Breast Cancer Diagnosis Using Deep Learning Algorithm

Naresh Khuriwal
Department of Computer Engineering
Poornima University Jaipur
Jaipur, India
naresh.khuriwal89@gmail.com,

Dr Nidhi Mishra
Department of Computer Engineering
Poornima University Jaipur
Jaipur, India
nidhi.mishra@poornima.edu.in

***ABSTRACT* -- The cancer is the most dangerous diseases in the world, its mainly effective for women. So,our prime target must be curing the cancer through scientific investigation and the second main target should be early detection of cancer because the early detection of cancer can be helpful for remove the cancer completed. After reviewed 41 papers we found that several techniques are available for cancer detection. In this paperwe proposedDeep Leaning algorithm neural network for diagnosed breast cancer using Wisconsin Breast Cancer database. The paper shows how we can use deep learning technology for diagnosis breast cancer using UCI Dataset. Because deep learning techniques almost used for high task objective Computer Vision, Image processing, Medical Diagnosis, Neural Language Processing. But in this paper, we are applying deep learning technology on the Wisconsin Breast Cancer Database and we have seen that is very beneficial for us for diagnosis breast cancer with accuracy 99.67%. This paper is divided in three parts first we have collect dataset and applied pre-processing algorithm for scaled and filter data then we have split dataset in training and testing purpose and generate some graph for visualization data. In last implement model on training dataset and achieved accuracy 99.67%. So, we have seen deep learning technology is a good way for diagnosis breast cancer with Wisconsin Breast Dataset. This database provides 569 rows and 30 features in the dataset. In this paper we have used 11 features for diagnosis breast cancer that we have got after pre-processing. But before train model we have applied some pre-processing algorithm like Label Encoder, Normalizerand StandardScalerfor scaled dataset then applied model and achieved accuracy. In this paper we also compare deep learning algorithm with other machine learning and seen our proposed system is proved best from others machine learning algorithm.**

Keyboards–Deep Learning, Convolutional Neural Network,Neural Network, Random Forest, Support Vector Machine, MachineLearning,WDBCDataset.

I. INTRODUCTION

Breast Cancer is the most common female cancer worldwide. All most 25% of all cancers with an estimated 1.67 million new cancer cases diagnosed in 2012. Breast cancer is the most common type of cancer among women and its incidence is increasing day by days. The life time risk of developing breast cancer in women is approximately 1/8 in USA, 1/ 12 in Europe, 1/40 in Asia (WHO 2008). As per the cancerindia.org one woman dies of cervical cancer every 8 minutes in India, around 2.5 million number of people living with cancer in India and over 7 lakh new cancer patients registered in every year. India also breast cancer is the most common cancer in women and account for 27% of all cancers. As per the previous study 1,44,937 new cases registered and deaths 70,218 in 2012. But in India the incidence rates begin to rise in the early thirties and peak at ages 50-60 years. According to the World Health Organization breast cancer was responsible for 502,000 deaths in 2005 alone and 1,301,867 of new cases of breast cancer resisted. Health case authorities continuously doing efforts to overcome this merciless disease in which one of the efforts is screening. By screening the breast cancer can be detected in early stages and thus the treatment can be more effective. Many other methods also available such as mammography, ultrasound, CT and MRI. Mammography is the most widely used screening method. In our proposed method we are using Break His[A1] dataset that provided Biopsy Images generated by Mammography method.The rest of the paper is organized as follows: Section 2 presents the related work that cover 10 author research work. Section 3 includes the architecture of the proposed work.Section 4 describes the methodology which is used for the proposed work. In section 5 we discuss the model implementation of proposed work.Section 6discusses the results and Section 7concludes the proposed work.

II. RELATED WORK

There are many deep learning and machine learning techniques are available for cancer detection and prediction. Some of most used deep learning techniques are for breast cancer diagnosis are Convolutional Neural Network, Recurrent Neural Network and some pre-trained model Alex Net, Google Net, VGG16, VGG19, ResNet. Some most used dataset also available for training and testing are Mammogram image, SEER, UCI, WBCD,The author Abdullah-Al Nahid developed Deep Learning Model utilizing a restricted Boltzmann machine that mainly used

ISBN: 978-1-5386-4119-4/18/$31.00 ©2018 IEEE

back propagation algorithm for classify histography images [1].The author Mohamad Mahmoud Al Rahhal used deep learning algorithm convolutional neural network for classification of histogram images and found 86.60% accuracy [2]. The author Moi Hoon Yap proposed three ultrasound-based method, a patch based LeNet, U-Net and transfer learning model FCN-Alex Net and achieved 98% accuracy [3]. The author Teresa Araujo proposed deep learning-based approach for classification of haematoxylin and eosin stained breast biopsy image using CNN algorithm and achieved 83.30% accuracy [4]. The author Hongchao Song presented the empirical mode decomposition-based feature extraction method and achieved 87% accuracy [5].The author Md. Milon Islam presented a novel modality for the breast cancer detection and discussed about two supervised machine learning algorithm and achieved 95.22% accuracy [6]. The author R.D.Ghongade proposed the RF random and RF-ELM classifier for determine the breast tumour from digital mammograms and achieved 98% accuracy [7]. The author Sudarshan Nayak presented the various machine learning techniques for determine the breast cancer from using 3D images and achieved 98% accuracy with SVM [8]. The author Yinchong Yang proposed encoder and decoder network approaches for create physician therapy for decision [9]. The author YohannesTesehay developed a weakly supervised computer-aided detection system that used for identify prostate cancer on mpMRI [10].

III. FLOW DIAGRAM OF PROPOSED WORK

The Fig1 shows the process flow diagram or proposed work. In this paper we used Wisconsin Breast Cancer Dataset that publicly available for download and used. In the second step we pre-processed the dataset. In the pre-processing we have used three different method for scaled and pre-process dataset. After that applied Deep Learning Neural Network and compute the accuracy. We have seen CNN achieved 99.67% accuracy.

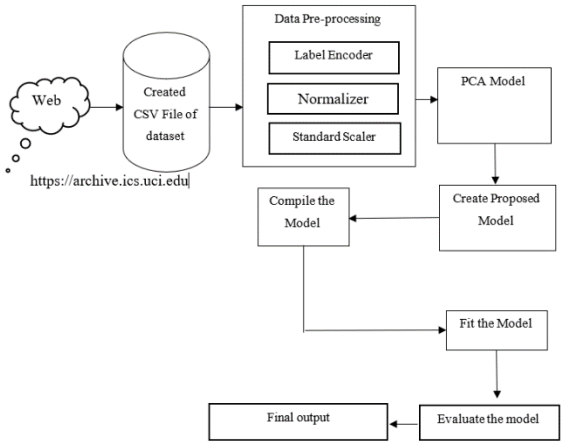

Fig. 1. Process Flow Diagram

IV. EXPERIMENT AND METHODOLOGY

In this paper we have used convolutional neural network for diagnosis breast cancer and also implemented same dataset on other machine learning algorithm such as Neural Network, Support Vector Machine, Random Forest. All paperconsists main two parts: predict models,Pre-processing data. In this paper we have used Wisconsin Breast Cancer Dataset that are publicly available for researcher this database is generated from biopsy images that having 569 rows and 30 columns of dataset. In the experiment part first, we have pre-processed the dataset and extracted useful feature using RFECV model then implementsome encoder and found final training dataset. For the trained the dataset we have used deep learning model convolutional neural network algorithm and achieved 99.67% accuracy.

V. PRE-PROCESSING DATA

As you know that data pre-processing is a data mining technique that used for filter data in a usable format. Because the real-world dataset almost available in different format. Its not available as per our requirement so its must be fitters the dataset in aunderstandable format Data pre-processing is a proven method of resolving such issues. Data pre-processing convert the dataset in to usable format for pre-processing we have used standardization method to pre-process the BCIdataset.

A. Label EncoderMethod:

Label Encoder is efficient tool for encoding the levels of the categorical features into numeric values. Label Encoder encode labels with value between 0 and classes -1. All our categorical features are encoded. In this paper we have classify malignant and Benign in diagnosis with 0 and 1. After encoding dataset we have applied neural network on these dataset and achieved accuracy. But still the achieved accuracy is not good for breast cancer diagnosis.

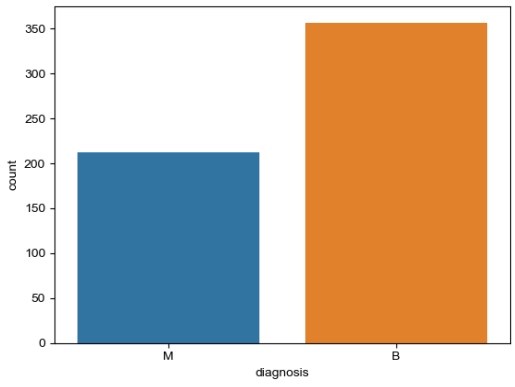

Fig. 2. Number of Malignant and Benign

B. *Normalizer Method*

Data normalization is the process of rescaling one and more attributes to the range of 0 to 1. This means that the largest value for each attribute is 1 and the smallest value is 0. Normalization is a good technique to use when you do not know the distribution of your data. After applied Label Encoder method your all dataset is converted into numeric dataset. Because Label Encoder encode all string label dataset into numeric dataset. We have known normalization process work on numeric dataset. So after normalized data we have applied neural network on that dataset and achieved accuracy. But again accuracy is not enough till now.So we have applied next process for pre-processing.

$$\frac{x_i}{\sqrt{x_i^2 + y_i^2 + z_i^2}} \qquad 1$$

C. *StandardScaler Method*

Data standardization is the process of rescaling one or more attributes so that they have a mean value of 0 and a standard deviation of 1. This technique is more affective if dataset attribute have distribution is Gaussian.After scaling data we have predict model and seen that accuracy is best comparing to another method.

$$\frac{x_i - \text{mean}(x)}{\text{stdev}(x)} \qquad 2$$

After data pre-processing we have got 11 unique features that having titled coupled relation to each other. The table 1 show all 11 features that we have used formodel trained.

TABLE1 Total number of selected features

Sr. No	Attributes	Group
1	Diagnosis	Class
2	Texture Mean	Mean
3	Area Mean	Mean
4	Concavity Mean	Mean
5	Symmetry Mean	Mean
6	Fractal Dimension Mean	Mean
7	Area Se	Standard Error
8	Fractal Dimension Se	Standard Error
9	Smoothness Worst	Worst
10	Concavity Worst	Worst
11	Symmetry Worst	Worst
12	Fractal Dimension Worst	Worst

D. *Principal Component Analysis*

Now, we perform a PCA on the standardized and the non-standardized dataset to transform the dataset into a 2-dimensional features subspace. In this paper we have applied PCA on both dataset without standardized dataset and on non-standardized dataset. Fig 3 shows these both dataset in scatterplot.

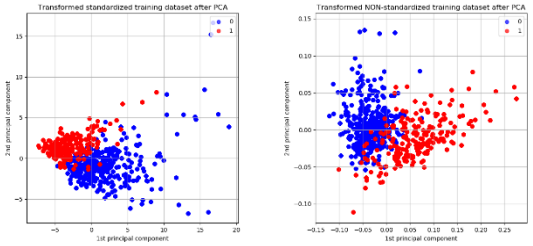

Fig.3Standardized and non-standardized dataset after PCA

PCA have feature subspace that maximizes the variance along the axes. The Eigen Vectors and eigenvalues of a covariance matrix represent the core of a PCA. The eigenvectors determine the direction to the new features space and eigenvalues determine their magnitude. PCA perform the Eigen decomposition on the covariance matrix E. The covariance between two features can calculate by formula.

$$\sigma_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - x_j)(x_{ik} - x_k) \qquad 3$$

Where x bar is the mean vector

The explained variance tells us how much information can be attributed to each of principal component. The plot in fig 4 clearly shows that most of the variance can be explained by the first principal component.

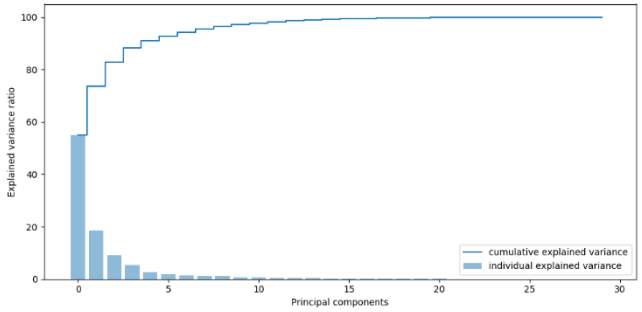

Fig. 4.Explained Variance Ratio

VI. MODEL IMPLEMENTATION

In this stage we have implementedDeep Learning Neural Network algorithm on UCI dataset.Neural network in work on a human biological method. In which we have to pass input and get output. But in these two layer some hidden layer are work and some additional process must be added before calculate final output. These additional processes are added unit of bias, add some of additional hidden layer, calculate some activation function then the final output generated. In this paper we have used following parameters for calculating and trained model.

TABLEII. USED PARAMETERS IN CNN MODEL

Number of input	**12**
Number of Neurons	12
Activation function	Sigmoid
Number of Epochs	20

$$X = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad 4$$

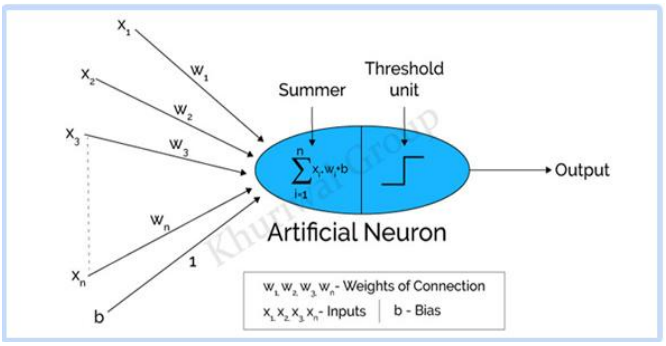

Fig 5. Shows Neural Network Functionality

When we want to use perceptron bias b = - threshold then perceptron rule rewritten. Fig show the basic functionality of the neural network. In which after the fixed position the neurons are fired that is called threshold condition.

TABLEIII. CNN MODEL SUMMARY

Layer Name	Type	Output Shape	Parameters
Convolutional Layer	Dense	12	496
Hidden Layer	Dense	8	136
Fully Connected Layer	Dense	1	9
Total	params:	641	
Trainable	params:	641	
Non-trainable	params:	0	

In this paper I have create the four-layer convolutional model in which first is input layer. Convolutional layer, hidden layer and fully connected layers. The total number of neurons is 12 for first input layers and 8 neurons for hidden layers and final 1 neurons in output layers. The table III shows the total number of parameters that used in the model.

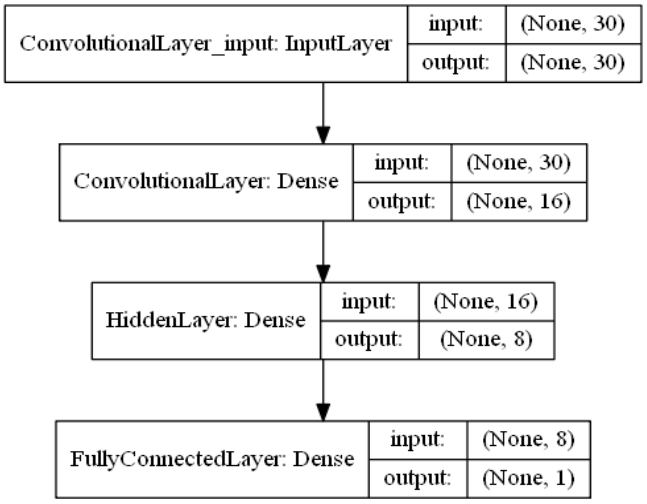

Fig 6. Deep Learning Model

Fig 6 shows when we have pass Wisconsin breast cancer database then this dataset enter as 30 features in the model after that some hidden layer work and finally output generated in two class malignant a Benign. In your model we have used 512 neurons and 600 epochs for trained the model. We have used 600 epochs because this a best and final position on which the accuracy will be stable and achieved best accuracy.

VII. RESULTS AND DISCUSSION

In this paper we proposed Deep Learning with Neural Network algorithm for diagnosis and detection of breast cancer. We have used standardization method for pre-processing breast cancer dataset then we have applied PCA algorithm. After applied PCA algorithm visualized non-transform and transform dataset after PCA. Then we have implemented neural network algorithm on these 11 features and achieved 99.67% accuracy.Wisconsin Breast Cancer Dataset havecontain 699 rows with features categories 30features.Achieved 99.67% accuracy is good from others compared machine learning algorithm.

TABLEIV. CNN MODEL RESULT

Epoch	Loss	Accuracy	Value of Loss	Value of Accuracy
1	0.0918	0.8965	0.0246	0.9724
2	0.0206	0.9746	0.0155	0.9821
3	0.0154	0.9803	0.0129	0.986

4	0.0134	0.9837	0.0117	0.9854
5	0.0119	0.9855	0.0107	0.9854
6	0.0108	0.9889	0.0099	0.9898
7	0.0099	0.9895	0.0095	0.9902
8	0.0091	0.9901	0.0089	0.9898
9	0.0086	0.9904	0.0083	0.9917
10	0.0079	0.9911	0.0077	0.9929
11	0.0074	0.9918	0.0071	0.9898
12	0.0068	0.9926	0.0065	0.9929
13	0.0062	0.9937	0.006	0.9914
14	0.0057	0.9945	0.0058	0.9944
15	0.0052	0.9949	0.0056	0.9944
16	0.0049	0.9959	0.0049	0.9944
17	0.0045	0.9961	0.0046	0.9967
18	0.0042	0.9964	0.0044	0.9967
19	0.0039	0.9969	0.0041	0.9967
20	0.0037	0.9969	0.004	0.9967
21	0.0035	0.997	0.0038	0.9967
22	0.0034	0.997	0.0037	0.9967
23	0.0033	0.997	0.0036	0.9967
24	0.0032	0.997	0.0035	0.9967
25	0.0032	0.997	0.0035	0.9967
26	0.0031	0.997	0.0035	0.9967
27	0.0031	0.997	0.0034	0.9967
28	0.003	0.997	0.0034	0.9967
29	0.003	0.997	0.0034	0.9967
30	0.003	0.997	0.0034	0.9967

TABLEV. ACCURACY OF PERPOSED WORK

Test Score	**0.003366**
Test Accuracy	0.9967
Accuracy	99.67%

TABLEVI. COMPARETIVE ANALISYS OF PROPOSED WORK WITH OTHER ALGORITHMS

Algorithm	**Precision**	**Recall**	**F-measure**
Neural Network	93.30	93.30	93.30
Nearest Neighbour	88.00	88.05	88.05
Random Forest	97.01	97.01	97.01
Support Vector Machine	95.01	61.01	74.01
Convolutional Network	99.60	99.50	99.60

The table VI shows the comparative analysis of proposed model with others machine learning algorithm. in this paper we have also trained WBCD dataset on four other machine learning algorithm and achieved accuracy that shows in table IV.

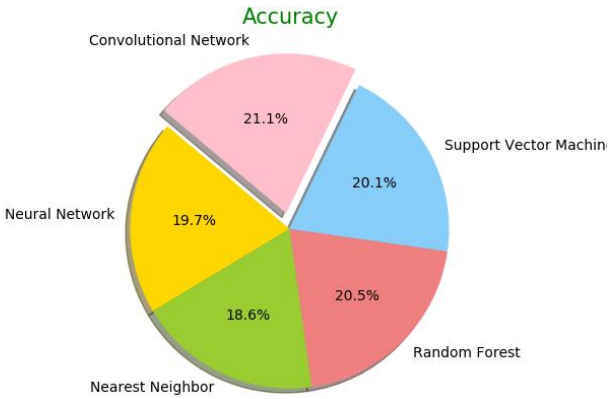

Fig 7. Accuracy Pie Chart

VIII. CONCLUSION AND FUTURE WORK

In this paper we proposed the deep learning method convolutional neural network that mostly used for classification of images dataset. After the implementation this method we have achieved 99.67% accuracy. In this paper we used only 12 features for diagnosis of cancer. In future we will try on images datasetwith Convolution Neural Network and will try toachieve best accuracy. Our work proved that Deep Learning neural network algorithm also effective for human vital data analyzation and we can do pre-diagnosis without any special medical knowledge.

References

[1] Abdullah-Al Nahid, Aaron Mikaelian and Yinan Kong, Histopathological breast-image classification with restricted Boltzmann machine along with backpropagation, Biomedical Research Volume 29, Issue 10, (2018).

[2] [24] Mohamad Mahmoud Al Rahhal, “Breast Cancer Classification in Histopathological Images using Convolutional Neural Network” International Journal of Advanced Computer Science and Applications (IJACSA), 9(3), 2018.

[3] M. H. Yap et al., "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks," in IEEE

Journal of Biomedical and Health Informatics, vol. 22, no. 4, pp. 1218-1226, July 2018.

[4] Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, et al. (2017) Classification of breast cancer histology images using Convolutional Neural Networks. PLoS ONE 12(6): e0177544. https://doi.org/10.1371/journal.pone.0177544.

[5] H. Song, A. Men and Z. Jiang, "Breast tumor detection using empirical mode decomposition features," in IEEE Access.

[6] M. M. Islam, H. Iqbal, M. R. Haque and M. K. Hasan, "Prediction of breast cancer using support vector machine and K-Nearest neighbors," 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, 2017, pp. 226-229.

[7] R. D. Ghongade and D. G. Wakde, "Detection and classification of breast cancer from digital mammograms using RF and RF-ELM algorithm," 2017 1st International Conference on Electronics, Materials Engineering and Nano-Technology (IEMENTech), Kolkata, 2017, pp. 1-6.

[8] S. Nayak and D. Gope, "Comparison of supervised learning algorithms for RF-based breast cancer detection," 2017 Computing and Electromagnetics International Workshop (CEM), Barcelona, 2017, pp. 13-14.

[9] Y. Yang, P. A. Fasching and V. Tresp, "Predictive Modeling of Therapy Decisions in Metastatic Breast Cancer with Recurrent Neural Network Encoder and Multinomial Hierarchical Regression Decoder," 2017 IEEE International Conference on Healthcare Informatics (ICHI), Park City, UT, 2017, pp. 46-55.

[10] Y. Tsehay et al., "Biopsy-guided learning with deep convolutional neural networks for Prostate Cancer detection on multiparametric MRI," 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, 2017, pp. 642-645.

[11] A. Alzubaidi, G. Cosma, D. Brown and A. G. Pockley, "Breast Cancer Diagnosis Using a Hybrid Genetic Algorithm for Feature Selection Based on Mutual Information," 2016 International Conference on Interactive Technologies and Games (ITAG), Notthingham, 2016, pp. 70-76.

[12] A. I. Pritom, M. A. R. Munshi, S. A. Sabab and S. Shihab, "Predicting breast cancer recurrence using effective classification and feature selection technique," 2016 19th International Conference on Computer and Information Technology (ICCIT), Dhaka, 2016, pp. 310-314.

[13] A. Osareh and B. Shadgar, "Machine learning techniques to diagnose breast cancer," 2010 5th International Symposium on Health Informatics and Bioinformatics, Antalya, 2010, pp. 114-120.

[14] A. Qasem et al., "Breast cancer mass localization based on machine learning," 2014 IEEE 10th International Colloquium on Signal Processing and its Applications, Kuala Lumpur, 2014, pp. 31-36.

[15] B. M. Abed et al., "A hybrid classification algorithm approach for breast cancer diagnosis," 2016 IEEE Industrial Electronics and Applications Conference (IEACon), Kota Kinabalu, 2016, pp. 269-274.

[16] B. M. Gayathri and C. P. Sumathi, "Comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, pp. 1-5.

[17] C. Deng and M. Perkowski, "A Novel Weighted Hierarchical Adaptive Voting Ensemble Machine Learning Method for Breast Cancer Detection," 2015 IEEE International Symposium on Multiple-Valued Logic, Waterloo, ON, 2015, pp. 115-120.

[18] D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), Ras Al Khaimah, 2016, pp. 1-4.

[19] D. T. Saleh, A. Attia and O. Shaker, "Studying combined breast cancer biomarkers using machine learning techniques," 2016 IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI), Herlany, 2016, pp. 247-251.

[20] H. R. Mhaske and D. A. Phalke, "Melanoma skin cancer detection and classification based on supervised and unsupervised learning," 2013 International conference on Circuits, Controls and Communications (CCUBE), Bengaluru, 2013, pp. 1-5. [3]