A Two-Stage Method of Cervical Cancer Detection using Hybrid Classification Methods

Orasa Patsadu
Faculty of Science and Technology
Rajamangala University of Technology
Krungthep, Thailand
orasa.p@mail.rmutk.ac.th

Srisuda Soranunsri
Faculty of Science and Technology
Rajamangala University of Technology
Krungthep, Thailand
srisuda.s@mail.rmutk.ac.th

Chananate Arthayukti
Faculty of Science and Technology
Rajamangala University of Technology
Krungthep, Thailand
chananate@mail.rmutk.ac.th

Abstract: *Cervical cancer is a primary cause of death in women. Human Papilloma Virus (HPV) virus is the main factor that causes cervical cancer. Therefore, risk reduction of cervical cancer is cervical cancer screening and vaccination against infection of HPV. The methods of cervical cancer screening are medical history, HPV virus detection, Pap smear detection, and small pieces of meat cutting. However, Pap smear detection is a popular technique for cervical cancer screening. In addition, the classification method is used to increase the performance of cervical cancer detection. Therefore, our research proposed method of cervical cancer detection using hybrid classification methods with risk factor data of cervical cancer and cervical cancer behavior risk data, which is divided into two stages: primary risk estimation of a chance cervical cancer and cervical cancer prediction. Firstly, the dataset was prepared to handle missing values and imbalanced data. Secondly, we select suitable attributes for model building. Finally, the dataset was used to build the model using hybrid classification methods by bagging with Random Forest. From our experiment, the model for primary risk estimation of a chance cervical cancer and cervical cancer prediction has an accuracy of 99.25% and 98.00%, respectively. The result shows that our proposed method accurately predicts cervical cancer compared to state-of-the-art methods. Therefore, it confirms that our proposed model can be applied to use for primary risk self-estimation of a chance of cervical cancer before deciding to meet a physician. Moreover, it helps the physician with diagnosis and treatment to reduce costs and the death rate.*

Keywords: *Cervical cancer, hybrid classification methods, two stage classification, machine learning, healthcare.*

Received July 23, 2024; accepted February 6, 2025
https://doi.org/10.34028/iajit/22/3/3

1. Introduction

Cervical cancer is considered to be the cause of death in women, inferior to breast cancer, and frequently found in ages 30-60 years [22]. The risk factors of a chance of cervical cancer are failure to undergo screening, syphilis, gonorrhea, taking birth control pills continuously for a long time, giving birth more than 4 times, and smoking. The leading cause of a chance cervical cancer is the HPV virus, which is often transmitted from sexual diseases. Therefore, to monitor disease, women aged 30 years and over should undergo screening to detect cervical cancer every 1 year and should be checked every 3 years in case of no risk [39]. However, the screening method should be performance, affordable, available, and easy to continuously follow because continuous screening increases the opportunity to defend and control the metastatic stage of the disease, including receiving prompt treatment. Therefore, if it has a high-performance system, which helps primary risk estimation that it has risk or no risk from demographic information, the user can acknowledge and immediately protect preliminary risk. Moreover, if the system can accurately detect cervical cancer, it helps undergo screening and diagnosis to reduce and control cervical cancer in the metastatic stage and to protect against illness, including decreasing death rate and medical costs and increasing good health and well-being. Our hypothesis is to propose a high-performance system for primary risk estimation of a chance of cervical cancer and cervical cancer prediction.

The cervical cancer diagnosis technique consists of several techniques, such as Pap smear, iodine test, biopsy, and colposcopy. In Thailand, several researchers also use cervical cancer diagnosis techniques through HPV testing, liquid-based cytology, and questionnaires [19, 21]. However, medical often use Pap smears to undergo screening for cervical cancer [5]. Therefore, our research uses cycology, indicating Pap smear to build a model for cervical cancer prediction [17]. Several researchers use classification methods to solve medical problems [24, 31, 36]. Furthermore, numerous researchers proposed methods for cervical cancer detection using classification methods [15, 20, 23, 26, 27, 28, 29, 33, 40, 45]. To increase the detection precision and reduce the detection error of individual classifiers, several researchers proposed hybrid classification methods for cervical cancer detection [1, 2, 6, 7, 13, 18, 30, 35, 37, 46]. Most of the previous research focuses on model building to predict cervical cancer. However, it still lacks primary risk self-assessment using demographic information to support decision-making to undergo screening since the first phase before the

metastatic stage to reduce the risk of a chance of cervical cancer and curability. Therefore, if the user can acknowledge the first phase risk before having symptoms, it will help reduce disease severity and death. In addition, it is beneficial for users to assess primary risks before deciding to meet a physician because of several causes, such as no time, embarrassed officers, and fear. Moreover, our proposed model benefits physicians in supporting diagnosis and further treatment. It became our motivation.

While the method for cervical cancer prediction is the main domain of research studies in this field, detection inaccuracy and the lack of primary risk estimation of the chance of cervical cancer are still shortcomings. Addressing the challenge, the main contributions are as follows:

1. We studied the pattern classification method for primary risk estimation of a chance of cervical cancer from demographic data and healthcare for self-assessment before deciding to meet the physician.
2. To enhance the accuracy of cervical cancer prediction by selecting appropriate attributes and using hybrid classification methods.

The remainder of this paper is organized as follows: Section 2 is related to work. Section 3 is our proposed method for cervical cancer detection. Section 4 shows results and discussions. The last section summarizes the main conclusion and future work.

2. Related Work

Several researchers propose methods for cervical cancer detection using classification methods. Mehmood *et al.* [27] proposed a classification method to detect cervical cancer. The process of cervical cancer detection is divided into 3 processes. The first process is data preparation using pearson correlation. The second process is feature selection using random forest. The last process is model building using random forest and shallow neural network. This model can detect cervical cancer with an accuracy of 93.6%.

Rahaman *et al.* [33] presented cervical cancer detection using a Hybrid Deep Feature Fusion (HDFF) and Late Fusion (LF) methods based on the SIPaKMeD Pap Smear dataset and Herlev datasets. The results show that HDFF can detect cervical cancer with higher accuracy than LF.

Lilhore *et al.* [23] proposed cervical cancer detection using the classification method. The first step is correlation analysis and Boruta analysis to select features. The second step is cervical cancer detection using Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and Boruta. From the experiment, Boruta analysis can detect cervical cancer with high accuracy when compared to other methods.

Mahmoud *et al.* [25] presented a method to detect cervical cancer using a quantum grasshopper optimization technique based on images of the cervical cells (Pap smear test). The quantum-based grasshopper computing algorithm with the fuzzy c-means algorithm and Convolutional Neural Network (CNN) are used to detect cervical cancer. The proposed method can detect cervical cancer with 98% accuracy.

Kruczkowski *et al.* [20] presented cervical cancer detection using the classification method. "RF, XGBoost, Naïve Bayes (NB), and CNN" are used to build the model. The proposed model detects cervical cancer with an accuracy of 95%. This model can support doctors to diagnose for further treatment.

Tanimu *et al.* [40] proposed a method to detect cervical cancer using Decision Tree. This paper uses a public dataset from the University of California, Irvine. The dataset is handled to imbalance class using a combination technique of oversampling data using SMTOE and undersampling data using Tomek's Link (SMOTETomek). Then, recursive feature elimination, least absolute shrinkage, and selection operator extract suitable attributes for model building. The result of detection is an accuracy of 98.72% and a sensitivity of 100% with recursive feature elimination and SMOTETomek methods. Also, Zhang *et al.* [45], Mukku and Thomas [28], Kalbhor and Shinde [15], Nandanwar and Dhonde [29], Masot *et al.* [26] use data mining techniques to detect cervical cancer.

To increase the performance of cervical cancer detection using an individual classifier, numerous researchers proposed methods for cervical cancer detection using hybrid classification methods. Shinde *et al.* [37] proposed a method to detect cervical cancer using Pap smear cytology images based on classification. There are 2 methods to classify cervical cancer. The first method uses principal component analysis and hybrid machine learning techniques (SVM, RF, and Neural Network (NN)) based on maximum votes. The second method uses feature fusion vectors and an Artificial Neural Network (ANN) based on a Convolutional Neural Network (CNN). From the experiment, the second method is high performance.

Chandran *et al.* [6] presented a method for cervical cancer detection using Deep Learning (DL) technique. There are 2 techniques: CNN (VGG19) and Colposcopy Ensemble Network (CYENET). Colposcopy images are used to prepare and build models. The result shows that CYENET can detect cervical cancer with 92.3% accuracy compared to CNN (VGG19).

Nikookar *et al.* [30] proposed cervical cancer detection using hybrid classification methods based on colposcopic images. The system can detect cervical cancer with sensitivity (96%) and specificity (94%).

Zorkafli *et al.* [46] proposed cervical cancer detection using a hybrid multi-layered perceptron and genetic algorithm. Cells are used to extract features for model building. The result shows that this model detects cervical cancer with high accuracy.

Ilyas and Ahmad [13] presented cervical cancer

detection using Ensemble methods. "DT, SVM, RF, K-Nearest Neighbor (KNN), naïve bayes, multilayer perceptron, J48 trees, and Logistic Regression" are used to build a model for cervical cancer detection with an accuracy of 94%.

Khanam and Mondal [18] proposed a model to detect cervical cancer using hybrid classification methods. Firstly, the dataset is prepared to adjust the imbalance class using Borderline-SMOTE. Then, the model is built using several classifiers. Hybrid classification methods (Random Forest, Support Vector Machine, ExtraTreeClassifier, XGBoost, and Bagging) achieve an accuracy of 94.4%. Also, Bhavani and Govardhan [2], Alam *et al.* [1], Razali *et al*. [35] proposed a method to detect cervical cancer using the same techniques as Khanam and Mondal [18].

Curia [7] proposed a method to detect cervical cancer using hybrid classification methods based on a dataset on the UCI repository. This model can detect cervical cancer with an accuracy of 94.5%.

From previous research studies, most research focuses on methods for cervical cancer prediction. However, preliminary risk estimation of a chance of cervical cancer using demographic data seems to appear in this field rarely. Nevertheless, demographic data is essential data for preliminary risk estimation of a chance cervical cancer to preliminary screening before deciding to meet a physician for diagnosis and treatment because it can reduce time and cost. We examined based on the advantages of demographic data and hybrid classification methods. Therefore, we proposed a new method for preliminary risk estimation of the chance of cervical cancer.

3. Methods

This paper consists of two stages of cervical cancer detection (stage 1: primary risk estimation of a chance cervical cancer and stage 2: cervical cancer prediction), as shown in Figure 1.

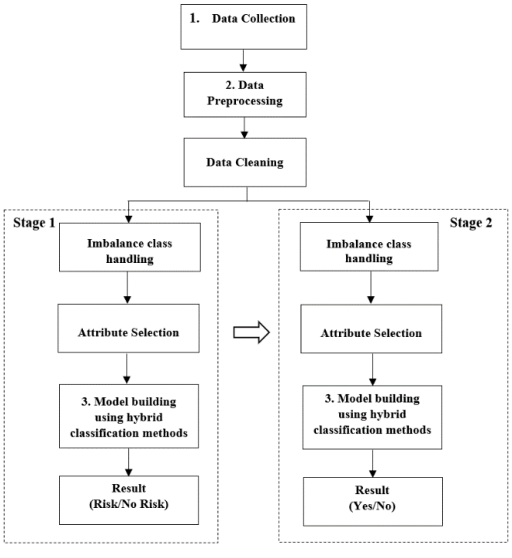

Figure 1. Stage of cervical cancer detection.

From Figure 1, each stage of cervical cancer detection consists of 3 parts such as data collection, data preprocessing, and model building. Each part is described in the next section.

3.1. Data Collection

In this paper, we use 2 datasets from the UC Irvine machine learning repository, which has several researchers using these datasets [7, 13]. The first dataset was collected from 'Hospital Universitario de Caracas' in Caracas, Venezuela of Kelwin *et al.* [16]. The dataset consists of demographic information, habits, and historical medical. There are 858 rows (32 attributes and 4 classes: Hinselmann, Schiller, Cytology, and Biopsy) [41]. For Hinselmann, it is a method for cervical cancer detection using a microscope to detect Schiller cells, which test the reaction between iodine and glycogen of the vaginal mucosa. The Schiller test is checked in the area of colored and not-colored iodine. Cytology is a method for cervical cancer detection by Cytology of the cervix. Biopsy is Biopsy diagnosis, which is a surgery process in separate cells or tissue to detect disease or the scope of disease [9, 10, 17, 34]. However, physicians often use methods for cervical cancer detection using a pap smear, which is a diagnosis of cytology of the cervix [17] model for cervical cancer prediction. In addition, to reduce the risk of overfitting, we use the second dataset (cervical cancer behavior risk dataset) to build the model for cervical cancer prediction. There are 72 rows and 20 attributes (19 attributes and 1 class (yes and no)), which have no missing values [42].

3.2. Data Preprocessing

To obtain high performance in cervical cancer detection, we prepare data by data cleaning, imbalance class handling, and appropriated attribute selection, which are explained in the next section.

3.2.1. Data Cleaning

From the exploration the first dataset, some attributes are missing values of more than 92%, such as Sexually Transmitted Diseases (STDs): Time since first diagnosis and STDs: Time since last diagnosis. Therefore, we remove two attributes. The remainder of the attribute, which is numeric and nominal, handles missing values. Numeric handles missing values using median [12]. Nominal handles missing values using mode [12], as shown in Equations (1) and (2).

$$Median = L + \left[\frac{\frac{N}{2} - F}{f_m}\right] I \quad (1)$$

L is the real lower limit of the median
F is the cumulative frequency of the lower limit of the median
fm is the cumulative frequency of the median
N is the total number

I is the width of class interaction

$$Mode = L_0 + 1\left(\frac{d_1}{d_1 + d_2}\right) \quad (2)$$

L_0 is a lower bound.
I is the width of the mode.
d_1 is the difference between the frequency of the mode and the previous class of mode.
d_2 is the difference between the frequency of mode and post-class of mode.

After, we handle missing values. We performed outlier detection using Boxplot [12]. We found that some attributes (STDs: HPV, Number of sexual partners, Smokes, Smokes (years), Hormonal Contraceptives, Hormonal Contraceptives (years), Age, First sexual intercourse (age), Number of pregnancies, and STDs: AIDS) have outliers, which has a difference distribution from the group. Therefore, we remove 4 rows. From missing value handling and outlier detection, there is a total dataset of 854 rows (30 attributes and 4 classes) for model building.

3.2.2. Imbalance Class Handling

Once we clean the data, the next step is exploring the balance of the dataset to enhance performance and reduce the bias of answers, which is performed as follows.

3.2.2.1. Imbalance Class Handling of Data for Primary Risk Estimation of a Chance Cervical Cancer (Stage 1)

We use the first dataset, which has an imbalance class, to build a model, as seen in Table 1.

Table 1. Number of class to build model for primary risk estimation of a chance cervical cancer.

Class label	Number
Risk	103
No risk	751

From Table 1, no risk data has more than risk data at approximately 7.29%. To enhance the model's performance and reduce overfitting, we handle the imbalance class using adaboost [12]. AdaBoost focuses on adjusting the weight of the class with a smaller number of samples to handle an imbalanced class [12].

3.2.2.2. Imbalance Class Handling of Data for Cervical Cancer Prediction (Stage 2)

We use both datasets to build a model. As seen in Table 2, they still have an imbalanced class.

Table 2. Number of classes of both datasets to build a model for cervical cancer prediction.

Class label	Citology	Cancer
1 (yes)	43	21
0 (no)	811	51

From Table 2, this research uses the class label of citology from the first dataset to build a model for cervical cancer prediction because this method has been known to screen cervical cancer [6]. From the data explore, healthy data (no) has more than cervical cancer data (yes) at approximately 18.86%. For the second dataset, healthy data (no) has more than cervical cancer data (yes) at approximately 2.43%. Therefore, we use the same technique as stage 1 to handle the imbalance class of both datasets [37].

3.2.3. Appropriated Attribute Selection

When we handle an imbalance class, we select the appropriate attributes for the model building to obtain a high-performance model, which is described as follows.

3.2.3.1. Appropriated Attribute Selection for Model Building of Primary Risk Estimation of a Chance Cervical Cancer (Stage 1)

From previous research studies, factors indicate the risk of a chance of cervical cancer, as seen in Table 3.

Table 3. Factors to indicate risk of a chance of cervical cancer.

Sonkong [39]	Predawan and Sompeewong [32]	Wesabi *et al.* [43]	Sindiani *et al.* [38]	Deng *et al.* [8]
age, education, occupation, income/mounth, marital status, number of pregnancies, age of first pregnancies, history of hormonal contraception, using hormonal contraception pills for a long time, smoking, and disorder of reproductive function	age, number of sexual partners, age at 1st sexual coitus, number of parturition, hormonal contraception, IUD	age, first sexual intercourse, number of pregnancies, smokes, hormonal contraceptives, and STDs:genital herpes	age, number of sexual partners, first sexual intercourse, number of pregnancies, smoking, hormonal contraceptives per years, IUD, STD, condylamotosis, pericon dylomotosis, genital herpes, AIDS, STD-No of diagnosis	number of sexual partners, number of pregnancies, age, age of first sexual intercourse, and hormonal contraceptives

From Table 3, several demographic factors affect the risk of a chance of cervical cancer. In addition, sexually transmitted diseases, which consist of chlamydia, gonorrhea, syphilis, and Human Immunodeficiency Virus (HIV)/AIDS, also affect the risk of a chance of cervical cancer [3, 4]. From previous research studies, several factors, which include age, number of sexual partners, age of first sexual intercourse, age of first pregnancies, using hormonal contraception pills for a long time, smoking, number of pregnancies, and sexually transmitted diseases, indicate the primary risk of a chance cervical cancer. Therefore, our research selects these factors to build a model for primary risk estimation of a chance of cervical cancer, as seen in Table 4.

From Table 4, several factors affect the risk of a chance of cervical cancer. These factors are assigned class labels by considering criteria from the medical

articles, as seen in Table 5 [3, 4].

Table 4. Appropriated factors to build a model for primary risk estimation of a chance cervical cancer.

Attribute	Attribute value
STDs:HPV	Not infected, Infect
Number of sexual partners	One, more than 6
Smokes	No smoking, smoking
Smokes (years)	Less than equal to 10 years, more than 10 years
Hormonal contraceptives	No, Yes
Hormonal contraceptives (years)	Less than equal to 5years, more than 5 years
Age	Less than equal to 40, more than 40
First sexual intercourse (age)	Less than equal to 15, more than 15
Number of pregnancies	One person, more than one person
STDs:AIDS	No, Yes

From Table 5, the criteria of risk indication are if STDs: HPV is infection, number of sexual partners has more than 6 persons, smokes=yes, smokes (years) has more than 10 years, hormonal contraceptives=yes, hormonal contraceptives (years) has more than 5 years, age has more than 40 years, first sexual intercourse (age) has less than 15 years, number of pregnancies has more than 1 person, and STDs: AIDS=yes. The result is "risk". Criteria of no risk indication are if STDs: HPV is no infection, the number of sexual partners has less than equal to 6 persons, smokes=no, smokes (years) has less than equal to 10 years, hormonal contraceptives = no, hormonal contraceptives (years) have less than equal to 5 years, age has less than equal to 40 years, first sexual intercourse (age) has more than equal to 15 years, number of pregnancies has less than equal to 1 person, and STDs: AIDS=no. The result is "no risk". From data preparation, there are 10 attributes and 1 class (risk and no risk) for the model building of primary risk estimation of a chance cervical cancer.

Table 5. Criteria of determination of risk and no risk of a chance cervical cancer.

Risk estimation criteria		
Condition	**No risk**	**Risk**
STDs:HPV	No infection	Infection
Number of sexual partners	≤ 6 persons	> 6 persons
Smokes	No	Yes
Smokes (years)	≤ 10 years	> 10 years
Hormonal contraceptives	No	Yes
Hormonal contraceptives (years)	≤ 5 years	> 5years
Age	≤ 40 years	> 40 years
First sexual intercourse (age)	≥ 15 years	< 15 years
Number of pregnancies	≤ 1 person	> 1 person
STDs: AIDS	No	Yes

3.2.3.2. Appropriated Attribute Selection for Model Building of Cervical Cancer Prediction (Stage 2)

For the method of appropriate attribute selection to build a model for cervical cancer prediction, we are a preliminary experiment with 3 methods. Firstly, we select attributes according to the research of Ilyas and Ahmad [13], which uses several attributes such as Dx: Cancer, Dx: HPV, Dx, STDs: genital herpes, STDs: HIV, STDs, Dx: CIN, STDs: number of diagnoses, STDs (number), STDs: vulvo-perineal condylomatosis, STDs: condylomatosis, Hormonal Contraceptives (years), Smokes (years), IUD, age, number of pregnancies, STDs: syphilis, IUD (years), STDs: time since first diagnosis, smokes, smokes (packs/year), STDs: time since last diagnosis, hormonal contraceptives, STDs: vaginal condylomatosis, STDs: HPV, STDs: pelvic inflammatory disease, STDs: Hepatitis B, STDs: molluscum contagiosum, first sexual intercourse (age), number of sexual partners, STDs: cervical condylomatosis, and STDs: AIDS. From the experiment, the result shows that the accuracy of cervical cancer prediction is 94.00%. Secondly, the attributes, which consist of smokes (years), smokes, IUD (years), IUD, hormonal contraceptives, hormonal contraceptives (years), number of pregnancies, first sexual intercourse, age, and number of sexual partners, are selected according to the research of Khanam and Mondal [18]. This method can detect cervical cancer with an accuracy of 94.40%. Finally, to reduce overfitting and increase the model's performance, we experimented with appropriate attribute selection using Recursive Feature Elimination (RFE) [11]. RFE is attribute selection with Wrapper method to select appropriate attributes with the model performance using ranking [11]. The result of attribute ranking is shown in Table 6.

Table 6. Appropriate attributes selection using RFE with ranking.

No ranking	Attributes
1	Age
3	Number of sexual partners
4	First sexual intercourse
5	Num of pregnancies
7	Smokes
10	Smokes (years)
15	Smokes (packs/year)
2	Hormonal Contraceptives
12	Hormonal Contraceptives (years)
11	IUD
6	IUD (years)
26	STDs
8	STDs (number)
49	STDs:condylomatosis
47	STDs:cervical condylomatosis
27	STDs:vaginal condylomatosis
25	STDs:vulvo-perineal condylomatosis
16	STDs:syphilis
46	STDs:pelvic inflammatory disease
48	STDs:genital herpes
28	STDs:molluscum contagiosum
17	STDs:AIDS
9	STDs:HIV
29	STDs:Hepatitis B
19	STDs:HPV
39	STDs: Number of diagnosis
13	Dx:Cancer
44	Dx:CIN
18	Dx:HPV
14	Dx
21	behavior_sexualRisk
35	behavior_eating
40	behavior_personalHygine
32	intention_aggregation
41	intention_commitment
45	attitude_consistency
43	attitude_spontaneity
42	norm_significantPerson
24	norm_fulfillment
36	perception_vulnerability
20	perception_severity
30	motivation_strength
38	motivation_willingness
31	socialSupport_emotionality
33	socialSupport_appreciation
34	socialSupport_instrumental
37	empowerment_knowledge
22	empowerment_abilities
23	empowerment_desires

From Table 6, our proposed method can detect cervical cancer with an accuracy of 98.00%. Therefore, the competitive result of the appropriate attribute selection method for cervical cancer prediction is shown in Table 7.

Table 7. Competitive result of the appropriate attribute selection method for cervical cancer prediction.

Attributes	Result
Ilyas and Ahmad [13]	94.00%
Khanam and Mondal [18]	94.40%
Our proposed method	98.00%*

*high accuracy

From Table 7, we selected appropriate attributes to build a model for cervical cancer prediction using RFE with ranking. There are 49 attributes and 1 class of cancer (yes and no).

3.3. Model Building

Once we obtain appropriate attributes for model building, each model building is explained as follows.

3.3.1. Model Building for Primary Risk Estimation of a Chance Cervical Cancer (Stage 1)

In model building for primary risk estimation of a chance cervical cancer, we use several classification methods such as Multilayer Perceptron (MLP), KNearest Neighbors (KNN), RF, and DT [12, 44]. We set the parameter as follows.

For the MLP model, MLP [12] is one of the principles of Artificial Neural Network (ANN) for classification in multilayer. MLP consists of input, output, and hidden layer nodes. MLP is a popular method because it can classify data into several types and is suitable for complex work, as shown in Equation (3).

$$y = f(W^T x) = f(\sum_{i=1}^{N} W_i x_i + b) \quad (3)$$

f is the activation function
N is the number of neurons
W is ANN model weights
b is a bias vector

We experiment to build a performance model for primary risk estimation of a chance of cervical cancer using MLP. From the experiment, we set the parameter as seen in Figure 2.

From Figure 2, we set 10 input nodes such as STDs: HPV, Number of sexual partners, Smokes, Smokes (years), hormonal contraceptives, hormonal contraceptives (years), age, first sexual intercourse (age), Number of pregnancies, and STDs: AIDS, 6 hidden layer nodes and 2 outputs (Risk/No risk), momentum=0.5, and learning rate=0.2.

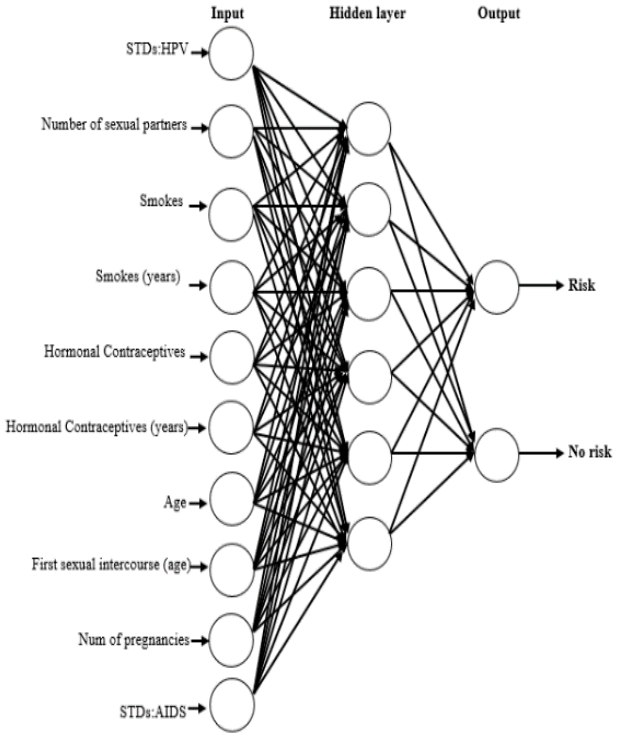

Figure 2. Model for primary risk estimation of a chance cervical cancer using MLP.

For the KNN model, KNN [12] is an easy and uncomplexity classification method using the similarity of data and the *K* value. If the answer is similar to the K value, the answer is the *K* value. The distance computation uses Euclidean distance, as shown in Equation (4).

$$dist(p,q) = \sqrt{\sum_{k=1}^{n} (p_k - q_k)^2} \quad (4)$$

We set parameters with several cases. From the experiment, the best model is a parameter setting with K=3, the NearestNeighborsearch algorithm (linearnnsearch with euclidean distance), and inputs are the same as with the MLP model.

For the RF model, RF [37] is a pervasive method for the solution of Regression and classification developed from DT. RF consists of many trees to enhance classification performance. Therefore, we use RF by hybrid many DT, as shown in Equation (5).

$$F(x)=1\ n\ P\ i\in[n]\ Ti(x). \quad (5)$$

Random Forest by an ensemble of decision trees $F=\{Ti\}i\in[n]$.

We experiment by parameter setting with a number of random selection of 10 attributes, number of iterations (100), max depth (10), and batch size (10).

Finally, we build a model for primary risk estimation of the chance of cervical cancer using DT [12]. DT is a classification method. It is suitable for continuous and discrete data. The decision tree includes the inner node, branch, and leaf node. We set input and output similar to the previous model. We use the J48 algorithm and set

batch size (100), seed (5), and confidence factor (0.5).

To increase the performance of our proposed method, we build the model using hybrid classification methods, which consist of several methods such as vote ensemble, bagging, and RF [12, 44]. Each technique has different advantages and disadvantages. We experiment to select an appropriate model for primary risk estimation of a chance of cervical cancer. The result shows that bagging using Random Forest has a high accuracy when compared to other methods. We set the parameter of bagging using Random Forest and numiterations=100, batch size (100), and seed (5) to select the best answer. Bagging [12] is a classification method with data random from a dataset, which is divided into several sets and the same classification method, as seen in Figure 3.

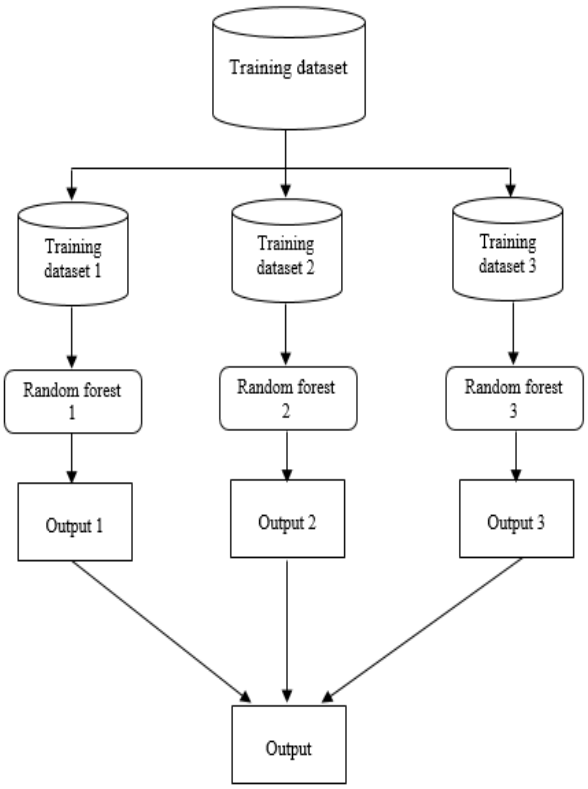

Figure 3. Primary risk estimation of a chance of cervical cancer by Bagging using Random Forest.

3.3.2. Model Building for Cervical Cancer Prediction (Stage 2)

In model building for cervical cancer prediction, we also use the same classifier as a model of stage 1. We set the parameter as follows.

The MLP model has 3 layers (49 input nodes, as explained in section 3.2.3.2, 26 hidden layer nodes, and 2 output nodes (Yes and No)). We set the learning rate at 0.3, momentum (0.2), batch size (20), and training time (500), as shown in Figure 4.

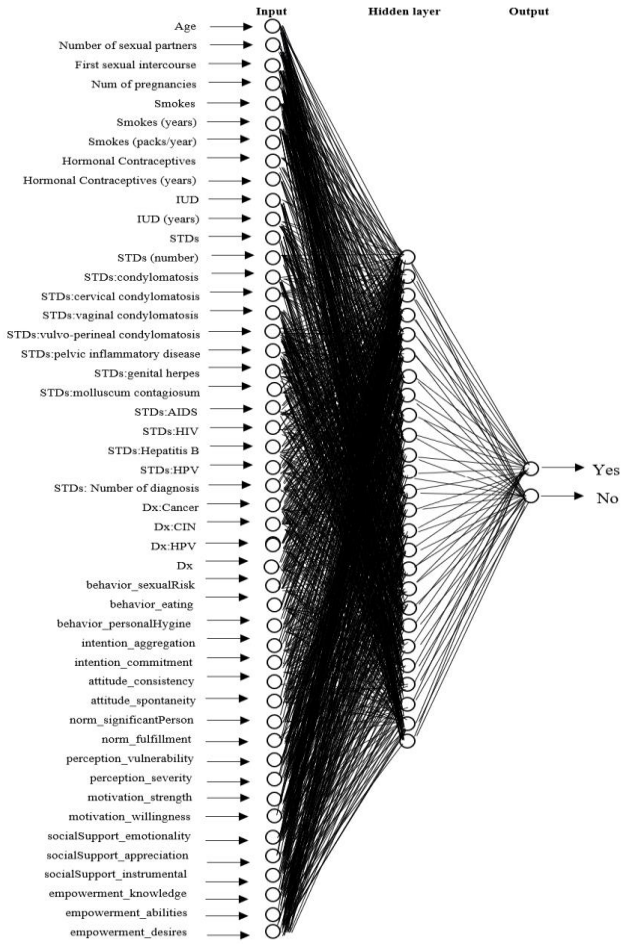

Figure 4. Model building for cervical cancer prediction using MLP.

For the KNN model, we set K=3 and the NearestNeighbor algorithm (LinearNNsearch with Euclidean distance), which uses attributes similar to the MLP model.

For the RF model, we experiment with setting an appropriate parameter by randomly selecting an attribute of 49 attributes, number of iterations (100), max depth (10), and batch size (10).

For the DT model, we set the input and output similarly to the previous model. We use the J48 algorithm and set batch size (100), seed (5), and confidence factor (0.5).

Finally, we use hybrid classification methods, which set parameters by bagging using Random Forest and numiterations =100, batch size (100), seed (5), the same as the model of stage 1.

3.3.3. Performance Measurement of Model

Performance measurement of the model uses several evaluations such as accuracy, precision, recall, and F1 score as shown in Equations (6) to (9), which compute to confusion matrix [12, 44] as shown in Table 8.

Table 8. Confusion matrix.

Actual/ Prediction	True	False
True	TP	FP
False	FN	TN

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1\ score = \frac{2XPrecisionXRecall}{Precision + Recall} \quad (9)$$

4. Results and Discussion

In this section, we describe the result of our proposed model. In addition, we demonstrate a case study to show the effectiveness of the proposed model. The dataset, which is handled in the imbalance class, as explained in section 3.2.2, is used to build the model and evaluate the effectiveness of the model with 10-fold cross-validation. The dataset was divided into the training dataset and the testing dataset. The performance evaluation of the model uses 4 measurements (accuracy, precision, recall, and F1 score), as shown in Table 9.

From Table 9, the result of primary risk estimation of a chance cervical cancer using DT achieved high accuracy with 98.80% compared to other classifiers. Next, the result of preliminary risk estimation of a chance cervical cancer using RF is an accuracy of 98.42%. Finally, the result of preliminary risk estimation of a chance cervical cancer using MLP and KNN is an accuracy of 96.85% and 95.54%, respectively.

Table 9. Result of primary risk estimation of a chance cervical cancer (stage 1) and cervical cancer prediction (stage 2) using each classifier.

	MLP		KNN		RF		DT	
	Stage 1	**Stage 2**	**Stage 1**	**Stage 2**	**Stage 1**	**Stage 2**	**Stage 1**	**Stage 2**
Accuracy	96.85%	95.00%	95.54%	93.00%	98.42%	96.00%	98.80%	97.00%
Precision	96.85%	95.00%	95.54%	92.00%	98.42%	95.00%	98.80%	96.00%
Recall	96.85%	95.00%	95.54%	93.00%	98.42%	96.00%	98.80%	97.00%
F1 score	96.85%	95.00%	95.54%	92.00%	98.42%	96.00%	98.80%	97.00%

For cervical cancer prediction, the result of cervical cancer prediction using DT also is high accuracy with 97.00% compared to other classifiers. Next, RF had an accuracy of 96.00%. MLP had an accuracy of 95.00%. KNN also had the lowest accuracy of 93.00%. From the result of the experiment of both models, DT has accuracy more than other methods, because the prediction model using DT is independent of data distribution. Therefore, DT can accurately detect cervical cancer in several cases. However, we found that some cases are errors. For example, the user is young and does not smoke, but the model predicted that the user had a risk of a chance of cervical cancer. To enhance the performance of primary risk estimation of a chance of cervical cancer and cervical cancer prediction using an individual classifier, we use hybrid classification methods by bagging using Random Forest. The result is shown in Table 10.

Table 10. Result of primary risk estimation of a chance cervical cancer (stage 1) and cervical cancer prediction (stage 2) using hybrid classification methods.

	Our proposed method		Other datasets [14]	
	Stage 1	**Stage 2**	**Stage 1**	**Stage 2**
Accuracy	99.25%	98.00%	99.20%	97.75%
Precision	99.25%	98.00%	99.20%	97.75%
Recall	99.25%	98.00%	99.20%	97.75%
F1 score	99.25%	98.00%	99.20%	97.75%

From Table 10, the result of primary risk estimation of a chance cervical cancer and cervical cancer prediction using hybrid classification methods achieved accuracies of 99.25% and 98.00%, respectively. Therefore, the result shows that hybrid classification methods outperformed individual classification. For example, the user is young. The user has many children, but the user does not have many partners and is not smoking. Our proposed model can detect accuracy. Therefore, we use our proposed model to develop a prototype system for primary risk estimation of a chance cervical cancer and cervical cancer prediction.

Moreover, we applied SHAP to interpret the feature importance in our proposed model for primary risk estimation of a chance cervical cancer and cervical cancer prediction, as shown in Figures 5 and 6.

From Figure 5, hormonal contraceptives (years) is the most influential feature for the first model, with a positive SHAP value indicating that the subject tends to have a chance of cervical cancer. Next, other features impacted the primary risk estimation of a chance of cervical cancer by the distribution of SHAP values.

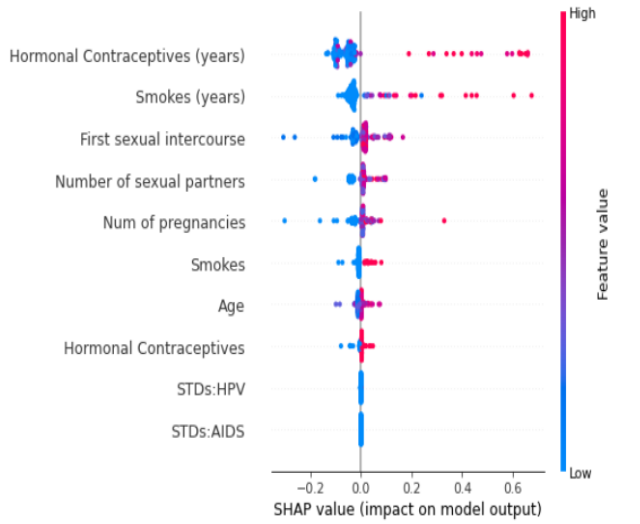

Figure 5. Relationship between features and SHAP value of primary risk estimation of a chance cervical cancer.

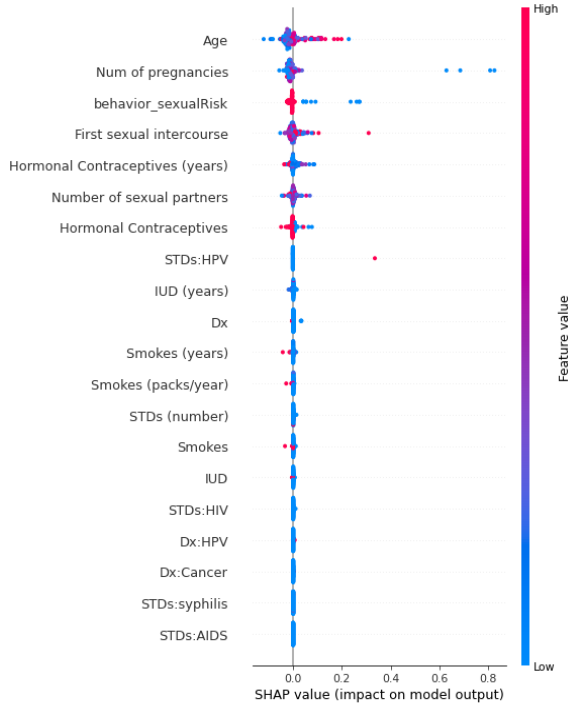

Figure 6. Relationship between features and SHAP value of cervical cancer prediction.

The second model from Figure 6 shows that age is the most influential feature, which indicates that the subject was likely to develop cervical cancer. Other features also had an impact on predicting cervical cancer, as illustrated by the distribution of SHAP values.

In addition, to ensure our proposed model generalizes well to real-world data, we use a cervical cancer dataset from Kaggle [14]. There are 835 rows and 36 attributes to test the accuracy of our proposed model. The result shows that primary risk estimation of a chance cervical cancer and cervical cancer prediction achieved accuracies of 99.20% and 97.75%, respectively. Therefore, it confirms that our proposed model is still accurate even though other data were used for testing. Our proposed model can be applied for real use.

Furthermore, we compared our proposed method with the state-of-the-art methods on the same example dataset. The result shows that our proposed method exceeds the research of Ilyas and Ahmad [13] and Khanam and Mondal [18], as seen in Table 11. Our proposed method can predict accuracy in several cases. For example, the user uses hormonal contraceptives and uses them for a long time, having many sex partners, smoking, and having aged. Therefore, it confirms that our proposed method is also accurate in cervical cancer prediction. We can apply our proposed model to practicality, which is useful for users and physicians in preliminary screening, diagnosis, and further treatment. Moreover, it conforms to our hypothesis, which proposed a high-performance system for primary risk estimation of a chance of cervical cancer and cervical cancer prediction. In addition, previous research focuses on model building for cervical cancer prediction, but it also lacks a model for primary risk estimation of a chance cervical cancer to help users with a preliminary assessment before deciding to meet a physician. However, our research has both models of primary risk estimation of a chance cervical cancer and cervical cancer prediction.

Table 11. Comparison between our proposed method and state-of-the-art cervical cancer prediction reported in the literature using the same dataset.

	Method	Accuracy	Precision	Recall	F1 score
Ilyas and Ahmad [13]	Ensemble classification method based on majority vote	94.00%	97.00%	97.00%	96.00%
Khanam and Mondal [18]	Stacking ensemble method (RF, SVM, ExtraTreeClassifier, XGBoost, and Bagging)	94.40%	94.80%	93.70%	93.00%
Our proposed method	bagging using random forest	98.00%	98.00%	98.00%	98.00%

From Table 11, we compared the result of our proposed method with the state-of-the-art methods. The result shows that our proposed method gains accuracy than other methods, which has increased accuracy, as shown in Figure 7.

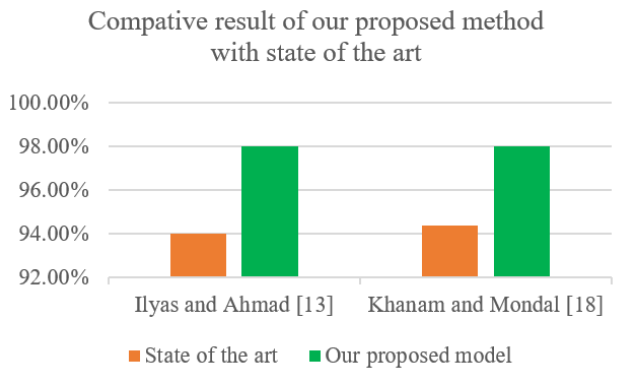

Figure 7. Competitive result of our proposed method with state-of-the-art methods.

From Figure 7, our proposed method outperformed two previous studies. Predictions of these two previous studies also had errors. Therefore, our proposed model is efficient for application in real-life situations because it is precise, reliable, affordable, and easy to use. It benefits clinical adaptability for decision support in physician diagnosis, which is an assistant tool for detecting disease and timely treatment. In addition, our proposed method uses a processing time of no more than 1 second compared to previous research. Our proposed method was developed with Python and can detect cervical cancer in real-time on an intel core i5 Central Processing Unit (CPU) @1.7 GHz and 16 GB RAM processing platform.

In addition, to explain the work method and the advantage of our proposed system, we demonstrate a work simulation with test data records randomly selected from the dataset for primary risk estimation of a chance cervical cancer in stage 1 (Figure 8).

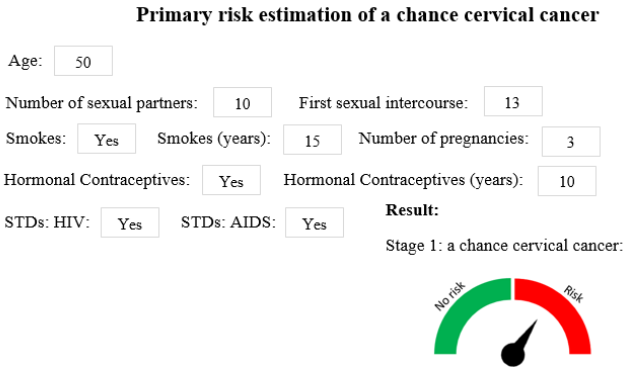

Figure 8. Case study of primary risk estimation of a chance cervical cancer.

From Figure 8, the system uses processing time no more than 1 second, when the subject fills in data into the system. The result shows that the subject, whose data record was randomly selected from the dataset, as state has a risk of a chance of cervical cancer. Therefore, the result of the prediction can describe that the subject has the opportunity of cervical cancer. This information is beneficial to acknowledging the primary risks and healthcare issues before deciding to meet a physician. In addition, it reduces the cost of medical checkups and the burden on medical staff. Moreover, it can also be used as a tool to follow and monitor a continuous risk of a chance of cervical cancer. To confirm the result of stage 1, the subject should hurry to meet the physician for a thorough diagnosis again. The data is sent to stage 2 to predict cervical cancer, as seen in Figure 9.

From Figure 9, when data are filled in the system, the system also uses processing time no more than 1 second. The system informs that the patient has a chance of cervical cancer, which uses a red label to show the result to the physician. It confirms that the patient has a chance of cervical cancer, which conforms to the result from Stage 1. This information benefits physicians in making decisions and supporting diagnosis and further treatment. In addition, it also reduces the time for physicians to diagnose patients and helps patients acknowledge primary risk from the beginning to minimize disease progression and reduce the death rate.

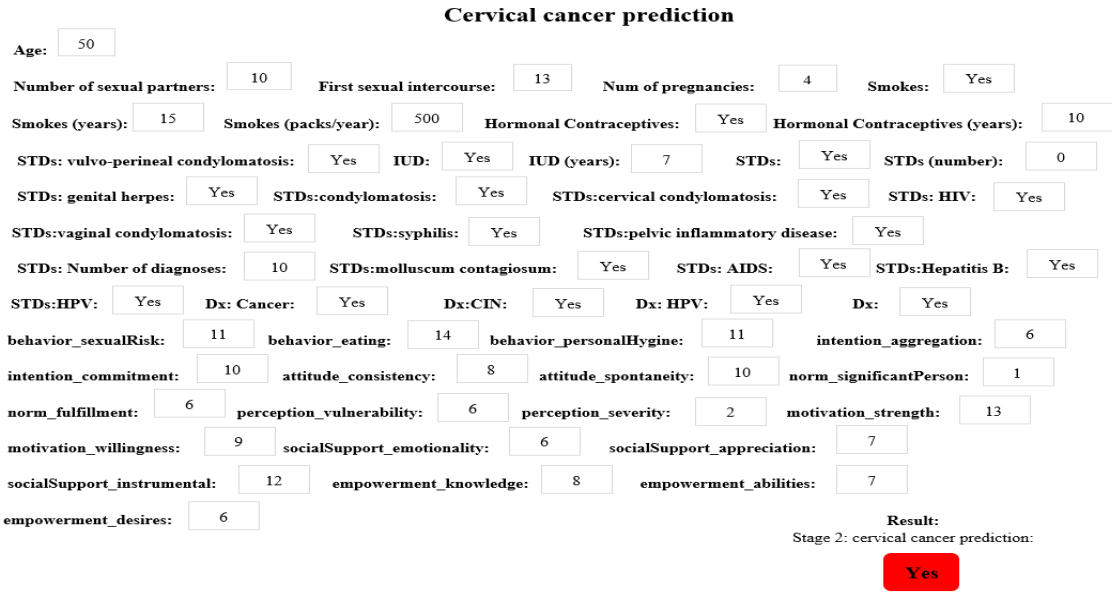

Figure 9. Case study of cervical cancer prediction.

5. Conclusions and Future Works

This paper proposes two stages of primary risk estimation of a chance cervical cancer and cervical cancer prediction using hybrid classification methods. A risk dataset of a chance cervical cancer and cervical cancer behavior risk data are used to build model. MLP, KNN, RF, and DT are used to build models. The result shows that the primary risk estimation of a chance of cervical cancer using DT achieves an accuracy of 98.80%. Then, the primary risk estimation of a chance of cervical cancer using RF is an accuracy of 98.42%. Lastly, MLP and KNN have accuracy of 96.85% and 95.54%, respectively. When the user has a risk of a chance cervical cancer, we predict cervical cancer to confirm disease detection. The result shows that cervical cancer prediction using DT also achieved an accuracy of 97.00%. Cervical cancer prediction using RF, MLP, and KNN has an accuracy of 96.00%, 95.00%, and 93.00%, respectively. To enhance the performance of primary risk estimation of a chance cervical cancer and cervical cancer prediction, we build a model using hybrid classification methods by bagging using Random Forest. The result shows that primary risk estimation of a chance cervical cancer achieves an accuracy of 99.25%, and cervical cancer prediction has an accuracy of 98.00%. Our proposed method outperforms state-of the-art methods. Therefore, our proposed method can be applied to predict cervical cancer in real situations.

For future work, we plan to develop a system for primary risk estimation of a chance cervical cancer and cervical cancer prediction to use in real situations. We expect that the system will assist in reducing the number

of patients who go to the hospital because every year, patients will be screened to monitor symptoms, especially the risk group who have to monitor continuous symptoms.

References

[1] Alam T., Khan M., Iqbal M., Wahab A., and Mushtaq M., "Cervical Cancer Prediction through Different Screening Methods Using Data Mining," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, pp. 388-396, 2019. DOI:10.14569/IJACSA.2019.0100251

[2] Bhavani C. and Govardhan A., "Cervical Cancer Prediction Using Stacked Ensemble Algorithm with SMOTE and RFERF," *Materials Today: Proceedings,* vol. 8, no. 3, pp. 3451-3457, 2023. https://doi.org/10.1016/j.matpr.2021.07.269

[3] Cervical Cancer, https://www.medparkhospital.com/content/cervical-cancer, Last Visited, 2024.

[4] Cervical Cancer, https://www.rama.mahidol.ac.th/cancer_center/th/protfolio/knowledge/gyne/servix, Last Visited, 2024.

[5] Cervical Cancer, Last Visited, 2024.

[6] Chandran V., Sumithra M., Karthick A., George T., Deivakani M., Elakkiya B., Subramaniam U., and Manoharan S., "Diagnosis of Cervical Cancer based on Ensemble Deep Learning Network using Colposcopy Images," *BioMed Research International*, vol. 2021, pp. 1-15, 2021. DOI:10.1155/2021/5584004

[7] Curia F., "Cervical Cancer Risk Prediction with Robust Ensemble and Explainable Black Boxes Method," *Health and Technology*, vol. 11, pp. 875-885, 2021. https://doi.org/10.1007/s12553-021-00554-6

[8] Deng X., Luo Y., and Wang C., "Analysis of Risk Factors for Cervical Cancer Based on Machine Learning Methods," *in Proceedings of the 5th IEEE International Conference on Cloud Computing and Intelligence Systems*, China, pp. 631-635, 2018. DOI:10.1109/CCIS.2018.8691126

[9] Galgano M., Castle P., Atkins K., Brix W., Nassau S., and Stoler M., "Using Biomarkers as Objective Standards in the Diagnosis of Cervical Biopsies," *The American Journal of Surgical Pathology*, vol. 34, no. 8, pp. 1077-1087, 2010. DOI:10.1097/PAS.0b013e3181e8b2c4

[10] Guvenc G., Akyuz A., and Açikel C., "Health Belief Model Scale for Cervical Cancer and Pap Smear Test: Psychometric Testing," *Journal of Advanced Nursing*, vol. 67, pp. 428-437, 2011. DOI:10.1111/j.1365-2648.2010.05450.x

[11] Guyon I., Weston J., Barnhill S., and Vapnik V., "Gene Selection for Cancer Classification using Support Vector Machines," *Machine learning*, vol. 46, pp. 389-422, 2002. https://doi.org/10.1023/A:1012487302797

[12] Han J., Kamber M., and Pei J., *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers, 2012.

[13] Ilyas Q. and Ahmad M., "An Enhanced Ensemble Diagnosis of Cervical Cancer: A Pursuit of Machine Intelligence towards Sustainable Health," *Digital Object Identifier*, vol. 9, pp. 12374-12388, 2021. DOI:10.1109/ACCESS.2021.3049165

[14] Kaggle, https://www.kaggle.com/datasets/ranzeet013/cervical-cancer-dataset, Last Visited, 2024.

[15] Kalbhor M. and Shinde S., "ColpoClassifier: A Hybrid Framework for Classification of the Cervigrams," *Diagnostics*, vol. 13, no. 1103, pp. 1-16, 2023. https://doi.org/10.3390/diagnostics13061103

[16] Kelwin F., Jaime C., and Jessica F., Cervical Cancer (Risk Factors), 2017. https://doi.org/10.24432/C5Z310, Last Visited, 2024.

[17] Kerkar R. and Kulkarni Y., "Screening for Cervical Cancer: An Overview," *Journal Obstet Gynecol India*, vol. 56, no. 2, pp. 115-122, 2006. https://jogi.co.in/storage/files/02_ra_screening-1.pdf

[18] Khanam F. and Mondal R., "Ensemble Machine Learning Algorithms for the Diagnosis of Cervical Cancer," *in Proceedings of the International Conference on Science and Contemporary Technologies*, Bangladesh, pp. 1-5, 2021. DOI:10.1109/ICSCT53883.2021.9642612

[19] Khomphaiboonkij U., Sreamsukcharoenchai N., Pitakkarnkul S., Rittiluechai K., and Tangjitgamol S., "Knowledge of Thai Women in Cervical Cancer Etiology and Screening," *PLoS ONE*, vol. 20, no. 1, pp. 1-12, 2023. https://doi.org/10.1371/journal.pone.0318184

[20] Kruczkowski M., Drabik-Kruczkowska A., Marciniak A., Tarczewska M., Kosowska M., and Szczerska M., "Predictions of Cervical Cancer Identification by Photonic Method Combined with Machine Learning," *Scientific Reports*, vol. 12, no. 3672, pp. 1-11, 2022. https://doi.org/10.1038/s41598-022-07723-1

[21] Kulpornsirikul M., Srison U., Laosooksathit W., and Tantitamit T., "Prevalence of Cervical High-risk Human Papilloma Viral Infection in Pregnant Women at HRH Maha Chakri Sirindhorn Medical Center Thailand," *Thai Journal of Obstetrics and Gynaecology,* vol. 32, no. 2, pp. 128-136, 2024. https://doi.org/10.14456/tjog.2024.14

[22] Laowahutanont P., Chaiwerawattana A., and Imsamran W., "Screening Guidelines Diagnose and Treat Cervical Cancer," *Kosit Press Co Ltd*,

2018.

[23] Lilhore U., Poongodi M., Kaur A., Simaiya S., Algarni A., Elmannai H., Vijayakumar V., Tunze G., and Hamdi M., "Hybrid Model for Detection of Cervical Cancer Using Causal Analysis and Machine Learning Techniques," *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 1-17, 2022. https://doi.org/10.1155/2022/4688327

[24] Madhukar B., Bharathi S., and Ashwin M., "Classification of Breast Cancer using Ensemble Filter Feature Selection with Triplet Attention Based Efficient Net Classifier," *The International Arab Journal of Information Technology*, vol. 21, no. 1, pp. 17-31, 2024. https://doi.org/10.34028/iajit/21/1/2

[25] Mahmoud H., AlArfaj A., and Hafez A., "A Fast Hybrid Classification Algorithm with Feature Reduction for Medical Images," *Applied Bionics and Biomechanics*, vol. 2022, pp. 1-11, 2022. https://doi.org/10.1155/2022/1367366

[26] Masot J., Perejon F., Saavedra L., and Morales M., "A Lightweight xAI Approach to Cervical Cancer Classification," *Medical and Biological Engineering and Computing*, vol. 62, pp. 2281-2304, 2024. https://doi.org/10.1007/s11517-024-03063-6

[27] Mehmood M., Rizwan M., ML M., and Abbas S., "Machine Learning Assisted Cervical Cancer Detection," *Frontiers in Public Health*, vol. 9, pp. 1-14, 2021. DOI:10.3389/fpubh.2021.788376

[28] Mukku L. and Thomas J., "Hybrid Decision Fusion based Multimodal Ensemble Framework for Cervical Cancer Detection," *International Journal of Scientific Research in Computer Science, Engineering, and Information Technology*, vol. 9, no. 1, pp. 17-24, 2023. https://doi.org/10.32628/CSEIT239018

[29] Nandanwar P. and Dhonde S., "A Novel Approach to Cervical Cancer Detection Using Hybrid Stacked Ensemble Models and Feature Selection," *International Journal of Electrical and Electronics Research*, vol. 11, no. 2, pp. 582-589, 2023. DOI: 10.37391/ijeer.110246

[30] Nikookar E., Naderi E., and Rahnavard A., "Cervical Cancer Prediction by Merging Features of Different Colposcopic Images and Using Ensemble Classifier," *Journal of Medical Signals and Sensors*, vol. 11, no. 2, pp. 67-78, 2021. DOI:10.4103/jmss.JMSS_16_20

[31] Patsadu O., Watanapa B., Dajpratham P., and Nukoolkit C., "Fall Motion Detection with Fall Severity Level Estimation by Mining Kinect 3D Data Stream," *The International Arab Journal of Information Technology*, vol. 15, no. 3, pp. 378-388, 2018.

[32] Predawan S. and Sompeewong P., "Cervical Cancer Screening Using Data Mining Technique," *Chonburi Hospital Journal*, vol. 45, no. 3, pp. 199-208, 2020.

[33] Rahaman M., Li C., Yao Y., Kulwa F., Wu X., Li X., and Wang Q., "Deepcervix: A Deep Learning-based Framework for the Classification of Cervical Cells using Hybrid Deep Feature Fusion Techniques," *Computers in Biology and Medicine*, vol. 136, pp. 104649, 2021. https://doi.org/10.1016/j.compbiomed.2021.104649

[34] Ramaraju H., Nagaveni Y., and Khazi A., "Use of Schiller's Test versus Pap Smear to Increase Detection Rate of Cervical Dysplasias," *International Journal of Reproduction, Contraception, Obstetrics and Gynecology*, vol. 5, pp. 1446-1450, 2017. https://doi.org/10.18203/2320-1770.ijrcog20161302

[35] Razali N., Mostafa S., Mustapha A., Wahab M., and Ibrahim N., "Risk Factors of Cervical Cancer using Classification in Data Mining," *Journal of Physics: Conference Series*, vol. 1529, pp. 1-8, 2020. DOI:10.1088/1742-6596/1529/2/022102

[36] Sheta A., El-Ashmawi W., and Baareh A., "Heart Disease Diagnosis Using Decision Trees with Feature Selection Method," *The International Arab Journal of Information Technology*, vol. 21, no. 3, pp. 427-438, 2024. https://doi.org/10.34028/iajit/21/3/7

[37] Shinde S., Kalbhor M., and Wajire P., "DeepCyto: A Hybrid Framework for Cervical Cancer Classification by using Deep Feature Fusion of Cytology Images," *Mathematical Biosciences and Engineering*, vol. 19, no. 7, pp. 6415-6434, 2022. DOI:10.3934/mbe.2022301

[38] Sindiani A., Alshdaifal E., and Alkhatib A., "Investigating Cervical Risk Factors that Lead to Cytological and Biopsy Examination," *MED ARCH*, vol. 74, no. 4, pp. 294-297, 2020. DOI:10.5455/medarh.2020.74.294-297

[39] Sonkong K., "Factors Associated with Cervical Cancer Screening among Women Aged 30-60 Years in the Taktok Sub-District Health Promotion Hospital in Tak Province," *Journal of Medicine and Health Sciences*, vol. 28, no. 3, pp. 61-74, 2021.

[40] Tanimu J., Hamada M., Hassan M., Kakudi H., and Abiodun J., "A Machine Learning Method for Classification of Cervical Cancer," *Electronics*, vol. 11, no. 463, pp. 1-23, 2022. https://doi.org/10.3390/electronics11030463

[41] UCI Machine Learning Repository, https://archive.ics.uci.edu/dataset/383/cervical+cancer+risk+factors, Last Visited, 2024.

[42] UCI Machine Learning Repository, https://archive.ics.uci.edu/dataset/537/cervical+cancer+behavior+risk, Last Visited, 2024.

[43] Wesabi Y., Choudhury A., and Won D.,

"Classification of Cervical Cancer Dataset," *in Proceedings of the Institute of Industrial and Systems Engineers Annual Conference*, Orlando, pp. 1-6, 2018.

[44] Witten I., Frank E., and Hall M., *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, 2011.

[45] Zhang L., Zhu Y., Song Y., Han Y., Sun D., Qin S., and Gao Y., "Intelligent Diagnosis of Cervical Cancer Based on Data Mining Algorithm," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1-9, 2021. DOI:10.1155/2021/7690902

[46] Zorkafli M., Osman M., Isa I., Ahmad F., and Sulaiman S., "Classification of Cervical Cancer Using Hybrid Multi-layered Perceptron Network Trained by Genetic Algorithm," *Procedia Computer Science*, vol. 163, pp. 494-501, 2019. https://doi.org/10.1016/j.procs.2019.12.132

Orasa Patsadu received her B.B.A. in Information System from Rajamangala University of Technology Krungthep, Thailand, in 2007, M.Sc. in Software Engineering from King Mongkut's University of Technology Thonburi, Thailand, in 2010, and Ph.D. in Computer Science from King Mongkut's University of Technology Thonburi, Thailand, in 2016. Currently she is working as Lecturer in Computer Science, Faculty of Science and Technology, Rajamangala University of Technology Krungthep, Bangkok, Thailand. Her current research interests are in the area of Data mining, Data Science, Software Engineering, and Visualization.

Srisuda Soranunsri received her B.Sc. in Computer Science from Ramkhamhaeng University, Thailand, in 1993 and M.S.Tech.Ed. in Computer and Information Technology from King Mongkut's University of Technology Thonburi, Thailand, in 2002. Currently she is working as Lecturer in Computer Science, Faculty of Science and Technology, Rajamangala University of Technology Krungthep, Bangkok, Thailand. Her current research interests are in the area of Database, Image Processing, and Software Engineering.

Chananate Arthayukti received her B.B.A. in Information System from Rajamangala Institute of Technology, Thailand, in 1993 and M.Sc. in Computer Science from Chulalongkorn University, Thailand, in 2003. Currently she is working as Lecturer in Computer Science, Faculty of Science and Technology, Rajamangala University of Technology Krungthep, Bangkok, Thailand. Her current research interests are in the area of Database, Image Processing, and Software Engineering.