JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DETECTION OF BREAST CANCER USING DEEP NEURAL NETWORK (DNN)

[1]Mrs.G.Sharmila, [2]B.Swathi, [3]S.Soundarya, [4]S.Subaranjani

[1]Assistant Professor, Department of Computer Science & Engineering,

[2,3,4] UG Scholar, B.Tech, Department of Computer Science & Engineering,

[1,2,3,4] Manakula Vinayagar Institute of Technology, Puducherry, India.

ABSTRACT: Breast cancer is the most common type of cancer in women and is fatal if not detected and treated early. The cancer cell spreads to other parts of the body and progresses to the next stage if left untreated. This paper presents a machine learning model for the automatic diagnosis of breast cancer. The method used in this paper is DNN as the classification model and a random forest for feature selection. The proposed method detects breast cancer using the machine learning technique Deep Neural Network (DNN) and classification algorithms are used for image and text classifications. We discuss the types of breast cancer tumors and the challenges of early detection to prevent greater harm. Mammography image classification is performed using multi-class image classification. Random Forest (RF) classifiers are used to handle noisy multidimensional data in text classification. The RF model consists of a series of decision trees, each of which is trained on random subsets of features. To derive the most efficient machine learning model, BUI datasets are utilized. The latter approach is particularly interesting because it fits into the growing trend towards personalized predictive medicine. In composing this evaluation, we carried out thorough analysis of the many kinds of machine learning techniques, the types of data involved and the effectiveness of these methods in predicting cancer and prognosis.

INDEX TERMS- Healthcare system, machine learning, breast cancer, deep neural network, cancer diagnosis.

INTRODUCTION

Breast cancer is one of the most dangerous and prevalent cancers among women, causing the deaths of large numbers of women worldwide. Breast cancer is a type of cancer that begins at breast level. It can begin in either or both breasts. It is important to understand that the majority of breast sizes are benign and non-cancerous (malignant). Benign breast sizes are abnormal outgrowths which do not go beyond the breast. They are not fatal, but certain types of benign breast sizes may increase your risk of breast cancer. Breast cancer can spread when cancer cells penetrate the blood or lymphatic system and then spread to other parts of the body. More than one type of screening can be combined for best results, each with distinct advantages and disadvantages. There are several methods of breast cancer screening, such as ultrasound, magnetic resonance imaging (MRI), mammography, thermogram and then computed tomography (CT).The Mammogram is a type of X-Ray that is used to detect cancer in women with a range of age [50-70] years and also it can detect tiny tumors and get the accurate results [10]. Mammography may be used to detect breast cancer in women without symptoms. This type of mammogram is called tracing mammogram. Screening mammography usually involves two or more radiographic images of each breast. X-rays often show tumors that cannot be felt. Screening mammograms also make it possible to detect microcalcifications (tiny calcium deposits), which sometimes indicate the presence of breast cancer.

Mammograms can also be used to detect breast cancer if a lump or other signs or symptoms of the disease are found. This type of mammogram is diagnostic mammograms.

Computer-Aided Detection (CAD) systems to assist as a second reader with the radiologists' decisions to decrease both the false negatives or the false positives. The development and use of CAD systems have increased since 2001 and especially in the period from 2004 to 2008, which has increased by 91% either by private labs or hospitals [11]. Through the usage of CAD, it has been proved how it improves the process of cancer detection at earlier stages by decreasing the false positive and negative rates [12]–[15]. Not only this but also its use reduces the consumed time required by the radiologists to check the screened mammograms [16]. CADs are usually developed to localize the lesions existing in the screened breast mammogram [17], [18]. Mammography is considered one of the most common ways that can be applied to detect breast cancer in an earlier stage. The main reason behind making mammography the preferable choice for women is that it needs a very low dose of x-rays which makes it somehow safe unlike Magnetic Resonance Imaging (MRI) that needs a large dose of magnets and radio waves. Therefore, our proposed approach opts to

work with mammography datasets to detect existing cancers. Although many developments have been carried on to enhance the existing CADs for better detection accuracy, there are many challenges still acting as a barrier. The main problems in the already existing CADs are represented in the missing capability of applying them in the real-life systems used by hospitals or labs due to their long time for detection. Some CADs perform the detection in real-time but with lower accuracy. Hence, the good detection performance and the fast execution is somehow a trade-off problem. Digitized mammograms, on the other hand, tend to be large and deep pixels, requiring almost published documents to scale them

ready to form a specific pattern. The resizing results in missing some of the important information that may exist in the screened mammograms and consequently it may affect the obtained accuracy [19]. A Deep Neural Network (DNN) is an ANN with many hidden layers between the input and output layers. Like surface ANNs, DNNs can model complex nonlinear relationships. The main purpose of a neural network is to take a series of inputs, perform increasingly complex calculations on them, and send them to solve real-world problems like classification. We just feed the neural networks. Neural networks are widely used in supervised learning and reinforcement learning problems. These networks are based on a series of interconnected layers. Image classification is a classical problem of image processing, computer vision and machine learning fields. The Random Forest (RF) classifiers are suitable for dealing with the high dimensional noisy data in text classification. An RF model comprises a set of decision trees each of which is trained using random subsets of features.

I. LITERATURE REVIEW

In [1] The system will be already trained by giving various datasets initially so that when the user enters the input data it goes through the attributes and its values will be validated and it gives the result. As it has both classification and regression strategies it gives the best accuracy which got outcomes expresses that Naïve Bayes is the calculation with best classification precision of 97%, next RBF Network is the calculation best an exactness of 96.77% in classification, also J48 is the third with a classification precision of 93.41%.SVM have the best execution with Associate in nursing exactness of 97.07%. Limitations in this paper are although it is a common distance measure, Euclidean distance is not scale in-variant which means that distances computed might be skewed depending on the units of the features and then the cosine similarity is that the magnitude of vectors is not taken into account. The proposed model in [2] presents a comparative study of different machine learning algorithms, for the detection of breast cancer. Performance comparison of the machine learning algorithms techniques has been carried out using the Wisconsin Diagnosis Breast Cancer data set. In this paper, it has been observed that the accuracy of KNN is found to be 95.90%, the accuracy of RF equals 94.74% and accuracy of Naive Bayes equals 94.47%. And it is found that KNN is the most effective in detection of breast cancer as it has the best accuracy, precision and F1 score over the other algorithms. Thus supervised machine learning techniques will be very supportive in early diagnosis and prognosis of a cancer type in cancer research. Limitations in this paper is does not work well with large datasets: In large datasets, the cost of calculating the distance between the new point and each existing point is huge which degrades the performance of the algorithm. In [3] paper they proposed Deep Learning with Neural Network algorithm for diagnosis and detection of breast cancer. They have used the Wisconsin Breast Cancer Dataset and also implemented machine learning algorithms such as Neural Network, Support Vector Machine, Random Forest. In this paper they used only 12 features for diagnosis of cancer and achieved 99.67% accuracy. Their work proved that the Deep Learning neural network algorithm is also effective for human vital data analysis and can do pre-diagnosis without any special medical knowledge. A further drawback of data mining is its imperfect accuracy. In BC classification [4] It presents two different classifiers: the Naive Bayes (NB) classifier and the K-nearest Neighbor (KNN) classifier for breast cancer classification. In this, they propose a comparison between the two new implementations and evaluate their accuracy using cross validation. After an accurate comparison between our algorithms, we noticed that KNN achieved a higher efficiency of 97.51%, however, even NB has a good accuracy at 96.19 %, if the dataset is larger, the KNN's time for running will increase. Limitations in this paper are naive bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases. In KNN for large datasets, the cost of calculating the distance between the new point and each existing point is huge which degrades the performance of the algorithm. In disanosis using ML method [5] The main problem of the project is to detect breast cancer based on a set of features calculated from a digitized image of the Fine Needle Aspiration (FNA) of a breast mass from a patient. Classic machine learning models including Logistic Regression, Nearest Neighbor, Support Vector Machine, etc. are tested on the Breast Cancer Wisconsin dataset. In conclusion, by comparing with the 78% precision of the mammogram, the seven traditional models achieved a higher precision of about 95% for the breast. This paper demonstrates that machine learning models can be used for an automatic diagnosis for breast cancer. Limitations in this paper is when there is a large dataset, a linear SVM takes less time to train and predict compared to a Kernelized SVM in the expanded feature space. Kernelized SVM could overfit generating more complex trained SVM models when compared to a linear SVM. In Thermal imaging approaches[6] Building a screening method that does not cause body tissue damage (non-invasive) and does not involve physical touch is challenging. Thermography has been proposed as an early detection screening method. The asymmetrical thermal distribution on breast thermo grams can be evaluated using computer-assisted technology. A risk-free screening method using thermography could then be proposed for self-breast screening method at an early stage. The study was used in this article are image processing, convolutional neuronal networks, deep learning. Although thermography can show changes in heart and vascular features, it does not show how the breast has changed. It can detect changes that are not cancerous, and a person would need to have a mammogram to clarify the results. In [7] new breast cancer identification method by using machine learning algorithms and clinical data. A supervised (Relief algorithm) and unsupervised (Autoencoder, PCA algorithms) techniques have been used for related features selection from a data set and then these selected features have been used for training and testing of classifier support vector machines for accurate and on time detection of breast cancer. Additionally, k fold cross validation method has been used for model validation and best hyper parameters selection. One drawback of feature extraction is that the new features generated are not interpretable by humans. Limitations in this paper are feature selection is the problem of choosing a small subset of features that ideally is necessary and sufficient to describe the target concept. In [8] presented an efficient and automated approach to segment masses in the mammograms. They used hierarchical clustering to isolate the salient area followed by extraction of features to reject false detection. Applied their method to two popular publicly available datasets (mini-MIAS and DDSM). A total of 56 images from the mini-mias database and 76 images from DDSM were randomly selected. Results are explained in terms of ROC (Receiver Operating Characteristics) curves and compared with other state-of- the-art techniques. de-noising is applied after the top-hat morphological operator. Furthermore, image gray-level was enhanced by wavelet transform and a Wiener filter. And finally, the

segmentation method was employed using multiple thresholding, wavelet transform, and genetic algorithm. They used a manual process to reduce the false positives generated by genetic algorithms. SVM takes a long training time on large datasets. In [9] principal component analysis (PCA) was used to identify valuable parts of the data and further reduce the dimensions of the data. The cumulative proportion of the top five major components was 99.89%. The multilayer perceptron network (MLP) method was then used to extract characteristics included in the data, and the structure of the network was designed for the exploration of how data developed as the dimensions increased or decreased. Limitations in this paper are low interpretability of principal components. Principal components are linear combinations of the features from the original data, but they are not as easy to interpret. For example, it is difficult to tell which are the most important features in the dataset after computing principal components and currently, one ofthe biggest limitations to transfer learning is the problem ofnegative transfer.

II. PROPOSED METHODOLOGY

Our proposed system implements detection of breast cancer using one of the machine learning types which is Dnn (Deep Neural Network). A deep neural network (DNN) can be considered as stacked neural networks, i.e., networks composed of several layers. Deep neural networks (DNNs) have achieved unprecedented success in computer vision. Deep neural networks offer a lot of value to statisticians, particularly in increasing accuracy of a machine learning model and to handle unstructured data, unlabeled data, but also non-linearity as well. Breast cancer cells usually form a tumor that can often be seen on a mammogram or ultrasound or felt as a lump. Breast cancer is most common in women, but men also can get breast cancer . Breast cancer cells can spread to other parts of the body and grow there, too. When cancer cells do this, it's called metastasis. Machine learning detects the lumps or tumors in the breast which causes the cancer. Breast cancer is a disease in which malignant (cancer) cells grow in the breast tissues. A tumor is a mass of diseased tissue. There are two types of breast tumors: non-cancerous, ''benign,'' and cancerous, or ''malignant.'' Cancer starts in the cells that are the basic building blocks in the breast or other body parts that make up tissue. Occasionally the way of cell outgrowth goes wrong, and new cells form or old or damaged cells would not die as they do when the body does not need them. Any new breast, lump, or breast changes should then be monitored by a health care professional experienced in the diagnosis of breast disease that is commonly a sign of breast cancer. Deep neural networks are used for virtual x-ray recognition and also text datasets like symptoms of breast cancer .The image classification is a classical problem of image processing, computer vision and machine learning.

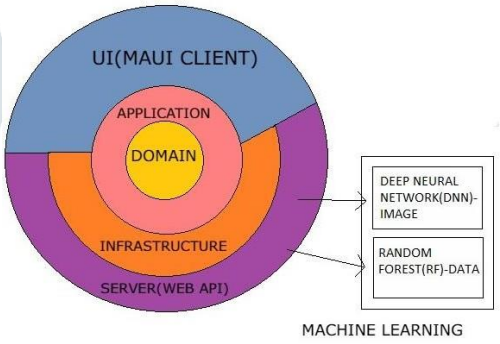

Fig.1. System architecture

Using dependency inversion throughout the project, depending on abstractions (interfaces) and not the implementations, allows us to switch out the implementation at runtime transparently. We are depending on abstractions at compile-time, which gives us strict contracts to work with, and we are being provided with the implementation at runtime. Testability is very high with the Onion architecture because everything depends on abstractions. The infrastructure layer has access over the application. The Server has access of the infrastructure, application and Domain. The UI has access of Infrastructure and Domain. It does not have full access on data whereas communication is done via API. Classification between the objects is an easy task for humans but it has proved to be a complex problem for machines. The raise of high capacity computers, the availability of high quality and low priced video cameras and the increasing need for automatic video analysis has generated an interest in object classification algorithms. The dataset BUI is uploaded and then the data preprocessing process starts using the dataset images. Data Cleaning: Data Cleaning is particularly done as part of data preprocessing to clean the data by filling missing values, smoothing the noisy data, resolving the inconsistency, and removing outliers. Data Integration: Data Integration is one of the data preprocessing steps that are used to merge the data present in multiple sources into a single larger data store like a data warehouse. Data Integration is needed especially when we are aiming to solve a real-world scenario like detecting the presence of nodules from CT Scan images. The only option is to integrate the images from multiple medical nodes to form a larger database. Data Transformation: Once data clearing has been done, we need to consolidate the quality data into alternate forms by changing the value, structure, or format of data using the Data Transformation strategies. Data Reduction: The size of the dataset in a data warehouse can be too large to be handled by data analysis and data mining algorithms. Data augmentation is performed. Which is used to address both the requirements, the diversity of the training data,

and the amount of data. Besides these two, augmented data can also be used to address the class imbalance problem in classification tasks. Data augmentation can be effectively used to train the DL models in such applications. Some of the simple transformations applied to the image are; geometric transformations such as Flipping, Rotation, Translation, Cropping, Scaling, and color space transformations such as color casting, Varying brightness, and noise injection. Deep learning algorithms train machines by learning from examples. Using the deep neural network, the output is predicted whether the condition is malignant tumor or benign tumor or normal.

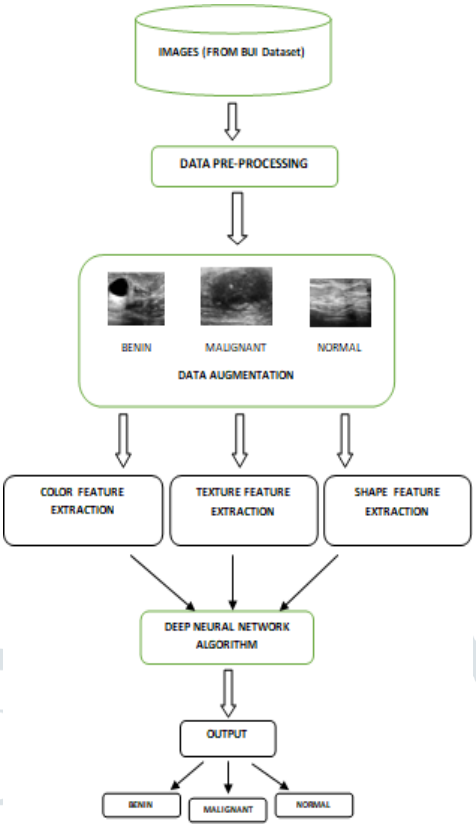

Fig.2. System workflow

Using the deep neural network, the output is predicted whether the condition is malignant tumor or benign tumor or normal. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of over fitting. The performance analysis of the proposed (label) of the whole image. In Training Set is the set of data that is used to train and make the model learn the hidden features patterns in the data. The validation set is a set of data, separate from the training set that is used to validate our model performance during training and testing set. The validation data set provides an unbiased evaluation of a model fit on the training data set while tuning the model's hyperparameters. The test set is a separate set of data used to test the model after completing the training. Below is the graph representation of all the values observed from training, testing and validating sets of image dataset using Deep Neural Network as well as text data using Random Forest algorithm.

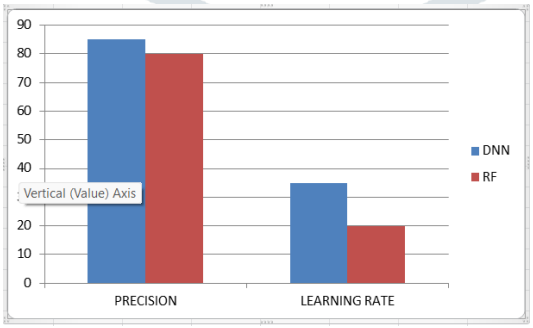

Fig.3. Graph representation

The Deep neural network algorithm presented the best performance by resulting in the great training & testing set with AUC of 85% and the Random forest regression presented the good performance by resulting in the training & testing set with AUC of 80%. These findings show that automatic deep learning methods can be readily trained to attain high accuracy on heterogeneous platforms, and hold tremendous promise for improving clinical tools to reduce false positive and false negative screening image results.

ALGORITHMS	PRECISION	TRAINING TIME
Deep Neural Network	85%	35
Random Forest Classifier	80%	20

Fig.4. Accuracy percentage table

The above table shows that Precision slot of DNN will be 85% with training period of 35 and precision slot of RF will be 80% with training period of 20.
lgorithm have analyzed the performance of detecting a type of breast tumor whether it is a malignant or benign tumor. Datasets are trained and then validated which is then tested to predict the output. We develop a deep neural network algorithm that can accurately detect breast cancer on screening patients X-rays using an "end-to-end" training approach that efficiently leverages training datasets with either complete clinical annotation or only the cancer status

CONCLUSION

We have proposed a method for breast cancer detection using machine learning techniques. The RF model consists of a series of decision trees, each of which is trained with random subsets of features. A deep neural network (DNN) classifier is used to classify a tumor into benign or malignant breast cancer groups using different image

classification classes. Text classification is performed using a random forest, RF predictions are obtained by majority voting of the predictions of all trees in the forest. This study systematically assesses the model's power to reduce the early detection rate of patients and assists physicians in clinical practice. It should be noted that all the results obtained refer only to the database, this can be seen as a limitation of our work, so in future work it is necessary to check the use of the same algorithms and methods on the other databases to confirm this Results from the data in this database and in our future work, we plan to apply our machine learning and other algorithms to sets of larger datasets with more disease classes using new parameters to achieve greater accuracy.

III. REFERENCES

[1] Dr.A.Sivasangari, Mamatha Sai Yarabarla, Lakshmi Kavya Ravi, "Breast Cancer Prediction via Machine Learning" International Conference on trends in Electronics and Informatics, 2019.

[2] Shubham Sharma, Archit Aggarwal, Tanupriya Choudhury," Breast Cancer Detection Using Machine Learning Algorithms" International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), 2020.

[3] Naresh Khuriwal, Dr Nidhi Mishra, "Breast Cancer Diagnosis Using Deep Learning Algorithm" International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2018.

[4] Meriem AMRANE, Saliha OUKID, Ikram GAGAOUA, Tolga ENSARl, "Breast Cancer Classification Using Machine Learning" Institute of Electrical and Electronics Engineers, 2018.

[5] Guanqing Li, Jiawen Zhang, Zhiyuan Lou, "Breast Cancer Disanosis Using Machine Method" Femto-ST Sciences & Technologies, 2019.

[6] Roslidar, Aulia Rahman, Rusdha Muharar, Muhammad Rizky Syahputra, Fitri Arnia, Maimun Syukr , Biswajeet Pradhan And Khairul Munad, "A Review on Recent Progress in Thermal Imaging and Deep Learning Approaches for Breast Cancer Detection" Ministry of Education and Culture of the Republic of Indonesia under 2020.

[7] Amin Ul Haq, Jian Ping Li , Abdus Saboor, Jalaluddin Khanl, Samad Wali, Sultan Ahmad, Amjad Ali, Ghufran Ahmad Khan, "Detection of Breast Cancer Through Clinical Data using Supervised and Unsupervised Feature Selection Techniques" the National Natural Science Foundation of China under Grant 61370073, in part by the National High Technology Research and Development Program of China under Grant , 2021. Sajida Imran, Bilal Ahmed Lodhi, and Ali Alzahrani "Unsupervised Method to Localize Masses in Mammograms "Deanship of Scientific Research at King Faisal University (Nasher Track), 2021.

[8] Huan-Jung Chiu, Tzuu-Hseng S. Li, and Ping-Huan Kuo, "Breast Cancer–Detection System Using PCA, Multilayer Perceptron, Transfer Learning, and Support Vector Machine" Ministry of Science and Technology, Taiwan, 2020.

[9] M. Etehadtavakol and E. Y. K. Ng, "Breast thermography as a potential non-contact method in the early detection of cancer: A review," Journal of Mechanicsin Medicine and Biology, vol. 13, no. 02, p. 1330001, 2013.

[10] V. M. Rao, D. C. Levin, L. Parker, B. Cavanaugh, A.
J. Frangos, and J. H. Sunshine, ‘‘How widely is computer- aided detection used in screening and diagnostic mammography?’’ J. Amer. College Radiol., vol. 7, no. 10, pp. 802–805, Oct. 2010.

[11] M. Sato, M. Kawai, Y. Nishino, D. Shibuya, N. Ohuchi, and T. Ishibashi, ‘‘Cost effectiveness analysis for breast cancer screening: Double reading versus single + CAD reading,’’ Breast Cancer, vol. 21, no. 5, pp. 532–541,2014.

[12] X. Bargalló, G. Santamaría, M. del Amo, P. Arguis, J. Ríos, J. Grau, M. Burrel, E. Cores, and M. Velasco, ‘‘Single reading with computeraided detection performed by selected radiologists in a breast cancer screening program,’’ Eur. J. Radiol., vol. 83, no. 11, pp. 2019–2023, Nov. 2014.

[13] E. L. Henriksen, J. F. Carlsen, I. M. Vejborg, M. B. Nielsen, and C. A. Lauridsen, ‘‘The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: A systematic review,’’ Acta Radiol., vol. 60, no. 1, pp. 13–18, Jan. 2019.

[14] C. Romero, A. Almenar, J. M. Pinto, C. Varela, E. Muñoz, and M. Botella, ‘‘Impact on breast cancer diagnosis in a multidisciplinary unit after the incorporation of mammography digitalization and computer-aided detection systems,’’ Amer. J. Roentgenol., vol. 197, no. 6, pp. 1492–1497, Dec. 2011.

[15] T. Onega, E. J. Aiello Bowles, D. L. Miglioretti, P. A.Carney, B. M. Geller, B. C. Yankaskas, K. Kerlikowske, E.
A. Sickles, and J. G. Elmore, ‘‘Radiologists’ perceptions of computer aided detection versus double reading for mammography interpretation,’’ Academic Radiol., vol. 17, no. 10, pp. 1217–1226, Oct. 2010.

[16] G. Hamed, M. Abd El-Rahman Marey, S. El-Sayed Amin, and M. F. Tolba, ‘‘The mass size effect on breast cancer detection using 2-levels of evaluation,’’ in Proc. Int.

onf. Adv. Intell. Syst. Inform. Cham, Switzerland:Springer, 2020, pp. 324–335.

[17] M. A. S. A. Husaini, M. H. Habaebi, S. A. Hameed,
M. R. Islam, and T. S. Gunawan, ‘‘A systematic review of breast cancer detection using thermography and neural networks,’’ IEEE Access, vol. 8, pp. 208922–208937, 2020.

[19]G.Hamed, M. Marey, S.Amin, and M. Tolba, ‘‘Comparative study and analysis of recent computer aided diagnosis systems for masses detection in mammograms,’’ Int. J. Intell. Comput. Inf. Sci., vol. 21, no.1, pp. 33–48, Feb.2021.