Intelligent cervical cancer detection: empowering healthcare with machine learning algorithms

Uma Yadav[1], Vipin D. Bondre[2], Shweta V. Bondre[1], Bhakti Thakre[3], Poorva Agrawal[4], Shruti Thakur[5]
[1]School of Computer Science and Engineering, Ramdeobaba University, Formerly Shri Ramdeobaba College of Engineering and Management, Nagpur, India
[2]Department of Electronics and Telecommunication Engineering, Yeshwantrao Chavan College of Engineering, Nagpur, India
[3]Department of Computer Science and Engineering (Cyber Security), St. Vincent Pallotti College of Engineering and Technology, Nagpur, India
[4]Symbiosis Institute of Technology, Nagpur Campus Symbiosis International (Deemed University), Nagpur, India
[5]Department of Computer Science and Engineering, G H Raisoni College of Engineering, Nagpur, India

Article Info

Article history:

Received Apr 4, 2024
Revised Jul 17, 2024
Accepted Jul 26, 2024

Keywords:

Artificial neural networks
Cervical cancer
Logistic regression
Machine learning
Random forest
Support vector machine

ABSTRACT

Cervical cancer remains a significant global health issue, particularly in underdeveloped nations, where it contributes to high mortality rates. Early detection is critical for improving treatment outcomes and survival rates. This study employs machine learning (ML) algorithms to predict cervical cancer risk using a dataset from the University of California at Irvine (UCI), which includes demographic and clinical attributes such as age, sexual history, smoking habits, and medical history. After applying data preprocessing techniques, several classification algorithms, including logistic regression (LR), support vector machine (SVM), random forest (RF), decision tree, adaptive boosting (AdaBoost), and artificial neural networks (ANN), were trained and evaluated. The models were assessed using classification metrics such as precision, recall, and F1 score. Among the models, the ANN demonstrated the highest accuracy, achieving a score of 0.95. In addition, correlation analysis revealed significant relationships between various risk factors, providing insights into cervical cancer mechanisms and potential preventive measures. The study highlights the potential of ML in improving cervical cancer detection and patient outcomes, suggesting that advanced ML techniques can be valuable tools in healthcare research and clinical applications.

Corresponding Author:

Uma Yadav
School of Computer Science and Engineering, Ramdeobaba University
Formerly Shri Ramdeobaba College of Engineering and Management
Nagpur 440013, India
Email: uma.yadav12@gmail.com

1. INTRODUCTION

Worldwide, the health of women is at risk from the lethal illness known as cervical cancer, and it can be challenging to identify its early symptoms [1]. Cervical tumor, which has abundant negative things, is solitary of the most severe illnesses people might have. Women's ignorance of the value of early detection contributes to the higher mortality rate of uterine cancer. It grounds damage to the cervix's yawning tissues and can ultimately spread to further portions of the body, such as the lungs, vagina, and liver which can mark the condition more stimulating [2]. Brain tumour diagnosis, breast cancer recognition, cervical cancer recognition, physical activity recognition, COVID detection, thermal sensation recognition, and the evaluation of dementia patients' cognitive health all make use of machine learning (ML) and deep learning (DL). It is more efficient than conventional diagnostic methods thanks to developments in the healthcare sector. Every year, 493,000 new cases of cervical cancer are reported in health reports issued by "Global cancer statistics", 15% of which are female cancer patients. With an 83% mortality rate, this illness is primarily prevalent in underdeveloped nations. prominent in African nations, such as Uganda, which has 65% of confirmed cases and ranks 14th among the countries with the highest frequency of cervical cancer. The twenty-five percent common reason of sickness and death in females worldwide is cervical malignant expansion [3]. The pervasiveness of the human papillomavirus (HPV) is associated to the progression of cervical tumor. The worldwide load of cervical cancer has been condensed as a consequence of initial screening, creating the sickness preventable.

Cervical tumor can occur if the HPV contagion is not treated. The HPV is the most repeated transferable agent in cervical cancer since it can progress into neoplasms. Neoplastic progress is the term used to describe the uncontrolled growth of cervical tumor cells and the proliferation of anomalous cells as a result of a malevolent stage [4]. The cervical tumor dataset assimilated for inspection contains severance, omitted values, and noise. Mining approaches are deliberated one of the utmost encounters and noteworthy grounds of learning in prescription due to the rising prominence of fitness problems. Numerous data quarrying approaches may be used to promptly intend further research and health care, which can save lives, particularly in the case of cervical cancer. The initial phase in all data mining processes is preprocessing, which accounts for 80% of the information.

Earlier studies have used a variety of standard ML-based strategies, like k-nearest neighbours (KNN), K-means clustering, and random forest (RF), to diagnose cervical malignance [5]–[7]. A medical decision sustenance network for malicious growth was created in 2017 by Wui and Hao, employing a acquaintance based approach in combination with a rudimentary set of assumptions and calculations using transmissible genomic procedures in a soft computing prototypical [8]. Artificial neural networks (ANN) have the traits of self-learning, parallelism, and sensitivity to interior disappointment. These qualities are utilised by the proposed paradigm. Though the coarse set theory's evidence dispensation powers and the strong, similar, and destructive search are features of genetic algorithms (GAs), it still has noteworthy computation disputes.

Gray level concurrence matrix (GLCM), support vector machines (SVM), KNN, spatial fuzzy cluster algorithms, convolutional neural networks (CNN), C5.0, RF, and hierarchical cluster algorithms for feature mining, separation, and cell classification were discussed by Ashok and Aruna [9] and estimated in terms of distinctive factors like dataset capacity, drawbacks, and precision. However, it works best with small datasets. For the purpose of diagnosing cervical cancer, Tseng *et al.* [10] have applied data extraction, image processing, and ML approaches. Each picture comprises a mix of smoothness and form features. Reciprocal data assortment, progressive forward assortment, and unsystematic section assortment are all given the best topographies. Using the SVM, cervical growth pictures are categorised. To select the best tool for cervical growth diagnosis, different assortment methods are associated. SVM scuffles because there is no probabilistic rationalization for cataloguing, which consequences in tremendously strict categorizations.

In Makassar, Indonesia, Anuraga *et al.* [11] made an effort to investigate the factors that influence cancer patients' survival. The specimens used in this study comprised up to 38 cancer patients. They use sample information training and the RF to treasure tree fusion information. The primary concern with this method is that it only has a 50% accuracy rate. Alyafeai and Ghouti [12] developed a fully unified cervical tumor recognition and screening pipeline from cervical images in 2020. Two deep neural system learning representations are now being established for programmed cervical tumour analysis and identification. The first test has a recognition precision of 0.68 in relations of unification connection estimation since it can identify the cervix region 1,000 times quicker than the most advanced data-driven simulations. The second model uses self-extracted traits to detect cervical tumours. Two simple models with an emphasis on CNNs are used to train such features. William *et al.* [13] automated the cervical cancer diagnosis process using images from Pap smears in an exertion to inferior the possibility of fault.

One of the paramount predominant tumors in females globally is cervical tumor. Several readings on cervical tumor have been conceded out recently, retaining cutting-edge approaches that deal early-stage forecast. Early forecast has promoted from the use of ML [14]. Therefore, absence of information, absence of admittance to assets and health amenities, and the expenditure of appearing regular studies in some nations are the main explanations for this infection among female populaces [15]. The efficiency of readings and the invention of specific patient information have both increased thanks to ML. When forecasting clients' preference in numerous service situations, one investigator [16] used ML, text quarrying, and econometric algorithms to identify which core and improved qualities and sentiments are most important. In addition, this research illustrates the significance of unremitting eminence enhancement in the enactment of ML procedures from a health precaution organization and organization info technology opinion by recounting numerous ML procedures and consuming ML procedures to analyse healthcare data [17]. This study found procedures that are more appropriate for scientific convention in the grouping of positive and negative cervical tumor. These procedures can be used to detect cervical tumor. In order to measure the investigative effectiveness of DL procedures in detecting infections in medicinal imaging, DL has demonstrated a major influence on fitness and medicinal imaging [18]. The following are the goals of this investigation: i) to use ML procedures to analyse and categorise cervical tumor in direction to help medical professionals in appropriately detecting the infection, and ii) to conclude the associations between the attributes that are record probable to cause cervical tumor.

The remaining portion is organized as follows. Section 2 outlines the suggested approach alongwith dataset description. The outcome and discussion are examined in section 3. Section 4 concludes with a discussion of the findings and future research.

2. PROPOSED METHOD

Predictive model selection (PMS), training technique, and data preprocessing are the several categories under which the suggested research approach is divided [19]. The proposed study's architectural diagram is shown in Figure 1. It is obvious from Figure 1 that the architecture has been divided into three parts since the model used in this investigation completes some crucial duties at respective level. How to clean up the dataset so that ML may use it is covered in the data preprocessing section. The PMS section demonstrates the kind of predictive model chosen for this study's cervical cancer prediction. The training methods section outlines the requirements for model training. Finally, using the Python computer language, we developed the platform to offer a comprehensive pipeline for cervical cancer prediction. The technique used in this study is more suitable for categorising cervical cancer diagnoses, both positive and negative, for clinical application. The techniques logistic regression (LR), decision tree (DT), KNN, SVM, RF, gradient boosting (GB), adaptive boosting (AdaBoost), and extreme gradient boosting (XGBoost) can be used to classify cervical cancer [20]. The next sections outline the progression and effects.

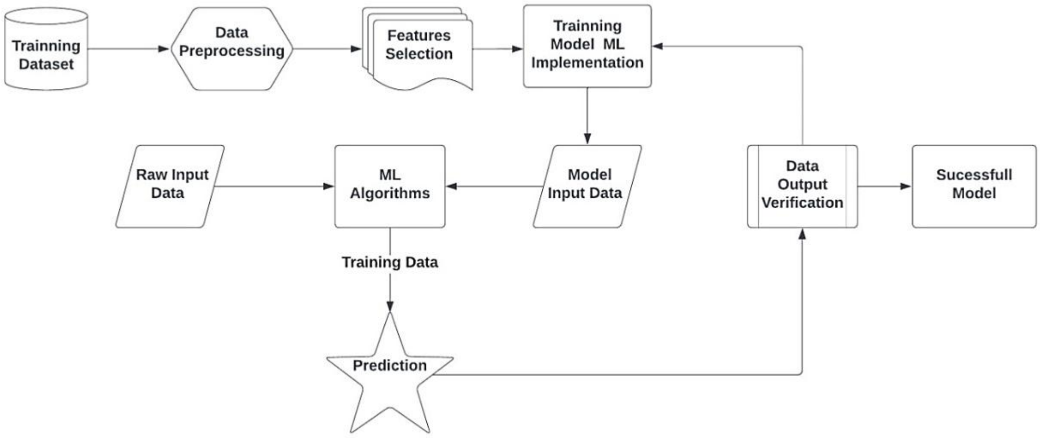

Figure 1. Proposed architecture to detect cervical cancer

2.1. Datasets description

The University of California at Irvine's (UCI) database has made the dataset available. For 858 occurrences with 36 attributes each, the collection included demographic data, patient histories, current practises, and procedures. Since incomplete examples were chosen to avoid addressing privacy concerns, the dataset lacks a number of elements that are common to healthcare record frameworks [21]. The dataset features and the kind of assessment for each attribute are shown in Table 1.

2.2. Data preprocessing

Data cleaning, data alteration, and data reduction are the three categories into which data preparation is classified. Data preparation is essential since it has a direct bearing on a project's success. When characteristics or feature values comprise noise, missing data, redundant data, or outliers, this is referred to as data impurity [21]. From this dataset, we have eliminated outliers and missing values. To turn the data into formats appropriate for the mining process, the data transformation step is maintained. This study includes idea hierarchy creation, attribute selection, normalization, and discretization. Analysis gets more challenging when dealing with a significant volume of data, especially when the dimension of the data is vast. To address this, a data reduction strategy is used in this study. It aims to decrease the price of data processing and storage while increasing storage efficiency. We selected the dimension reduction method since it is an additional helpful method for reducing overfitting in ML prototypes. We have used the principle component analysis (PCA) method for that.

Table 1. Dataset depiction

Sr. No.	Attribute	Category	Sr. No.	Attribute	Category
1	Age	Int	19	STDs:pelvic inflammatory disease	Bool
2	Total sexual buddies	Int	20	STDs:genital herpes	Bool
3	First sensual intercourse	Int	21	STDs:molluscum contagiosum	Bool
4	Total pregnancies	Int	22	STDs:AIDS	Bool
5	Smokes	Bool	23	STDs:HIV	Bool
6	Smokes (years)	Int	24	STDs:Hepatitis B	Bool
7	Smokes (packs/year)	Int	25	STDs:HPV	Bool
8	Hormonal Contraceptives	Bool	26	STDs: Number of diagnosis	Int
9	Hormonal Contraceptives (years)	Int	27	STDs: Time since first diagnosis	Int
10	IUD	Bool	28	STDs: Time since last diagnosis	Int
11	IUD (years)	Int	29	Dx:Cancer	Bool
12	STDs	Bool	30	Dx:CIN	Bool
13	STDs (number)	Int	31	Dx:HPV	Bool
14	STDs:condylomatosis	Bool	32	Dx	Bool
15	STDs:cervical condylomatosis	Bool	33	Hinselmann	Bool
16	STDs:vaginal condylomatosis	Bool	34	Schiller	Bool
17	STDs:vulvo-perineal condylomatosis	Bool	35	Citology	Bool
18	STDs:syphilis	Bool	36	Biopsy	Bool

2.3. Model selection

SVM, LR, decision tree classifier (DTC), RF, AdaBoost, GB, XGBoost, and KNN are just a few of the ML cataloguing techniques that have been employed in the model selection. This segment has emphasized a few of the procedures that, when applied to the chosen study dataset, produced an acceptable degree of accuracy. As a result, the following subsections show the theoretical meaning of these algorithms.

2.3.1. Decision tree

The regression and classification tree procedure, frequently recognized as the DT, can be used to handle issues involving both regression and classification. The term "tree" was further to the DT's name because of how much it be similar to a tree's outlets. Similar to how a tree instigates at its base, a DT also begins at the root knot. The outlets of this tree division out from the root knot under various decision circumstances; these knots are known as decision knots and become leaf knots if a choice is made.

2.3.2. Support vector machine

The SVM technique may be applied to regression and classification issues. SVMs are, nonetheless, relatively common for small to medium-sized classification datasets of reasonably complicated kinds. This technique uses a hyperplane to divide the data points, and the kernel chooses the shape of the hyperplane. In many instances, a normal scatter plot with numerous attributes cannot distinguish between double or additional data classes. A key component of an SVM that can transform two-dimensional data into three-dimensional data and hence distinguish between various kinds is its kernel [22].

2.3.3. Random forest

The use of several learners in ensemble learning improves model performance. Another form of ensemble learning is RF. The likelihood that outliers will have an impact on the results is decreased by using the RF bagging technique. Both categorical and continuous data benefit from this. Although datasets do not need to be scaled, complicated models demand more processing power the more learners there are. Voting is used in this method to determine the outcome. Ensemble learning is the term used for such an algorithm. Numerous trees or bushes make up random woods. The number of DTs in RFs is comparable to the quantity of trees in a forest. The choice that the majority of trees make is regarded as the ultimate one.

2.3.4. Adaptive boosting

By pooling the information of several weak learners, the AdaBoost strategy develops a prevailing learner. In this case, each and every weak learner makes use of the same response, also known as a training data. The relevancy of each primary response or portion of training data is equal. Following the first weak learner, who is given increased weight for the forecasts prepared by the first weak learner, comes the second weak learner, who is then assigned responsibility for rectifying the wrong predictions produced by the first weak learner. As a result, albeit with more weight, the mistakes that the second weak learner made in its forecasts are conceded on to the third weak learner.

2.3.5. Gradient boosting

The GB technique also makes use of the consecutive ensemble learning methodology. Loss optimization enables weak learners to strengthen over time and become better than they were. The third weak learner, for instance, excels over the second, the fourth weak learner over the first, and so on. As the weedy learner's periodicity growths, the model's degree of inaccuracy decreases, and it develops a tougher learner. The GB method performs excellently when dealing with regression-type problems. AdaBoost differs from GB in that the error is steadily decreased by adjusting the weight of the incorrect prediction samples. GB optimizes both the loss function and individual losses. Additionally, there is a decrease in inaccuracy. Apiece weak learner modifies its alternate weak learner prototypical to improve this loss function, making the subsequent weak learner stronger than the prior one. Loss function optimization, a weak learner, and an additive model make up GB.

3. RESULTS AND DISCUSSION

The analysis of cervical data presents a comprehensive understanding of the intricate relationships between various attributes and their impact on cervical cancer detection. Through correlation analysis and classification reports generated by ML algorithms, significant insights have been gleaned, shedding light on the efficacy of predictive models in this critical domain. This discussion will delve into the findings derived from the correlation chart and classification reports, elucidating the implications for cervical cancer research and the potential clinical applications of these results.

3.1. Cervical data analysis

The correlation chart is presented in Figure 2. Correlation describes the association between two or more attributes [23]. These elements could be characteristics of the raw records we use to estimate our target attribute. A mathematical technique called correlation is used to assess how one attribute changes or moves in relative to another. It provides information on the strength of the correlation among the two attributes. It is a bivariate analytic portion that establishes how multiple attributes are related to one another [24]. Outcome the association is important in cervical investigation since it allows for the identification of key elements by showing how each attribute is related to the others. There may be a positive correlation between two traits (attributes). The association between two features (attributes) can also be negative. This suggests that the other attribute(s) decrease as the value of the first attribute increases. However, there is no connection if the significance of one attribute rises or falls while the value of the other attributes does not. Figure 2 provides an illustration of the linkages. Figures 3 and 4 shows the total measurement for pregnancies, age, sexual partners, and smokes as well as a link between the amount of pregnancies and the biopsy results. The uterus lowest and thinnest part is called the cervix. A channel is formed leading to the vaginal entrance. There are numerous techniques for doing cervical surgeries. It is clear from Figure 4 that there is a correlation between a biopsy and pregnancy; however, it varies from time to time.

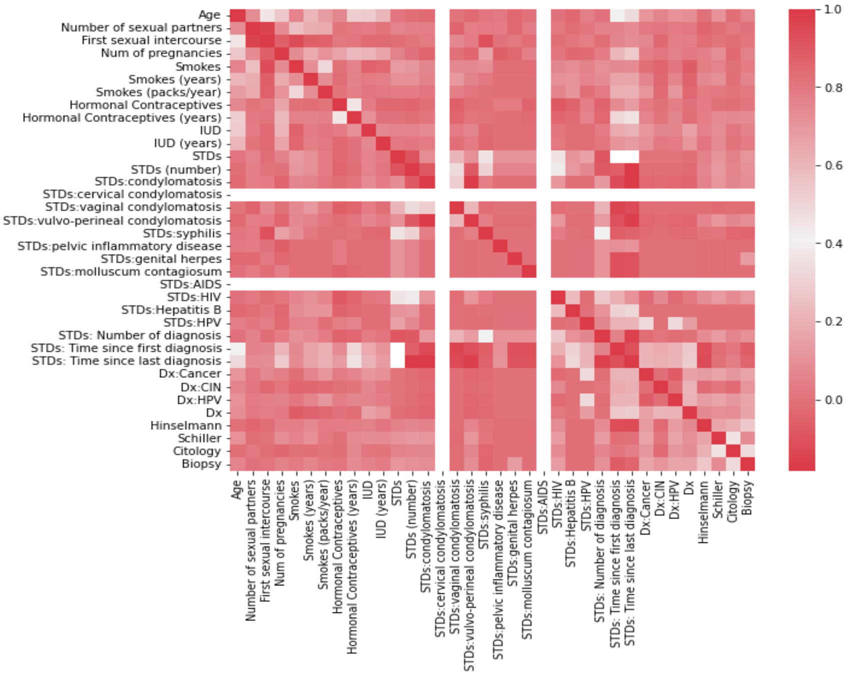

Figure 2. Corelation of input attributes

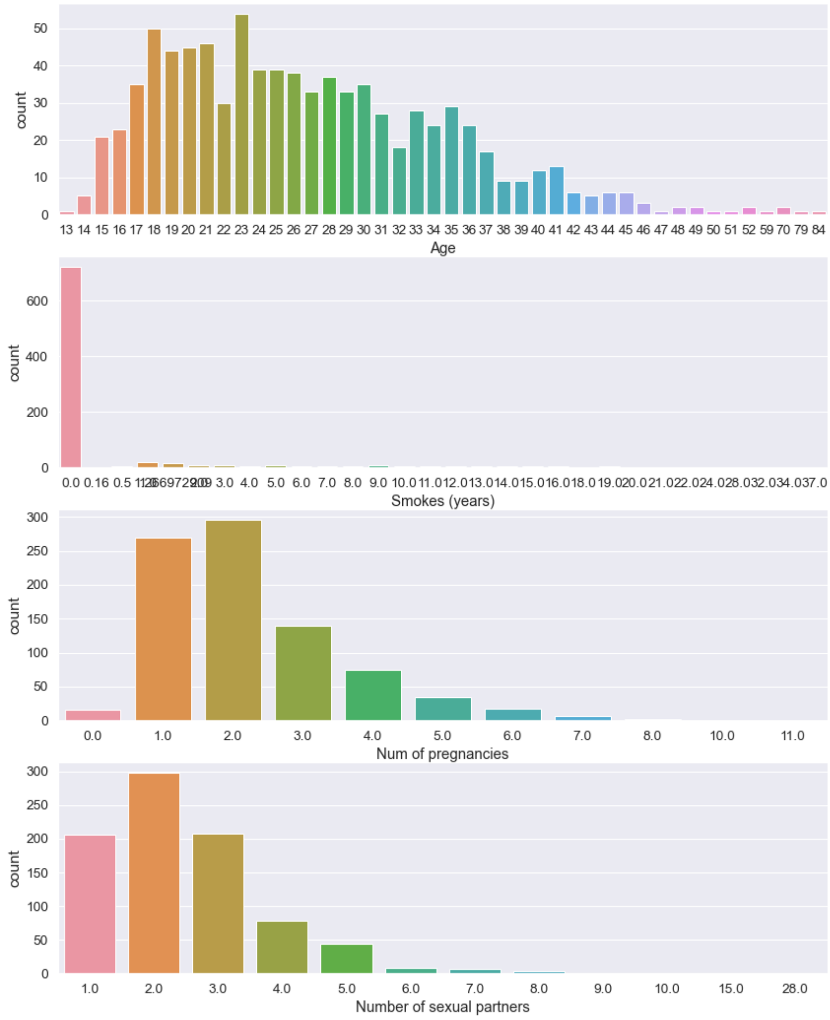

Figure 3. Total measurement in relations of the age, number of smokes, pregnancies, and sexual partners

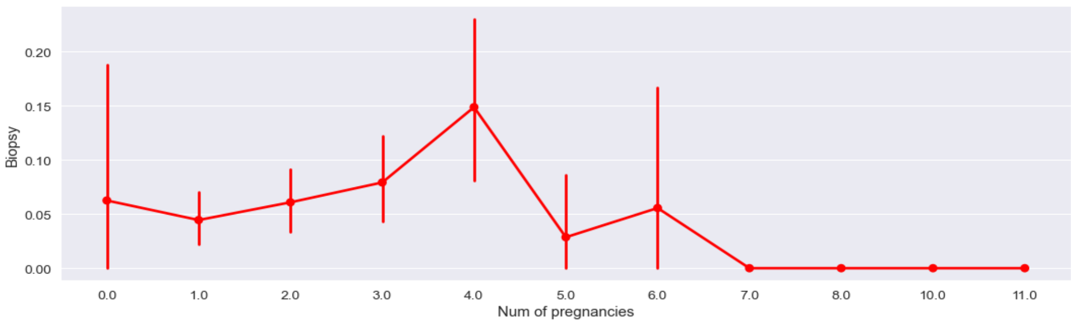

Figure 4. Visual representation of the link among the number of pregnancies and the biopsy

3.2. Result analysis

Applying a classification report will estimate how accurate the predictions made by the classification algorithms will be. The report provides per-class examples of the recall, accuracy, and F1 measure of the important classification metrics. The metrics are calculated using true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [25]. Table 2 shows the classification reports for numerous classical ML methods. F1 score is the weighted mean of recall and precision; this total takes into account both FN and FP; precision is the fraction of the prototype's exact positive estimates to the entire correct and incorrect positive guesses; and recall is the fraction of being able to predict positive as positive. A classification statement has been included in the Table 2, where "0" denotes a negative classification and "1" a positive classification.

The following in (1) through (4) are utilized to produce the classification report [25]:
Precision: The correlation between the prototypes accurate positive estimate and the aggregate (accurate and incorrect) positive estimate. It is said as (1):

$$Pr = \frac{Tp}{Tp+Fp} \tag{1}$$

Recall: The proportion of accurate to incorrect predictions serves as a measure of positivity. The following is how it is expressed in mathematical representation.

$$Rc = \frac{Tp}{Tp+Fn} \tag{2}$$

Accuracy: This term refers to the total number of accurately anticipated events. It is specified as (3).

$$Acc = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \tag{3}$$

F1 scores: Compared to the accuracy measure, this harmonic average of recall and precision gives a more precise assessment of the number of misclassification cases. It may be written as (4).

$$F1 = \frac{2*Pr*Rc}{Pr+Rc} \tag{4}$$

With the appropriate evaluation matrices, we have interpreted several algorithms. According to Table 2 findings, LR, SVM [22], DT, RF, AdaBoost, and ANN have all produced the highest classification scores. For the GB technique, XGBoost offers a higher amount of regularisation. In XGBoost, radical regularization (L1 and L2) is used to improve prototypical generalisation. XGBoost overtakes the GB method in relations of performance. ANN achieved an accuracy of 0.95. ANN training accuracy and loss plot against Epoch is shown in Figure 5.

Table 2. Cervical cancer classification report

Algorithm	Purpose	Positive class (1)			Negative class (0)			Accuracy
		Pr	Rc	F1	Pr	Rc	F1	
Logistic	Cervical cancer detection	1.0	1.0	1.0	1.0	1.0	1.0	1.0
SVM		1.0	1.0	1.0	1.0	1.0	1.0	1.0
RF		1.0	1.0	1.0	1.0	1.0	1.0	1.0
DT		1.0	1.0	1.0	1.0	1.0	1.0	1.0
AdaBoost		1.0	1.0	1.0	1.0	1.0	1.0	1.0
ANN		1.0	0.35	0.55	0.91	1.0	0.95	0.91

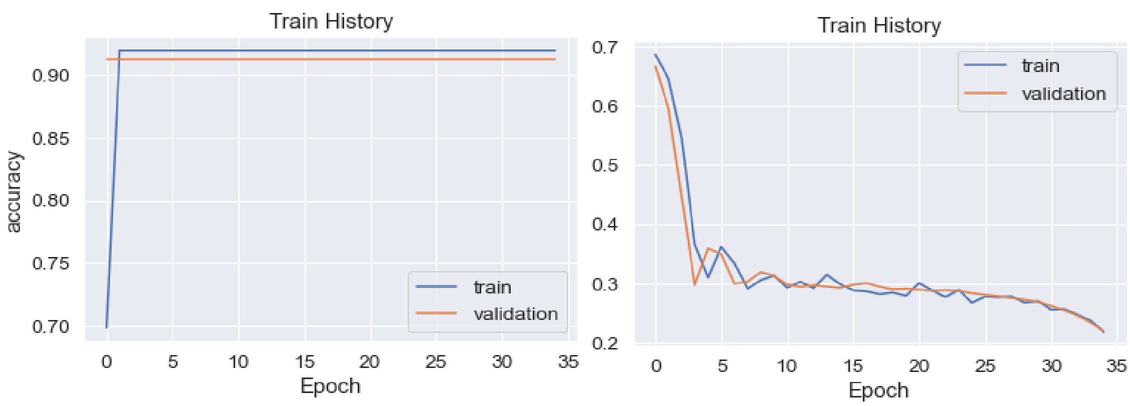

Figure 5. Training accuracy and loss plot against Epochs

4. CONCLUSION

In conclusion, this study demonstrates the effectiveness of ML techniques in predicting cervical cancer risk and identifying key risk factors associated with the disease. By analyzing a comprehensive dataset comprising demographic and clinical attributes, we show that ANN, in particular, exhibit high accuracy in predicting cervical cancer risk. The correlation analysis highlights significant associations between various risk factors, providing valuable insights for future research and clinical practice. Our findings suggest that ML-based approaches have the potential to enhance cervical cancer diagnosis and management, ultimately contributing to improved patient outcomes. Moving forward, further research is needed to validate the performance of ML models in diverse patient populations and explore the integration of additional data sources for more comprehensive risk assessment. Overall, our study contributes to advancing the field of cervical cancer research and underscores the importance of harnessing ML techniques for disease prediction and management.

REFERENCES

[1] X. Yang, M. Da, W. Zhang, Q. Qi, C. Zhang, and S. Han, "Role of Lactobacillus in cervical cancer," *Cancer Management and Research*, vol. 10, pp. 1219–1229, 2018, doi: 10.2147/CMAR.S165228.

[2] T. Mukama, R. Ndejjo, A. Musabyimana, A. A. Halage, and D. Musoke, "Women's knowledge and attitudes towards cervical cancer prevention: A cross sectional study in Eastern Uganda," *BMC Women's Health*, vol. 17, no. 9, Jan. 2017, doi: 10.1186/s12905-017-0365-3.

[3] U. Yadav, A. K. Sharma, and D. Patil, "Review of automated depression detection: social posts, audio and video, open challenges and future direction," *Concurrency and Computation: Practice and Experience*, vol. 35, no. 1, 2023, doi: 10.1002/cpe.7407.

[4] C. M. Martin, K. Astbury, L. McEvoy, S. O'Toole, O. Sheils, and J. J. O'Leary, "Gene expression profiling in cervical cancer: identification of novel markers for disease diagnosis and therapy.," *Methods in molecular biology*, vol. 511, pp. 333–359, 2009, doi: 10.1007/978-1-59745-447-6_15.

[5] F. Asadi, C. Salehnasab, and L. Ajori, "Supervised algorithms of machine learning for the prediction of cervical cancer," *Journal of Biomedical Physics and Engineering*, vol. 10, no. 4, pp. 513–522, 2020, doi: 10.31661/jbpe.v0i0.1912-1027.

[6] M. A. Devi, S. Ravi, J. Vaishnavi, and S. Punitha, "Classification of cervical cancer using artificial neural networks," *Procedia Computer Science*, vol. 89, pp. 465–472, 2016, doi: 10.1016/j.procs.2016.06.105.

[7] G. Sun, S. Li, Y. Cao, and F. Lang, "Cervical cancer diagnosis based on random forest," *International Journal of Performability Engineering*, vol. 13, no. 4, pp. 446–457, 2017, doi: 10.23940/ijpe.17.04.p12.446457.

[8] W. Wu and H. Zhou, "Data-driven diagnosis of cervical cancer with support vector machine-based approaches," *IEEE Access*, vol. 5, pp. 25189–25195, 2017, doi: 10.1109/ACCESS.2017.2763984.

[9] B. Ashok and P. Aruna, "Comparison of feature selection methods for diagnosis of cervical cancer using SVM classifier," *International Journal of Engineering Research and Applications*, vol. 6, no. 1, pp. 94–99, 2016.

[10] C. J. Tseng, C. J. Lu, C. C. Chang, and G. D. Chen, "Application of machine learning to predict the recurrence-proneness for cervical cancer," *Neural Computing and Applications*, vol. 24, no. 6, pp. 1311–1316, 2014, doi: 10.1007/s00521-013-1359-1.

[11] G. Anuraga, J. W. Fernanda, and Pebrianty, "Random forest prognostic factor in colorectal cancer," *Journal of Physics: Conference Series*, vol. 1217, no. 1, 2019, doi: 10.1088/1742-6596/1217/1/012098.

[12] Z. Alyafeai and L. Ghouti, "A fully-automated deep learning pipeline for cervical cancer classification," *Expert Systems with Applications*, vol. 141, 2020, doi: 10.1016/j.eswa.2019.112951.

[13] W. William, A. Ware, A. H. Basaza-Ejiri, and J. Obungoloch, "Cervical cancer classification from Pap-smears using an enhanced fuzzy C-means algorithm," *Informatics in Medicine Unlocked*, vol. 14, pp. 23–33, 2019, doi: 10.1016/j.imu.2019.02.001.

[14] S. Geeitha and M. Thangamani, "Integrating HSICBFO and FWSMOTE algorithm-prediction through risk factors in cervical cancer," *Journal of Ambient Intelligence and Humanized Computing*, Springer, vol. 12, no. 3, pp. 3213–3225, Jun. 2020, doi: 10.1007/s12652-020-02194-6.

[15] A. Ghoneim, G. Muhammad, and M. S. Hossain, "Cervical cancer classification using convolutional neural networks and extreme learning machines," *Future Generation Computer Systems*, vol. 102, pp. 643–649, 2020, doi: 10.1016/j.future.2019.09.015.

[16] U. Yadav and A. K. Sharma, "A novel automated depression detection technique using text transcript," *International Journal of Imaging Systems and Technology*, vol. 33, no. 1, pp. 108–122, 2023, doi: 10.1002/ima.22793.

[17] U. D. Yadav and P. S. Mohod, "Adding persuasive features in graphical password to increase the capacity of KBAM," *2013 IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology, ICE-CCN 2013*, pp. 513–517, 2013, doi: 10.1109/ICE-CCN.2013.6528553.

[18] L. Akter, F. Al-Islam, M. M. Islam, M. S. Al-Rakhami, and M. R. Haque, "Prediction of cervical cancer from behavior risk using machine learning techniques," *SN Computer Science*, vol. 2, no. 3, 2021, doi: 10.1007/s42979-021-00551-6.

[19] H. T. Endale *et al.*, "MiRNA in cervical cancer: diagnosis to therapy: systematic review," *Heliyon*, vol. 10, no. 3, 2024, doi: 10.1016/j.heliyon.2024.e24398.

[20] A. A. Swanson and L. Pantanowitz, "The evolution of cervical cancer screening," *Journal of the American Society of Cytopathology*, vol. 13, no. 1, pp. 10–15, 2024, doi: 10.1016/j.jasc.2023.09.007.

[21] R. Weegar and K. Sundström, "Using machine learning for predicting cervical cancer from Swedish electronic health records by mining hierarchical representations," *PLoS ONE*, vol. 15, no. 8, Aug. 2020, doi: 10.1371/journal.pone.0237911.

[22] S. V. Khedikar and U. Yadav, "Detection of disease from radiology," *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS),* Coimbatore, India, 2017, pp. 1-4, doi: 10.1109/ICIIECS.2017.8276174.

[23] U. Yadav and A. K. Sharma, "Review on automated depression detection from audio visual clue using sentiment analysis," *Proceedings of the 2nd International Conference on Electronics and Sustainable Communication Systems, ICESC 2021*, pp. 1462–1467, 2021, doi: 10.1109/ICESC51422.2021.9532751.

[24] S. Mathivanan, D. Francis, and S. Srinivasan, "Enhancing cervical cancer detection and robust classification through a fusion of deep learning models," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-61063-w.

[25] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

BIOGRAPHIES OF AUTHORS

Uma Yadav is an accomplished academic with a Ph.D. in Computer Science and Engineering from GHRU Amravati, India. She currently serves as an Assistant Professor at Ramdeobaba University (formerly Shri Ramdeobaba College of Engineering and Management), where she specializes in data science, machine learning, cloud computing, big data management, and data mining. With over 14 years of professional experience, she has made significant contributions to both academia and industry, authoring and co-authoring 20+ research papers in refereed journals and conferences. In addition to her research publications, she has also contributed several book chapters in areas such as networking, robotics, reinforcement learning, and biomedical technologies. She can be contacted at email: uma.yadav12@gmail.com.

Dr. Vipin D. Bondre received his B.Tech. from North Maharashtra University Jalgaon and M.Tech. degrees from Rashtrasant Tukadoji Mahararaj Nagpur University, Nagpur in 2005 and 2011 respectively. He has been awarded Ph.D. in wireless communication in 2020 from RTMNU, Nagpur University. He is an Assistant Professor at Yeshwantrao Chavan College of Engineering, Nagpur, an autonomous institute. His research interests include wireless communication and embedded systems. He has authored or co-authored more than 25 research papers in refereed journals and conferences. He can be contacted at email: vipin.bondre@ycce.edu.

Shweta V. Bondre holds a Ph.D. in Computer Science and Engineering from GHRU Amravati, India, and is currently an Assistant Professor at Ramdeobaba University (formerly Shri Ramdeobaba College of Engineering and Management). With over 14 years of experience, she has developed expertise in data science, machine learning, DBMS, big data management, and data mining. Her extensive contributions to academia and industry are reflected in her authorship of 20+ research papers published in esteemed journals and conferences. In addition to her research work, she has contributed several book chapters on topics ranging from networking and robotics to reinforcement learning and biomedical technologies. She can be contacted at email: bondresv@rknec.edu.

Bhakti Thakre received her B.Tech. and M.Tech. degrees from Nagpur university in 2006 and 2014, respectively. She is an Assistant Professor at St. Vincent Pallotti College of Engineering and Technology. Her research interests include data science, deep learning, and cyber security. She has authored or co-authored 10 research papers in refereed journals and conferences. She can be contacted at email: bthakre11@gmail.com.

Dr. Poorva Agrawal obtained her Ph.D. degree in Computer Science and Engineering from Symbiosis International (Deemed University), Pune, India in 2020. She has been engaged in research and teaching for more than 12 years. At present she is working as Senior Assistant Professor in Department of Computer Science and Engineering at Symbiosis Institute of Technology Nagpur, Symbiosis International (Deemed University) Pune, MH, India. She has presented more than 30 papers in international, national journals, and conferences. Her research interests include artificial intelligence, machine learning, data science, and computer vision. She can be contacted at email: poorva.agrawal@sitnagpur.siu.edu.in.

Dr. Shruti Thakur obtained her Ph.D. degree in Computer Science and Engineering from Sant Gadge Baba Amravati University Completed Ph.D. on March 22, 2024 and Assistant Professor in G H Raisoni College of Engineering Nagpur. She has more than 12 years' experience in teaching and research. Published more than 25 research papers in national, international journals, conferences, 1 book chapter in Scopus, and 2 copyright. She can be contacted at email: shruti.thakur@raisoni.net.