Smart Health: Using Machine Learning to Analyze, Clustering, and Predict Cervical Cancer

1st Kezia Artha Monica  
*Computer Science Department Semarang Campus*  
*Bina Nusantara University*Jakarta, 11480, Indonesia  
kezia.monica@binus.ac.id

4th Adhe Lingga Dewi  
*Computer Science Department Semarang Campus*  
*Bina Nusantara University*Jakarta, 11480, Indonesia  
adhe.dewi@binus.ac.id2nd Jesaya Anthony Dhalimartono  
*Computer Science Department Semarang Campus*  
*Bina Nusantara University*Jakarta, 11480, Indonesia  
jesaya.dhalimartono@binus.ac.id

5th Bambang Dwi Widjanarko  
*Computer Science Department Semarang Campus*  
*Bina Nusantara University*Jakarta, 11480, Indonesia  
[bwijanarko@binus.edu](mailto:bwijanarko@binus.edu)3rd Vincent Immanuel Rahardjo  
*Computer Science Department Semarang Campus*  
*Bina Nusantara University*Jakarta, 11480, Indonesia  
vincent.rahardjo@binus.ac.id

*Abstract*—This study aims to explore and compare the effectiveness of various machine learning algorithms in early detection of cervical cancer. Using diverse datasets, we conducted analysis, clustering, and prediction with algorithms such as Support Vector Machine, Logistic Regression, Gaussian Naive Bayes, Decision Tree, dan K-Nearest Neighbour. The research method includes data preprocessing, model training and tuning, model evaluation, and result interpretation. Data was collected through online surveys, retrospective medical data, and interviews with experts. Our goal is to identify the most optimal algorithm in detecting cervical cancer with high accuracy, precision, and recall. This research can contribute to the development of more effective prevention and treatment strategies for cervical cancer. With a holistic approach and comprehensive methodology, we hope this study can provide a better understanding of the role of machine learning in healthcare and yield clinically relevant results.

Keywords—Machine Learning Algorithms, Early Detection, Cervical Cancer, Model Evaluation, Data Processing

# Introduction

Cell division that occurs abnormally [1] by destroying cells and surrounding tissues [2] can be defined as cancer. One type of cancer that is a leading cause of death among women is cervical cancer [3]. Cervical cancer is also referred to as cervical carcinoma as it occurs in the lower part of the uterus connecting the uterus to the vagina [4]. According to [5], cervical cancer ranks fourth worldwide as the leading cause of death in women. In 2018, the WHO estimated that there were 570,000 women (about half the population of Hawaii) worldwide diagnosed with cervical cancer, and approximately 54% of them, or 300,000 women (about half the population of Wyoming), died from this disease [6].

According to the WHO, about 99% of cervical cancer cases are caused by human papillomaviruses (HPV) infection, with the percentage reaching 99.7% according to [7]. HPV is a DNA virus responsible for causing several epithelial lesions and cancers [8] that later develop into carcinoma [9]. Besides cervical cancer, HPV is also a cause of laryngeal [10], oral [11], lung [12], and anogenital [13] cancers. Factors contributing to the high risk of HPV include sexual activity (age of first sexual intercourse and number of sexual partners) [14], smoking [15], oral contraceptive use for more than 5 years [16], and exposure to radiation and UV light [17]. As the most common HPV-related cancer, cervical cancer mostly occurs in several low-income countries [18] due to the lack of screening processes for pre-cervical cancer and limited treatment options [19].

If left untreated, cervical cancer will spread to nearby organs near the cervix. Cancer-related deaths can be prevented by modifying or avoiding major risk factors and implementing various prevention strategies such as HPV vaccination [20] and pap tests [21], as well as treatments including surgery [22], radiation [23], and chemotherapy [24]. Screening is a strategy to identify individuals with certain suggestive cervical cancer abnormalities or pre-cervical cancer who do not yet have any symptoms so that they can be promptly diagnosed and treated. According to the WHO, primary prevention processes include HPV vaccination and secondary prevention includes screening, which will prevent the majority of cervical cancer cases. Rapid vaccination and twice-lifetime cervical screening in all countries could prevent approximately 14.4 million cases of cervical cancer over the next 50 years [9]. Delay in the treatment process may have irreversible effects on the patient's condition as chronic diseases progress. The use of appropriate methods and early detection processes can increase the potential for successful cervical cancer treatment. Therefore, it is crucial to use alternative methods that are more easily implemented and work with different datasets to produce more reliable predictions.

The aggregation of data and learning algorithms such as artificial intelligence (AI) and machine learning can revolutionize and aid the field of medicine, including in diagnosing a disease [25]. Machine learning can have a significant impact on solving complex problems by learning patterns in data, thus enabling the resolution or prediction of specific observed data. Through classification, clustering, and diagnosis processes, machine learning algorithms can be developed to diagnose cancer, specifically cervical cancer. Intelligent algorithm methods used in detecting cervical cancer include Support Vector Machine (SVM) [26], Convolutional Neural Network [26], Fuzzy Logic [27], Naive Bayes [28], Decision Tree [29], Artificial Neural Network [30], etc. In this study, researchers investigate and compare the abilities of the abilities of the forementioned algorithms in classifying, analyzing, and predicting cervical cancer. This comparison process is conducted to determine which algorithm is most effective in detecting cervical cancer more accurately.

# Literature Review

Cervical cancer, or cervical carcinoma, is one type of cancer that affects women worldwide. Human papillomavirus (HPV) infection is the primary cause of cervical cancer, with approximately 99% of cases attributed to HPV infection [7]. Risk factors include sexual activity, smoking, oral contraceptive use, and exposure to radiation and UV light [14][17]. Prevention through HPV vaccination and regular screening has proven to be effective [20].

The use of machine learning algorithms, such as Support Vector Machine (SVM), Logistic Regression, Gaussian Naive Bayes, Decision Tree, and K-Nearest Neighbour (KNN), has been an interesting research subject in diagnosing and predicting cervical cancer [26]. Previous studies have shown that machine learning algorithms can provide accurate and fast results in diagnosing cervical cancer [25].

Support Vector Machine (SVM) has become one of the commonly used models in cervical cancer analysis. SVM allows for the classification of complex data and can handle both linear and non-linear data [26]. Previous research has shown the success of SVM in predicting cervical cancer diagnosis with high accuracy [31]. Logistic Regression is a statistical model often used in cervical cancer analysis. Although simple, Logistic Regression has good interpretability and can provide information about the relationship between predictor variables and cervical cancer diagnosis [32]. Previous studies have shown that Logistic Regression can provide satisfactory results in predicting cervical cancer risk [33]. Gaussian Naive Bayes is a probabilistic model commonly used in cervical cancer classification. Although it makes simple assumptions about data distribution, Gaussian Naive Bayes can provide satisfactory results in predicting cervical cancer diagnosis [28]. Decision Tree is an intuitive and easily interpretable model in cervical cancer analysis. Decision Tree allows for clear separation between different classes and can provide decision rules easily understood by medical practitioners [29]. Previous research has shown that Decision Tree is effective in classifying patients based on cervical cancer risk [34]. K-Nearest Neighbour (KNN) is a model based on clustering data based on proximity to nearest neighbors. KNN has been used in cervical cancer analysis to identify patterns in complex data [35]. Previous research indicates that KNN can provide reliable results in classifying medical data related to cervical cancer [28].

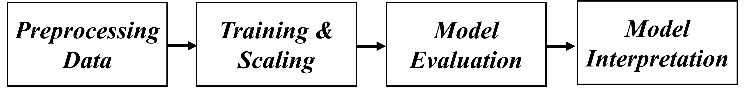
By using various models such as SVM, Logistic Regression, Gaussian Naive Bayes, Decision Tree, and KNN in cervical cancer analysis, it is hoped that this research can provide better insights into prediction, early detection, and treatment of cervical cancer. This will help reduce cervical cancer mortality globally and improve the quality of life for affected patients [18].

# Metodology Research

The aim of this research is to investigate and compare the capabilities of various machine learning algorithms in detecting cervical cancer, thereby identifying the most optimal machine learning algorithm. Accurate and precise results are obtained by conducting multiple trials using different datasets and algorithms to find the optimal algorithm across diverse types of datasets.

## Research Methodology

This research aims to develop an accurate and interpretable Machine Learning (ML) model for detecting cervical cancer. The research methodology consists of four main steps, which can be illustrated in a block diagram as shown in Figure 1 below.



1. Example of a figure caption. (*figure caption*)

### Research Methodology

The first step is to perform data preprocessing to address missing data and ensure all features are on the same scale. Missing data is handled using mean imputation for numerical data and mode imputation for categorical data. Features with a prominent level of missingness (above 70%) will be dropped and not used in this research. The data is then standardized to ensure all features are on the same scale.

### Training and Scaling

The data is divided into training and testing sets. The ML model is trained on the training set with hyperparameters optimized through cross-validation. The ML model is then scaled to the testing set to evaluate its performance. Several steps are performed during training and scaling, such as SMOTE to address data imbalance, data splitting to ensure each subset (training, testing, and validation) represents the entire dataset, and feature scaling to standardize the data. The classification stage then involves using methods such as Gaussian Naive Bayes, Decision Tree, Logistic Regression, SVM, and KNN.

### Model Evaluation

The performance of the ML model is evaluated using metrics such as accuracy, precision, recall, and F1-score. Cross-validation is performed to ensure robust results. The performance of the ML model is compared with other existing models to demonstrate its superiority.

### Model Interpretation

The ML models are analysed to understand how each model makes predictions. This research methodology is designed to develop ML models that can be used for early and accurate detection of cervical cancer. Interpretable models allow for a better understanding of cervical cancer risk factors and can aid in developing more effective prevention and treatment strategies. In this model interpretation, Python is used due to the ease of using available libraries for interpreting the models.

## Data Gathering Techniques

The data collection is sourced from datasets available on the UCI Machine Learning Repository, which provides accurate datasets that can be processed to find the most efficient algorithm for detecting cervical cancer.

## Respondet

The respondents or informants in this study are individuals with relevant knowledge or experience in cervical cancer. They may include cervical cancer patients, individuals with family members or friends suffering from cervical cancer, or medical experts and researchers with in-depth knowledge of the disease, who have contributed their data to the UCI Machine Learning Repository.

## Data Analytic Techniques

Qualitative data analysis will be conducted using a thematic or narrative analysis approach to identify patterns or significant findings from interviews or focus group discussions. Meanwhile, quantitative data analysis will involve statistical methods by comparing the cross-validation results of four machine learning models (Gaussian Naive Bayes, Decision Tree, SVM, and KNN). Additionally, the results of accuracy, sensitivity, specificity, precision, and F-measure testing will be compared to evaluate the effectiveness of the machine learning models in detecting cervical cancer. The formulas, calculation methods, and descriptions of each ML evaluation method can be found in Table 1 below.

1. Formulas

|  |  |  |
| --- | --- | --- |
| **Name** | **Formulas** | **Description** |
| Accuracy | (TP + TN)/(P+N) | The accuracy in predicting both healthy and unhealthy patients. |
| Sensitivity | TP/(TP+FN) | The percentage of sick individuals correctly identified as having the disease. |
| Specisication | TN/(FP+TN) | The percentage of healthy individuals correctly diagnosed as healthy. |
| Precision | TP/(TP+FP) | Positive Predictive Value |
| F-measure | (2 x recall x precision) / (recall+precision) | The harmonic means that combines Precision and Recall. |

## Testing

For testing, we use a licensed dataset from hospital to prove the credibility of this Machine Learning. We test it to prove that the mentioned Machine Learning most effective model to detect cervical cancer, is indeed true. The step that we needed to do is like training the model, but we don’t need to split the dataset into test and train, because the dataset that we use is already in test form.

# Result and Discussion

The focus of this research is to find the optimal algorithm in predicting cervical cancer by involving various types of datasets with different clustering and testing several algorithms on all types of datasets. From the obtained results, it will then be analyzed to find the most optimal algorithm. The data in this study consists of several samples with 20 division categories as shown in Table 2 below.

1. Atribut Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Atribut** | **Tipe variabel** | **Atribut** | **Tipe variabel** | **Atribut** | **Tipe variabel** |
| behavior\_sexualRisk | integer | attitude\_spontaneity | Integer | motivation\_willingness | Integer |
| behavior\_eating | integer | norm\_significantPerson | Integer | socialSupport\_emotionality | Integer |
| behavior\_personalHygine | integer | norm\_fulfillment | Integer | socialSupport\_instrumental | Integer |
| intention\_aggretion | integer | perception\_vulnerability | Integer | empowerment\_knowledge | Integer |
| intention\_commitment | integer | perception\_severity | Integer | empowerment\_abilities | Integer |
| attitude\_consistency | integer | motivation\_strength | Integer | empowerment\_desires | Integer |
| ca\_cervix | integer |  |  |  |  |

## Scaling and Addresing Imbalance Data

When plotting, there is no empty data, so there is no need to fill in missing data. The next step is scaling, at this stage it is done to balance the data using the SMOTE method. The SMOTE method stands for Synthetic Minority Over-sampling Technique, which aims to address data imbalance by synthesizing data in smaller or minority quantities. With better data distribution, it is hoped that the performance of machine learning can be improved. The data distribution before and after SMOTE sampling can be seen in Table 3 below.

1. Data Distributions

|  |  |  |
| --- | --- | --- |
|  | **ca\_cervix** | |
| **Class 0** | **Class 1** |
| Original class count | 51 | 21 |
| After SMOTE sampling | 51 | 51 |

After performing SMOTE sampling, the number of classes in classes 0 and 1 becomes balanced. This will avoid data redundancy.

## Cross Validation

Cross-validation is a technique used to measure the performance of a machine learning model by dividing the data into non-overlapping subsets, then training the model on several subsets and testing its performance on the remaining subset. The results of the cross-validation conducted can be seen in Table 4. After balancing and removing ambiguity from the available data, we can observe the effectiveness of ML methods for cervical cancer data based on accuracy, sensitivity, specificity, precision, and F-Measure.

1. Test Result

|  |  |
| --- | --- |
| **Methods** | **Value cross-validation** |
| GNB (Gaussian Naive Bayes) | 0.9125 |
| SVM | 0.975 |
| Decision Tree | 0.95 |
| LR (Logistic Regression) | 0.975 |
| KNN | 0.95 |
| RFC (Random Forest Classifier) | 0.9375 |

## Result

The results of these five evaluation methods can be seen in Tables 5 below. The research has achieved its goal of finding the most accurate algorithm or method for detecting cervical cancer. The study was conducted using cervical cancer datasets available from the UCI Machine Learning repository. Machine Learning methods used include Gaussian Naïve Bayes, Decision Tree, Logistic Regression, Random Forest Classifiers, KNN, and SVM to assess the effectiveness of each method and conclude which method is most accurate in detecting cervical cancer.

1. Test Result

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Classifier /Performance** | **GNB** | **Decision Tree** | **LR** | **RFC** | **KNN** | **SVM (rbf)** |
| Accuracy | 0.904761 | 0.904761 | 0.904761 | 0.952380 | 0.857142 | 0.952380 |
| Sensitivity | 0.8 | 0.8 | 0.9 | 0.9 | 0.8 | 0.9 |
| Specification | 1.0 | 1.0 | 0.909090 | 1.0 | 0.909090 | 1.0 |
| Presicion | 1.0 | 1.0 | 0.9 | 1.0 | 0.888889 | 1.0 |
| *F-measure* | 0.8888889 | 0.888889 | 0.9 | 0.947368 | 0.842105 | 0.947368 |

Testing processes used scales 0 and 1 to measure whether the test results were good or not. The closer the value is to 1, the better the test results, and vice versa. From the accuracy test results shown in Tables 5 and 6, it can be seen that the best value is obtained by the RFC (Random Forest Classifiers) method and SVM (rbf) with an accuracy value of 0.95238 out of 1. The best sensitivity value is obtained by the Logistic Regression, RFC (Random Forest Classifiers), SVM (rbf) methods with a sensitivity value of 0.9 out of 1. The specificity value of each method is good, which is 1 except for the Regression and KNN methods because their specificity values are only 0.909090 out of 1. The precision value of each method is good, which is 1 except for the Logistic Regression and KNN methods because their precision values are 0.9 and 0.888889, respectively. The best F-measure value is obtained by the SVM (linear), SVM (rbf), and RFC (Random Forest Classifiers) methods with an F-measure value of 0.947368 out of 1. Based on the results obtained by comparing the performance of accuracy, sensitivity, specificity, precision, and F-measure, it can be concluded that the RFC (Random Forest Classifiers) and SVM (rbf) methods are the best Machine Learning methods for processing cervical cancer data. The evaluation results of these three methods show values close to 1 in all performance aspects, indicating that these three methods are the most optimal.

## Testing

For further elaboration of our paper, we did some tests to prove the credibility of the chosen Machine Learning method for processing cervical cancer data. After we do the test, this is the result that we can see below in Table 6. To understand the table, we need to compare the train dataset with the tested dataset and match the yes and the no. YES, means they have cervical cancer, and NO means they do not have cervical cancer. The more matches we found, the YES accurate the Machine Learning model. From the train we found that the most effective model is RFC and SVM (rbf), so in the tables below, those three models should have the most matched results. To make it easier to understand, the highlighted row is the one that’s mismatched.

1. Result Testing

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data Tested | GNB | | Decision Tree | | KNN | | LR | | Random Forest | | SVM | |
| Test | Train | Test | Train | Test | Train | Test | Train | Test | Train | Test | Train |
| 1 | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| 2 | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| 3 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 4 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 5 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 6 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 7 | YES | NO | NO | NO | YES | NO | NO | NO | NO | NO | YES | NO |
| 8 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 9 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 10 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 11 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 12 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 13 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 14 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 15 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 16 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 17 | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| 18 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 19 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| 20 | NO | NO | NO | NO | NO | NO | YES | NO | NO | NO | NO | NO |
| 21 | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |

Based on Table 6, testing using Gaussian Naïve Bayes, it appears that one thing has been missed, this proves that GNB is not the most efficient model for predicting cervical cancer. Testing using Decision Tree, we can see that there’s no one missed, it does prove that DT is the most efficient model to predict cervical cancer. Test using K- Nearest Neighbor, we can see that there’s one missed, it does prove that KNN is not the most efficient model to predict cervical cancer. Test using Logistic Regression, we can see that there’s one miss, it does prove that Logistic Regression is not the most efficient model to predict cervical cancer. Test using Random Forest Classifier, we can see that there’s no one missed, it does prove that GNB is the most efficient model to predict cervical cancer. Test using Support Vector Machine, we can see that there’s one miss, it does prove that Support Vector Machine is not the most efficient model to predict cervical cancer. For further elaboration of our paper, we did some tests to prove the credibility of the chosen Machine Learning method for processing cervical cancer data. From the test above, the Decision Tree and RFC are the most efficient ones.

## Trial

To prove our training and test data, we do a trial with an input according to the features that needed to be filled. This process will be done with the help of the user that will fill in the required features through our website, and then the website will generate the result, whether there is any cervical cancer present or not based on the findings. The website will still display the prediction using all 5 methods, but we will highlight the RFC part to show that it is the most accurate answer. The trial data can be seen in table 7 below.

1. Result Testing

|  |  |  |  |
| --- | --- | --- | --- |
| **Atribut** | **User** | | |
| **1** | **2** | **3** |
| behavior\_sexualRisk | 8 | 10 | 15 |
| behavior\_eating | 9 | 7 | 1 |
| behavior\_personalHygine | 10 | 4 | 10 |
| intention\_aggretion | 2 | 2 | 8 |
| intention\_commitment | 4 | 6 | 9 |
| attitude\_consistency | 7 | 8 | 5 |
| attitude\_spontaneity | 8 | 9 | 2 |
| norm\_significantPerson | 7 | 5 | 10 |
| norm\_fulfillment | 5 | 3 | 15 |
| perception\_vulnerability | 3 | 1 | 4 |
| perception\_severity | 6 | 5 | 2 |
| motivation\_strength | 7 | 7 | 8 |
| motivation\_willingness | 0 | 8 | 9 |
| socialSupport\_emotionality | 2 | 3 | 10 |
| socialSupport\_appreciation | 5 | 7 | 11 |
| socialSupport\_instrumental | 7 | 3 | 14 |
| empowerment\_knowledge | 4 | 5 | 13 |
| empowerment\_abilities | 5 | 9 | 7 |
| empowerment\_desires | 6 | 2 | 2 |
| The prediction by RFC | False | False | False |
| The prediction by LR | False | False | False |
| The prediction by GNB | False | False | False |
| The prediction by Decision Tree | False | False | False |
| The prediction by KNN | False | False | False |

# Result and Discussion

This research was conducted to identify the most efficient Machine Learning algorithm in detecting cervical cancer to enable early intervention. The algorithms used include Support Vector Machine (SVM), Logistic Regression, Gaussian Naive Bayes, Decision Tree, and K-Nearest Neighbour (KNN), Random Forest Classifiers (RFC) with data processing conducted in the Python programming language. The steps involved in data processing prior to analysis include data preprocessing, training, and scaling, evaluating models, and model interpretation. In the data preprocessing stage, the data is first checked for any missing values and processed accordingly, followed by standardization using SMOTE sampling. The training and scaling stage involves cross-validation to balance the data, thus avoiding redundancy and enabling more accurate data processing. Subsequently, during evaluating models, the ML algorithms are compared based on their accuracy, sensitivity, specificity, precision, and F-measure. From these five evaluation models, the algorithm with results closest to 1 in all evaluation metrics is selected. Upon calculation, it is concluded that RFC (Random Forest Classifiers) and SVM (rbf) are the best ML methods for processing cervical cancer data, as these three methods consistently show results closest to 1 across all evaluation metrics, indicating their optimality.

After training with ML model, we do some tests, and from the test, we find that Decision Tree and RFC are the most efficient ones, but from the training, its RFC and SVM. Why does this happen? It happens because SVM is more compatible for processed data. So, we need to process the data first before inputting using the SVM model. SVM is also quite vulnerable to overfitting, which can be mis predict new data. For Decision Tree, is dependent on the root, so if the root changes, the validity also changes. RFC is way more stable than the other and performs a consistent analysis whether it’s trained/ processed data or new data. After trial with user input through stream lit web, we can generate the result according to 5 methods. Since the most accurate is RFC, the user should look into RFC and check is there an urgency to check into the hospital or not. From that we can conclude that RFC is the best model to predict cervical cancer.

##### References

1. Brown, J. S., Amend, S. R., Austin, R. H., Gatenby, R. A., Hammarlund, E. U., & Pienta, K. J. (2023). Updating the definition of cancer. Molecular Cancer Research, 21(11), 1142-1147.
2. Nenclares, P., & Harrington, K. J. (2020). The biology of cancer. Medicine, 48(2), 67-72.
3. Fowler JR, Maani EV, Dunton CJ, et al. Cervical Cancer. [Updated 2023 Nov 12]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK431093/>
4. Esmaeilzadeh, A. A., & Nasirzadeh, F. (2023). Uterus Cancer. Eurasian Journal of Chemical, Medicinal and Petroleum Research, 2(5), 63-83.
5. Arbyn, M., Weiderpass, E., Bruni, L., de Sanjosé, S., Saraiya, M., Ferlay, J., & Bray, F. (2020). Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. The Lancet Global Health, 8(2), e191-e203.
6. Cohen, P. A., Jhingran, A., Oaknin, A., & Denny, L. (2019). Cervical cancer. The Lancet, 393(10167), 169-182.
7. Okunade, K. S. (2020). Human papillomavirus and cervical cancer. Journal of Obstetrics and Gynaecology, 40(5), 602-608.
8. Luria L, Cardoza-Favarato G. Human Papillomavirus. [Updated 2023 Jan 16]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK448132/>
9. Soheili, M., Keyvani, H., Soheili, M., & Nasseri, S. (2021). Human papillomavirus: A review study of epidemiology, carcinogenesis, diagnostic methods, and treatment of all HPV-related cancers. Medical journal of the Islamic Republic of Iran, 35, 65.
10. Nocini, R., Molteni, G., Mattiuzzi, C., & Lippi, G. (2020). Updates on larynx cancer epidemiology. Chinese Journal of Cancer Research, 32(1), 18.
11. Betz, S. J. (2019). HPV-related papillary lesions of the oral mucosa: a review. Head and neck pathology, 13(1), 80-90.[12] Rezaei, M., Mostafaei, S., Aghaei, A., Hosseini, N., Darabi, H., Nouri, M., ... & Moghoofei, M. (2020). The association between HPV gene expression, inflammatory agents and cellular genes involved in EMT in lung cancer tissue. BMC cancer, 20, 1-11.
12. Rezaei, M., Mostafaei, S., Aghaei, A., Hosseini, N., Darabi, H., Nouri, M., ... & Moghoofei, M. (2020). The association between HPV gene expression, inflammatory agents and cellular genes involved in EMT in lung cancer tissue. BMC cancer, 20, 1-11.
13. Tripathi, A., & Sahu, U. (2022). An overview of HPV: Causes, symptoms, and clinical manifestations. Immunopathology, Diagnosis and Treatment of HPV Induced Malignancies, 1-19.
14. Markowitz, L. E., & Schiller, J. T. (2021). Human papillomavirus vaccines. The Journal of infectious diseases, 224(Supplement\_4), S367-S378.
15. Alotaibi, M., Valova, V., HÄnsel, T., Stromberger, C., Kofla, G., Olze, H., ... & Coordes, A. (2021). Impact of Smoking on the Survival of Patients with High-risk HPV-positive HNSCC: A Meta-analysis. in vivo, 35(2), 1017-1026.
16. Asthana, S., Busa, V., & Labani, S. (2020). Oral contraceptives use and risk of cervical cancer—A systematic review & meta-analysis. European Journal of Obstetrics & Gynecology and Reproductive Biology, 247, 163-175.
17. Godar, D. E. (2021). UV and reactive oxygen species activate human papillomaviruses causing skin cancers. Challenges in Sun Protection, 55, 339-353.
18. Jalil, A. T., & Karevskiy, A. (2020). The cervical cancer (CC) epidemiology and human papillomavirus (HPV) in the middle east. International Journal of Environment, Engineering and Education, 2(2), 7-12.
19. Daponte, N., Valasoulis, G., Michail, G., Magaliou, I., Daponte, A. I., Garas, A., ... & Daponte, A. (2023). HPV-based self-sampling in cervical cancer screening: an updated review of the current evidence in the literature. Cancers, 15(6), 1669.
20. Rees, C. P., Brhlikova, P., & Pollock, A. M. (2020). Will HPV vaccination prevent cervical cancer? Journal of the Royal Society of Medicine, 113(2), 64-78.
21. Ahmed, N. U., Rojas, P., & Degarege, A. (2020). Knowledge of cervical cancer and adherence to Pap smear screening test among female university students in a multiethnic institution, USA. Journal of the National Medical Association, 112(3), 300-307.
22. Zhao, H., He, Y., Yang, S. L., Zhao, Q., & Wu, Y. M. (2019). Neoadjuvant chemotherapy with radical surgery vs radical surgery alone for cervical cancer: a systematic review and meta-analysis. OncoTargets and therapy, 1881-1891.
23. Mayadev, J. S., Ke, G., Mahantshetty, U., Pereira, M. D., Tarnawski, R., & Toita, T. (2022). Global challenges of radiotherapy for the treatment of locally advanced cervical cancer. International Journal of Gynecologic Cancer, 32(3).
24. Gadducci, A., & Cosio, S. (2020). Neoadjuvant chemotherapy in locally advanced cervical cancer: review of the literature and perspectives of clinical research. Anticancer research, 40(9), 4819-4828.
25. Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. Journal of ambient intelligence and humanized computing, 14(7), 8459-8486.
26. Aurelia, J. E., Rustam, Z., & Wirasati, I. (2021). Cervical cancer classification using convolutional neural network-support vector machine. TELKOMNIKA (Telecommunication Computing Electronics and Control), 19(5), 1605-1611.
27. Pramanik, R., Biswas, M., Sen, S., de Souza Júnior, L. A., Papa, J. P., & Sarkar, R. (2022). A fuzzy distance-based ensemble of deep models for cervical cancer detection. Computer Methods and Programs in Biomedicine, 219, 106776.
28. Setiawan, Q. S., Rustam, Z., & Pandelaki, J. (2021, December). Comparison of Naive Bayes and support vector machine with grey wolf optimization feature selection for cervical cancer data classification. In 2021 international conference on decision aid sciences and application (DASA) (pp. 451-455). IEEE.
29. Mehmood, M., Rizwan, M., & Abbas, S. (2021). Machine learning assisted cervical cancer detection. Frontiers in public health, 9, 788376.
30. Fekri-Ershad, S., & Ramakrishnan, S. (2022). Cervical cancer diagnosis based on modified uniform local ternary patterns and feed forward multilayer network optimized by genetic algorithm. Computers in Biology and Medicine, 144, 105392.
31. Xie, L., Chu, R., Wang, K., Zhang, X., Li, J., Zhao, Z., ... & Kong, B. (2020). Prognostic assessment of cervical cancer patients by clinical staging and surgical-pathological factor: a support vector machine-based approach. Frontiers in oncology, 10, 1353.
32. Sachan, P. L., Singh, M., Patel, M. L., & Sachan, R. (2018). A study on cervical cancer screening using pap smear test and clinical correlation. Asia-Pacific journal of oncology nursing, 5(3), 337-341.
33. Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database, 2020, baaa010.
34. Karim, E., & Neehal, N. (2019, May). An empirical study of cervical cancer diagnosis using ensemble methods. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT) (pp. 1-5). IEEE.
35. Jahan, S., Islam, M. S., Islam, L., Rashme, T. Y., Prova, A. A., Paul, B. K., ... & Mosharof, M. K. (2021). Automated invasive cervical cancer disease detection at early stage through suitable machine learning model. SN Applied Sciences, 3, 1-17.