Women Healthcare Mobile App-An Approach to Predict Early Stage of Cervical Cancer

Naveen R Chanukotimath[1], Asha K[2,*], Keerthi Prasad G[3], Manjula G M[4]

[1,2,3] Dept. of IS & E, GMIT, Davangere, Karnataka, India.
[4] Dept. of Mathematics, GMIT, Davangere, Karnataka,
*Corresponding author. Email: ashak@gmit.ac.in

ABSTRACT

Most of the woman nowadays is ending up their life at middle age between 35-50 years, reason they are suffering from Cervical related cancer tumours. Many women are unaware of having cervical related cyst in the early stages. A survey was conducted on classifiers such as Decision Tree, Multilayer Perceptron, and Nave Bayes, with True Positive Rate, False Positive Rate, Precision, and Recall being measured, and an Android Mobile App was built to forecast the risk of having a Cervical associated cyst in its early stages. For training and testing the classifiers, we have used cervical dataset from University of California at Irvine Machine Learning Repository. Cervical Dataset consists of 858 records containing 32 attribute values and 4 diagnosis class value. For Cervical cancer prediction only 21 attributes and 1 biopsy class value is considered. The proposed Android Mobile App capable of predicting risk of a woman is affected by cervical related cyst in early stages. A normal woman can identify chances of having Cervical Cancer at finger tips with the proposed Android Mobile App.

Keywords: Android Mobile App, Cervical, Decision Tree, Evaluation, Multilayer Perceptron, Naïve Bayes.

1. INTRODUCTION

The majority of women in the globe are affected by cervical cancer. Cervical cancer grows in the cervix's cells, the tube that connects the uterus and the vaginal canal. Cervical cancer is most common type of cancer, which is caused by the Human Papilloma Virus (HPV). When the immune system is exposed to HPV, it normally fights it off. However, in a small number of people, the virus can persist for years, contributing to the transformation of some cervical cells into malignant cells. Because early-stage cervical cancer frequently has no signs or symptoms, medical practitioners must be vigilant in identifying and diagnosing the disease. Cervical cancer symptoms do not appear till the later stages of the disease.

The suggested approach will be able to predict whether a person will get cervical cancer. For classification, the suggested approach uses the Decision Tree C4.5 (J48), Multilayer Perceptron, and Nave Bayes classifiers. The training and testing datasets are taken from the regular UC Irvine Machine Learning repository. The attributes are listed in alphabetical order, separated by commas. Question marks "?" are used to denote attribute values that are unknown. "Dx (Diagnosis):

Cancer, Dx: CIN (Cervical Intraepithelial Neoplasia), Dx: HPV, Dx: Biopsy" is the diagnosis class. The Dataset's Attribute Names and Possible Values are listed in Table 1.

Table 1. Dataset Details for Cervical Cancer

Attribute Name	Possible Values
age	13 to 84
Sexual partners in number	1 to 28
The first sexual encounter	10 to 32
The total number of pregnancies	1 to 11
Smokes	0 to 1
Smokes (years)	0 to 37
Smokes (packs/year)	1 to 37
Hormonal Contraceptives	0 to 1
Hormonal Contraceptives (years)	1 to 30
Intrauterine Device	0 to 1
Intrauterine Device (years)	0 to 19
Sexually Transmitted Diseases	0 to 1
Sexually Transmitted Diseases (number)	0 to 4
Sexually Transmitted Diseases :condylomatosis	0 to 1
Sexually Transmitted Diseases: cervical condylomatosis	0

Sexually Transmitted Diseases: vaginal condylomatosis	0 to 1
Sexually Transmitted Diseases: Perineal condylomatosis (vulvo)	0 to 1
Sexually Transmitted Diseases: syphilis	0 to 1
Sexually Transmitted Diseases: Inflammation of the pelvis	0 to 1
Sexually Transmitted Diseases: herpes genital	0 to 1
Sexually Transmitted Diseases: molluscum contagio	0 to 1
Sexually Transmitted Diseases: Acquired Immune Deficiency Syndrome	0
Sexually Transmitted Diseases: Human Immunodeficiency Virus infection	0 to 1
Sexually Transmitted Diseases: Hepatitis B	0 to 1
Sexually Transmitted Diseases: Human Papillpoma Virus	0 to 1
Sexually Transmitted Diseases: The total number of diagnoses	0 to 37
Sexually Transmitted Diseases: The amount of time since the initial diagnosis	1 to 22
Sexually Transmitted Diseases: Time since last diagnosis	1 to 22
Diagnosis: Cancer	0 to 1
Diagnosis: Cervical Intraepithelial Neoplasia	0 to 1
Diagnosis: Human Papillpoma Virus	0 to 1
Diagnosis	0 to 1
Hinselmann	0 to 1
Schiller	0 to 1
Cytology	0 to 1
Biopsy	0 to 1

2. RELATED WORKS

Kayalvizhi and Kanimozhi [1] They used machine learning methods like Logistic Regression, Random Forest Model, J48, and Support Vector Machine to perform the tests on the dataset. They found that accuracy is less for Logistic Regression, SVM and Decision Tree but they obtained good precision and recall for Decision Tree. They have obtained reasonable accuracy and recall rate for Random Forest Model with both original data and resampled data.

Asadi *et al.* [2] gathered data from 145 patients with 23 attributes, which were then analysed using Algorithms for machine learning classification such as Support Vector Machine, QUEST, C&R tree, MLP, and RBF are some of the acronyms used. Accuracy, sensitivity, precision, and area under the curve were among the parameters used to evaluate these methods. The precision, sensitivity, specificity, and AUC of Quest and

C&R trees were 95.55, 90.48, 100, and 95.20, 95.45, 90.00, 100, and 91.50, RBF 95.45, 90.00, 100, and 91.50, SVM 93.33, 90.48, 95.83, and 95.80, and MLP 90.90, 90.00, 91.67, and 91.50, respectively.

Chanukotimath and Sunil Kumar [3] has found the best algorithm Decision tree that yields accurate results. Proposed algorithm can be used in the future for predicting Thyroid Diseases at the earliest to take preventive measures.

Alsmariy *et al.* [4] proposed a classification model called SMOTE (The synthetic minority oversampling technique)-voting-PCA by combining three classifiers: Decision tree, logistic regression and random forest obtained accuracy of 97% for the prediction of cervical cancer according to certain evaluation measures.

Patra [5] the proposed RBF classifier has an accuracy of 81.25 percent and is considered to be an efficient classifier technique for lung cancer data prediction, according to the comparison technique.

Debelee *et al.* [6] The proposed deep learning methods were applied in three different ways: Transfer learning by freezing some layers of the deep learning network, and architecture change to reduce the number of parameters in the network. Furthermore, researchers from academic and medical institutions in industrialized countries have investigated the use of deep learning in imaging systems for the diagnosis of various cancer situations.

Singh and Sharma [7] they deployed six classifiers for predicting the stage of cervical cancer. By considering the false-positive rate, f-measure and precision, the decision tree classifier confirms the required stage prediction.

Parikh and Menon [8] Three models for cervical cancer data have been suggested, and their performance has been compared. All three versions are considered to be satisfactory. The k-nearest-neighbor model, on the other hand, has better accuracy, precision, recall, and AUC value.

Shetty and Shah [9] proposed neuro-fuzzy classifier to diagnose thyroid diseases, and said that other major diseases can be predicted by training the classifier with appropriate dataset and improves the performance of medical diagnosis system.

Kourou *et al.* [10] proposed a model for recognizing development and treating cancerous conditions.

3. METHODOLOGY

Figure 1 depicts the proposed the structure of our implemented system. The first step is to select insightful characters from the WEKA libraries, and then divide the datasets into two forms of training, training and validation. To construct classifier models, the training

and validation sets are combined to create classifier models. These classifier models are used to pick the test set patterns. The classifier models are then validated using various algorithms such as Random forest, Naive bayes, Decision tree, Logical regression, Logistic model trees, and Support vector machines, before being used to predict the outcome [11-15].
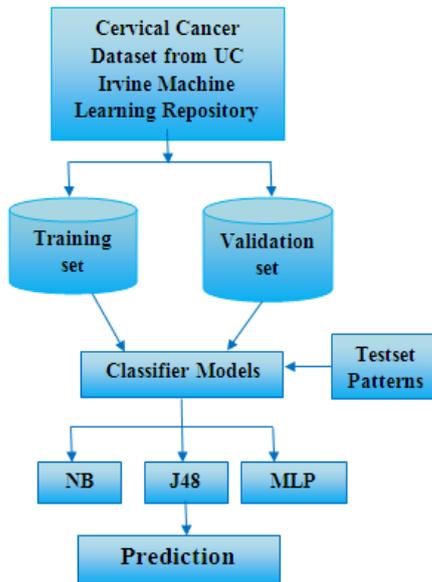

Figure 1 Architecture of the System

Figure 2 depicts the diagrammatical flowchart for the System that has been proposed. We begin with the dataset obtained from the UCI repository; since the data set contains various attribute values such as numeric, alphabetical characters, and some special symbols. The classifiers J48, NB, and MLP were unable to handle such data values, the dataset was numeric to nominal converted [16]. The application is designed in two ways,

1. Dataset Split Ratio: In this case the classifiers are tested with respect to the parameters including accuracy, precision, rates of true positives, true negatives, false positives, and false negatives with the split ratio of 0.6 (60 percent preparation and 40 percent for testing) 0.7 and 0.9.

2. Supplied Test Set: In this case the risk of having cervical cancer can be predicted by feeding the application with the symptoms or different attribute values of the user.

true positive rate, true negative rate, false positive rate and false negative rate
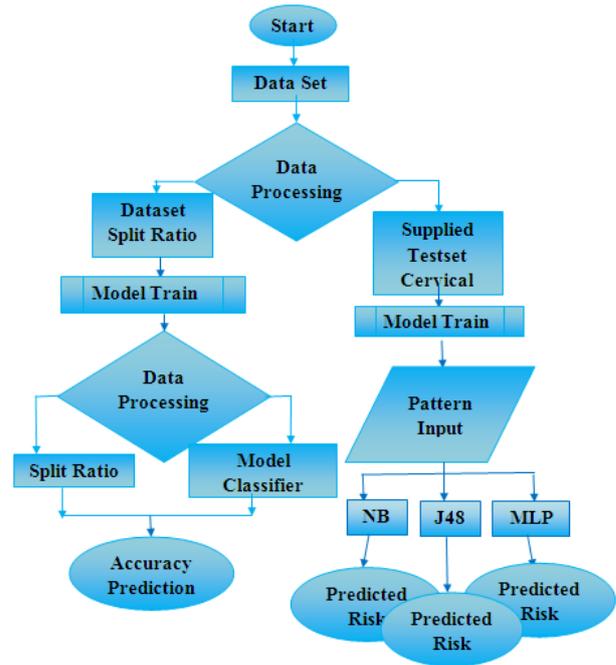

Figure 2 Flow Chart for the System

The data flow diagram for the Proposed System is shown in Figure 3, which depicts the data flow in the system. By providing the application with the user's symptoms or different attribute values, the risk of cervical cancer can be predicted [17-19].
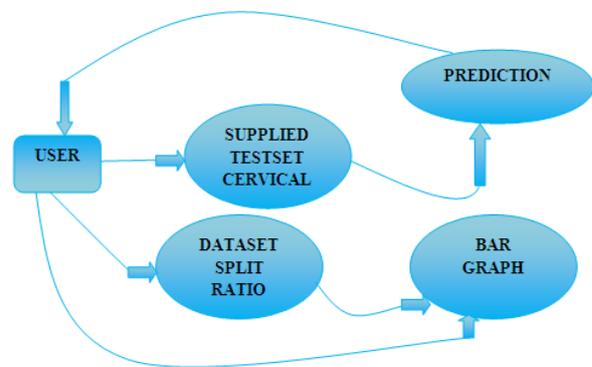

Figure 3 Dataflow Diagram of the System

4. IMPLEMENATION

The proposed prediction system is implemented as Mobile Application in Android [20]. To implement this system, we have used WEKA with Java in Android API JAR libraries for classifiers. In Graphical User Interface (GUI) there is an option to choose the dataset file, select classifiers Naïve Bayes, C4.5 (J48) Decision Tree and Multilayer Perceptron. Figure 4 shows the GUI of the system.
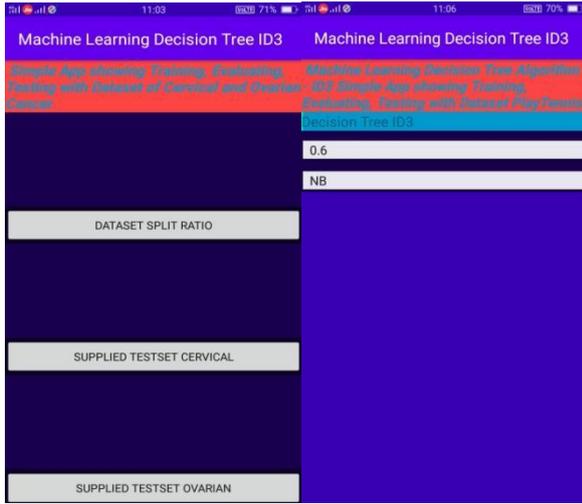
Figure 4 GUI of the Implemented System.

In this system we have provided feature to partition the Dataset into Training Set and Testing Set with Split Ratio like 0.6, 0.7, 0.8, 0.9 and selecting classifiers like NB, MLP and J48. Figure 5 shows GUI for the specifying Split Ratio and Classifier Selection [21-24].

To carry out experiments with the classifiers, the framework also has options for Training, Evaluation, Testing, and Analyze. Measured parameters can be shown in a variety of ways. Precision, TP Rate, FP Rate Results in the Table can be recalled, predictions about the Testing Set can be printed in the Text Region, and the results can be evaluated using a Bar Graph. Figure 6 GUI showing results of Accuracy for Naïve Bayes classifier, C4.5 (J48) Decision Tree classifier and Multilayer Perceptron classifier, and Figure 7 shows the evaluation summary for C4.5 (J48) Decision Tree classifier [25-29].
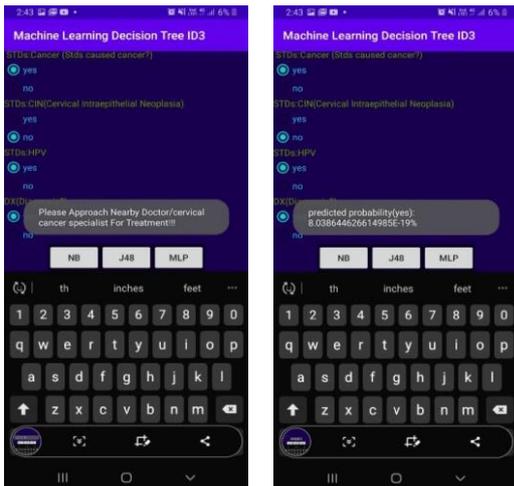
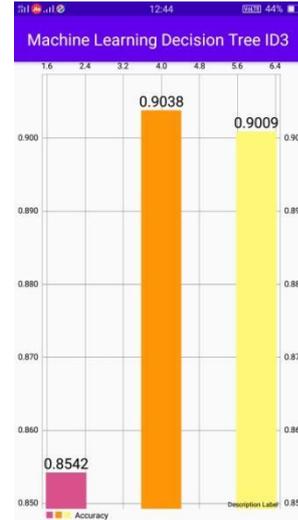

Figure 5 GUI for the specifying Split Ratio.

Figure 6 Accuracy of NB, J48 & MLP for Cervical Cancer
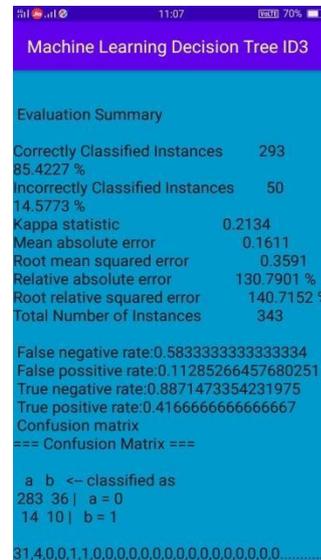

Figure 7 Evaluation summary for C4.5 (J48) Decision Tree classifier.

5. RESULTS AND ANALYSIS

The criteria Accuracy, Correctly Classified Instances, and Incorrectly Classified Instances have all been considered. The total number of instances is compared to these variables. The result shows that of C4.5 (J48) Decision Tree is 94.18% better than Naïve Bayes 84.88%, and Multilayer Perceptron classifiers 90.69% for split ration 0.9. For Split ratio 0.8, C4.5 (J48) Decision Tree is 92.44% better than Naïve Bayes 83.72%, and Multilayer Perceptron classifiers 91.27% and for spilt ratio of 0.6 again C4.5 (J48) Decision Tree is 90.38% better than Naïve Bayes 85.42%, and Multilayer Perceptron classifiers 90.19%.

Table 2 shows the results obtained for Accuracy of all three classifiers and Figure 8 shows the graph of

Accuracy obtained for Correct and Incorrect classification, against total number of instances.

Although accuracy is a widely used metric, it is ineffective when used to assess the performance of an unbalanced dataset. It's the proportion of the overall number of right predictions to the total number of predictions [3]. Equation (1) is used to calculate the Accuracy.

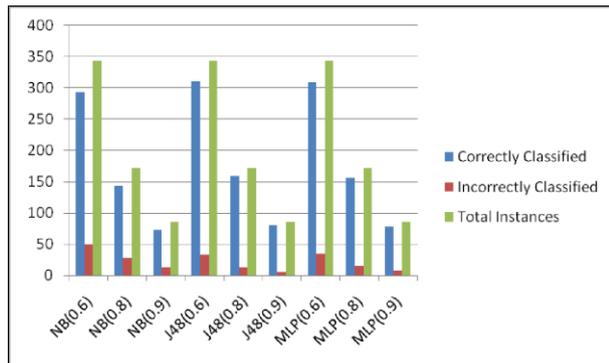$$Total\ Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Figure 8 Graph of Accuracy obtained for Correct and Incorrect classification, against total number of instances

Results obtained for FNR, FPR, TNR and TPR (Equations 2, 3, 4 and 5) are shown in Table 3. Figure 9 shows the graph of the results obtained for various parameters. Again the performance of C4.5 (J48) Decision Tree is better.

Confusion matrix: It is a table that gives information regarding both actual and expected classes and is used to measure performance for two or more classes. [4]

- True Positive (TP) relates to the number of records that are positive and are accurately classified.
- False Positive (FP) is the term used to describe the number of negative samples that were mistakenly recognized as positive records.
- True Negative (TN) relates to the number of records that are negative and accurately categorized.
- False Negative (FN) is a term that refers to the number of positive samples that were mistakenly classified as negative records.

$$TPR = TP/TP + TN \tag{2}$$
$$FPR = FP/FP + TN \tag{3}$$
$$TNR = TN/TN + FP \tag{4}$$
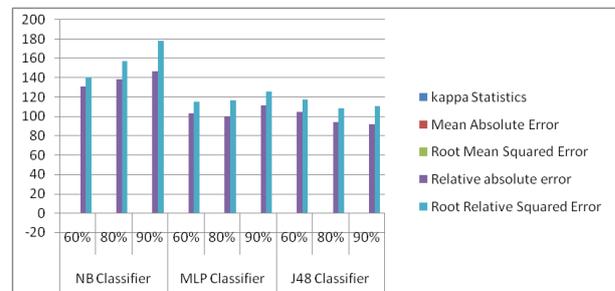$$FNR = FN/FN + TP \tag{5}$$

Figure 9 Graph obtained for various parameters shown in Table 3

Table 2. Results obtained for Accuracy by different classifiers

Classifiers	NB Classifier			MLP Classifier			J48 Classifier		
Split Ratio	60%	80%	90%	60%	80%	90%	60%	80%	90%
Correctly Classified Instances	293	144	73	309	157	78	310	159	81
Incorrectly Classified Instances	50	28	13	34	15	8	33	13	5
Total Number of Instances	343	172	86	343	172	86	343	172	86
Accuracy (%)	85.42	83.72	84.88	90.10	91.27	90.69	90.38	92.44	94.18

Table 3. Results obtained for various parameters

Classifiers	NB Classifier			MLP Classifier			J48 Classifier		
Split Ratio	60%	80%	90%	60%	80%	90%	60%	80%	90%
FNR	0.583	0.636	0.500	0.875	0.818	1.000	0.958	1.000	1.000
FPR	0.113	0.130	0.134	0.041	0.037	0.048	0.313	0.012	0.012
TNR	0.887	0.869	0.866	0.959	0.963	0.951	0.969	0.988	0.988
TPR	0.417	0.364	0.500	0.125	0.182	0.000	0.042	0	0

Table 4. Results obtained different classifiers

Classifiers	NB Classifier			MLP Classifier			J48 Classifier		
Split Ratio	60%	80%	90%	60%	80%	90%	60%	80%	90%
kappa Statistics	0.213	0.1464	0.1767	0.099	0.165	-0.049	0.138	0.0201	-0.019
Mean Absolute Error	0.16	0.17	0.16	0.13	0.12	0.12	0.13	0.11	0.10
Root Mean Squared Error	0.36	0.38	0.38	0.29	0.28	0.27	0.30	0.27	0.23

Relative absolute error	130.79	138.74	146.94	102.90	99.94	111.75	104.33	94.13	91.83
Root Relative Squared Error	140.72	156.86	178.44	115.30	116.42	125.74	117.15	108.36	110.41

We have also calculated values like Kappa statistic, Mean absolute error, Root mean squared error, Relative absolute error, and Root relative squared error. The output of the C4.5 (J48) Decision Tree is superior once again. The findings are shown in Table 4 and Figure 10.

Kappa statistic: A statistic for comparing Observed Accuracy to Expected Accuracy is the Kappa statistic. The kappa statistic is used to assess not only a single classifier, but also multiple classifiers.

Kappa = (observed accuracy minus expected accuracy) / (expected accuracy minus observed accuracy) (1 - expected accuracy)

Mean Absolute Error (MAE): is a metric for determining the accuracy of forecasts or projections in relation to actual results. The following is how the mean absolute error is calculated: As the name implies, the mean absolute error is the average of absolute errors, where is the prediction and is the true value.

$$mae = \frac{\sum_{i=1}^{n} abs\,(y_i - \lambda(x_i))}{n}$$

(6)

Relative Squared Error: It normalizes the total squared error by dividing it by the simple predictor's total squared error. The error is reduced to the same dimensions as the quantity being predicted by calculating the square root of the RSE.

Relative Absolute Error: In the same manner that the relative squared error is compared to a simple predictor, the relative absolute error is compared to the average of the real values.

Root Relative Squared Error: The RRSE is a measure of how reliable a simple predictor would have been if it had been utilized in the first place. This straightforward predictor is simply the average of the actual values. As a result, the relative squared error divides the total squared error by the simple predictor's total squared error.

The equation is used to measure the RRSE, Ei of a distinctive model I. (7).

$$E_i = \sqrt{\frac{\sum_{j-i}^{n} (P_{ij} - T_j)^2}{\sum_{j-i}^{n}(T_j - \overline{T})^2}}$$

(7)

P(ij) is the value that an individual model predicts for record j. Tj is the sample case j target value, which is determined using the equation (8).

$$\overline{T} = \frac{1}{n}\sum_{j-i}^{n} T_j$$

(8)

The numerator must equal 0 and Ei must equal 0 for a perfect match. As a result, the Ei index ranges from 0 to infinity, with 0 representing perfection.
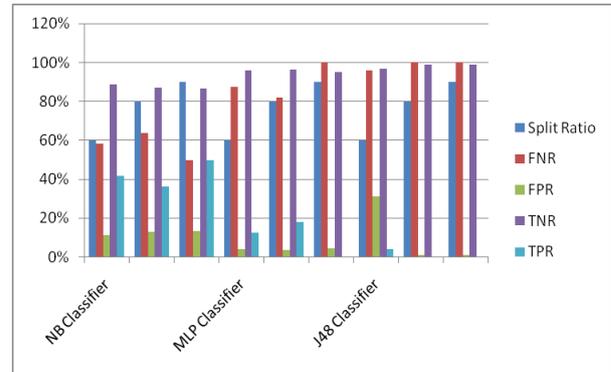

Figure 10 Graph obtained for the results shown in Table 4.

6. CONCLUSION

We experimented with the Nave Bayes, J48, and Multilayer Perceptron classifiers for the prediction of Cervical Cancer in this article. We have got good results with the C4.5 (J48) Decision Tree. We used the parameters Accuracy, Instances that were correctly classified and instances that were mistakenly classified were compared to the total number of instances. Precision, Recall, and the Kappa statistic to evaluate results. Absolute errors can take many forms, including root mean squared error, relative absolute error, and root relative squared error. The overall result is better with the C4.5 (J48) Decision Tree. In future we want evaluate the proposed system performance with various classifiers and preprocessing techniques.

REFERENCES

[1] Kayalvizhi. K. R, N Kanimozhi "Prediction of Cervical Cancer using Machine Learning and Deep Learning Algorithms" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 Issue-6, October 2020, pp.426-430, URL: https://www.ijtsrd.com/papers/ijtsrd33378.pdf

[2] Asadi F, Salehnasab C, Ajori L "Supervised Algorithms of Machine Learning for the Prediction of Cervical Cancer", Journal of Biomedical Physics and Engineering, 2020 Aug; 10(4): 513–522. Published online 2020 Aug 1, doi: 10.31661/jbpe.v0i0.1912-1027.

[3] Naveen R Chanukotimath, Dr. Sunil Kumar B S, "Assessment of Classifiers with Machine Learning based Thyroid Disease Prediction System",

International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-5, January 2020.

[4] Riham Alsmariy , Graham Healy , Hoda Abdelhafez "Predicting Cervical Cancer using Machine Learning Methods", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 7, 2020.

[5] Radhanath Patra "Prediction of Lung Cancer Using Machine Learning Classifier" International Conference on Computing Science, Communication and Security COMS2 2020: Computing Science, Communication and Security pp 132-142

[6] Taye Girma Debelee , Samuel Rahimeto Kebede, Friedhelm Schwenker, Zemene Matewos Shewarega "Deep Learning in Selected Cancers' Image Analysis—A Survey" Journal of Imaging 2020, 6, 121; doi:10.3390/jimaging6110121

[7] Jaswinder Singh, Sandeep Sharma, "Prediction of Cervical Cancer Using Machine Learning Techniques", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 11 (2019) pp. 2570-2577

[8] Dhwaani Parikh, Vineet Menon "Machine Learning Applied to Cervical Cancer Data" Mathematical Sciences and Computing, 2019, 1, 53-64 Published Online January 2019 in MECS (http://www.mecs-press.net) DOI: 10.5815/ijmsc.2019.01.05

[9] Akshatha Shetty, Vrushika Shah, "Survey of Cervical Cancer Prediction Using Machine Learning: Comparative Approach", 2018, 9th International Conference on Computing Communication and Networking Technology" (ICCCNT).

[10] Konstantina Kourou, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V, Dimitrios I "Machine learning applications in cancer prognosis and prediction" Computational and Structural Biotechnology Journal Volume 13, 2015, Pages 8-17

[11] Subramani, Prabu, K. Srinivas, R. Sujatha, and B. D. Parameshachari. "Prediction of muscular paralysis disease based on hybrid feature extraction with machine learning technique for COVID-19 and post-COVID-19 patients." *Personal and Ubiquitous Computing* (2021): 1-14.

[12] Bhuvaneswary, N., S. Prabu, K. Tamilselvan, and K. G. Parthiban. "Efficient Implementation of Multiply Accumulate Operation Unit Using an Interlaced Partition Multiplier." *Journal of Computational and Theoretical Nanoscience* 18, no. 4 (2021): 1321-1326.

[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] Pham, Dung V., Giang L. Nguyen, Tu N. Nguyen, Canh V. Pham, and Anh V. Nguyen. "Multi-topic misinformation blocking with budget constraint on online social networks." *IEEE Access* 8 (2020): 78879-78889.

[15] Naeem, Muhammad Ali, Tu N. Nguyen, Rashid Ali, Korhan Cengiz, Yahui Meng, and Tahir Khurshaid. "Hybrid Cache Management in IoT-based Named Data Networking." *IEEE Internet of Things Journal* (2021).

[16] Le, Ngoc Tuyen, Jing-Wein Wang, Duc Huy Le, Chih-Chiang Wang, and Tu N. Nguyen. "Fingerprint enhancement based on tensor of wavelet subbands for classification." *IEEE Access* 8 (2020): 6602-6615.

[17] WEN WU "Data-Driven Diagnosis of Cervical Cancer with Support Vector Machine- Based Approaches", Department of Blood Transfusion, Jinan Military General Hospital, Jinan, China Year: 2017.

[18] Parameshachari, B. D., H. T. Panduranga, and Silvia liberata Ullo. "Analysis and computation of encryption technique to enhance security of medical images." In *IOP Conference Series: Materials Science and Engineering*, vol. 925, no. 1, p. 012028. IOP Publishing, 2020.

[19] Rajendran, Ganesh B., Uma M. Kumarasamy, Chiara Zarro, Parameshachari B. Divakarachari, and Silvia L. Ullo. "Land-use and land-cover classification using a human group-based particle swarm optimization algorithm with an LSTM Classifier on hybrid pre-processing remote-sensing images." *Remote Sensing* 12, no. 24 (2020): 4135.

[20] Subramani, Prabu, Ganesh Babu Rajendran, Jewel Sengupta, Rocío Pérez de Prado, and Parameshachari Bidare Divakarachari. "A block bi-diagonalization-based pre-coding for indoor multiple-input-multiple-output-visible light communication system." *Energies* 13, no. 13 (2020): 3466.

[21] Xiaoyu Deng, Yan Luo, Cong Wang "Analysis of Risk Factors for Cervical Cancer Based on Machine Learning Methods", School of Automation, Beijing University of Posts and Telecommunications, Beijing, 100876, China.

[22] K. Yu, L. Tan, X. Shang, J. Huang, G. Srivastava and P. Chatterjee, "Efficient and Privacy-Preserving Medical Research Support Platform Against COVID-19: A Blockchain-Based Approach", IEEE Consumer Electronics Magazine, doi: 10.1109/MCE.2020.3035520.

[23] Z. Guo, Y. Shen, A. K. Bashir, M. Imran, N. Kumar, D. Zhang and K. Yu, "Robust Spammer Detection

Using Collaborative Neural Network in Internet of Thing Applications", IEEE Internet of Things Journal, vol. 8, no. 12, pp. 9549-9558, 15 June15, 2021, doi: 10.1109/JIOT.2020.3003802.

[24] L. Tan, H. Xiao, K. Yu, M. Aloqaily, Y. Jararweh, "A Blockchain-empowered Crowdsourcing System for 5G-enabled Smart Cities", Computer Standards &Interfaces, https://doi.org/10.1016/j.csi.2021.103 517

[25] Dhwaani Parikh, Vineet Menon "Machine Learning Applied to Cervical Cancer Data", RMIT University, 124 La Trobe St, Melbourne VIC 3000.

[26] K. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi and T. Sato, "Deep-Learning-Empowered Breast Cancer Auxiliary Diagnosis for 5GB Remote E-Health," IEEE Wireless Communications, vol. 28, no. 3, pp. 54-61, June 2021, doi: 10.1109/MWC.001.2000374.

[27] K. Yu, L. Tan, S. Mumtaz, S. Al-Rubaye, A. Al-Dulaimi, A. K. Bashir, F. A. Khan, "Securing Critical Infrastructures: Deep Learning-based Threat Detection in the IIoT", IEEE Communications Magazine, 2021.

[28] K. Yu, Z. Guo, Y. Shen, W. Wang, J. C. Lin, T. Sato, "Secure Artificial Intelligence of Things for Implicit Group Recommendations", IEEE Internet of Things Journal, 2021, doi: 10.1109/JIOT.2021.3079574.

[29] Upasana "Handle Imbalanced Classification Problems in machine learning", Consultant of Data & Analytics in KPMG.