	Real	130.90	220.53	192.49	15.26
9	Nucleus Green	Real	86.73	189.93	140.19	20.40
10	Cytoplasm Green	Real	101.04	233.32	212.31	17.73
11	Nucleus Blue	Real	134.55	252.36	224.29	23.03
12	Cytoplasm Blue	Real	158.72	254.99	252.52	8.45
13	Nucleus Intensity1	Real	119.54	204.19	168.78	16.93
14	Cytoplasm Intensity1	Real	148.37	234.55	219.11	11.89
15	Nucleus Intensity2	Real	102.31	192.33	150.28	17.43
16	Cytoplasm Intensity2	Real	132.82	230.40	210.97	14.58
17	Nucleus Saturation	Real	43.38	133.39	86.09	15.23
18	Cytoplasm Saturation	Real	32.39	107.81	54.68	12.83

C. Feature Selection and Ranking Step

One of the main objectives of this research is to identify the best classifier to handle the CC dataset when using all features, 75% of the features and 50% of the features. Therefore, the step of feature selection and ranking is crucial to this research.

Three different techniques have been used to rank the features. These techniques are InfoGainAttributeEval [23], ClassifierAttributeEval [23] and GainRatioAttributeEval [23]. All these techniques have been trained on the considered dataset using WEKA [23]. WEKA is short for Waikato Environment for Knowledge Analysis. WEKA is an open-source software that is used widely in data analysis in the domains of data mining and machine learning.

Regarding InfoGainAttributeEval, this technique evaluates the worth of an attribute by measuring the information gain with respect to the class. The ClassifierAttributeEval technique evaluates the worth of an attribute by using a user-specified classifier. Finally, the GainRatioAttributeEval technique depends on the gain ratio to evaluate the worth of an attribute with respect to the considered class. More information regarding these attribute evaluators and other feature-ranking techniques can be found in [23].

Table 2 depicts the ranking of the features after applying the three previously mentioned ranking techniques on the considered dataset.

Table 2. Feature-selection evaluation step using three attribute evaluators.

No.	Attribute	Ranking Using InfoGainAttributeEval	Ranking Using ClassifierAttributeEval	Ranking Using GainRatioAttributeEval
1	Cytoplasm Area	1	8	4
2	Cytoplasm Green	2	11	3
3	Cytoplasm Perimeter	3	4	7

4	Cytoplasm Intensity2	4	13	2
5	Cytoplasm Intensity1	5	15	1
6	Cytoplasm Blue	6	17	10
7	Cytoplasm Saturation	7	1	11
8	Cytoplasm Grey	8	5	5
9	Cytoplasm Red	9	9	6
10	Nucleus Perimeter	10	2	8
11	Nucleus Area	11	18	9
12	Nucleus Saturation	12	6	12
13	Nucleus Red	13	3	13
14	Nucleus Grey Level	14	7	14
15	Nucleus Intensity1	15	16	16
16	Nucleus Green	16	10	15
17	Nucleus Intensity2	17	14	17
18	Nucleus Blue	18	12	18

Table 3 depicts the features of the dataset after ranking. Features have been ranked using the summation of the ranks of the three considered ranking techniques. The feature with the least sum is ranked first and the feature with the highest sum is ranked last.

Table 3. Attributes' ranking using three attribute evaluators.

Order	Attribute	Ranking Using InfoGainAttributeEval	Ranking Using ClassifierAttributeEval	Ranking Using GainRatioAttributeEval	Sum
1	Cytoplasm Area	1	8	4	13
2	Cytoplasm Perimeter	3	4	7	14
3	Cytoplasm Green	2	11	3	16
4	Cytoplasm Grey Level	8	5	5	18
5	Cytoplasm Intensity2	4	13	2	19
6	Cytoplasm Saturation	7	1	11	19
7	Nucleus Perimeter	10	2	8	20
8	Cytoplasm Intensity1	5	15	1	21
9	Cytoplasm Red	9	9	6	24
10	Nucleus Red	13	3	13	29
11	Nucleus Saturation	12	6	12	30
12	Cytoplasm Blue	6	17	10	33
13	Nucleus Grey Level	14	7	14	35
14	Nucleus Area	11	18	9	38
15	Nucleus Green	16	10	15	41
16	Nucleus Intensity1	15	16	16	47
17	Nucleus Intensity2	17	14	17	48
18	Nucleus Blue	18	12	18	48

Based on Table 3, the classifiers considered in this research are trained on three versions of CC dataset. The first version consists of all features (18 features). The second version consists of the best ranked 75% of the features (12 features). The third version consists of the best ranked 50% of the features (9 features). The considered classifiers are evaluated based on their performance on the three versions and using several evaluation metrics.

D. Evaluation of the Considered Classifiers

The main objective of this research is to early predict CC using machine-learning techniques as accurately as possible. Therefore, many classifiers should be considered to identify the best one. Hence, eighteen different classifiers have been considered and extensively evaluated. These eighteen classifiers belong to six well-known learning strategies.

From Bayes learning strategy, the following three classifiers have been considered: BayesNet [21], NaiveBayes [24] and NaiveBayesUpdateable [24]. The function-learning strategy has been represented through three classifiers: Logistic [25], SMO [26] and SimpleLogistic [27]. The Lazy learning strategy has been represented also using three classifiers: Instance-based Learning (IBL) [28], KStar [29] and Locally Weighted Naive Bayes (LWNB) [30].

For Meta learning strategy, the following classifiers have been considered: AdaBoostM1 [31], LogitBoost [32] and MultiClassClassifier [23]. Also, three different classifiers have been used to represent the Rule-based learning strategy. These classifiers are: DecisionTable [32], JRip [34] and PART [35]. Finally, the Tree learning strategy has been represented by RandomTree [23], RandomForest [36] and J48 [37] classifiers.

The previously mentioned classifiers have been evaluated using four different evaluation metrics: Accuracy, Precision, Recall and F1-Measure (F1-Score), using the following equations.

$$Accuracy = (TP + TN) / (P+N) \tag{6}$$

$$Precision = TP / (TP + FP) \tag{7}$$

$$Recall = TP / (TP + FN) \tag{8}$$

$$F1\text{-}Measure = 2 * (precision * recall) / (precision + recall) \tag{9}$$

where:

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative. P and N are total positive and negative classes.

Table 4 depicts the evaluation results of the eighteen classifiers grouped by learning strategy and using the Accuracy metric. The evaluation considers all features, 75% of the features and 50% of the features, respectively.

Table 4. Evaluation results using the Accuracy metric.

Learning Strategy	Classifier	All Features	75 % of Features	50 % of Features
Bayes	BayesNet	82.000	82.600	82.200
	NaiveBayes	81.200	81.600	82.000
	NaiveBayesUpdateable	81.200	81.600	82.000
	Average	81.467	81.933	82.067
Functions	Logistic	**91.400**	86.000	86.600
	SMO	84.800	84.000	84.200
	SimpleLogistic	87.800	85.600	84.200
	Average	88.000	85.200	85.000
Lazy	IBL	86.200	85.800	86.200
	KStar	88.600	89.000	86.200
	LWNB	85.400	85.400	85.400
	Average	86.730	86.733	85.933
Meta	AdaBoostM1	84.800	84.800	84.800
	LogitBoost	89.000	88.400	87.000
	MultiClassClassifier	**91.400**	87.200	87.200
	Average	**88.400**	86.733	86.333
Rules	DecisionTable	85.200	85.200	85.200
	JRip	87.800	86.600	82.800
	PART	88.000	86.000	86.200
	Average	87.000	85.933	84.733
Trees	RandomTree	84.000	85.800	87.000
	RandomForest	**91.400**	**89.800**	**88.400**
	J48	88.000	87.200	87.400
	Average	87.800	**87.600**	**87.600**

According to Table 4, RandomForest showed the best results considering all features, 12 features and 9 features. Logistic and MultiClassClassifier showed an identical result to RandomForest when considering all features. Moreover, Tree as a learning strategy showed the best result with 12 and 9 features, while Meta learning strategy showed the best performance when considering all features.

It is worth mentioning that NaiveBayes and NaiveBayesUpdateable showed an identical performance on the three datasets (all features' dataset, 75% of the features' dataset and 50% of the features' dataset).

Table 5 depicts the evaluation results of the eighteen classifiers grouped by learning strategy and using the Precision metric. The evaluation considers using all features, 75% of the features and 50% of the features. From Table 5, it can be clearly seen that LWNB classifier showed the best performance considering the Precision metric on the three considered cases (all features, 75% of the features, 50% of the features).

Considering learning strategies, Meta as a learning strategy showed the best performance on the three considered cases. Also, Lazy learning strategy showed an identical result to Meta learning strategy when considering 75% of the features. It is worth mentioning that NaiveBayes and NaiveBayesUpdateable showed an identical performance on the three datasets (all features' dataset, 75% of the features' dataset and 50% of the features' dataset).

Table 5. Evaluation results using the Precision metric.

Learning Strategy	Classifier	All Features	75 % of Features	50 % of Features
Bayes	BayesNet	0.838	0.838	0.842
	NaiveBayes	0.828	0.839	0.841
	NaiveBayesUpdateable	0.828	0.839	0.841
	Average	0.831	0.839	0.841
Functions	Logistic	0.914	0.860	0.866
	SMO	0.837	0.952	0.949
	SimpleLogistic	0.878	0.859	0.842
	Average	0.876	0.890	0.886
Lazy	IBL	0.861	0.860	0.865
	KStar	0.883	0.886	0.859
	LWNB	**0.978**	**0.978**	**0.978**
	Average	0.907	**0.908**	0.901
Meta	AdaBoostM1	0.967	0.967	0.967
	LogitBoost	0.889	0.881	0.869
	MultiClassClassifier	0.912	0.873	0.872
	Average	**0.923**	**0.908**	**0.903**
Rules	DecisionTable	0.847	0.847	0.840
	JRip	0.883	0.869	0.835
	PART	0.878	0.857	0.871
	Average	0.869	0.858	0.849
Trees	RandomTree	0.839	0.852	0.878
	RandomForest	0.907	0.888	0.885
	J48	0.884	0.880	0.879
	Average	0.877	0.873	0.881

Table 6 depicts the evaluation results of the eighteen classifiers grouped by learning strategy and using the Recall metric. The evaluation considers using all features, 75% of the features and 50% of the features.

Table 6. Evaluation results using the Recall metric.

Learning Strategy	Classifier	All Features	75 % of Features	50 % of Features
Bayes	BayesNet	0.820	0.826	0.822
	NaiveBayes	0.812	0.816	0.820
	NaiveBayesUpdateable	0.812	0.816	0.820
	Average	0.815	0.819	0.821
Functions	Logistic	**0.914**	0.860	0.866
	SMO	0.848	0.840	0.842
	SimpleLogistic	0.878	0.856	0.842
	Average	0.880	0.852	0.850
Lazy	IBL	0.862	0.858	0.862

	KStar	0.886	**0.890**	0.862
	LWNB	0.854	0.854	0.854
	Average	0.867	0.867	0.859
Meta	AdaBoostM1	0.848	0.848	0.848
	LogitBoost	0.890	0.884	0.870
	MultiClassClassifier	**0.914**	0.872	0.872
	Average	**0.884**	0.867	0.863
Rules	DecisionTable	0.852	0.852	0.852
	JRip	0.878	0.866	0.828
	PART	0.880	0.860	0.862
	Average	0.870	0.859	0.847
Trees	RandomTree	0.840	0.858	0.870
	RandomForest	**0.914**	0.888	**0.884**
	J48	0.880	0.872	0.874
	Average	0.878	**0.873**	**0.876**

From Table 6, RandomForest showed the best performance on all features' dataset and 50% features' dataset. KStar showed the best result on the dataset with 75% of the features. Also, Logistic and MultiClassClassifier showed the best results on all features' dataset along with RandomForest Classifier.

Regarding to the best learning strategy, as can be seen from Table 6, Trees showed the best performance on the dataset with 75% of the features and the dataset with 50% of the features, while Meta learning strategy showed the best performance on the dataset with all features.

It is worth mentioning that NaiveBayes and NaiveBayesUpdateable showed an identical performance on the three datasets (all features' dataset, 75% of the features' dataset and 50% of the features' dataset).

Table 7 depicts the evaluation results of the eighteen classifiers grouped by learning strategy and using the F1-Measure (F1-Score) metric. The evaluation considers using all features, 75% of the features and 50% of the features. According to Table 7, LWNB classifier has a superior constant performance compared with the other seventeen classifiers. LWNB achieved the best results on all features' dataset, 75% of the features' dataset and 50% of the features' dataset.

Considering the learning strategy, Meta as a learning strategy showed the best performance on the dataset with all features, the dataset with 75% of the features and the dataset with 50% of the features. Also, Lazy learning strategy showed the best performance on the dataset with 50% of the features.

It is worth mentioning that NaiveBayes and NaiveBayesUpdateable showed an identical performance on the three datasets (all features' dataset, 75% of the features' dataset and 50% of the features' dataset).

Table 7. Evaluation results using the F1-Measure metric

Learning Strategy	Classifier	All Features	75 % of Features	50 % of Features
Bayes	BayesNet	0.823	0.822	0.821
	NaiveBayes	0.818	0.823	0.828
	NaiveBayesUpdateable	0.818	0.823	0.828
	Average	0.820	0.823	0.826
Functions	Logistic	0.914	0.860	0.866
	SMO	0.824	0.947	0.945
	SimpleLogistic	0.878	0.857	0.841
	Average	0.872	0.888	0.884
Lazy	IBL	0.862	0.859	0.864
	KStar	0.885	0.888	0.860
	LWNB	**0.962**	**0.962**	**0.962**
	Average	0.903	**0.903**	0.895

Meta	AdaBoostM1	0.958	0.958	0.958
	LogitBoost	0.888	0.881	0.869
	MultiClassClassifier	0.913	0.870	0.871
	Average	**0.920**	**0.903**	**0.899**
Rules	DecisionTable	0.843	0.844	0.839
	JRip	0.879	0.867	0.831
	PART	0.879	0.858	0.865
	Average	0.867	0.856	0.845
Trees	RandomTree	0.840	0.855	0.874
	RandomForest	0.904	0.888	0.884
	J48	0.882	0.875	0.876
	Average	0.875	0.873	0.878

Table 8 summarizes the results obtained from Table 4 to Table 7 by identifying the best classifier with respect to the considered metric and the number of features being used.

Table 8. Summarization of the best classifier with respect to evaluation metric and number of the considered features.

Metric	All Features	75 % of Features	50 % of Features
Accuracy	Logistic MultiClassClassifier RandomForest	RandomForest	RandomForest
Precision	LWNB	LWNB	LWNB
Recall	Logistic MultiClassClassifier RandomForest	KStar	RandomForest
F1-Measure	LWNB	LWNB	LWNB

According to Table 8, LWNB classifier is the best classifier among all considered classifiers. LWNB classifier achieved the best performance six times. RandomForest classifier is the second best classifier, since it achieved the best performance five times. LWNB classifier is the optimal choice when there is a need to optimize Precision and F1-Measure metrics. RandomForest classifier is the best choice when there is a need to optimize Accuracy and Recall metrics. Moreover, Logistic and MultiClassClassifier showed an excellent performance when considering all features with Accuracy and Recall metrics.

Table 9 depicts the best learning strategy with respect to the considered evaluation metric and the number of features being used. Table 9 summarizes the results from Table 4 to Table 7.

Table 9. Summarization of the best learning strategy with respect to evaluation metric and number of the considered features.

Metric	All Features	75 % of Features	50 % of Features
Accuracy	Meta	Trees	Trees
Precision	Meta	Meta Lazy	Meta
Recall	Meta	Trees	Trees
F-Measure	Meta	Meta Lazy	Meta

From Table 9, It is obvious that Meta as a learning strategy is the dominant strategy. Meta showed the best performance considering the four evaluation metrics. Trees learning strategy is the second best learning strategy and Lazy learning strategy is the third best strategy according to Table 9.

In general, medical datasets like the dataset considered in this research usually suffer from the problem of imbalance class distribution. For example, in the CC dataset, the dominant class is the “Normal” class with a frequency equal to 376. For “LSIL” class, the frequency is 79, while the frequency of

“HSIL” class is 45, as mentioned previously. One of the main characteristics of the optimal classifier is the ability to handle the problem of imbalance class distribution.

Therefore, it has been decided to evaluate the eighteen classifiers considered in this research based on how accurate they can predict the least frequent, but most significant, classes (LSIL and HSIL). True Positive (TP) metric has been used to accomplish this task. TP metric calculates the percentage at which the classifier correctly predicts the positive classes.

Table 10 depicts the evaluation results of the eighteen considered classifiers using the TP metric and grouped by the learning strategy. It is worth mentioning that for the TP metric, the higher the value, the better the performance of the classifier.

Table 10. Evaluation results with respect to the TP metric for “HSIL” and “LSIL” classes using all features.

Learning Strategy	Classifier	HSIL	LSIL
Bayes	BayesNet	0.200	0.747
	NaiveBayes	0.600	0.405
	NaiveBayesUpdateable	0.600	0.405
	Average	0.467	0.519
Functions	Logistic	**0.711**	0.734
	SMO	0.022	0.861
	SimpleLogistic	0.667	0.620
	Average	0.467	0.738
Lazy	IBL	0.600	0.557
	KStar	0.578	0.633
	LWNB	0.000	**0.899**
	Average	0.393	0.696
Meta	AdaBoostM1	0.000	0.848
	LogitBoost	0.422	0.747
	MultiClassClassifier	0.689	0.722
	Average	0.370	**0.772**
Rules	DecisionTable	0.156	0.772
	JRip	0.489	0.734
	PART	0.578	0.658
	Average	0.408	0.721
Trees	RandomTree	0.422	0.532
	RandomForest	0.489	0.810
	J48	0.578	0.658
	Average	**0.496**	0.667

According to Table 10, Logistic classifier is the best classifier to predict the class label “HSIL” with a TP rate equal to 0.711, while LWNB is the best classifier to predict the class label “LSIL” with a TP rate equal to 0.899.

Considering the learning strategy, Trees is the most suitable learning strategy to predict the class label “HSIL”, while Meta is the most appropriate learning strategy to predict the class labels “LSIL”.

Since no classifier can be the dominant classifier for dealing with the problem of imbalance class distribution, it is highly recommended to adopt an ensemble model to overcome this serious problem. Based on the results of this research, it is recommended to include LWNB, RandomForest and Logistic in any future proposed ensemble models.

4. Conclusion and Future Work

In this paper, a dataset consisting of 500 images related to CC has been collected from different hospital and specialized medical centers. Also, eighteen different classifiers which belong to six learning strategies have been trained on the collected dataset and evaluated. The evaluation of the classifiers considered four evaluation metrics with respect to all features in the dataset, 75% of the features and 50% of the features. The results revealed that LWNB classifier has achieved the best performance in general. RandomForest showed the second best performance. Also, considering the learning strategy, Meta learning strategy showed the best overall performance compared with the other five strategies. Moreover, Logistic and LWNB classifiers are the best choice to deal with the problem of imbalance class distribution, which is very common in medical diagnostic datasets. Based on the results of this research, the main recommendation for future work is to adopt an ensemble model that consists of LWNB, RandomForest and Logistic classifiers to achieve high performance in the early prediction of CC.

REFERENCES

[1] M. A. Abu-Lubad, A. J. Dua'a, G. F. Helaly et al., "Human Papillomavirus as an Independent Risk Factor of Invasive Cervical and Endometrial Carcinomas in Jordan," Journal of Infection and Public Health, vol. 13, no. 4, pp. 613-618, 2022.

[2] B. Obeidat, I. Matalka, A. Mohtaseb et al., "Prevalence and Distribution of High-risk Human Papillomavirus Genotypes in Cervical Carcinoma, Low-grade and High-grade Squamous Intraepithelial Lesions in Jordanian Women," European Journal of Gynaecological Oncology, vol. 34, no. 3, pp. 257-260, 2013.

[3] S. E. Jordan, M. Schlumbrecht, S. George et al., "The Moore Criteria: Applicability in a Diverse, Non-trial, Recurrent Cervical Cancer Population," Gynecologic Oncology, vol. 157, no. 1, pp. 167-172, 2022.

[4] M. Al Qadire, K. M. Aldiabat, E. Alsrayheen et al., "Public Attitudes toward Cancer and Cancer Patients: A Jordanian National Online Survey," Middle East Journal of Cancer, vol. 13, DOI: 10.30476/mejc.2020.86835.1381, 2020.

[5] A. I. Khasawneh, F. F. Asali, R. M. Kilani et al., "Prevalence and Genotype Distribution of Human Papillomavirus among a Sub-population of Jordanian Women," International Journal of Women's Health and Reproduction Sciences, vol. 9, no. 1, pp. 17-23, 2021.

[6] R. Alazaidah, M. A. Almaiah and M. Al-luwaici, "Associative Classification in Multi-label Classification: An Investigative Study," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 7, no. 2, pp. 166 - 179, 2021.

[7] M. Al-luwaici, A. K., Junoh, W. A. AlZoubi., R. Alazaidah and W. Al-luwaici, "New Features Selection Method for Multi-label Classification Based on the Positive Dependencies among Labels," Solid State Technology, vol. 63, no. 2s, pp. 9896-9909, 2020.

[8] R. Alazaidah, F. A. Ahmad, M. F. M. Mohsin and W. A. AlZoubi, "Multi-label Ranking Method Based on Positive Class Correlations," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 6, no. 4, pp. 377-391, 2020.

[9] M. Alluwaici, A. K. Junoh and R. Alazaidah, "New Problem Transformation Method Based on the Local Positive Pairwise Dependencies among Labels," Journal of Information & Knowledge Management, vol. 19, no. 1, ID. 2040017, 2020.

[10] R. Alazaidah, F. K. Ahmad and M. F. M. Mohsin, "Multi Label Ranking Based on Positive Pairwise Correlations among Labels," The International Arab Journal of Information Technology, vol. 17, no. 4, pp. 440-449, 2020.

[11] B. J. Priyanka, "Machine Learning Approach for Prediction of Cervical Cancer," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 8, pp. 3050-3058, 2021.

[12] Q. M. Ilyas and M. Ahmad, "An Enhanced Ensemble Diagnosis of Cervical Aancer: A Pursuit of Machine Intelligence towards Sustainable Health," IEEE Access, vol. 9, pp. 12374-12388, 2021.

[13] J. Wahid and H. F. A. Al-Mazini, "Classification of Cervical Cancer Using Ant-miner for Medical Expertise Knowledge Management," Proc. of the Knowledge Management Int. Conf. (KMICe), Miri Sarawak, Malaysia, 25 –27 July 2018.

[14] I. Khoulqi and N. Idrissi, "Cervical Cancer Detection and Classification Using MRIs," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 8, no. 2, pp. 141-158, 2022.

[15] K. Fernandes, D. Chicco, J. S. Cardoso and J. Fernandes, "Supervised Deep Learning Embeddings for the Prediction of Cervical Cancer Diagnosis," PeerJ Computer Science, vol. 4, e154, DOI: 10.7717/peerj-cs.154, 2018.

[16] M. F. Ijaz, M. Attique and Y. Son, "Data-driven Cervical Cancer Prediction Model with Outlier Detection and Over-sampling Methods," Sensors, vol. 20, no. 10, ID. 2809, 2020.

[17] V. Mishra, S. Aslan and M. M. Asem, "Theoretical Assessment of Cervical Cancer Using Machine Learning Methods Based on Pap-Smear Test," Proc. of the 9th IEEE Annual Information Technology, Electronics and Mobile Communication Conf. (IEMCON), pp. 1367-1373, Vancouver, Canada, 2018.

[18] R. Vidya and G. M. Nasira, "Predicting Cervical Cancer Using Machine Learning Techniques - An Analysis," Glob. J. Pure Appl. Math, vol. 12, no. 3, 2016.

[19] N. Al Mudawi and A. Alazeb, "A Model for Predicting Cervical Cancer Using Machine Learning Algorithms," Sensors, vol. 22, no. 11, ID. 4132, 2022.

[20] N. A. M. Isa, S. A. Salamah and U. K. Ngah, "Adaptive Fuzzy Moving K-means Clustering Algorithm for Image Segmentation," IEEE Trans. on Consumer Electronics, vol. 55, no. 4, pp. 2145-2153, 2009.

[21] C. Zhang and P. Wang, "A New Method of Color Image Segmentation Based on Intensity and Hue Clustering," Proc. of the 15th IEEE Int. Conf. on Pattern Recognition (ICPR-2000), vol. 3, pp. 613-616, Barcelona, Spain, 2000.

[22] N. Mustafa, N. A. M. Isa, M. Y. Mashor and N. H. Othman, "Capability of New Features of Cervical Cells for Cervical Cancer Diagnostic System Using Hierarchical Neural Network," IJSSST, vol. 9, no. 2, pp. 56-64, 2008.

[23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA Data Mining Software: An Update," ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10-18, 2009.

[24] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifier," Proc. of the 11th Conf. on Uncertainty in Artificial Intelligence (UAI1995), pp. 338-345, San Mateo, 1995.

[25] S. Le Cessie and J. C. Van Houwelingen, "Ridge Estimators in Logistic Regression," Journal of the Royal Statistical Society: Series C (Applied Statistics), vol. 41, no. 1, pp. 191-201, 1992.

[26] J. Platt, "Using Analytic QP and Sparseness to Speed Training of Support Vector Machines," Advances in Neural Information Processing Systems, vol. 11, 1998.

[27] N. Landwehr, M. Hall and E. Frank, "Logistic Model Trees," Machine Learning, vol. 59, no. 1, pp. 161-205, 2005.

[28] D. W. Aha, D. Kibler and M. K. Albert, "Instance-based Learning Algorithms," Machine Learning, vol. 6, no. 1, pp. 37-66, 1991.

[29] J. G. Cleary and L. E. Trigg, "K*: An Instance-based Learner Using an Entropic Distance Measure," Proc. of the 12th Int. Conf. on Machine Learning, pp. 108-114, Tahoe City, California, July 9–12, 1995.

[30] E. Frank, M. Hall and B. Pfahringer, "Locally Weighted Naive Bayes," Proc. of the 19th Conf. on Uncertainty in Artificial Intelligence, pp. 249-256, arXiv:1212.2487, 2003.

[31] Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm," Proc. of the 13th Int. Conf. on Int. Conf. on Machine Learning (ICML'96), vol. 96, pp. 148-156, 1996.

[32] J. Friedman, T. Hastie and R. Tibshirani, "Additive Logistic Regression: A Statistical View of Boosting," The Annals of Statistics, vol. 28, no. 2, pp. 337-407, Stanford University, 1998.

[33] R. Kohavi, "The Power of Decision Tables," Proc. of the European Conf. on Machine Learning (ECML), pp. 174-189, Springer, Berlin, Heidelberg, 1995.

[34] W. W. Cohen, "Fast Effective Rule Induction," Proc. of the 12th Int. Conf. on Machine Learning, pp. 115-123, Tahoe City, California, 1995.

[35] E. Frank and I. H. Witten, "Generating Accurate Rule Sets without Global Optimization," Proc. of the 15th Int. Conf. on Machine Learning (ICML '98), pp. 144–151, 1998.

[36] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[37] J. R. Quinlan, C4. 5: Program for Machine Learning, Morgan Kaufmann Publishers, Inc., 1993.

ملخص البحث:

تهدف هذه الورقة الى استغلال الإمكانات العالية لتقنيات تعلّم الآلة من أجل الكشْف المبكّر عن الإصابة بسرطان عُنُق الرَّحم. حيث يتم استخدام ثلاث طرق لاختيار السِّمات وترتيبها لتحديد السِّمات الأكثر أهمية التي تُساعد في عملية التّشخيص. كذلك تمّ استخدام ثمانية عشر مُصَنِّفاً تتبع لِسِتّ استراتيجيات تعلّم بحيث تمّ تدريبها وتقييمها مقابل بيانات أولية تتكون من 500 صورة. من جهةٍ أخرى، جرى استقصاء مشكلة عدم التّوازن في توزيع الأصناف. وبينت النتائج أنّ مُصنِّف LWNB ومُصنِّف RandomForest حقّقا أفضل أداء بشكلٍ عام وباعتماد أربعة مقاييس للتّقييم كان مُصنِّف LWNB ومُصنِّف Logistic هما الأفضل من حيث معالجة مشكلة عدم التوازن في توزيع أصناف البيانات. ويمكن القول إنّ الاستنتاج النهائي الذي يُمكن الخروج به في هذا البحث هو أنّ استخدام نموذج مُجمَّع يتألف من عدّة مصنِّفات (مثل مُصنِّفات LWNB و RandomForest و Logistic) هو الحلّ الأمثل للتّعامل مع المشكلات المرتبطة بموضوع البحث، وهو الكشْف المبكّر عن الإصابة بسرطان عُنُق الرَّحِم.