JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CLASSIFICATION OF CERVICAL CANCER PREDICTION USING NEURAL NETWORK AND ML ALGORITHMS

[1]Mrs.S.Karthika, [2]Dr.K.Juliana Gnanaselvi ,

[1]Research Scholar(Ph.D. – Full time) , [2] Assistant Professor in Computer Science
[1]Department of Computer Science,
Rathinam College of Arts and Science, Eachanari., Coimbatore, India

Abstract: The fourth most prevalent chronic illness among women is cervical cancer. The skin cells and mucous film cells in the vaginal area are affected over time by disease. The World Health Organization (WHO) uses the general term "malignant growth" to refer to a group of diseases that can affect any area of the body and are incredibly harmful. Therefore, it will be highly beneficial to offer a model with exceptional precision and high accuracy for identifying at the right stage of contamination. In many medical imaging applications, artificial neural networks (ANN) play a significant role. An ANN is used to categorize the normal and abnormal cells in the cervix region of the uterus in order to discover cervical cancer cells. Because cervical cancer develops without any signs, it is particularly difficult to detect. An artificial neural network is used to distinguish between normal, abnormal, and malignant cells, producing more accurate findings than manual screening techniques like the Pap smear test and Liquid Cytology Based (LCB) test. The ANN employs a number of topologies to quickly and precisely identify cervical cells. When aberrant cervical cells are discovered in time, there is a higher chance that cervical cancer will be treated. Because manual identification requires trained pathologists, takes a long time, and is prone to error, automated methods of detecting aberrant cervical cells have been developed. Convolutional Neural Network Algorithms and machine learning algorithms will be used in this research study to produce a reliable method for diagnosing the disease as well as a full grasp of the risk factors connected to cervical cancer.

***Index Terms* – Artificial Neural Networks, Cervical Cancer, Pap Smear Test, Machine Learning, LCB.**

I. INTRODUCTION

Cervical cancer is caused by persistent infections in the skin and mucous membrane cells. The fact that this cancer doesn't show any signs when it first appears is one of its worrying characteristics. Although vaginal death is a serious side effect, cervical cancer usually develops slowly and only shows symptoms in its later stages. As a result, early symptoms are usually invisible. Treatment for cervical illness is only possible if it is discovered at an underlying stage. Developments in artificial intelligence (AI) and remote connectivity have made it possible for us to build an accurate, dynamic clinical analytic framework.

It operates even more efficiently and doesn't need human contact. Hospital screening tests had previously revealed cervical cancer. But it's possible that medical professionals—particularly nurses and doctors—haven't gotten the required training, which leads to them making bad choices when it comes to cancer prediction. Many of the artificial intelligence algorithms now available cannot predict cervical cancer with high accuracy. Therefore, having a high-accuracy prediction model to help identify at the right stage of infection will be very helpful. The research project intends to diagnose cervical cancer using neural network models such as Convolutional Neural Networks (CNN) and Artificial Neural Nets, and classification algorithms such as SVM and RF. The UC Irvine (UCI) ML repository has the cervical cancer risk factors dataset, which consists of 36 variables for 858. We first apply a preprocessing strategy called missing values removal and then min-max scaling under normalization to show the data and preprocess the raw data.

This research provides a valuable model for improving the effectiveness of machine learning algorithms and classification methodologies for cervical cancer diagnosis, which could lead to a reduction in death rates. In determining whether or not individuals genuinely had cervical cancer, it also focused on the model's sensitivity and overall accuracy. Cervical cancer diagnosis can be aided by the study's findings for medical professionals and cancer researchers. After that, they can begin treating the ailment, increasing the likelihood that the patient will fully recover.

There is potential for improving patient outcomes and reducing the death rate from cervical cancer by using machine learning algorithms in the diagnosis process. Our research provides a valuable framework for applying machine learning algorithms and classification techniques to improve the prediction accuracy of cervical cancer. It also highlights the general soundness and sensitivity of the model. A woman's chance of acquiring cervical cancer can be ascertained by looking at the risk factors and indicators included in the cervical cancer dataset. Features include age, number of pregnancies, number of sexual partners, smoking, time between last diagnosis and first diagnosis, and medical history. Data can be analyzed using EDA to uncover broad trends. It is possible that algorithms based on machine learning will aid in the detection of cervical cancer. These algorithms can be trained using a substantial dataset of patient data, which includes test results, medical history, and demographic data. These data can be analyzed by machine learning algorithms, which can identify trends and correlations that suggest cervical cancer. One of the primary advantages of using machine learning algorithms for cervical cancer detection is that they can be trained to identify patterns and correlations that may not be immediately apparent to human analysts. Additionally, machine learning algorithms can process vast amounts of data quickly and efficiently, making them ideal for analyzing large datasets.

II. LITERATURE REVIEW

Mitra P, Mitra S, and Pal [1] present a way for establishing a cross variety decision really strong organization in a fragile figuring perspective for identifying the different stages of cervical sickness. In order to hybridize data-based subnetwork module construction with Genetic Algorithms (GAs) for hereditary computations, Interactive Dichotomizer 3 (ID3) and unforgiving set speculations are employed. The coordination updates above show the extent of the gathering score, network size, and planning time as compared to a normal multi-layer perceptron.

Sunny Sharma [2] The C5.0 computation is utilized in the paired tree construction process. It has developed from its standard C4.5 computation. C5.0 generates decision trees from tagged data (preparing set). In most cases, it uses the Max Gain technique to choose the best-upgraded property from the available data in order to address organizing challenges.C5.0 is quicker, more memory-efficient, and allows multithreading during execution when compared to C4.5.

The Constrained Conditional Model (CCM), according to Soumya M., Sneha, and Arun Vinod [3], is a network in which each point represents the frequency of the occurrences the pixel is viewed by then. This technique is used by the Gray Level Co-occurrence Matrix [GLCM] to analyze the highlights.

In 2022, Surendiran R, Thangamani M, Monisha S, and Rajesh P [4] report on the use of deep learning and machine learning methods for the prediction of cervical cancer. The authors used a collection of patient data that included demographic information and medical history to train and evaluate various machine learning models, including logistic regression, decision trees, random forests, and support vector machines. They also employed deep learning methods for the same objective, including long short-term memory (LSTM) networks and convolutional neural networks (CNNs). In comparison to the other models, the LSTM network fared the best, as seen by its accuracy of 92.31%. The study's overall findings emphasize the promise of deep learning and machine learning methods as a tool for cervical cancer early detection.

Ashok and Aruna[5] discussed and evaluated a variety of machine learning (ML) techniques for feature extraction, segmentation, and cell classification, including support vector machines (SVM), gray level concurrence matrix (GLCM), KNN, convolutional neural networks (CNN), spatial fuzzy cluster algorithms, RF, C5.0, and hierarchical cluster algorithm. Typical parameters included dataset volume, drawbacks, and precision. Still, it performs optimally on small datasets.

Image processing, data extraction, and machine learning approaches have been applied by Chih-Jen Tseng and Chi-Jie Lu [6] to diagnose cervical cancer. From each image, the texture and shape features are retrieved together. For sequential forward selection, random section selection, and reciprocal information selection, optimal properties are selected. Using the SVM, images of cervical carcinoma are categorized. In order to ascertain the appropriate mechanism appropriate for the diagnosis of cervical cancer, several selection techniques are compared. SVM's shortcoming is its inability to provide a probabilistic justification for classification, which results in extremely strict classifications.

Anuraga et al.[7] made an effort to investigate the factors influencing the prognosis of cancer patients in Makassar, Indonesia. Up to 38 cancer patients were present in the specimens used in this study. They combine sample data training with the RF to find tree merger data. This technique's primary issue is that it only attains 50% accuracy. Bandyopadhyay and Nasipuri [8]concentrated on segmenting pre-processed pictures using K-Means clustering and performing Herlev analysis. From the segmented nucleus, shape attributes, validated ground truth values, and IOU segmentation results are extracted. With the assistance of the RF Classifier, the nucleus is identified based on form features, and a comparison is made with the other classifiers.

William et al[9]. automated the cervical cancer diagnosis procedure using Pap-drug images in an effort to lower the likelihood of error. To improve the image, local adaptive histogram equalization was applied.

Alyafeai and Ghouti [10] develop a fully integrated pipeline for cervical image-based cervical cancer detection and screening in 2020. Two deep neural network-learning models are part of the current pipeline for automated cervical tumor diagnosis and identification. In terms of union intersection (IoU) estimation, the first test achieves a detection precision of 0.68 by detecting the cervix area 1,000 times faster than the most advanced data-driven simulations. In the second model, cervical cancers are identified using self-extracted characteristics. Two lightweight co-evolutionary neural network (CNN)-focused models are utilized to learn such features.

III. TECHNIQUES FOR DEEP LEARNING AND MACHINE LEARNING

Choosing the right techniques and algorithms for a dataset is a crucial first step in creating a reliable and accurate model. This section on Machine Learning Algorithms examines various approaches to handling the dataset and talks about the Neural Network and Machine Learning Algorithms that were used.

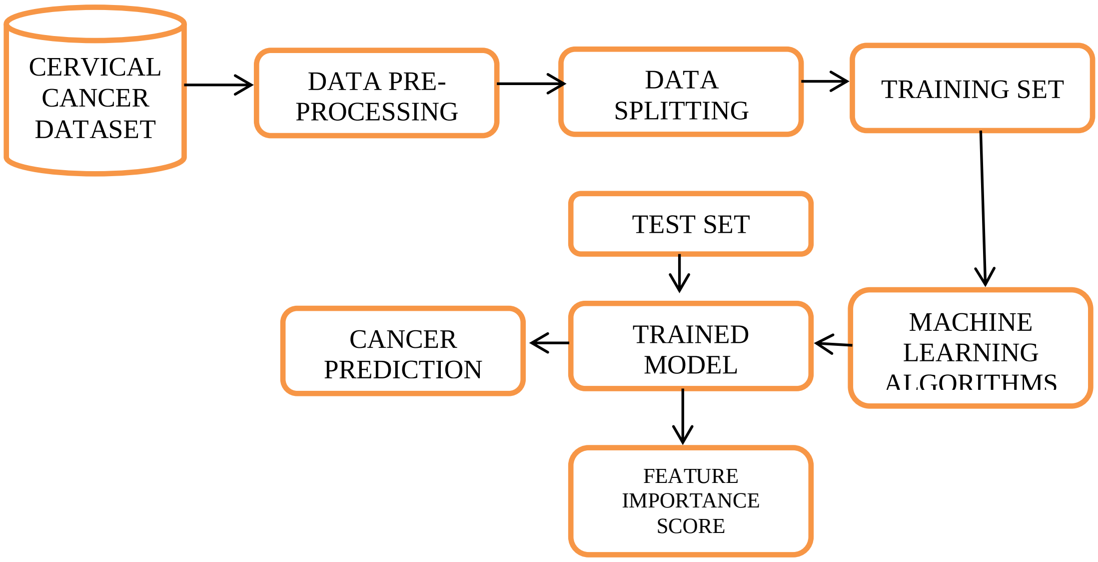

Figure 1. Flow Diagram of Machine Learning Algorithms

3.1Support Vector Machine Algorithm

SVM is a kind of ensemble machine learning technique that can be used for regression and classification issues. Characterization problems are a common application for it. In an SVM computation, each data unit object is represented as a point in n-dimensional geometry, where n is the number of features you have and the estimate of each component is the estimate of a certain arrangement. Next, we find the hyper-plane that splits the two classes so that we can carry out the order. A single person's perception's coordinates are known as the booster vector. The best way to differentiate between the two classes (hyper-plane and line) is via the SVM classifier. This work SVC() Function is used to generate an SVM model, which is the first step in the creation of an SVM classifier.

3.2 Random Forest

RF is a decision tree predictor in which each tree in the forest is dependent on the values of a random vector and is picked evenly and at random. The generalization error of a forest of tree classifiers is influenced by the strength of each individual tree in the forest as well as by their associations. When it comes to noise resistance, they are more robust. This supervised classification approach, which is used for prediction, is believed to be superior to decision trees because there are more trees in the forest.

RF develops uncorrelated combinations or multiple decision trees based on bootstrap aggregation (bagging) technique [20] using Classification and Regression Tree (CART) technique [19]. Learning the proper classification of some dependent variables (y) and some independent variables (x), as well as the relationship between them, is the aim of the CART technique. In RF, every tree creates an individual decision tree by randomly choosing a portion of the information. To ensure that every tree reaches a leaf node without pruning, RF repeatedly divides the chosen random subset from the root node to a child node.

Every tree separately classifies the characteristics and the target variable, then casts a vote to determine which tree class will be the winner. The final overall classification is determined by RF using the majority of the trees that have voted. The following steps can be used to describe how RF is constructed:
Step 1: From the dataset, create N bootstrap samples.
Step 2: Every node selects a random sample of size m attributes, where $m < M$. (M stands for the total number of characteristics).
Step 3: Builds a split with the m attributes chosen in Step 2 and determines the k node by utilizing the optimal split point. ("k" denotes the following node).
Step 4: Continue splitting the tree until it reaches only one leaf node and is finished.
Step 5: Each bootstrapped individually is used to train the algorithm.
Step 6: Gathers the prediction data from the (n) trained trees by using the trees' classification voting.
Step 7: Builds the final RF model using the features that received the most votes.

3.3 Feature Selection Techniques

There are two feature selection techniques that our model employs: Using Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA). Without lowering the model, these two dimensional reduction approaches can be utilized to condense large quantities of features into smaller amounts of features execution. The essential information is still included in the remaining features, just as it is in the whole features dataset. Using the formula for each technique, the features are prioritized and weighted based on their significance.

PCA is a statistical mathematical method that defines the feature orientation by utilizing the eigenvector. The primary goal of principal component analysis (PCA) is to map the n-dimension feature space into the k-dimension, or principle component, when $k < n$. The covariance matrix is constructed, and the eigenvectors and eigen values are then determined using the result. Since it demonstrates the strongest correlation between the attributes of the data set, the eigen vector with the highest eigen value is selected as the primary component of the cervical cancer dataset. To select the most important data and exclude the least important, the eigen values are arranged in ascending order. This implies a reduction of the highest level data to a lesser dimension. To determine the distribution of the data in the data collection, compute the variance.

IV. Neural Network Algorithms

An essential component of many medical imaging applications is the artificial neural network (ANN). An artificial neural network (ANN) is used in the cervix region of the uterus to identify normal and abnormal cells in order to detect cervical cancer cells.

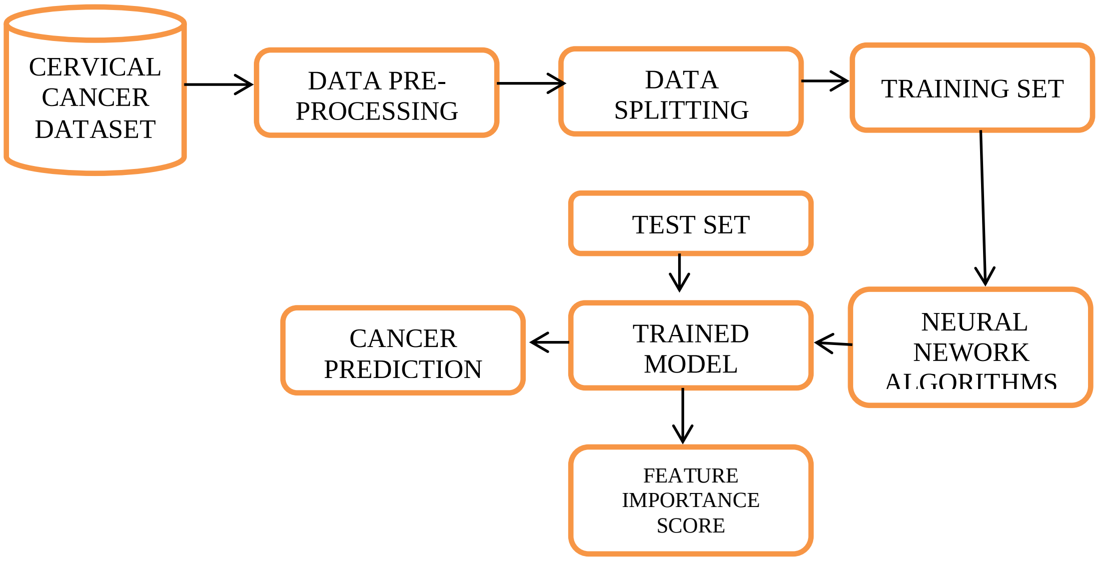

Figure 2. Flow Diagram of Neural Network Algorithms

4.1 Artificial Neural Networks

Neural networks (NNs), another name for artificial neural networks (ANNs), are computational models inspired by the organic neural systems present in human minds. An artificial neural network (ANN) is composed of several artificial neurons, which are connected units or nodes that mimic neurons throughout the learning process. a brain based on biology. Similar to brain synapses, each link has the ability to transmit a signal to other neurons. a synthetic neuron that accepts and processes inputs before forwarding them to associated neurons. The "signal" at the load causes the signal intensity at a connection to change, therefore it might be authentic. Neurons may have a threshold that, if crossed by the mixed signal, permits the delivery of a symbol. Layers are a common arrangement for neurons. The different operations and transformations on the inputs are carried out by distinct layers. Signals originate from the potentially after repeatedly iterating over the layers, from the first primary layer to the second main layer (the output layer). An ANN model with layers, a "relu" activation function, and the creation and compilation of models are produced at this stage.

The pre-processing, filtering, and feature extraction processes involved in digital image processing are carried out on sample data sets that are gathered. The input image is stored in an artificial neural network (ANN). The LVQ method is used to calculate the coefficient mean value of the extracted image, which is utilized to classify the normal and abnormal cell with a 90% accuracy result in cervical cell classification for cancer diagnosis. Artificial neural network (ANN) techniques are proposed by Fatemeh Hoda Moghimi et al. for application in health clinics. The ANN maps key components and reasoning using a multi-layered perceptron. An ANN's design comprises of a single input layer, one output layer, and an infinite number of hidden layers.
Normal and abnormal cells are classified using the three classifiers. The required features for the classification results are extracted from the input image. With the input, output, and hidden layers, a three-layered artificial neural network (ANN) is employed. Compared to other classifiers, the employed data set yields more accurate findings because it has been trained. N. Mustafa et al.5 extract novel characteristics from cervical cells and suggest an artificial neural network technology. The input image is taken from the Pap smear slides, and the intensity levels of the red, blue, green, and perimeter colors are extracted. This aids the ANN in differentiating between malignant and normal cervical cells.

4.2 Convolutional Neural Networks

In this study, we present a cervical cancer cell prediction model and a deep machine learning classification system based on neural networks. It is common practice to use the CNN model for prediction and classification. The cell pictures were used to train a CNN model to extract attributes that were deep learned. CNN is made to convert picture input into a numerical value, or output variable. They have proven to be so effective that they are currently the go-to method for any prediction problem that requires the input of picture data. This level generates a CNN model with an input layer, 13 hidden layers, and an output layer using "uniform" as the kernel initializer and "relu" as the activation function. After that, the models will be made and assembled.

V. CLASSIFICATION REPORTS

Table 1 CLASSIFICATION REPORT

		PRECISION	RECALL	F1-SCORE	SUPPORT
SVM	0	1.00	0.97	0.98	161
	1	0.69	1.00	0.81	11
RF	0	0.96	0.98	0.97	161
	1	0.57	0.36	0.44	11
ANN	0	0.99	0.96	0.97	161
	1	0.60	0.91	0.71	11
CNN	0	0.98	0.98	0.98	161
	1	0.64	0.64	0.64	11

Table 1, shows the Precision, Recall, F1 score and Support of the SVM, RF, CNN and ANN models.

VI. CONCLUSION

The World Health Organization (WHO) reports that 80 percent of instances of cervical cancer occur in underdeveloped nations. By figuring out the risk factors for cervical cancer, the cure ratio—the percentage of female patients that survive the disease can be raised. The method generates precise results for the prediction of cervical cancer cells by using artificial intelligence (AI) algorithms to the collected data sets. When compared to other algorithms, the algorithms exhibit higher accuracy. According to society, this method is especially helpful in identifying patients who may have cervical cancer. Following the construction of models such as SVM, RF from ML, CNN, and Neural Networks, we obtained 97% accuracy for SVM, 95.3% accuracy for CNN with high precision, 94% accuracy for RF, and 95.2% accuracy for ANN with normal precision. This research concluded that women between the ages of 29 and 45 are the ones who are most likely to develop cervical cancer. Thus, the system is set up to identify cancer in women at an early stage and maybe save their lives.

This paper demonstrates how artificial intelligence (AI) can be used to extract the underlying cause of cervical cancer from numerical data. This system can be used in a number of settings, such as linguistic research, hospitals for prompt cancer diagnosis, user mobile applications where users can take a survey to learn about the presence of cancer, etc.

We will continue to explore further approaches to deal with the imbalanced problem in the future. We will also use a number of categorization procedures, including ensemble methods, to enhance model performance.

REFERENCES

[1] Alyass, A., Turcotte, M., & Meyre, D. (2015). From big data analysis to personalized medicine for all: challenges and opportunities. *BMC medical genomics*, *8*(1), 1-12.

[2] Sharma, S., & Gupta, S. (2016). Decision tree approach in machine learning for prediction of cervical cancer stages using WEKA. *International Journal of Recent Trends in Engineering & Research*, *2*(8)

[3] Gogate, U., & Bakal, J. (2018). Healthcare monitoring system based on wireless sensor network for cardiac patients. *Biomedical & Pharmacology Journal*, *11*(3), 1681.

[4] Bhattacharjee, S., Ray, D., Saha, D., & Sobya, D. Classifying Pap Smear Images with an Advanced Composite Random Forest Model.

[5] Ashok, B., & Aruna, P. (2016). Comparison of Feature selection methods for diagnosis of cervical cancer using SVM classifier. *Int. J. Eng. Res. Appl*, *6*, 94-99.

[6] Tseng, C. J., Lu, C. J., Chang, C. C., & Chen, G. D. (2014). Application of machine learning to predict the recurrence-proneness for cervical cancer. *Neural Computing and Applications*, *24*, 1311-1316.

[7] Anuraga, G., & Fernanda, J. W. (2019, May). Random forest prognostic factor in colorectal cancer. In *Journal of Physics: Conference Series* (Vol. 1217, No. 1, p. 012098). IOP Publishing.

[8] Bandyopadhyay, H., & Nasipuri, M. (2020, February). Segmentation of pap smear images for cervical cancer detection. In *2020 IEEE Calcutta Conference (CALCON)* (pp. 30-33). IEEE.**[9]** William, W., Ware, A., Basaza-Ejiri, A. H., & Obungoloch, J. (2019). Cervical cancer classification from Pap-smears using an enhanced fuzzy C-means algorithm. *Informatics in Medicine Unlocked*, *14*, 23-33.

[10] Alyafeai, Z., & Ghouti, L. (2020). A fully-automated deep learning pipeline for cervical cancer classification. *Expert Systems with Applications*, *141*, 112951.

[11] Khamparia, A., Gupta, D., Rodrigues, J. J., & de Albuquerque, V. H. C. (2021). DCAVN: Cervical cancer prediction and classification using deep convolutional and variational autoencoder network. *Multimedia Tools and Applications*, *80*, 30399-30415.

[12] Chen, H., Liu, J., Wen, Q. M., Zuo, Z. Q., Liu, J. S., Feng, J., ... & Xiao, D. (2021). CytoBrain: cervical cancer screening system based on deep learning technology. *Journal of Computer Science and Technology*, *36*, 347-360.

[13] Sun, G., Li, S., Cao, Y., & Lang, F. (2017). Cervical cancer diagnosis based on random forest. *International Journal of Performability Engineering*, *13*(4), 446.

[14] Wu, W., & Zhou, H. (2017). Data-driven diagnosis of cervical cancer with support vector machine-based approaches. *IEEE Access*, *5*, 25189-25195.

[15] Vidya, R., & Nasira, G. M. (2016). Prediction of cervical cancer using hybrid induction technique: A solution for human hereditary disease patterns. *Indian Journal of Science and Technology*, *9*(30), 1-10.

[16] Mitra, P., Mitra, S., & Pal, S. K. (2000). Staging of cervical cancer with soft computing. *IEEE Transactions on Biomedical Engineering*, *47*(7), 934-940.

[17] Khamparia, A., Gupta, D., Rodrigues, J. J., & de Albuquerque, V. H. C. (2021). DCAVN: Cervical cancer prediction and classification using deep convolutional and variational autoencoder network. *Multimedia Tools and Applications*, *80*, 30399-30415.

[18] Chen, H., Liu, J., Wen, Q. M., Zuo, Z. Q., Liu, J. S., Feng, J., ... & Xiao, D. (2021). CytoBrain: cervical cancer screening system based on deep learning technology. *Journal of Computer Science and Technology*, *36*, 347-360.

[19] Chirenje, Z. M., Rusakaniko, S., Kirumbi, L., Ngwalle, E. W., Makuta-Tlebere, P., Kaggwa, S., ... & Makoae, L. (2001). Situation analysis for cervical cancer diagnosis and treatment in east, central and southern African countries. *Bulletin of the World Health organization*, *79*(2), 127-132.

[20] Sarwar, A., Ali, M., Suri, J., & Sharma, V. (2015, March). Performance evaluation of machine learning techniques for screening of cervical cancer. In *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 880-886). IEEE.

[21] Lee, C. K. H., Tse, Y. K., Ho, G. T. S., & Chung, S. H. (2021). Uncovering insights from healthcare archives to improve operations: an association analysis for cervical cancer screening. *Technological Forecasting and Social Change*, *162*, 120375.

[22] Punjani, D. N., & Atkotiya, K. H. (2017). Cervical cancer prediction using data mining. *Int J for Res Appl Sci Eng Tech*, *5*, 1856.**[23]** Sun, G., Li, S., Cao, Y., & Lang, F. (2017). Cervical cancer diagnosis based on random forest. *International Journal of Performability Engineering*, *13*(4), 446.

[24] Wu, W., & Zhou, H. (2017). Data-driven diagnosis of cervical cancer with support vector machine-based approaches. *IEEE Access*, *5*, 25189-25195.

[25] Geeitha, S., & Thangamani, M. (2021). Integrating HSICBFO and FWSMOTE algorithm-prediction through risk factors in cervical cancer. *Journal of Ambient Intelligence and Humanized Computing*, *12*, 3213-3225.

[26] Sun, G., Li, S., Cao, Y., & Lang, F. (2017). Cervical cancer diagnosis based on random forest. *International Journal of Performability Engineering*, *13*(4), 446.

[27] Punjani, D. N., & Atkotiya, K. H. (2017). Cervical cancer prediction using data mining. *Int J for Res Appl Sci Eng Tech*, *5*, 1856.

[28] Kumar, R., Srivastava, R., & Srivastava, S. (2015). Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features. *Journal of medical engineering*, *2015*.

[29] Bora, K., Chowdhury, M., Mahanta, L. B., Kundu, M. K., & Das, A. K. (2017). Automated classification of Pap smear images to detect cervical dysplasia. *Computer methods and programs in biomedicine*, *138*, 31-47.

[30] Jha, M., Gupta, R., & Saxena, R. (2021, November). Cervical cancer risk prediction using XGboost classifier. In *2021 7th International Conference on Signal Processing and Communication (ICSC)* (pp. 133-136). IEEE.