International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 11 (2019) pp. 2570-2577

Prediction of Cervical Cancer Using Machine Learning Techniques

Jaswinder Singh[1], Sandeep Sharma[2]

[1]*Department of Computer Science, Guru Nanak Dev University, Amritsar, Punjab, India.*

[2]*Department of Computer Science & Engineering, Guru Nanak Dev University, Amritsar, Punjab, India*

Abstract

Cervical cancer is one of the most deadly diseases in the world among women. It is caused by long term infection in skin cells and mucous membrane cells of the genital area. The worrisome aspect of this cancer lies in the fact that it does not show any symptoms when it starts occurring. This paper proposes a prediction model for diagnosing the correct stage of infection. The proposed model utilizes UCI data repository and machine learning classifiers for prediction. The data from sensors undergoes pre-processing where the feature extraction and validations updates the repository. Our proposed model considers ten features (refer **Figure 5**) of cervical cancer relating to four stages (refer **Table 1**). The pre-processed data is then made available to physician for verification followed by training of machine learning classifiers. We have deployed six classifiers for this task. The outcome of decision tree classifier confirms the appropriate stage prediction in terms of false-positive rate, f-measure, and precision.

Keywords: Cervical Cancer, Machine Learning, Sensors, Cancer Stages, Prediction, Infection, Comparative Analysis.

1. INTRODUCTION

Cervical cancer has been a major cause for death worldwide from the last few decades. It is a third main type of cancer after the lungs and breast cancer among women. It is only cured when detected and treated at its early stage. The stage is a description of cancer's growth and spread in the body of a woman. Physicians find out the stage of cancer after diagnostic test. By knowing the correct stage of cervical cancer physician can follow a right kind of treatment. There are four stages of a cervical cancer developed/Suggested by IFOG (International Federation of Obstetrics and Gynecology). Doctors evaluate the tumor and its spread to lymph nodes or to other parts of the body and then assign the stage. The following are the stages along with their description given in **Table 1**.

The studies show that there are six basic symptoms of cervical cancer, viz., abnormal bleeding between or after menopause, abnormal vaginal discharge, tiredness and weight loss, pain in the pelvis and abdomen and irritation while urinating. There are two main causes behind the infection of cervical cancer, viz., long term infection and infection through human papillomavirus (HPV).

Table 1. Stages by IFOG

STAGE (STG)	DESCRIPTION
STG-I	Cancer cells are starting growing and reach till tissues of cervix lining.
STG-IA	Detected under microscopy. No Lymph nodes are affected by cervical cells yet.
STG-IA1	Growth of cervical cells till 3mm of depth and 7mm of length in cervix lining.
STG-IA2	Cancerous region grown till 5mm in depth and less than 7mm in length.
STG-IB	Tumor larger than 1A2 and lesion in cervix is seen.
STG-IB1	Tumor spread ≤ 4cm
STG-IB2	Tumor larger than 4cm but no lymph nodes are affected yet.
STG-II	Cancer cells occupy nearby areas of the vagina and cervix, but still inside the pelvic region.
STG-IIA	Spread till tissues next to cervix.
STG-IIA1	Tumor spread ≤ 4cm
STG-IIA2	Tumor spread > 4cm
STG-IIB	Tumor/Cist spread to parametrical region.
STG-III	Spread till pelvic wall causes kidney swelling.
STG-IIIA	Affect lower third wall of the vagina.
STG-IIIB	Grown to pelvic wall and causes kidney malfunctioning.
STG-IVA	Spread to bladder or rectum and reach till Lymph nodes.
STG-IVB	Spread to other parts of the body and affecting their respective functioning.

The development of cancer cells and tumors is due to the presence of abnormal cells at the wall of the cervix. HPV affects skin cells and mucous cells, leading to their abnormal behavior which may lead to pre cancer and gradually twin into cervical cancer stage-I. Certain groups of women like women who smoke who have a weak immune system, who are infected with HIV and/or who have undergone some organ transplant are at higher risk than others to get infected with cervical cancer. According to NCBI around 96000 women get infected with cervical cancer per year in India and around 40% of them die of it. Most of these women are aged between 30 and 69. Gynecologists after physical examination or pap tests smear results; feel that the chances of cervical cancer are higher when swelling around womb, inside vagina. They examine the change in the tissues through a colposcopy (kind of magnifying glass). Depending upon the severity of infection observed in the physical examination other examination methods, viz., ultrasound, MRI (magnetic resonance imaging), X-ray and CT (computed tomography) or needed to examine the infection at deeper layers. There are many submissions to examine the cervical cancer and its stages through sensor based techniques involving wireless sensor networks and machine learning. The following are few recent works submitted in the same direction

2. RELATED WORK

One of the recent works of the healthcare monitoring system through wireless sensor networks has been done by (Gogate U. and Bakel. J., 2018). They have deployed bio sensors to measure temperature, heart rate and oxygen level of the body on Arduino Nano broad. The patient's data obtained from these sensors are made available to an IOT application named Think-speak for processing. After processing the data several times, the predictions can generate alerts for the physician to manage their patients. They have tested their proposed method on the cardiac patients and found it 95% accurate in the prediction of disease causes and heart-functioning-disorders [3].

Another submission by (Rahmani A. M. et. al., 2015) introduces the smart e health gateway for ubiquitous health monitoring system using IOT. Their idea of developing smart-homes and smart-hospitals gains popularity in present time under the realm of the internet explosion. They have shown that the deployment of smart gateways lead e-health can lift up several issues of health care systems, viz., reliability, scalability, energy efficiency, interoperability and accuracy [13].

Hassanalieragh M. et al., 2015 highlight some issues and challenge for IOT in healthcare systems. These challenges arise when heavily networked sensors gathered patients information, these sensors are either worn by patients on his/her body or they can be deployed in the vicinity around him/her. Continues data retrieval, its aggregation and its mining afterwards may lead to false implications. Health monitoring, prognosis, prevention and cure must be dealt with proper care by e-health monitoring system. The modalities like heterogeneity of gathered data necessitates effective customization for both patient-centric as well as disease-centric [4].

Alyass A. et al., 2015 raised challenges along with their opportunities to mark the importance of automated systems for personalized medicines. One of the biggest bottleneck they observed was a representation of omics data. The data associated with individuals DNA (genomics), transcribed RNA (transcriptomics), epigenomics, metabolomics, and proteomics etc. The computational quantification of these omics data is the major bottlenecks in the mining and prediction phases of a smart health monitoring system based on IOT [1]. Fan Y. J. et al., 2014 has coined the concept of a smart rehabilitation system based IOT. Their proposed model generates automated rehabilitation strategy based on both disease ontology and resource ontology. The availability of medical resources is analyzed by the proposed hierarchical system and human machine interaction helped strategy optimization [5]. Hussein A. S. et al., 2012 proposed a recommender system for diagnosis of chronic diseases they have exploited data mining techniques viz., random forest, J48 algorithm, decision trees and REP tree algorithms followed by 10 fold cross validations [6].

Janga S. C. and Edupuganti M. M. R, 2014 presented personalized medicine based on wireless networks and IOT. They have proposed a system for evaluation of omics data and integrated patient-physician interaction to evolve the appropriation in doses of drugs [12]. Sharma S. et al., 2016 implemented decision tree approach for prediction of cervical cancer. The same idea may be extended to furnish a smart health care system; the predictions probably guide system to take preventive or curative measures [2].

Recently, Gogate U. et al., 2018 implemented healthcare monitoring system for illness of cardiac patients [3]. On the cervical cancer, Snijders (P. J. F et al., 2006) reported detailed concepts and their clinical implications. These implications relate the degradation of patients to the cervical stages. A healthcare monitoring system could utilize the vicious impacts of these implication to provide a customized care to patients in a better way [11]. G. Jayalalitha et al., 2009 introduced another method for grading cervical cancer by exploiting the information related to intensity and texture of cancer cells [10]. Rahmadwati et al., 2010 gave a classification of cervical cancer based on evaluation of histology images. They have taken data from Indonesian hospital and perform feature classification on the basis of four typical features along with their three categories [9].

Another aid to do work on cervical cancer was carried forward by (Allwin S. et al., 2010) where; they have deployed a proposed modal on decision support system to classify the pathology for the cervical cancer. On the same lines, a healthcare monitoring system can be constructed to claim the IOT based handling of patients [8].

3. PROPOSED A FRAMEWORK

The architecture for an IOT based health care system used for detection and prevention of cervical cancer is given in **Figure 1.** The cancer related information is taken from body worn or deployed sensors. This information includes the result from the scan of the cancer affected area and blood cells. The features include color, size, shape and texture of cancer cells. Also a high resolution scan image of the cancer affected part is sent to the data repository. Here, our proposed model includes predefined contextual information to be compared by the physician through smart interface feature extraction module is implemented in Weka and python 3.6 for extracting concerned features and also the validation being carried out manually by the physician. The six machine learning algorithms viz., Naive Bayes, Function based SMO, Lazy Learner, Meta-heuristic, Rule based Decision Trees and

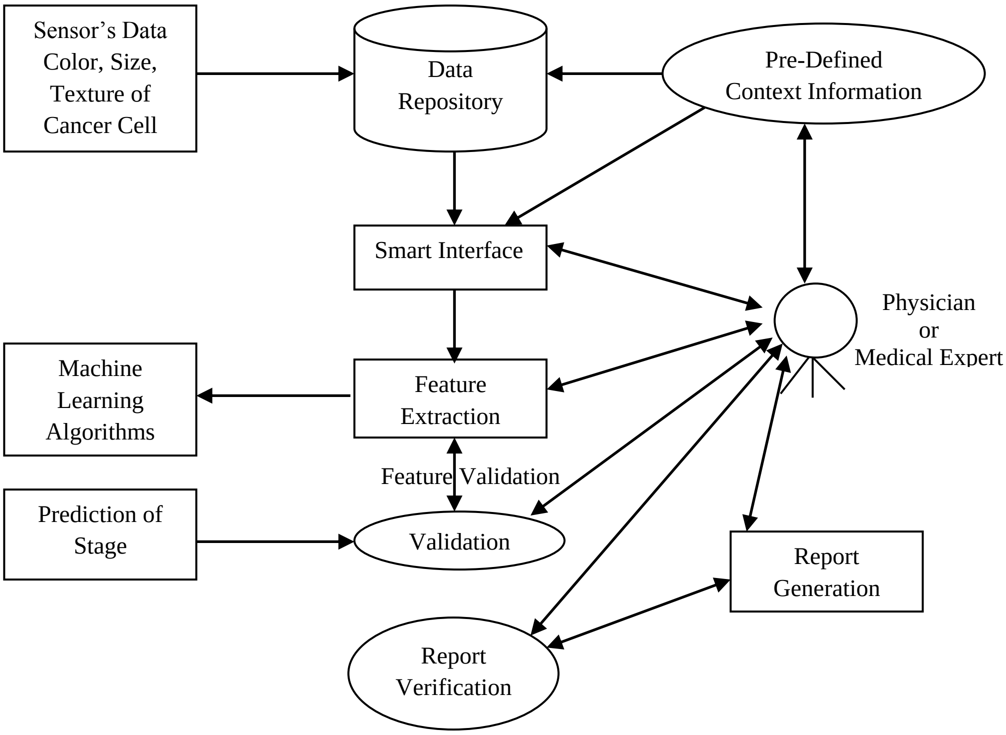

Figure 1. Proposed Framework

Logistic Regression are trained on acquired data set. The data set is manually annotated by physicians and researchers, according to the symptoms and observed values of the features. The prediction of stage of cervical cancer is followed by the ten folds cross validation before the final report generation and report validation.

4. PROPOSED METHODOLOGY

The proposed methodology of our model as shown in **Figure 2,** the very first step includes initiation of scan for infected person using sensors. The sensors are applied to the finger tips and feet of patient to retrieve blood pressure, blood sugar level heart rate, and respiratory rate. Also scan below abdomen till thighs will help the proposed system to scan for cervical cancer. The second step requests the data respiratory for pre-defined contextual information to be validated against the collected results from the first step. The collected results are kept in a test set, whereas the training set is built from predefined dataset from respiratory. The following third step trains classification models using a training set followed by a prediction of cancer stage using test set. The threshold values predicted by machine learning classifiers validate the test set. The health care provider module validates them before report generation. This round provides tenfold cross validation and corrects stage prediction following which reports generation module is guided.

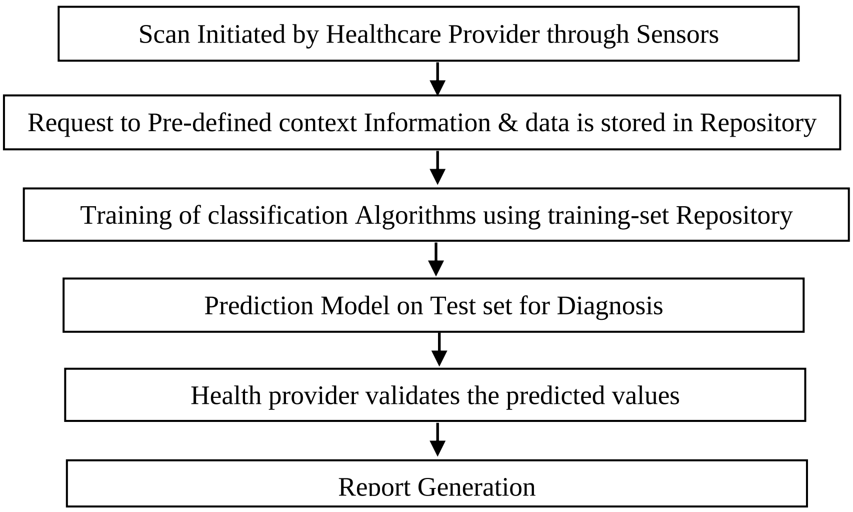

Figure 2. Proposed Methodology

5. PROPOSED SYSTEM FLOWCHART

We have designed a framework for healthcare monitoring system for cervical cancer affected patients and implemented our proposed method on Weka and Python 3.6. Our proposed method measures significant parameters for patient using different sensors installed in the vicinity of the patient (Figure 3.). The parameters include temperature, blood pressure, heart rate, respiratory rate; ECG x-rays scan, pulse rate, abdomen scan, and high resolution imaging for HPV. The data from sensors is in analog form an inbuilt modem on Arduino board will convert it into digital form. The collected digital data is shifted to the data repository using wireless communication. We have deployed Wi-Fi relay module ESP8266 for sending sensor data to the repository at the server using wireless communication. We have used IOT API "REST" generated from dream factory software, which is an open source IOT application development solution.

The data from server is made available to physicians through REST channels. In case of any emergency situation, authorized users will get alerts through this API.

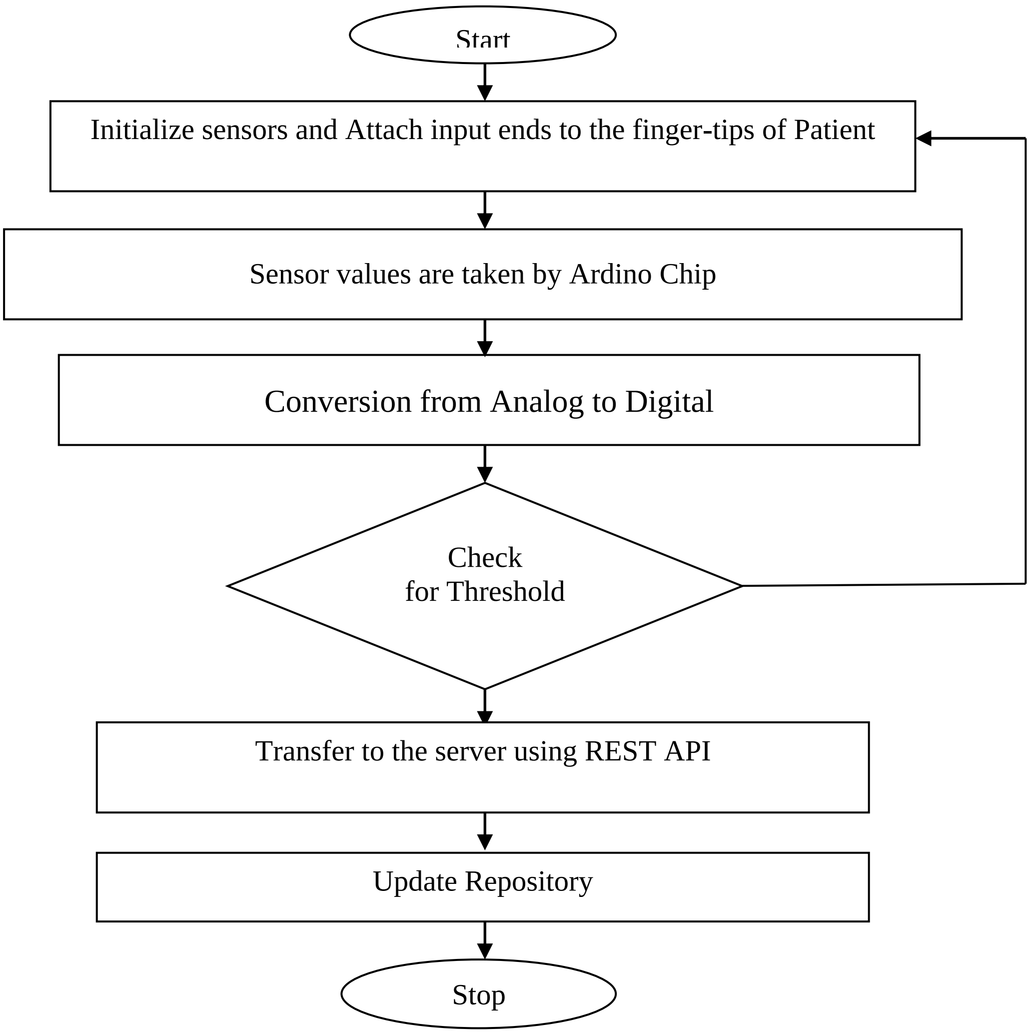

Figure 3. Proposed system Flowchart

6. RESULTS AND DISCUSSIONS

The results obtained during experimentation task from the different classifiers are expressed in this section. The detailed results along with their significance are mentioned below.

Different ML classifiers such as Naïve-Bayes (M1), Functions-based-Logistic-SMO (M2), Lazy-based-LWL (M3), Meta-based-Iterative-Classifier-Optimizer (M4), Rule-based –Decision-Table (M5), and Trees-based-Decision-Stump (M6)graphical have been deployed on the data set to predicting the stages of the cervical cancer. The comparative analysis of these ML techniques for the prediction of stage for cervical cancer is done on the basis of different analysis parameters such as True positive rate, False Positive rate, F-measure, MCC (Matthews-Correlation-Coefficient), Precision. The individual tableau for these metrics is shown in the respective order from the **Table 2** to Table 6.

The **Table 2** expresses the comparative analysis of aforementioned ML techniques based on the true positive rate and the results of the "tree based Decision stump" ML classifier are leading among the others in every epoch and the maximum positiveness of 0.770 is shown when compared to other ML classifiers. This indicates that the diagnosis of infected women in Stage-2b is the starting point for any critical damage of body organs.

Table 2. Comparative Analysis of ML Techniques for Cervical cancer stage prediction based on True Positive Rate

Cervical Cancer Stages	M1	M2	M3	M4	M5	**M6**
STG-1a	0	0	0	0	0	**0**
STG-1b	0.174	0.362	0.449	0.362	0.333	**0.493**
STG-2a	0.353	0	0	0	0	**0**
STG-2b	0.488	0.65	0.575	0.675	0.488	**0.770**
STG-3a	0	0	0	0	0	**0**
STG-3b	0.372	0.186	0.116	0.256	0.163	**0.100**
STG-4a	0	0	0	0	0	**0**
STG-4b	0	0	0	0	0	**0**

The results, based on False positive rate for the prediction of cervical cancer stages using above mentioned ML techniques are shown in **Table 3** and the obtained results show that the False Positive rate for "Tree Based Decision Stump (M6)" is least among the other ML techniques. Hence the proposed model utilizes this classifier for the training of classification model before predicting the stage of cervical cancer.

Table 3. Comparative Analysis of ML Techniques for Cervical cancer stage prediction based on False Positive Rate

Cervical Cancer Stages	M1	M2	M3	M4	M5	M6
STG-1a	0	0	0	0	0.021	0
STG-1b	0.101	0.315	0.321	0.256	0.369	0.198
STG-2a	0.227	0	0	0.005	0.015	0
STG-2b	0.363	0.573	0.548	0.541	0.51	0.318
STG-3a	0.004	0	0	0	0	0
STG-3b	0.191	0.046	0.077	0.093	0.093	0
STG-4a	0	0	0	0	0	0
STG-4b	0	0	0	0	0	0

Based on the F-measure, the results are tabled in **Table 4** that again shows that "Decision Stump(M6)" classifier is running ahead among the six classifiers for the prediction of stage for cervical cancer. The higher value of F-measure for the stage-2b reflects that majority of cases fall under the infection where the Tumor/Cist spread to parametrical regions of the vagina.

Table 4. Comparative Analysis of ML Techniques for Cervical cancer stage prediction based on F-Measure

Cervical Cancer Stages	M1	M2	M3	M4	M5	M6
STG-1a	0	0	0	0	0	0
STG-1b	0.245	0.34	0.403	0.365	0.299	0.444
STG-2a	0.261	0	0	0	0	0
STG-2b	0.443	0.468	0.434	0.493	0.392	0.499
STG-3a	0	0	0	0	0	0
STG-3b	0.333	0.267	0.159	0.306	0.206	0
STG-4a	0	0	0	0	0	0
STG-4b	0	0	0	0	0	0

The comparative analysis of different ML Techniques for cervical-cancer-stage-prediction based on the MCC (**Table 5**) and Precision (**Table 6**) shows that the "Decision Stump (M6)" is again the best classifier and performing its prediction task in a better way as compare to other ML classifier by achieving the maximum MCC as 0.185 and maximum precision as 0.405. This precision in stage-1b reflects that women at high risk of cervical cancer have cancerous region grown till 5mm in depth and less than 7mm in length (refer **Table 1**).

Table 5. Comparative Analysis of ML Techniques for Cervical cancer stage prediction based on MCC

Cervical Cancer Stages	M1	M2	M3	M4	M5	M6
STG-1a	0	0	0	0	-0.01	0
STG-1b	0.101	0.045	0.121	0.107	-0.034	0.185
STG-2a	0.103	0	0	-0.027	-0.046	0
STG-2b	0.12	0.074	0.026	0.128	-0.021	0.081
STG-3a	-0.013	0	0	0	0	0
STG-3b	0.168	0.209	0.054	0.192	0.088	0
STG-4a	0	0	0	0	0	0
STG-4b	0	0	0	0	0	0

Table 6. Comparative Analysis of ML Techniques for Cervical cancer stage prediction based on Precision

Cervical Cancer Stages	M1	M2	M3	M4	M5	M6
STG-1a	0	0	0	0	0	0
STG-1b	0.414	0.321	0.365	0.368	0.271	0.405
STG-2a	0.207	0	0	0	0	0
STG-2b	0.406	0.366	0.348	0.388	0.328	0.396
STG-3a	0	0	0	0	0	0
STG-3b	0.302	0.471	0.25	0.379	0.28	0
STG-4a	0	0	0	0	0	0
STG-4b	0	0	0	0	0	0

The graphical representation of different ML classifiers from M1-M6, which has been deployed on the dataset to predict the stages of the cervical cancer based of different analysis parameters such as True positive rate, False positive rate, F-measure, MCC (Matthews-correlation-coefficient), Precision is shown in **Figure 4.**

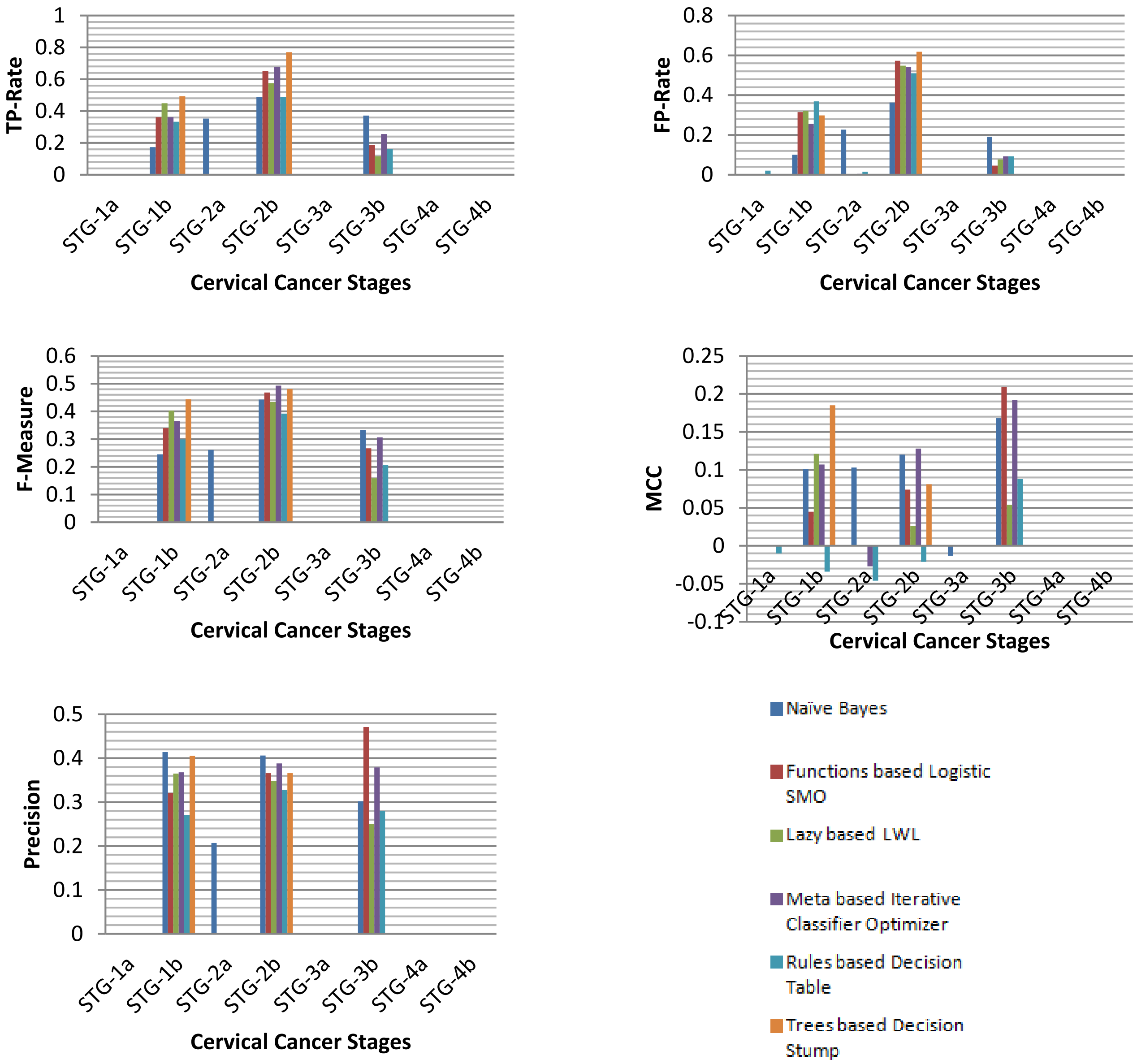

Figure 4. Comparative Analysis of different ML Techniques

The Pearson correlation between various attributes such as ClinDiameter, UterineBody, Relpelvic, RelAbdo, RelDistant, NodePET, Status, MRIVol, RelPrimary, RelSupraclav has been obtained and depicted in the **Figure 5** and elaborated as RelDistant, RelAbdo, Status, RelPrimary & Histology playing very important role in the prediction.

Attribute Correlation

	NodePET	ClinDiameter	MRIVol	UterineBody	Status	RelPrimary	Relpelvic	RelAbdo	RelSupraclav	RelDistant
NodePET	1	0.25	0.214	0.313	0.303	0.04	0.19	0.311	0.252	0.229
ClinDiameter	0.25	1	**0.615**	0.329	0.201	0.137	0.204	0.087	0.04	0.143
MRIVol	0.214	0.615	1	0.311	0.237	0.228	0.258	0.13	0.003	0.128
UterineBody	0.313	0.329	0.311	1	0.187	0.073	0.167	0.186	0.127	0.152
Status	0.303	0.201	0.237	0.187	1	0.563	0.629	**0.702**	0.357	**0.684**
RelPrimary	0.04	0.137	0.228	0.073	0.563	1	0.704	0.326	-0.008	0.219
Relpelvic	0.19	0.204	0.258	0.167	**0.629**	**0.704**	1	0.538	0.202	0.332
RelAbdo	0.311	0.087	0.13	0.186	0.702	0.326	0.538	1	0.338	0.478
RelSupraclav	0.252	0.04	0.003	0.127	0.357	-0.008	0.202	0.338	1	0.305
RelDistant	0.229	0.143	0.128	0.152	0.684	0.219	0.332	0.478	0.305	1

Figure 5. Attribute Correlation

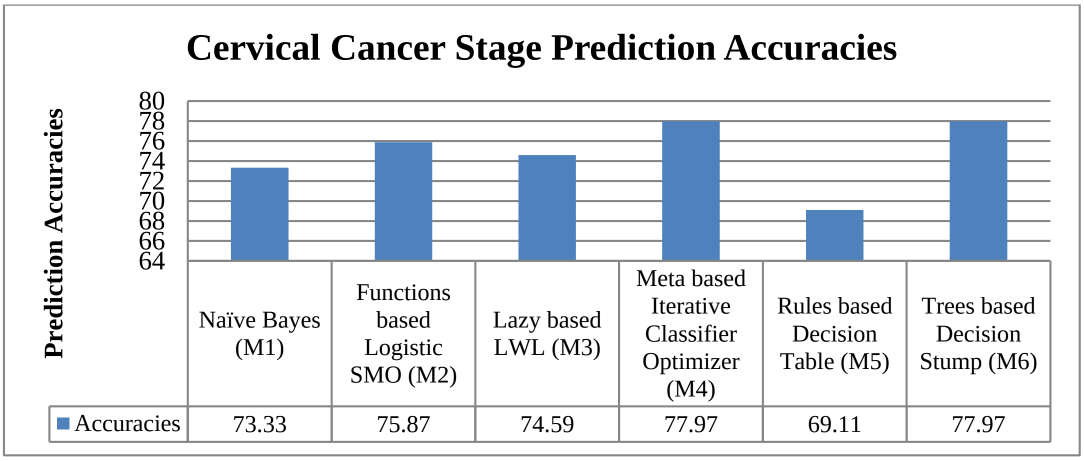

Figure 6. Accuracies Comparison for Cervical Cancer Stage Prediction

The **Figure 6** depicts the comparative analysis of accuracies achieved during the experimental task of stage prediction of cervical cancer using various already said ML techniques and the results shows that the Meta based Iterative classifier optimizer and the tree based decision stump (M6) machine learning classifier achieved the maximum prediction accuracy such as 77.97%, which is highest among the other ML classifier. This performance shows the capability of Tree and Meta based ML classifiers in the area of prediction as well as classification.

Another comparative performance of cervical cancer stage prediction is shown in the **Table 7,** which shows the confusion matrix for tree based Decision stump classifier stage wise.

Table 7. Confusion Matrix for Decision Tree based Classifier
Confusion Matrix for Decision Tree based Classifier

	a	b	c	d	e	F	g	h		Classification
A	0	1	0	0	0	0	0	0	|	a = STG-1a
B	0	54	0	15	0	0	0	0	|	b = STG-1b
C	0	5	27	2	0	0	0	0	|	c = STG-2a
D	0	19	0	61	0	0	0	0	|	d = STG-2b
E	0	1	0	1	7	0	0	0	|	e = STG-3a
F	0	4	0	0	0	39	0	0	|	f = STG-3b
G	0	0	0	1	0	0	0	0	|	g = STG-4a
H	0	0	0	0	0	0	0	0	|	h = STG-4b

7. CONCLUSION

The proposed model for prediction of stage for cervical cancer deploys wireless sensors for data collection from the patient. The collected analog data updates the data repository and then converted to its digital form using Ardino chip. Six machine learning techniques are trained with the updated repository. The five metrics, viz., true-positive-rate, false-positive-rate, f-measure, mcc, and precision are evaluated for these classifiers. The prediction of stage for cervical cancer is done on test-set. The training accuracies are compared and validated against six classifiers, where the decision-tree classifier leads in performance. The predictions confirm that the stage-2b is occurring more than 1b, 3a and 3b. Our proposed model distinctly predicts the correct stage for cervical cancer. Nonetheless, the proposed model marks significant performance in terms of accuracy of prediction, but the future work will optimize the performance by considering other features related to social, cultural, and eating habits of a patient.

REFERENCES

[1] M. T. a. D. M. Akram Alyass, "From big data analysis to personalized medicine for all: challenges and opportunities," *BMC Medical Genomics,* vol. 8, no. 33, 2015.

[2] S. G. Sunny Sharma, "Decision Tree approach in Machine Learning for Prediction of Cervical Cancer Stages using WEKA," *International Journal of Recent Trends in Engineering & Research,* vol. 2, no. 8, 2016.

[3] U. Gogate and J. Bakal, "Healthcare Monitoring System Based on Wireless Sensor Network for Cardiac Patients," *Biomedical and Pharmacology Journal,* vol. 11, no. 3, p. September, 2018.

[4] M. Hassanalieragh, A. Page, T. Soyata, G. Sharma, M. Aktas, G. Mateos, B. Kantarci and S. Andreescu, "Health Monitoring and Management Using Internet-of-Things (IoT) Sensing with Cloud-Based Processing: Opportunities and Challenges," in *2015 IEEE International Conference on Services Computing,* New York, NY, USA, 2015.

[5] Y. J. Fan, Y. H. Yin, L. D. Xu, Y. Zeng and F. Wu, "IoT-Based Smart Rehabilitation System," *EEE Transactions on Industrial Informatics,* vol. 10, no. 2, pp. 1568-1577, 2014.

[6] W. M. O. X. L. Asmaa S. Hussein, "Accurate and Reliable Recommender System for Chronic Disease Diagnosis," in *The First International Conference on Global Health Challenges,* 2012.

[7] InformedHealth.org, "Cervical cancer: Overview," 14 Dec 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK279259/. [Accessed Jan 2019].

[8] S. Allwin, S. P. K. Kenny and V. Manian, "Classification of stages of maligancies using textron signatures of a cervical cyto image," in *2010 IEEE International Conference on Computational Intelligence and Computing Research*, Coimbatore, India, 28-29 Dec. 2010.

[9] Rahmadwati, G. Naghdy, M. Ross, C. Todd and E. Norachmawati, "Classification Cervical Cancer Using Histology Images," in *2010 Second International Conference on Computer Engineering and Applications*, Bali Island, Indonesia, 12 April 2010.

[10] G. Jayalalitha and R. Uthayakumar, "Recognition of Cervical Cancer Based on Fractal Dimension," in *2009 International Conference on Advances in Recent Technologies in Communication and Computing*, Kottayam, Kerala, India, 17 November 2009.

[11] P. J. Snijders, R. D. Steenbergen, D. A. Heideman and C. J. Meijer, "HPV-mediated cervical carcinogenesis: concepts and clinical implications," *J Pathol,* vol. 208, no. 2, pp. 152-164, Jan, 2016.

[12] S. C. Janga and M. M. R. Edupuganti, "Systems and Network-Based Approaches for Personalized Medicine," *Current Synthetic and Systems Biology,* vol. 2, no. 3, p. 1000e109, 2014.

[13] N. K. T. T. N. G. J. G. B. N. P. L. H. T. Amir-Mohammad Rahmani, "Smart e-Health Gateway: Bringing intelligence to Internet-of-Things based ubiquitous healthcare systems," in *12th Annual IEEE Consumer Communications and Networking Conference (CCNC),* 2015.

[14] A. S. o. C. O. (ASCO), "Cervical Cancer: Stages," Cancer.Net Editorial Board, July 2017. [Online]. Available: https://www.cancer.net/cancer-types/cervical-cancer/stages. [Accessed 01 Jan 2019].

[15] C. Balleyguier, E. Sala, D. C. T, A. Bergman, B. Brkljacic, F. Danza, R. Forstner, B. Hamm, R. Kubik-Huch, C. Lopez, R. Manfredi, J. McHugo, L. Oleaga, K. Togashi and K. Kinkel, "Staging of uterine cervical cancer with MRI: guidelines of the European Society of Urogenital Radiology," *Eur Radiol.,* vol. 21, no. 5, pp. 1102-1110, 2011.