OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

A SURVEY ON BREAST CANCER DIAGNOSIS BASED ON MACHINE LEARNING TECHNIQUES

Kishore Mishra[1], Geeta Tiwari[2]

Assistant Professor, Computom Institute of Information Technology and Management, Jaipur[1]

email : kmishra16@gmail.com

Associate Professor, Computom Institute of Technology and Management, Jaipur[2]

Abstract: ***The oldest recognised cancer type in humans is Breast Cancer. In approximately 1600 BC, Egypt recorded the oldest cancer detection and definition. Although this problem has subsequently been evaluated and investigated for the purpose of avoiding serious consequences, it remains one of the deadliest diseases ever, with breast cancer mortality in the U.S alone exceeding 40,000 in 2012. In most cancer-affected women, cancer of breast is the leading cause of mortality. With a low mortality rate, mammography is one of the most successful early detection and diagnostic methods for cancer of breast . Breast x-rays, or mammograms, are used to identify early symptoms of breast cancer.These X-ray images reduce human cyst detection errors, reduce diagnostic time, and increase diagnosis accuracy.This study provides an overview of strategies machin learningn for diagnosis the breast cancer & classification and may be divided into three primary stages: pre-processing, feature extraction and classification. There are many innovative methods and approaches for timely identification of breast cancer in modern medical science. Most of these approaches involve state-of-the-art technologies like medical image processing. Early detection is useful for doctors, which considerably increases cancer patients' survival rates. This research examines three of the most used ML approaches for the diagnosis & direction of breast cancer. The techniques are the SVM, Random Forestry (RF) and the Naive Bayes (NB). For each of the above strategies, the probability will be computed and the most accurate algorithm results will be obtained.***

Keywords: *Medical Diagnosis, Breast Cancer, Classification, Machine Learning .*

----------∴ ∴ ∴----------

I INTRODUCTION

"Breast cancer" is exceedingly common and, because of its fatality rate, it is regarded to be the second dangerous disease in the world. If the illness diagnoses occur prior to substantial physical changes in the body, the affected individual may survive. Physicians examine and propose biographies when mammograms detect anomalies. In the event of a mis-diagnosis, the recommendation of a radiologist is highly critical at this time; But it takes more time to manually diagnose cancer, and so a high level of accuracy requires machine learning. Male and female cancer of the breast is prevalent, however it mainly affects the female. In the last day, 90% of women are more afflicted and 50% die. Statistics on the cancer of the United States showed that only 40,000 breast cancer fatalities in one year in the US occurred. The statistics have shown that this condition is quite common in women and a lot of work has been done and done to control such a deadly disease. There is no new study on breast cancer medicinal products and their roots date back to the 16th. Because of the lack of information and progress in the medical profession, the condition has continued to edge the human edge and has still been one of the deadliest diseases ever. A new diagnose mechanism is being introduced termed medical image processing, with recent advances in medical science and more particularly the participation of information-technology in medicine. The processing of medical images is not just restricted to cancer, but also helps in diagnosing various types of ailments, and this can be seen through statistics. With the use of imaging methods, tumour from an infected breast has become easier to detect and breast cancer diagnosis. Early detection can contribute to the correct

diagnosis and therapy that reduces the chance of most undesired illness outcome (death) [1].

In UAE[2] , 40 thirds of women patients with cancer of the breast are diagnosed. For several physicians, forecasting cancer progression accurately remains a difficult endeavour. The introduction of the latest 'medical technologies' & hence the massive lot of patient information have given impetus to the development of latest cancer prediction and cancer diagnosis approaches. Recurring breast cancer is one that returns after a time when the cancer cannot be diagnosed in the same or opposite breasts or bruising walls. While the diagnostic approach is largely supported by the information assessment gathered from patients and doctors, supportive instruments could be overlaid to help make the diagnosis easier. These tools are considered to remove errors of diagnostic & give a fastest way to analyse huge data sets.

"Machine Learning" is an (AI)Artificial Intelligence sub-field, which allows machines to learn through the knowledge of a given task without explicit programming. In recent decades, machine learning methodologies were widely used to construct predictive models to promote effective decision-making. These algorithms could be utilised in cancer analysis to discover entirely new patterns during a data collection, thereby forecasting whether or not a cancer is aggressive or benign. The results of these strategies will be assessed to enhance classification accuracy, reminder, accuracy and consequently the space below ROC. 1.2 Processing of data Recently, database processing has become a popular tool for data detection and extraction from enormous datasets with hidden patterns. The use of sophisticated knowledge manipulation tools is required to provide a solid dataset with unknown, valid patterns and linkages. We tend to use three strong data processing methods - SVM, Random Forest and Naive Bayes, a medium-sized known collection of thirty-five attributes and 198 patient cancer data. [3].

Types of breast cancer

Breast cancer types vary depending on the area and breast affected. Ductal carcinomas, in-vasive ductal carcinomas and invasive lobular carcinomas are among the most prevalent forms. Carcinomas tumours that begin in epithelial cells that line the body's organ and tissue [4].

In situ cancers

- "Ductal carcinoma in situ" is a pre & non- invasive cancers of breast, it is startedfrom inside of the milk ducts.
- "Lobular carcinoma in situ"is an area of growth of cell abnormality in lobules, Milk producing glad at the breast ducts end.
- Breast cancer that has progressed into the surrounding tissue is known as (infiltrating) invasive cancer of the breast.

Breast cancer common types

- Inflammatory breast cancer is an uncommon and severe type of cancer of breast that quickly spreads & grows.
- Nipple paget disease is a very rare form of breast cancer. It starts in the breast canals and advances to the nipple skin and subsequently to the arenola.
- Phyllodes tumours are a type of breast cancer that is rather uncommon. They form in the breast stroma (connective tissue). The majority are harmless, but some are cancerous (cancer).

MACHINE LEARNING TECHNIQUES

The ML-technical learning method is frequently classified into 2 primary groups, supervised and unattendedIn supervised learning, a collection of data examples is used to train the system and is labelled to ensure that the correct output is achieved. Uncontrolled learning, on the other hand, has no preexisting knowledge sets or conceptions of the desired outcome, implying that the objective is difficult to achieve.

Support Vector Machine:

The Vector Support Machine is a monitor for classifying the method for machine learning frequently used in cancer prediction. In all categories, SVM attempts to choose the most important instances. The names for these examples are support vectors. These classes are split into a (linear) function which divides as broadly as possible by such a support vector.

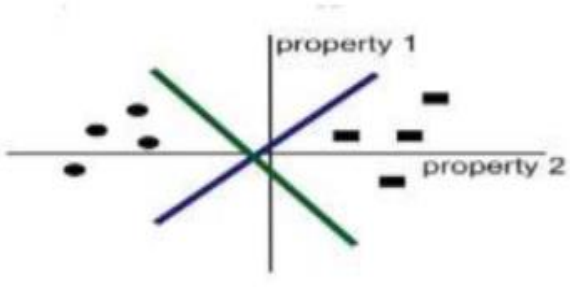

Fig. 1. SVM generated hyper-planes

Random Forest:

"RF" takes many decision-making trees to join the trees forest. An algorithmic technique is used for selecting a

random sample size N from the replacement data set or a random sample from the forecasts without substitution. The acquired information is subdivided. The remaining information is then deleted. Depending on the number of trees necessary, these operations are repeated several times. Finally, the trees sort the observations into one of several categories and count them. Cases are then classed on the basis that the decision tree is voted by a majority.

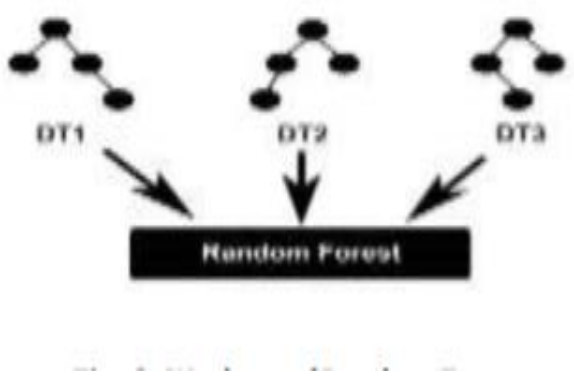

Fig. 2. Working of Random Forest

Naive Bayes:

Naive Bayes is a subfield of visual probabilistic models that serve to predict[3] and display knowledge in ambiguous areas. Naïve Bayes matches a widely used machine learning structure known as an acyclic directed graph (DAG). It contains of numerous node, each equivalent to the variable,& the nodes of edge are directly dependent on the nodes in the chart [3].

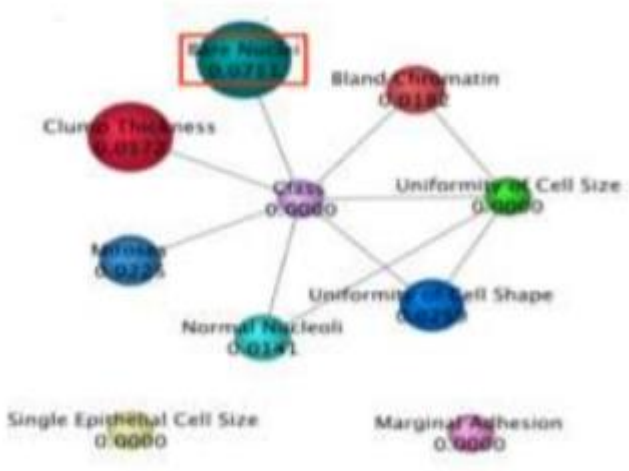

Fig. 3. Breast cancer attributes – DAG Model

II LITERATURE SURVEY

Arpit B. and Aruna T. [5] Proposed a GENN(GONN) for the classification of breast cancer (malignant and benign). By introducing novel crossover and mutation operators, they have optimised neural network design. They compared classification accuracy, sensitivity, specificity, confusion matrix, ROC, and AUC un-the GONN ROC curves to classical and classical models, using WBCD to evaluate their work. This method offers a good classification system. However, by utilising a bigger dataset than WBCD, GONN's efficiency for real-time breast cancer diagnosis can be improved.

Ashraf O. I.and Siti M. S. [6]Computer-based method for automatically classifying illness of breast cancer. To enhance both accuracy and network structure, a neural network based on the improved, non-dominated genetic sorting algorithm (NSGA-II) was applied to a multilayer perceptron (MLP). This study increases classification accuracy when compared to previous techniques. MLP, on the other hand, has the ability to remain in local minima.

Na L.et.al. [7] Classification intelligent model for the diagnosis of cancer of breast-cancer based on the “hybrid technique”. for selection,s of features: Gaining of simulated re-induced gene algorithm wrappers (IGSAGAW); Removal of the redundant feature irrelevance from the feature space; This procedure can increase the accuracy of classification and reduce calculation costs. The proposes the method is being tested on “Wisconsin Original Breast Cancer” (WBC) and Wisconsin Breast Cancer Disease (WBCD). The proposed method performs well and decreases the number of calculations required.

Nawel Z.et.al. [8] The design & implementation of “Computer Assisted Detection” for the classification,s of mammography images were presented. To reduce the dimensionality of the vector feature, the system uses a GA-based function approach for classification and a semi-supervised vector support machine (S3VM).Digital ScreeningMammography Database (DDSM) data set has been certified for experiments. The approach proposed improves precision.

Abdulkader H. et al. [9] Automated breast tissue classification method has been created. The BackPropagation Learning Algorithm is one of two machine learning methods used by the system (BPNN). The neural network's feeds (RBFN). Breast cancer tissue is divided into six types: carcinoma, fibro-adenoma, mastopathy, and others. To categorise six distinct breast tissues, researchers utilised electrical impedance spectroscopy (EIST). In terms of accuracy, lowest errors, maximum periods, or training time, the radial network outperformed the back propagation network. The proposed method increases accuracy while reducing training time.

Kemal P. et al. [10] Proposed mad normalisation hybrid approach, KMC based weighted feature and AdaBoost classification 1. Proposed Hybrid approach. The existence of breast cancer can be detected in three steps: The MAD normalisation approach was employed to normalise the dataset in the first stage. The weighting of normalised data was done in the second stage using k-means clustering. Weighting feature Lastly, a weighting data set classification was applied in the AdaBoostM1 classification. The coimbra (BCC) breast cancer dataset extracted from the study database UCImachine was used. The accuracy of this procedure is good. But it's a computer-consuming process.

Teresa A. j. et al. [11] The proposed classification approach for the use of Convolutionary Neural Networks for hematoxyl and eosintained breast biopsy pictures (CNNs). Normal tissue, benign lesions, carcinoma in situ, and aggressive malignancy are the four types of medically important tissue. The suggested CNN architecture aims to incorporate information from several scales. The model is used to classify photos from the BioImaging 2015 Brast Histological Classification Challenge, which included high-resolution unpressed, annotated HE Stain images.

Fabio A. et al. [12] BreaKHis, a public database of breast cancer pathologic images, was classified using a deep learning technique. They presented a strategy grounded on the abstraction of picture patches for CNN development and the combination of these finishing classification patches. This way, model modifications can be avoided, as well as more sophisticated and costly design. However, a lack of data results in costly studies.

Hiba A. et al. [13] The performance of four alternative classifiers was compared using the support vector machine , tree decision (C4.5),k Nearest neighbours (k-NN) & nave bays (NB).

III CONCLUSION

This study discusses several aspects that affect agriculture's overall production. In this article, we evaluate and discuss the outcomes of a variety of research carried out over a period in the subject of plant disease detection. This publication also gives a brief overview of many diseases that affect the growth and symptoms of plants. The processing of images from the acquisition to categorization takes five main processes. We concluded that a deep learning system delivers the greatest outcomes with less computer effort and less predictive time. We have also determined that the visual approach for the diagnosis of plant diseases produces more satisfactory results than non-visual systems. Digital imaging results are more efficient, faster and cheaper than manual plant analysis for the diagnosis of diseases.

REFERENCES

[1] Mussarat Yasmin, Muhammad Sharif and Sajjad Mohsin, “Survey Paper on Diagnosis of Breast Cancer Using Image Processing Techniques”, Research Journal of Recent Sciences, Vol. 2(10), October (2013), pp.88-98.

[2] “Breast cancer presentation delays among Arab and national women in the UAE, a qualitative study,” Y. Elobaid, T.-C. Aw, J. N. W. Lim, S. Hamid in Mar. 2016.

[3] Lingaraj N, Krishna Kumar S, Banu Priya K, R.M. Shiny, “SURVEY ON ANALYSIS OF BREAST CANCER PREDICTION”, International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 03, Mar 2020, pp.1281-1284.

[4] http://www.springer.com/lncs. Last accessed 1 February 2019

[5] Arpit, B., Aruna, T.: Breast Cancer Diagnosis Using Genetically Optimized NeuralNetwork Model. Expert Systems with Applications, Elsevier, pp. 1-15(2015).

[6] Ashraf, O. I. and Siti, M. S.: Intelligent breast cancer diagnosis based on enhancedPareto optimal and multilayer perceptron neural network. International Journalof Computer Aided Engineering and Technology, Inderscience, Vol. 10, No. 5, pp.543-556(2018)

[7] Na, L., Qi, E., Xu, M., Bo, G., Gui-Qiu, L.: A novel intelligent classification modelfor breast cancer diagnosis. Information Processing and Management, Elsevier, pp.609-623(2019).

[8] Nawel, Z., Nabiha, A., Nilanjan, D., and Mokhtar, S.: Adaptive Semi SupervisedSupport Vector Machine Semi Supervised Learning with Features Cooperation forBreast Cancer Classification. Journal of Medical Imaging and Health Informatics,American scientific publisher, pp. 53-62(2016).

[9] Abdulkader, H., John, B. I., Rahib, H.A.: Machine learning techniques for classifi-cation of breast tissue. In: 9th International Conference on Theory and Applicationof Soft Computing, Computing with Words and Perception, ICSCCW, pp. 402-410. Procedia Computer Science, Elsevier. Budapest, Hungary (2017).

[10] Kemal, P., and mit, .: A Novel ML Approach to Prediction of Breast Cancer:Combining of mad normalization, KMC based feature weighting and AdaBoostM1classifier. In: 2nd International Symposium on Multidisciplinary Studies and Inno-vative Technologies (ISMSIT),IEEE, pp. 1-4 (2018).

[11] Teresa, A., Guilherme, A., Eduardo, C., Jos, R., Paulo, A., Catarina, E., Antnio, P.,and Aurlio, C.: Classification

of breast cancer histology images using ConvolutionalNeural Networks. PLOS ONE, pp. 1-14 (2017).

[12] Fabio, A. S., Luiz, S. O., Caroline, P., and Laurent, H.: Breast Cancer Histopatho-logical Image Classification using Convolutional Neural Networks. In: Conference:International Joint Conference on Neural Networks (IJCNN), pp. 1-8(2016).

[13] Hiba, A., Hajar, M., Hassan, A. M., and Thomas, N.: Using Machine LearningAlgorithms for Breast Cancer Risk Prediction and Diagnosis. In: The 6th Interna-tional Symposium on Frontiers in Ambient and Mobile Systems (FAMS),ProcediaComputer Science, Elsevier, pp. 1064-1069 (2016).