ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

ISSN 2454-9940
www.ijasem.org
INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
Vol19, Issue 1,2025

A NEURAL NETWORK-BASED STRATEGY FOR THE EARLY DETECTION OF CERVICAL CANCER

[1]GEETHANJALI DUGYALA, [2]KAMTAM RAVINDRA, [3]RITESH PRASAD, [4]DANTHURI SANJAY, [5]BONTHA GOPINATH, [6]Dr. KAPILA RAMDAS, [7]Mr. A DANIEL PRANEET,
[12345]Student Department of DS, Narsimha Reddy Engineering College, Maisammaguda (V), Kompally, Secunderabad, Telangana-500100.
[6]Associate Professor, Department of CSE, Narsimha Reddy Engineering College, Maisammaguda (V), Kompally, Secunderabad, Telangana-500100.
[7]Assistant Professor, Department of Mechanical Engineering, Narsimha Reddy Engineering College, Maisammaguda (V), Kompally, Secunderabad, Telangana-500100.

Abstract—

An important global public health concern, cervical cancer impacts women all over the globe. Cervical cancer is a deadly illness, thus identifying those at risk early on is crucial. the general population about this illness in order to avoid it. Health care providers and those in danger may both benefit from early prediction using an ML model. Utilizing a dataset from the UCI ML repository, eleven supervised ML algorithms are used to predict early dangers of this illness in this work. To forecast the early dangers, we scour the ML models and estimate performance metrics like ROC-AUC, re-call, accuracy, and precision. The study's use of the Multi-Layer Perceptron (MLP) method with the default hyperparameters resulted in a prediction accuracy of 93.33 percent, according to the reasonable analysis that followed. The accuracy was 93.33% while using the hyperparameter tuning technique with Grid Search Cross Validation (GSCV), K-Nearest Neighbors (KNN), Decision Tree Classifier (DTC), Support Vector Machine (SVM), Random Forest Classifier (RFC), and Multi-Layer Perceptron (MLP).

*Keywords—Cervical Cancer, UCI repository, Analytical**Analysis, ML Techniques***

INTRODUCTION

Cancer of the cervix, or cervical cancer, is a major public health concern that impacts millions of women worldwide.

on a global scale, especially in less developed nations. Over 89% of the fatalities from this cause occur in non-developing countries, according to the World Health Organization [1]. Some 445,000 cases were found in 2012, with new cases accounting for over 83% of ALL [2]. Period irregularities, abnormal bleeding, and abnormal menstrual patterns are all signs of cervical cancer. Therefore, cervical cancer may be detected with a pap smear test, which also reduces the chance of mortality by almost 90% and cervical cancer by 60% to 90% [3]. Some major issues with this test include a lack of medical equipment, poor treatment, poor diagnostic repeatability, irresponsible maintenance, and the professionals giving the test becoming bored with it because of how it drones on and on [4]. Additionally, poor lifestyle choices might transmit the cancer-causing Human Papillomavirus (HPV). Condylomas, which release infectious virions, are an outward sign of high-risk HPV infection [5]. Although the number has decreased due to pap smear checks and HPV vaccination, the mortality toll is still too high [6]. Half of all cervical cancer occurrences in the US arise because people didn't check, and another 10% had never been screened before [7]. Consequently, many lives may be saved via the early detection of cervical cancer with the use of lifestyle information [8-10]. A significant input in this area may be achieved by gathering enough data, evaluating it, and discovering the underlying pattern [11–14]. Healthcare providers may now guarantee early diagnosis and treatment thanks to machine learning methods made possible by developments in data science [15–19]. Automation and computer-aided diagnostic systems have been the subject of much study [20–23], with the hope that they may shorten patients' screening times [24–27] and make diagnoses more straightforward [28–31]. To predict the danger based on actions, Sobar et al. used a classifier. Their best accuracy was 91.67%, achieved by combining two industry-standard methods [32]. Using Pap smear pictures and the SVM algorithm for categorization, Kashyap et al. [33] proposed a technique and achieved a 95% accuracy rate. Nevertheless, a Pap smear test and a classifier based on Fourier-Transform Infrared (FTIR) spectroscopy were used by Njoroge et al. to achieve a total accuracy of 72% [34]. Conversely, a

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol19, Issue 1,2025

model proposed by Fazal et al. [35] achieved a maximum accuracy of 99.5% by classifying the data using random forest (RF) classifiers and using DBSCAN and isolation forest as outlier removers. In addition, Wu et al. [36] found that SVM-PCA outperformed other models when they utilized three SVM-based approaches to identify and categorize four targets. Hyeon et al. created a model to detect cervical cell state in microscopic images by using a convolutional neural network and several machine learning classifiers. A maximum accuracy of 89.7 percent was achieved [37]. Eleven supervised machine learning models were used in this research, including DTC, MLP, RFC, KNN, SVM, CatBoost (CatB), Gaussian Naïve Bayes (GNB), Gradient Boosting Classifier (GradB), AdaBoost (AdaB), XG Boost (XGB), and XG Boost with Random Forest (XGBRF), using provided datasets. Possible significant implications for ehealthcare system development and computer-assisted diagnostics are highlighted by the results.

METHODOLOGY

Obtainable via the UCI ML repository, the dataset details risk factors for cervical cancer [38]. There are a total of 72 cases in the collection, with 19 characteristics, one of which is a goal column. With the exception of one, all of the characteristics have numeric values. Figure 1 shows the results of an exploratory data study that defined the association between features via the use of a correlation heatmap. The statistical information from the dataset is shown in Table I.

TABLE I. DIFFERENT ATTRIBUTES

No.	Numeric Attributes	Max-Min	Mean	Standard Deviation
1	behavior_sexualRisk	10-2	9.67	1.19
2	behavior_eating	15-3	12.79	2.36
3	behavior_personalHygine	15-3	11.08	3.03
4	intention_aggregation	10-2	7.90	2.74
5	attitude_consistency	15-6	13.35	2.37
6	intention_commitment	10-2	7.18	1.52
7	attitude_spontaneity	10-4	8.61	1.52
8	norm_significantPerson	5-1	3.13	1.85
9	norm_fulfillment	15-3	8.49	4.91
10	perception_vulnerability	15-3	8.51	4.28
11	perception_severity	10-2	5.39	3.40
12	motivation_strength	15-3	12.65	3.21
13	motivation_willingness	15-3	9.69	4.13
14	socialSupport_emotionality	15-3	8.09	4.24
15	SocialSupport_appreciation	10-2	6.16	2.90
16	socialSupport_instrumental	15-3	10.38	4.32
17	empowerment_knowledge	15-3	10.54	4.37
18	empowerment_abilities	15-3	9.32	4.18
19	empowerment_desires	15-3	10.28	4.48
20	ca_cervix	1-0	0.29	0.46

Based on exploratory data analysis, the dataset was split 80/20 between training and testing ML systems. Two methods were used in this study: one relied on default ML algorithm hyperparameters, while the second one tuned hyperparameters using Grid Search Cross-Validation using a 10-fold technique. The train dataset is used to train eleven supervised machine learning algorithms, including DTC, GNB, RFC, KNN, SVM, CatB, MLP, GradB, AdaB, XGB, and XGBRF. The two hyperparameter methods, the default and the tweaked, are then tested on the same dataset to determine their respective performance metrics. Figure 2 shows the whole algorithm used in the paper.

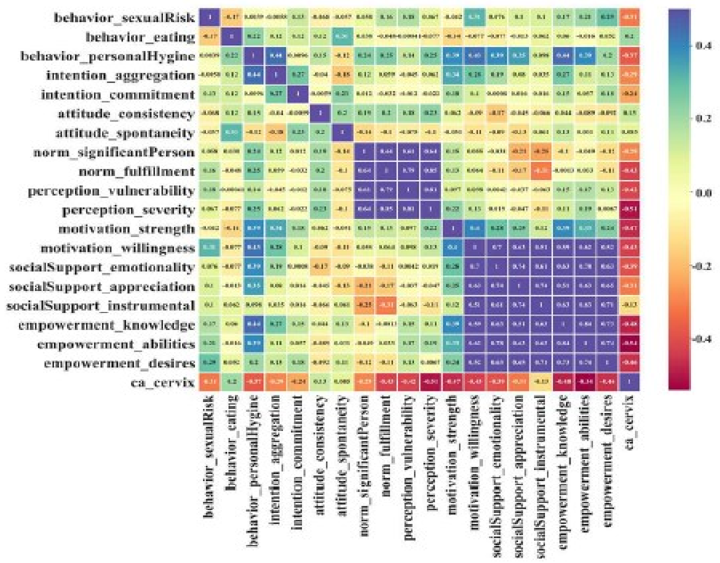

Fig. 1 Correlation heatmap of attributes

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol19, Issue 1,2025

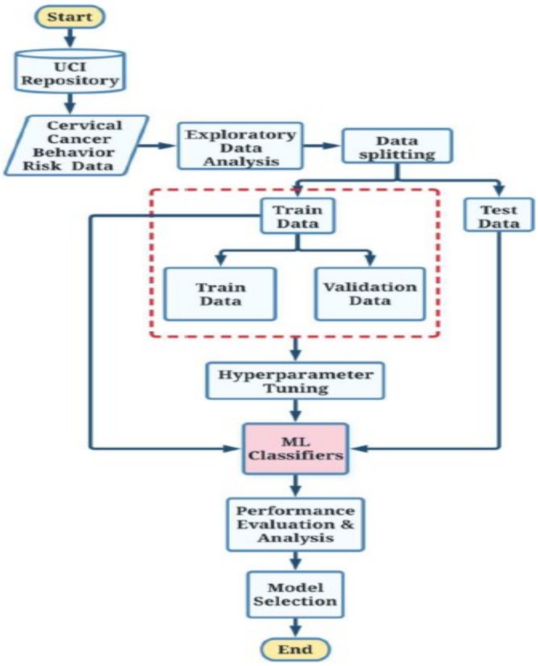

Fig. 2 Algorithm of the work

RESULTS & DISCUSSION

The first method involves training and testing ML classifiers using the default hyperparameters of ML algorithms, and then evaluating and presenting the performance metrics in Table III. Figure 3 shows the ROC and Table II shows the confusion matrices. Table III shows that when compared to other algorithms, the MLP algorithms perform the best. Results show that MLP has the highest F1-score (0.9524), re-call (1.000), accuracy (0.9333), and precision (0.9091) of all models. GNB, CatB, and GradB enhance recall, while RFC, SVM, and XGBRF maximize accuracy.

In addition, the highest ROC_AUC achieved was 0.9545 when the GradB algorithm was used.

In the second technique, ROC was used in Figure 4 and hyperparameter tuning was done using a 10-fold GSCV method, as shown in Table IV. Then, the performance assessments were conducted. Based on the performance measures shown in Table IV, the DTC, RFC, KNN, and SVM achieved better accuracy (0.9333) and precision (0.9091) than all other ML models. On top of that, GradB maximizes recall and accuracy, much as XGBRF and GNB. The results obtained from the trials are rather good. It is clear that just one method, MLP, achieved the highest level of performance (accuracy) while using the default hyperparameters. Hyperparameter adjustment using the GSCV method, however, resulted in superior performance for a number of algorithms (DTC, RFC, KNN, SVM, MLP). Additionally, the second method improved the overall efficiency of ML algorithms.

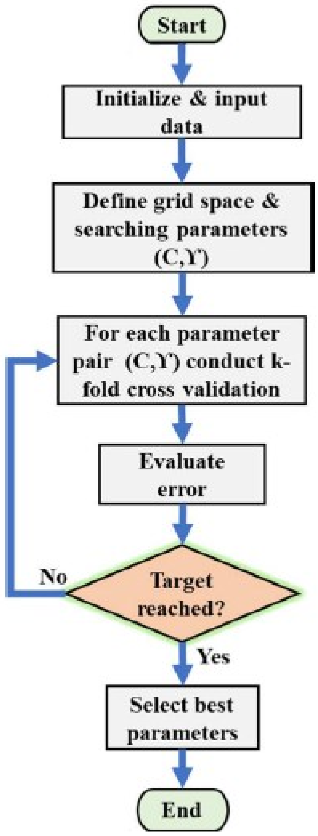

Fig. 3 Grid Search algorithm

TABLE II. CONFUSION MATRICES OF ML ALGORITHMS

Algorithms	Without Tuning				With Tuning			
	T_P	T_N	F_P	F_N	T_P	T_N	F_P	F_N
DTC	9	3	2	1	10	4	1	0
GNB	9	4	2	0	9	4	2	0
RFC	10	3	1	1	10	4	1	0
KNN	9	3	2	1	10	4	1	0
SVM	10	3	1	1	10	4	1	0
CatB	9	4	2	0	9	3	2	1
MLP	10	4	1	0	10	4	1	0
GradB	8	4	3	0	8	4	3	0
AdaB	9	3	2	1	9	3	2	1
XGB	9	3	2	1	9	3	2	1
XGBRF	10	3	1	1	10	3	1	1

TABLE III. ANALYSIS OF ML ALGORITHMS (WITHOUT TUNING)

ISSN 2454-9940
www.ijasem.org
Vol19, Issue 1,2025

	Accuracy	Precision	F-1	Re-call	ROC-AUC
DTC	.800	.818	.857	.900	.784
GNB	.867	.818	.900	1.000	.909
RFC	.867	.909	.909	.909	.829
KNN	.800	.818	.857	.900	.784
SVM	.867	.909	.909	.909	.829
CatB	.867	.818	.900	1.000	.909
MLP	.933	.909	.952	1.000	.864
GradB	.800	.727	.842	1.000	.954
AdaB	.800	.818	.857	.900	.784
XGB	.800	.818	.857	.900	.784
XGBRF	.867	.909	.909	.909	.829

TABLE IV. ANALYSIS OF ML ALGORITHMS (WITH TUNING)

	Accuracy	Precision	F-1	Re-call	ROC-AUC
DTC	.9333	.9091	.952	1.0	.9545
GNB	.8667	.8182	.900	1.0	.9091
RFC	.9333	.9091	.952	1.0	.9545
KNN	.9333	.9091	.952	1.0	.9545
SVM	.9333	.9091	.952	1.0	.9545
CatB	.8000	.8182	.857	.9	.7841
MLP	.9333	.9091	.952	1.0	.8636
GradB	.8000	.7273	.842	1.0	.9545
AdaB	.8000	.8182	.857	.9	.7841
XGB	.8000	.8182	.857	.9	.7841
XGBRF	.8667	.9091	.909	.9	.8295

This is an important finding from the study that could lead to a more reliable, efficient, and accurate model for predicting the risk of cervical cancer. In order to help healthcare providers get a head start and make faster diagnoses, this study provides an analytical methodology to show how various machine learning classifiers can be used for risk prediction.

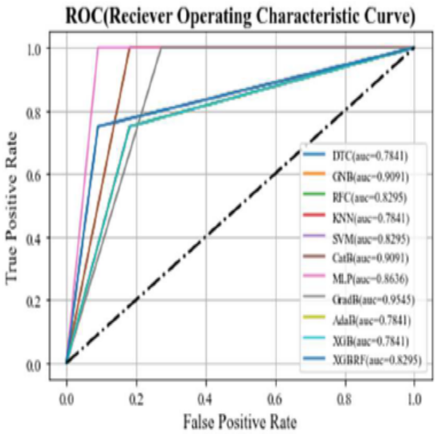

Fig. 3 ROC (without hyperparameter tuning)

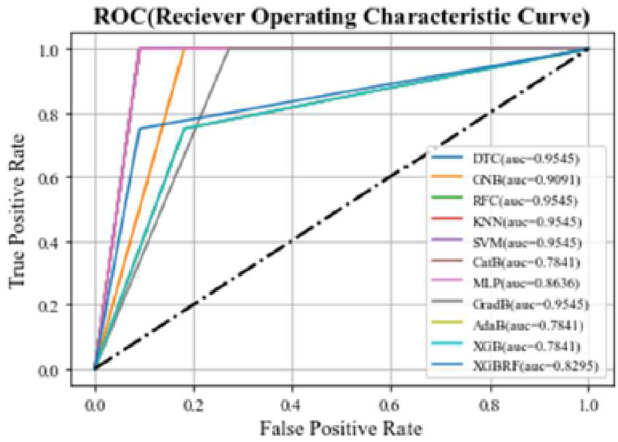

Fig. 4 ROC (with hyperparameter tuning)

This allows for the development of an intelligent support system and the introduction of an efficient healthcare management system, ensuring that individuals from all walks of life may enjoy effectively treating cancer.

CONCLUSION

For women all around the globe, cervical cancer is the biggest threat. Lessening the mortality toll from this illness is possible with early identification or risk prediction.In order to construct an accurate prediction model with the help of ML algorithms, a large amount of data is gathered and assessed. In order to foretell the threat of cervical cancer, this study evaluates eleven different supervised ML systems.The purpose of this research was to enhance the performance of classifiers by hyperparameter tuning using GSCV. The algorithms tested, including DTC, RFC, KNN, SVM, and MLP, achieved a maximum accuracy of 93.33%. The most important thing that came out of this study is that predictions are more accurate and consistent, which might help with computer-aided diagnosis and make it a better tool for doctors. Nevertheless, extensive testing is necessary prior to its use in a clinical environment. The creation of an e-healthcare system depends on the completion of many future breakthroughs in this field, and further data collection may help with that.

REFERENCES

[1]. "Cervical cancer." https://www.who.int/news-room/factsheets/detail/cervical-cancer (accessed Feb. 19, 2022).

[2]. [WHO, "Comprehensive Cervical Cancer Control," Geneva, pp. 366–378, 2014.

[3]. M. Safaeian, et al., "Cervical Cancer Prevention-Cervical Screening:Science in Evolution," Obstet.

Gynecol. Clin. North Am., 34(4), pp.739–760, 2007, doi: 10.1016/j.ogc.2007.09.004.

[4]. N. Colombo, et al. "Cervical cancer: ESMO clinical practiceguidelines for diagnosis, treatment and follow-up," Ann. Oncol., vol.23, no. SUPPL. 7, 2012, doi: 10.1093/annonc/mds268.

[5]. J. Doorbar, "Molecular biology of human papillomavirus infectionand cervical cancer," Clin. Sci., vol. 110, no. 5, pp. 525–541, 2006

[6]. I. C. Scarinci et al., "Cervical cancer prevention: New tools and oldbarriers," Cancer, vol. 116, no. 11, pp. 2531–2542, 2010

[7]. D. Saslow et al., "American Cancer Society, American Society forColposcopy and Cervical Pathology, and American Society forClinical Pathology screening guidelines for the prevention and earlydetection of cervical cancer," CA. Cancer J. Clin., vol. 62, no. 3, pp.147–172, 2012, doi: 10.3322/caac.21139.

[8]. J. Lu, et al. "Machine learning for assisting cervical cancer diagnosis:An ensemble approach," Futur. Gener. Comput. Syst., vol. 106, pp.199–205, 2020, doi: 10.1016/j.future.2019.12.033.

[9]. B. Nithya and V. Ilango, "Evaluation of machine learning basedoptimized feature selection approaches and classification methods forcervical cancer prediction," SN Appl. Sci., vol. 1, no. 6, 2019

[10]. R. Weegar and K. Sundström, "Using machine learning for predictingcervical cancer from Swedish electronic health records by mininghierarchical representations," PLoS One, 15(8), pp. 1–19, 2020

[11]. F. Faisal and M. M. Nishat, "An Investigation for EnhancingRegistration Performance with Brain Atlas by Novel Image InpaintingTechnique using Dice and Jaccard Score on Multiple Sclerosis (MS)Tissue," Biomed. and Pharm. J., vol. 12, no. 3, pp. 1249-1262, 2019

[12]. M. M. Nishat, F. Faisal, T. Hasan, M. F. B. Karim, Z. Islam and M.R. Kaysar, "An Investigative Approach to Employ Support VectorClassifier as a Potential Detector of Brain Cancer from MRIDataset," 2021 Int. Conf. on Electronics, Comm. and Inf. Tech.(ICECIT), 2021, pp. 1-4, doi: 10.1109/ICECIT54077.2021.9641168

[13]. M. M. Nishat et al., "Performance Investigation of Different BoostingAlgorithms in Predicting Chronic Kidney Disease," 2020 2nd Int.Conf. on STI 4.0, pp. 1-5, 2020,doi: 0.1109/STI50764.2020.9350440

[14]. M. A. A. R. Asif et al., "Performance Evaluation and ComparativeAnalysis of Different Machine Learning Algorithms in PredictingCardiovascular Disease," Engineering Letters, vol. 29, no. 2, pp. 731-741, 2021

[15]. M. M. Nishat et al., "A Comprehensive Analysis on DetectingChronic Kidney Disease by Employing Machine LearningAlgorithms," EAI Endorsed Transactions on Pervasive Health andTechnology, vol. 18, no. e6, 2021, doi: 10.4108/eai.13-8-2021.170671[16] M. R. Farazi, F. Faisal, Z. Zaman and S. Farhan, "Inpainting multiplesclerosis lesions for improving registration performance with brainatlas," 2016 Int. Conf. on Medical Engg., Health Info. and Tech.(MediTec), 2016, pp. 1-6, doi: 10.1109/MEDITEC.2016.7835363.

[16]. M. A. A. R. Asif et al., "Computer Aided Diagnosis of ThyroidDisease Using Machine Learning Algorithms," 2020 11th Int. Conf.on Electrical and Computer Engineering (ICECE), 2020, pp. 222-225, doi: 10.1109/ICECE51571.2020.9393054[18] M. M. Nishat and F. Faisal, "An Investigation of SpectroscopicCharacterization on Biological Tissue," 2018 4th Int. Conf. onElectricalEngg. and Information & Communication Tech.(iCEEiCT), 2018, pp. 290-295, doi: 0.1109/CEEICT.2018.8628081.

[17]. F. Faisal, M. M. Nishat, M. A. Mahbub, M. M. I. Shawon, and M. M.U. H. Alvi, "Covid-19 and its impact on school closures: a predictiveanalysis using machine learning algorithms," 2021 Int. Conf. onScience& Contemporary Technologies (ICSCT), 2021, pp. 1-6, doi:10.1109/ICSCT53883.2021.9642617

[18]. A. A. Rahman et al., "Detection of Epileptic Seizure from EEGSignal Data by Employing Machine Learning Algorithms withHyperparameter Optimization," 2021 4th Int. Conf. on Bio-Engineering for Smart Technologies (BioSMART), 2021, pp. 1-4, doi:10.1109/BioSMART54244.2021.9677770

[19]. A. A. Rahman, L. I. Khalid, M. I. Siraji, M. M. Nishat, F. Faisal andA. Ahmed, "Enhancing the Performance of Machine LearningClassifiers by Hyperparameter Optimization in Detecting AnxietyLevels of Online Gamers," 2021 24th Int. Conf. on Computer andInformation Technology (ICCIT), 2021, pp. 1-5, doi:10.1109/ICCIT54785.2021.9689911

[20]. T. Hasan et al., "Exploring the Performances of Stacking Classifier inPredicting Patients Having Stroke," 2021 8th NAFOSTED Conferenceon Information and Computer Science (NICS), pp. 242-247, 2021,doi: 10.1109/NICS54270.2021.9701526.

[21]. M. M. Nishat, T. Hasan, S. M. Nasrullah, F. Faisal, M. A. A. R. Asif,and M. A. Hoque, "Detection of Parkinson's Disease by EmployingBoosting Algorithms," 2021 Joint 10th Int. Conf. Informatics,Electronics& Vision (ICIEV) and 2021 5th Int. Conf. on Imaging,Vision& Pattern Recognition (icIVPR), pp. 1-7, 2021, doi:10.1109/ICIEVicIVPR52578.2021.9564108

[22]. Nishat, M. M., et al. "Detection of Autism Spectrum Disorder byDiscriminant Analysis lgorithm," BIM, pp. 473-482, Springer,Singapore, 2022 doi: 10.1007/978-981-16-6636-0_36