Grenze International Journal of Engineering and Technology, Jan Issue

A Survey of Deep Learning Assisted Prediction of Cervical Cancer: A Comparative Approach

N. Indumathi[1] and Dr. D. Shanthi[2]
[1-2]PSNA College of Engineering and Technology, Department of CSE, Dindigul, India
Email: indhujegan1996@gmail.com, dshan71@gmail.com

***Abstract*—There are more referrals and unneeded diagnostic tests as a result of current approaches for cervical cancer prediction. The purpose of this study was to create and assess a more reliable cervical cancer prediction model. According to the World Health Organization, cervical cancer is the fourth most common disease in women, accounting for 7.9 percent of all malignancies in women. Machine learning (ML) and deep learning (DL) are being used by more and more people and businesses to evaluate vast volumes of data and generate useful insights. Metaheuristics in the category of evolutionary algorithms include genetic algorithms (GAs). In large, challenging search areas, GAs can find the optimal or nearly optimal solutions, and they are frequently employed for search and optimization. For outcome prediction, statistical models, different kinds of medical imaging, and machine learning have all been utilized with encouraging outcomes. Machine learning has proven to be more effective than traditional statistical models at navigating the complexities of massive amounts of data and identifying prognostic indicators. However, the shortcomings of prediction studies and prediction models, such as over fitting, lack of interpretability, insufficient data, and simplification, suggest that further work is required to improve clinical outcome prediction and make it more accurate, reliable, and useful for clinical usage.**

***Index Terms*— Statistical Models, Prognostic Indicators, Machine Learning and Deep Learning, genetic algorithms.**

I. INTRODUCTION

One of the most common medical diseases affecting women worldwide, particularly in developing countries, is cervical illness. According to WHO statistics, 270 000 women died from cervical malignant growth in 2012, with 90% of these deaths occurring in low- and middle-income countries. In 2012, 445,000 new cases were recorded (around 84 percent of the new cases around the world). The pace of these passings was expected to increase by 25% in the absence of even the slightest hint of short activity. Early symptoms of cervical illness can include a regular monthly cycle, excessive bleeding, random bleeding, or spotting. The Pap (Papanicolaou) smear is a practically painless, quick, and simple cervical malignant growth screening test. Cells from a woman's cervix are gathered and spread out on a microscope slide during a pelvic exam for evaluation. The passing rate for cervical disease has decreased by 90% as a result of pap tests, and the number of cases has decreased by 60% to 90%. This test has a few flaws, including a weak clinical component, bad patient consistency, a frivolous conclusion, insufficient follow-up treatment, and carelessness on the part of the professionals in charge of it because it is so exhausting. There have been fewer fatalities from cervix malignant growth with the introduction of the pap smear test and the HPV vaccination to prevent infections in younger women, that is, those who are

Grenze ID: 01.GIJET.9.1.28

less than 18; however, the death rate in non-industrialized nations is still high. Cervical malignant growth develops in a woman's cervix. The lower, more tapered portion of a woman's uterus, or belly, is known as the cervix. The human papilloma virus, sometimes known as HPV, is an infection that can lead to disease and is transmitted through sex.

Most people have experienced HPV disease at some point, and this virus may go away on its own and turn out to be harmless. But cervical level moles or condylomas that produce contagious virions are also signs of useful contamination by high-risk HPV strains. According to a research, 10% of cervical illness instances in the US involve women who haven't had a personal test in the last five years, and nearly half of these cases involve women who have never had the condition evaluated.

There have been several initiatives to use simulated intelligence for task-related tasks, such as promoting the early eradication of cervical disease. Among the widely used computer-based intelligence techniques are Support Vector Machine, K-Nearest Neighbors, Irregular Woods Tree (RFT), Characterization and Backslide Tree (Truck), Fake Brain Organization, and others. In this audit, we separate the various artificial intelligence models used by professionals over the past few years based on the simulated intelligence calculations applied, dataset, preprocessing approaches, programming employed, and ultimately the accuracy arrived at. The final section of the paper refers to conclusions that have been reached from the coordinated review. The purpose of this research is to clearly illuminate the state of the art in cervical threatening development assumption employing significant learning in order to provide the framework for future assessment in this area. Numerous computations in AI have been adapted to deal with problems involving data, typically by identifying infinitely flexible personal plans. These estimates are enhanced to a crucial degree of precision either by changing the bounds of artificial intelligence or by eliminating irrelevant data through the incorporation of choice. A type of metaheuristic estimations known as inherited computations (GAs) was inspired by natural genetic components to select the best strategies. GAs can be used as the information's primary classifier, as a limit enhancer for base classifiers, or as a component selector. ML uses recently mentioned suggestive criteria as variables, analogous to morphology or surface, and requires factors pre-selected by people to carry out representation duties. However, DL eliminates what the structure deems critical components through the planning calculation itself without the presumptions included in earlier human judgements. Here, we switch from ML execution utilizing recently established characteristic variables to DL, which eliminates fresh real information that may be obscure to human subject matter experts. Future specialists will be able to choose almost with certainty which models are generally appropriate for their inspirations thanks to this examination, which will ultimately provide clinicians with systems ready to do conclusively helping the assurance of cervical dangerous development.

II. Related Work

Electronic thinking, a new branch of programming, focuses on recreating human information without requiring direct human engagement. In general, it is believed that preparation and contribution have a wide spectrum of understanding as their results. Therefore, it is difficult for a PC to gather information from the information that is available in the crucial sector. After the commencement of the electronic chaos, the gathering and recording of data in the area of clinical audit moved forward. In this way, one of the key points of contention about the application of artificial intelligence in the field of clinical science is resolved. Support vector machines are one of the computer-based intelligence techniques that are regularly used in the evaluation of clinical data and the clear confirmation of cervical hazardous development. The main issue declarations for which support vector machines have been utilized are backslide, representation into something like one class, and oddities' disclosure. Predicting the region of the hyper plane and managing a quadratic condition are the two crucial aspects of assist vector machines.

Starting from the beginning, this cycle was completed using impressive, up to that point, quadratic headway methods. This approach performed admirably for small instructional files containing about 1000 models. Since they are not intended to meet necessities, they have the advantage of needing less developer effort. They also provide astounding numerical accuracy. When managing larger datasets or when the component does not fit into the memory requirement, a more precise methodology that takes a wider range of parameters into account should be employed.

Cervical cancer has been identified as the third most deadly disease for women overall. The main cause of this illness is HPV, which promotes growth in the infected area. Prosperity experts advise perfect inoculation and screening to remove the infection. In about 99.7% of cases of cervical disease, HPV infection is the primary cause of this condition. Additionally, a normal 528,000 cases were recorded in 2012, making it one of the

common hazardous changes in females. The optimal goal and the precise ID of the illness are requirements of the stream risky development research.

III. Literature Review

The review of literature presents a brief description of the existing researches on machine learning techniques used for the detection of cervical cancer. The narration of the various algorithms and approaches used is presented in the review.

In 2014, Maciej Zieba et al. presented SVM, a tool for resolving unbalanced results. The suggested structure combined the benefits of using group classifiers with price-sensitive support vectors for skewed information. Additionally, a method for detaching choices from the assisted SVM was presented. The productivity of the proposed arrangement was then evaluated by comparing the presentation of the unequal information with various computations. Finally, a more advanced SVM was applied to predict post-operative outcomes in patients with lung cellular breakdown.

SVM with Counterfeit Brain Organizations and Decision-making Trees is noted for this circumstance as the accuracy indicator (92,85% exactness) in N. Picco et al. The persistence of prostate cancer growth is also examined in context, along with phony brain networks, decision trees, and tactical relapse. In the section, information on patients who were dying from colon cancer growth was contrasted with predictions of endurance and then some specific brains that were yet undetermined.

With an AUC of 0.91, Utah State College in the US used a faster district convolutional neural network (F-RCNN) in 2019 to automatically identify the cervical region in cervicography images and classify dysplasia and illness. [7] The exceptional area on issue finding, information driven observation, and control of digital actual frameworks in IEEE Access is where this paper was distributed. According to the authors, several information-driven methods have recently been proposed and put into practice in the field to achieve the best give analysis. These methods include head part analysis (PCA), molecule swarm enhancement (PSO), fluffy positivistic C-implies grouping, direct relapse (LR), fake brain organization (ANN), and support vector machine (SVM). The majority of the methods provided demonstrate successful execution in completing their tasks, and the majority of calculations provide astounding accuracy. However, often used CNNs need a large data set to prepare the models, which could be a test in the case of clinical data. In comparison to the conventional calculations, it is also worse in terms of time execution. Such computations provide excellent arrangement precision scores, with examples including the Example Based K-Closest Neighbor giving more than 98% and Arbitrary Tree, Irregular Timberland. We discovered that by continuously using different information mining procedures, research and clinical consideration might be swiftly suggested, leading to an attempt to try and save lives, particularly those of those who experience cervical malignant development. Information preparation is the main task, and it accounts for the great bulk of the whole information mining task. Preprocessing of the information will speed up model development and reduce computational time for case classification.

IV. Methods used in Machine Learning

A. Support Vector Machine (SVM)

SVM is one of the guided AI Characterization techniques that is widely used in the analysis and prediction of diseases. The classes were isolated by building a straight capability that extensively separated them using support vectors, and SVM capabilities by selecting fundamental instances from each class known as Help vectors. By considering the most ideal hyperplane, this straight classifier hopes to increase the minor distance—the distance between the choice hyperplane and the nearest piece of information—between them.

B. Random Forest (RF)

To organise a forest of trees, RF decides to group a number of desirable trees together. The argument put forth is that having a single option tree can either produce a straightforward model or one that is very certain. The recursive process used by RF involves selecting one arbitrary sample of size N from the dataset with substitution and another erratic example from the indicators without substitution for each association. The trees that organise perception into one classification are counted, and distinct situations are described in light of a larger part vote *over the choice trees.*

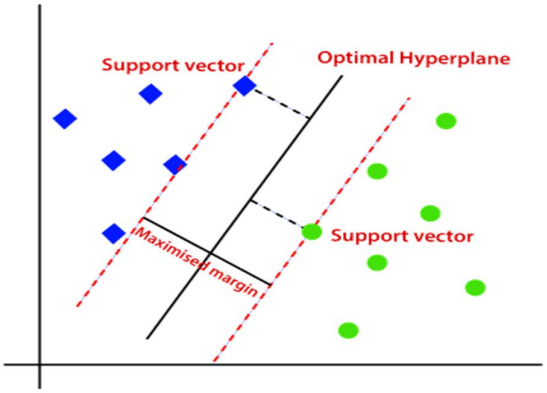

Fig 1: SVM classified with decision boundary and Hyper plane

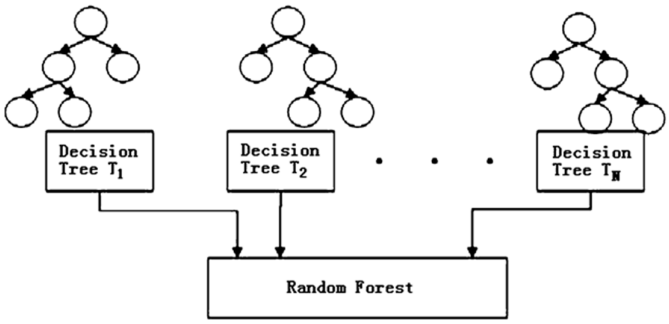

Fig 2: Random Forest with the bunch of Decision trees

C. Bayesian Networks (BN)

A branch of probabilistic graphical models called Bayesian Network is used for forecasting and information-representing ambiguous regions. Comparing the Bayesian Organization to the Coordinated Non-cyclic Chart, an AI framework that is frequently used (DAG). This diagram consists of several hubs, each of which is compared to an irregular variable. The hub edges discuss direct dependencies between the corresponding hubs in the diagram. In a network, a hub's guardians address their direct ancestors, and contingent likelihood is calculated using the likelihood.

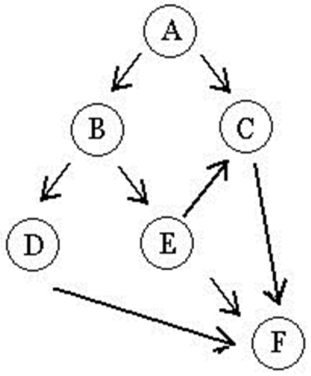

Fig 3. Bayesian Network with valid Probability

D. K-Nearest Neighbour

The K-Nearest Neighbor estimate is based on the comparative aspects of the relevant data. In this calculation, we made use of highlighting comparability, estimated the value of fresh data that would be of interest, and distributed the value based on how tightly paired the preparation informational gathering point was. It is used to

determine whether or not the patient has a malignant growth. An easy-to-use, regulated AI algorithm known as the k-closest neighbours (KNN) calculation can be used to address both order and relapse problems.

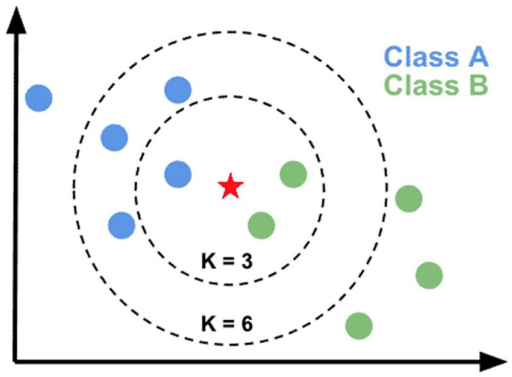

Fig 4. K-Nearest Neighbour algorithm for two different classes

E. Naïve Bayes algorithm (NB)

Naive Bayes is an order approach that relies on the Bayes hypothesis with indicator flexibility. The main justification for referring to this characterization procedures as "Credulous" is that this classifier anticipates that the presence of highlights in a particular class isn't connected with the presence of some other elements assuming that these elements are relied upon one another than these properties added to the likelihood of the class freely. It is "innocent" since it is rarely obvious. It is not challenging to hold and is very useful for large datasets.

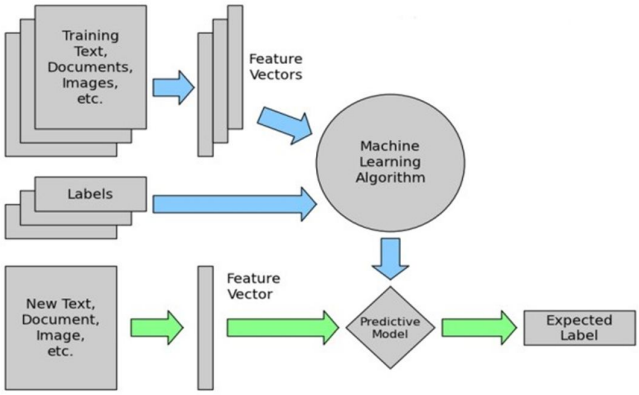

Fig 5. Patient record text classification using Naïve Bayes algorithm

F. Logistic Regression (LR)

A key order strategy is logistic regression, which has a place in the collection of linear classifier. It is comparable to linear and polynomial regression. It provides accurate and successfully deciphered results while still being rapid and relatively straightforward. As a result, it is also known as binary classification. This regression is not precisely the same as linear regression, but it is used to address the multiclass problem. With the aid of the Sigmoid function, the Logistic Regression model has the option of addressing what is occurring when the outcome can only have two possible values (0 and 1).
The following Table I represents the comparative approach of all machine learning and deep learning methods involved in the prediction of cervical cancer.

V. FINDINGS

When using artificial intelligence to predict illnesses, clinical data is essential. Using AI to identify anomalies in information requests, draw conclusions from the information that is available, and then go on to make predictions based on that information The following conclusions are reached by extrapolating research on existing projects in the field of cervical disease prediction using emerging AI technology. The analysts' AI

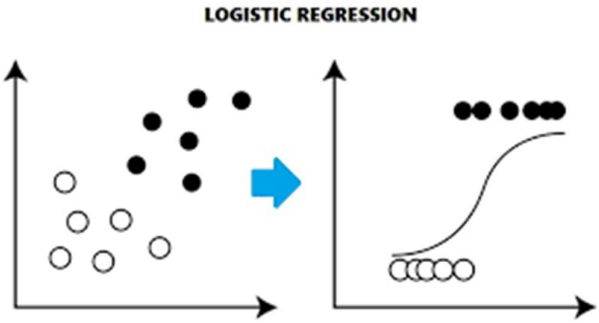

Fig 6. Logistic regression for the set of data points

TABLE I. COMPARISON THE VARIOUS SURVEYED APPROACHES

S.No	Method used	Dataset	Advantages	Disadvantages	Accuracy
1.	Inception Model	Herlev dataset	High Accuracy Good Universality Low Complexity	Need deep Network study to investigate cervical cells.	89%
2.	Transfer Learning based on pretrained dataset	Fujian Maternal child health Hospital Kaggle	More feasible and effective.	Only for Limited data	90%
3.	CNN-Extreme Learning Machine (ELM) based system	Herlev Dataset	Fast Feature Learning Easy Convergence Less randomized values	More Complexity in data Need more investigations	93%
4.	Gene assistance module voting strategy	Chinese hospital and Universitario decarcas, Venezuela	More Scalable and Practical	Only for Limited Datasets	92%
5.	Random forest and Adaboost algorithm	Radiotherapy dataset	Better treatment planning	Need to extract features Painful Treatment	94%
6	Colponet	Colposcopy image dataset	Better Accuracy Efficient Classification	Need to improve accuracy by extracting relevant features	93%
7.	CNN model	Papanicolaou Stained Cervical smear dataset	Better Sensitivity and Specificity	Reported 18% false negative images	90%
8.	Fourier Transform and Machine learning methods	Microscopic Image dataset	Fully automatic system Saving precious time for the microscopist	The level of Complexity is more.	94%
9.	CNN-SVM model	Herlev and One private dataset	Good Robustness Highest Accuracy	Need improvement to adjust parameter Need to hand crafted features	95%
10.	Stacked Autoencoder	UCI Database	Higher Accuracy Reduced data dimension	Training time is very high due to deducing the dimensions	94%

calculations for cervical malignant development clearly involve the Support Vector Machine [SVM] and its modifications the most. The primary argument for scientists' choice of the support vector machine can be attributed to how rigorous the calculations are. Incorporating data into robust models can protect them from irrelevant and perhaps misleading data. Help vector machines have the advantage of making it possible to demonstrate non-straight limits by using non-direct portions. Even if the preparation dataset used is biased, SVM addresses the issue of overfitting and summarizes the results in any case. The limitations of SVM include the results' lack of clarity and the lengthy computation required when using a large dataset. The clinical conclusion's dataset is severely unbalanced and typically consists of images. Information pre-handling practices and division techniques like these play a vital role in this regard. The information is pre-handled using techniques like k-implied grouping and gaussian channel. For the best execution, we need to choose the optimal combination of information preprocessing approach and arrangement computation. According to the need, the overview highlights the fact that a startling variety of datasets can be used for the diagnosis of malignant growth in the cervix region. Segmenting data based on Pap test, biopsy results, patient lifestyle propensities, and prior clinical records. The most recent development is the use of data from Fourier-Change Infrared (FTIR) spectroscopy. Since a significant percentage of the information in the finding is contained in picture design, there is some room

for employing and engaging in-depth learning in the field of disease expectation. Despite having huge potential due to the absence of large datasets, increased computing load, a propensity for overfitting, and the precise nature of model outcomes, it has not been utilized to its full potential. This leaves doors open for future research on a similar topic.

VI. Conclusion

Machine learning in medical data diagnosis is becoming one of the most promising and growing fields. Cervical cancer is one of the deadliest forms of cancer but can be treated if diagnosed in the early stages. Therefore, proposing a computerized approach for the diagnosis of cervical cancer is very imperative. The research papers from 2006-2017 have been reviewed in this survey paper and the findings based on the review has been stated. The various machine learning algorithms that have been used along with the features, datasets used and accuracy have been mentioned. A comparative approach has been followed. Also, the scope of deep learning in the prediction of cervical cancer using machine learning has been highlighted.

References

[1] Bargana Benair, Nagarajan “An Expert system for Predicting the cervical cancer using Data mining Techniques”, International Journal of Pure and Applied Mathematics(2018), Volume 118.

[2] Manoj Sharma Cervical Cancer prognosis using Genetic Algorithm and Adaptive Boosting approach”, Springer Publications(2019).

[3] Herlambang Brawijiya, Slamet Widodo, Samudi Improving the Accuracy of Neural Network with Genetic Algorithm for Cervical cancer prediction”, IEEE(2019).

[4] F.J. Shaikh, D.S. Rao,” Prediction of Cancer Disease using Machine learning Approach”, Elsevier materials today, 2021

[5] Roqia Saleem Awad Maabreh,1 Malik Bader Alazzam,"Machine Learning Algorithms for Prediction of Survival Curves in Breast Cancer Patients”, Hindawi Applied Bionics and Biomechanics, Volume 2021, Article ID 9338091, 2021.

[6] Ye Rang Park, Young Jae Kim,” Classification of cervical cancer using deep learning and machine learning approach”,Research Square, 2020

[7] Sohely Jahan· M. D. Saimun Islam · Linta Islam,” Automated invasive cervical cancer disease detection at early stage through suitable machine learning model”, Springer SN Applied Sciences journal, 2020.

[8] Dana Bazazeh and Raed Shubair,” Comparative Study of Machine Learning Algorithms for Breast Cancer Detection and Diagnosis”, IEEE journal, 2020.

[9] W. H. O. R. Health, W. H. O. C. Diseases, and H. Promotion, Comprehensive Cervical Cancer Control: A Guide to Essential Practice, World Health Organization, Geneva, Switzerland,2019.

[10] Cervical cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow up to Colombo and others Annals of Oncology, 2018.

[11] I. C. Scarinci et al., “Cervical Cancer Prevention: New Tools and Old Barriers Cancer,” vol. 116, no. 11, 2018.

[12] John Doorbar, “Molecular biology of human papilloma virus infection and cervical cancer”, Clinical Science, May01, 2018.

[13] D. Saslow et al., “American cancer society, American society for colposcopy and cervical pathology, and American society for clinical pathology screening guidelines for the prevention and early detection of cervical cancer,” CA-Cancer J. Clin., vol. 62,pp.147–172,2017.

[14] Igor Kononenko, “Machine learning for medical diagnosis: history, state of the art and perspective”, Artificial Intelligence in MedicineVolume23, Issue1, August 2017, Pages 89-109.

[15] Nasser H. Sweilam, A.A.Tharwat, N.K.Abdel Moniem, ”Support vector machine for diagnosis cancer disease: A comparative study”, Egyptian Informatics Journal, Volume 11, Issue2, December 2018, Pages 81-92.

[16] Y.Weng, C.Wu, Q.Jiang, W.Guo and C.Wang, "Application of support vector machines in medical data," 2016 4th International Conference on Cloud Computing and Intelligence Systems(CCIS),Beijing ,2017,pp.200-204.

[17] W.Wu and H.Zhou,"Data-DrivenDiagnosisofCervicalCancer With Support Vector Machine-Based Approaches," in IEEE Access, vol. 5,pp.25189-25195,2017.

[18] Y.E.Kurniawati, A.E.Permanasari and S.Fauziati,"Comparative study on data mining classification methods for cervical cancer prediction using pap smear results," 2017 1stInternational Conference on Biomedical Engineering(IBIOMED), Yogyakarta,2016,pp.1-5.

[19] Priyanka K Malli, Dr. Suvarna Nandyal, “Machine learning Technique for detection of Cervical Cancer using k-NN and Artificial Neural Network”, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), July-August 2017.

[20] R. Vidya1* and G. M. Nasira2, “Prediction of Cervical Cancer using Hybrid Induction Technique: A Solution for Human Hereditary Disease Patterns”, Indian Journal of Science and Technology, August 2017.

[21] D. Kashyap et al., "Cervical cancer detection and classification using Independent Level sets and multi SVMs," 2017 39thInternational Conference on Telecommunications and Signal Processing (TSP), Vienna,2017, pp.523-528.