Cancer Detection using Machine Learning

1st Nishchay Trivedi
Student Id: 1106924
Lakehead University

2nd Shiv Hansoti
Student Id: 1105615
Lakehead University

Abstract—**9.56 Million People in the world die due to cancer. It's the second most deathly factor. Cancer is classified as an assorted health issue involving multiple associated sub-types in various groups. Early diagnosis and prognosis of a cancer type have become a priority in cancer research because it will facilitate the potential medical care of patients. The significance of labeling patients with cancer into high-risk or low-risk groups led several academic teams to explore the application of solutions to machine learning (ML), including the biomedical and bioinformatics fields. Hence, these methods were used as a blueprint for the diagnosis and treatment of cancer disorders. Additionally, ML tools' ability to identify core features from diverse datasets shows their meaning. Different approaches such as Artificial Neural Networks (ANNs), Bayesian Networks (BNs), Support Vector Machines (SVMs), and Decision Trees (DTs) have been commonly used for the development of predictive models in cancer science, resulting in efficient and precise decision making. However though the use of ML approaches will enhance our knowledge of the progression of cancer, a sufficient degree of evidence is required to use these approaches in routine clinical practice. Here in this article, we describe an analysis of current innovations to ML and DL used throughout cancer progression modeling. The predictive models discussed are mostly based on multiple supervised ML techniques and even some unique traits of inputs and sample datasets. Given the increasing increase in the use of ML methods in cancer science, here are the latest publications that use these techniques as a tool for cancer risk or patient outcomes.**

I. Abbreviations

CFS Correlation based Feature Selection; AUC Area Under Curve; ROC Receiver Operating Characteristic ; SEER Surveillance, Epidemiology and End results Database; NSCLCNon-small Cell Lung Cancer;NCIcaArray National Cancer Institute Array Data Management System; ML Machine Learning; ANN,Artificial Neural Network; SVM Support Vector Machine ;DT Decision Tree ;BN Bayesian Network; SSL Semi-supervised Learning; TCGA The Cancer Genome Atlas Research Network ;HTT High-throughput Technologies; OSCC Oral Squamous Cell Carcinoma; BCRSVM Breast Cancer Support Vector Machine; PPI Protein–Protein Interaction ; GEO Gene Expression Omnibus; LCS Learning Classifying Systems;ES Early Stopping algorithm; DT Decision Tree

II. Introduction

Since a long time there have been several advances in the research for cancer prediction. Methods include screening at an early stage, so that it can be predicted before it starts causing harm to the human being. These methods are for the early prediction of cancer in the field of medicine. This is a very challenging task for the researchers as large amount of data has been collected. Therefore, Machine Learning methods are becoming a popular tool used by the researchers. These techniques can be used to identify several patterns and predictions from the complex data sets. In addition, the predictions can be further improved and the information like cancer type can also be found out.

Machine learning techniques are being used for cancer prediction and prognosis. And we here present a review of several studies being used for this task in the field of medicine. And in this review the features that are considered in order to determine the results are independent of any particular treatment. Moreover, the methods used to determine the cancer types would also be discussed along with their advantages and disadvantages. The data used by the methods with how it is processed for analysis will also be seen. First of all while seeing the data we can see that there is an integration of two types of data, i.e. clinical and genomic. While applying the Machine Learning methods we can see that it improves the accuracy of the model which predicts cancer susceptibility and recurrence. And by applying the ML techniques the accuracy of cancer prediction has been significantly improved by 15 to 20%.

There are several studies made in this area of research which were based on different strategies in order to discover or predict cancer at a very early stage. All these studies describe the different approaches being used cancer detection and identification. There is one drawback which is these models suffer from low sensitivity. As these models find it a bit difficult to differentiate between benign and malignant tumors. Microarrays have also been used for prediction of the cancer outcome and its potential as well as the limitations have been discussed. Gene signatures can be used for improving the ability for prognosis in cancer patients, but there is little to no progress for the application of microarrays in clinics. There is also a new method gene expression profiling which can be used but before using it in clinical practice more study needs to be done with larger data sets.

In this work, only the studies which used Machine Learning models in order for analyzing and predictions of cancer diagnosis are presented.

III. Literature Review

In the article of Image Processing techniques for analyzing CT scan images towards the early detection by Jeyaprakash

Vasanth Wason and Ayyappan Nagarajan, there are two commonly found types of lung cancer small and non-small cell lung cancer[1]. Most of the general symptoms found most widely are coughing of blood, pain in the chest, weight loss, shortness of breath, feeling weak, and losing appetite[2]. Early detection of the tumor improves the survival rate from 15 to 50[3]. And increasing this survival rate is a must. There are several methods like images generated by X-Rays, CT scans, and MRI's, which help in detecting lung cancer without surgery. Among all of these, a CT scan is the most commonly used method which produces 3D images of the lungs[3]. And the mortality rate reduction can be made whenever the disease is detected early. Early detection of cancer plays a vital role in preventing cancer cells from multiplying and spreading. Existing lung cancer detection techniques are not adequate for providing accuracy. Hence, it is of importance to develop new methods for the early detection of lung cancer. The performance of Multilayer and Neural Network classifier trained by 11 training algorithms with Independent Component Analysis feature extraction is known [1]. A MATLAB based software tool to process the cancer image pre-processing is available [2]. It is shown that image processing techniques are beneficial to detect tumor cells. A methodology based on average information parameters by utilizing image processing tools for lung cancer investigation is reported [4]. The authors revealed the real issue for the lung cancer diagnosis is the time constrictions for physical diagnosis. So they proposed a method that successfully rejected the null hypothesis test by implementing a standard statistical model. Miller et al. [5] used image pre-processing, image erosion, median filtering, thresholds, and feature extraction for image processing techniques to apply on CT images. The authors discussed the development of an image processing algorithm for lung cancer detection using CT Images. A neural network-based system for the computer-aided detection of lung nodules in chest radiograms is shown [6]. They represent an artificial neural network-based lung cancer detection system using CT images. Implementation and analysis of the image processing method for detecting lung cancer are described [7]. The authors use color attribute in the feature extraction stage to study lung cancer using binarization to predict cancer in its earlier stage. Therefore, it is of concern to describe the use of a Computer-Aided Diagnosis (CAD) system using Computed Tomography (CT) images to help in the early diagnosis of lung cancer (to differentiate between benign and malignant tumors).

In the literature, many works applied Artificial Intelligence (AI) techniques for breast cancer diagnosis to improve classification accuracy and time response [8]. This section gives some works related to the solution of medical breast cancer diagnosis using the machine and deep learning approaches.

Arpit B. and Aruna T. [9] proposed a genetically optimized neural network(GONN) for breast cancer classification (malignant and benign). They optimized the neural network architecture by introducing new crossover and mutation operators. To evaluate their work, they used WBCD. They compared the classificationaccuracy, sensitivity, specificity, confusion matrix, ROC curves, and AUC under ROC curves of GONN with classical model and classical Backpropagation model. This method presents a good accuracy classification. However, it can be improved by using a larger dataset than WBCD, feature extraction, to make it more efficient for real-time diagnosis of Breast Cancer.

Ashraf O. I. and Siti M. S. [10] proposed a computer-based method to automatically classify breast cancer disease. The method applied a multilayer perceptron (MLP) neural network based on enhanced non-dominated sorting genetic algorithm (NSGA-II) to optimize accuracy and network structure. Compared to other methods, this work improves classification accuracy. However, MLP can get stuck in local minima.

Na L. et al. [11] proposed an intelligent classification model for breast cancer diagnosis based on a hybrid feature selection approach. That gain directed simulated annealing genetic algorithm wrapper (IGSAGAW), to remove the redundant and irrelevant feature from the feature space and cost-sensitive support vector machine (CSSVM) learning algorithm. This process can help improve the classification accuracy and reduce the computational cost. The proposed method is applied to Wisconsin Original Breast Cancer (WBC) and WBCD to verify its effectiveness.

Nawel Z. et al. [12] presented a conception and implementation of Computer Assisted Detection (CAD) for mammogram images classification. The system is based on a GA-based features selection algorithm to reduce the dimensionality of the feature vector, and semi-supervised support vector machine (SVM)for classification. Experiments were validated on the Digital Database for Screening Mammography (DDSM) dataset. The proposed approach improved accuracy.

Abdulkader H. et al. [13] developed an automated system for the classification of breast tissue. The system uses two machine learning techniques: Feedforward neural network using the backpropagation learning algorithm (BPNN) and radial basis function network (RBFN). Breast cancer tissues were classified into six different tissues, Carcinoma, Fibro-adenoma, Mastopathy, Glandular, Connective, and Adipose tissue. Data were acquired using electrical impedance spectroscopy (EIS) method. The Radial basis function network outperforms the back-propagation network for classifying six different breast tissues in terms of accuracy, minimum error, maximum epochs, and training time. The proposed system improved accuracy and decreases training time. However, learning with ML Techniques for Breast Cancer Diagnosis: Literature Review neural networks can be weak in generalizing and can get stuck in local optima.

Haifeng W. et al. [14] designed an SVM-based ensemble learning model for breast cancer diagnosis. The proposed ensemble model includes two types of SVM structures, i.e., a C-SVM and a R-SVM, and six types of kernel functions. A weighted Area Under the Receiver Operating Characteristic Curve Ensemble (WAUCE) mechanism is proposed for model hybridization to import the expertise of different base classifiers on diagnostic tasks. The model was evaluated based on two datasets: the Wisconsin Breast Cancer (WBC) dataset

and the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, and one large dataset, the Surveillance, Epidemiology, and End Results (SEER) dataset. The proposed model increases diagnosis accuracy compared to other works based on single SVM. However, it is a computationally expensive method, and the training time is high.

Kemal P. et al. [11] proposed a hybrid approach based on mad normalization, KMC based feature weighting, and AdaBoost tM1 classifier. The detection of the presence of breast cancer is done in three steps: In the first step, the dataset was first normalized by the MAD normalization method. In the second step, k-means clustering (KMC) based feature weighting has been used for weighting the normalized data. Finally, the AdaBoost M1 classifier has been used to classify the weighted data set. The Breast Cancer Coimbra dataset (BCC) taken from the UCI machine learning database was used. This method shows good results in terms of accuracy. However, it is a computationally expensive method.

Teresa A. j. et al. [16] proposed a classification method of hematoxylin and eosinstained breast biopsy images using Convolutional Neural Networks (CNNs). They provide four classes of medical relevance: healthy tissue, benign lesion, in situ carcinoma, and invasive carcinoma. The proposed CNN architecture was designed in order to integrate information from multiple histological scales. The model is applied to image dataset composed of high resolution uncompressed and annotated HE stain images from the Bioimaging 2015 breast histology classification challenge.

Fabio A. et al. [17] used a deep learning approach to classify breast cancer histopathological images from BreaKHis, a public dataset. They proposed a method based on the extraction of image patches for training the Convolutional Neural Network (CNN) and the combination of these patches for final classification. This method allows for avoiding adaptations of the model, leading to more complex and computationally costly architecture. However, the lack of data provides costly expensive experiments. Hiba A. et al. [14] presented a comparison between the performances of four classifiers: Support Vector Machine (SVM), Decision Tree (C4.5), Naive Bayes(NB), and k Nearest Neighbors (kNN) on the Wisconsin Breast Cancer datasets.

IV. ML TECHNIQUES

While using the Machine learning techniques for analysis and prediction, always keep in mind that there are two phases of a learning process: 1) estimating unknown dependencies in a system and 2) use those estimated dependencies to predict the outputs for the system. The two main types of ML methods are supervised and unsupervised learning. Here in supervised learning, labeled examples are provided as training data. In contrast, there are no labeled examples in unsupervised learning, so there is no notion of the output during the learning process. Therefore in the unsupervised model, the model finds patterns or marks the groups by itself on the input data. That is not the case for supervised learning, where the classification process is used for the same. Furthermore, it categorizes the

data into classes. Then there are two other commonly used ML tasks, which are clustering and regression. Clustering here is an unsupervised task that finds out the categories by itself within the dataset. Moreover, each piece of data can be assigned to the already identified clusters based on characteristics. On the other hand, regression has a learning function that maps the data.

There is one more ML method that is widely used, semi-supervised learning, and as the name suggests, it is the combination of supervised and unsupervised learning methods. This means that it combines labeled and unlabeled data to construct an accurate learning model. Moreover, this method is preferred more for unlabeled datasets.

To determine the correct tool or method to be used, we need to know the exact type of data that we are going to use. Now, every data sample is described with several features. Each feature consists of different values. Several prepossessing steps need to be taken to make to data more suitable to analyze. These steps are mainly in order to clean up the data and hence improve the data quality. The data quality can be improved by removing the noise, outliers, missing data, duplicate data, etc. For preprocessing, some of the techniques used are feature selection, reduction of dimensions, and feature extraction. ML algorithms work better with low dimensionality. It helps in the removal of irrelevant features and is responsible for constructing more robust learning models. All of this is to achieve a good ML model, which can be used to perform classification, prediction, or similar kinds of tasks. Moreover, the most common among all of the methods is classification. A functional classification model should fit the training data set as well as classify all the instances accurately. We build our model is such a way that it has lower generalization error. The method used for analyzing the generalization error of the learning algorithm we are using is the bias-variance decomposition. Here the bias component measures the error rate, and the variance component is the second source of error over all the possible training and test sets.

After obtaining a classification model, we can see that the classifier's performance can be determined. Moreover, each proposed model's performance is measured by the terms specificity, sensitivity, accuracy, and area under the curve (AUC). Here the AUC is the measure of the model's performance, and it is based on the ROC curve, which plots between the axis sensitivity and 1 - specificity. In order to get reliable results, the training and testing datasets should be large enough. Furthermore, there are several methods used to evaluate the performance: random sampling, Cross-validation, Holdout method, and bootstrap method. A similar method to the holdout method is used in a random sampling method, but the only difference here is that the same method is used several times. In cross-validation, every sample is used the same number of times for training purposes and only once for the testing purpose. When discussing the holdout method, the dataset is divided into two different sets: training and testing. Moreover, in this method classification method is used to generate the model. The bootstrap method being the most

different divides samples into training and testing datasets. However, they are placed again into the same dataset after training.

The most recent approach of Machine Learning used for cancer prediction and prognosis is SVM. SVM stands for Support Vector Machines. SVM's can also give out probabilistic outputs. So an SVM can classify the tumors into benign or malignant classes based on the tumor size and patient's age. Moreover, there exists a decision boundary so that misclassification can be detected.

ANNs handle a variety of classification or pattern recognition problems. They are trained to generate an output as a combination of the input variables. Multiple hidden layers that represent the neural connections mathematically are typically used for this process. Even though ANNs serve as a standard gold method in several classification tasks [19], they suffer from certain drawbacks. Their generic layered structure proves to be time-consuming while it can lead to abysmal performance. Additionally, this specific technique is characterized as a "black-box" technology. Trying to determine how it performs the classification process or why an ANN did not work is almost impossible to detect. Fig. 3 depicts the structure of an ANN with its interconnected group of nodes.

DT follows a tree type of structure classification scheme in which the nodes represent input variables. Moreover, leaves represent decision outcomes. DT is one of the earliest and most prominent Machine Learning methods that have been broadly applied for classification purposes. Based on the DTs' architecture, it can be said that they are simple to interpret and "quick" to learn. When the tree is traversed for classification of a new sample, we can conjecture about its class. The decisions resulted from their specific architecture allow for adequate reasoning, which makes them an appealing technique.

SVMs are a more recent approach of ML methods applied in the field of cancer prediction/prognosis. The SVM maps the input vector to a feature space of high dimensionality. Then identifies the hyperplane, which separates the data points into two different classes. Furthermore, the marginal distance between the decision hyperplane and instances that are closest to the boundary is maximized. The resulting classifier can be used for the reliable classification of new samples. It is worth noting that probabilistic outputs can also be obtained for SVMs [20]. In SVM, hyperplane can be thought of as a decision boundary between the two clusters. The existence of a decision boundary allows for the detection of any misclassification produced by the method.

BN classifiers produce probability estimations rather than predictions. As their name reveals, they are used to represent knowledge coupled with probabilistic dependencies among the variables of interest via a directed a cyclic graph. BNs have been applied widely to several classification tasks as well as for knowledge representation and reasoning purposes.

In Image processing and segmentation techniques

- the image is processed by using filtering techniques.
- It segments the cancer nodule from the CT scan image. The segmentation method separates and identifies the touching objects in the image. This feature helps in the proper segmentation of cancer nodules when it is touching to other false nodules.
- features extraction where features like area, perimeter, centroid, diameter, eccentricity, and mean intensity are extracted from the image. These features are used as training features to develop the classifier.
- The classification module classifies the detected nodule as malignant or benign by using the trained classification method.

Bi-clustering mining is a useful tool to discover the column consistency patterns on the training data. The patterns frequently appearing in the tumors with the same label can be regarded as a potential diagnostic rule. The diagnostic rules are subsequently utilized to construct component classifiers of the Adaboost algorithm via a novel rules combination strategy. Finally, the AdaBoost learning is performed to discover effective combinations and integrate them into a robust classifier.

A Recurrent Neural Network (RNN) is the class of artificial neural networks. In there, the connections between the nodes form a directed graph alongside a temporal sequence. Moreover, this allows the RNN to exhibit temporal dynamic behavior. Derived from feed-forward neural networks, RNNs can use their internal state (memory) to process variable-length sequences of inputs.

Bidirectional Recurrent Neural Network (BRNN) connects two hidden layers of opposite directions to the same output. With this form of deep generative learning, the output layer can get information from the past (backward), and future (forward) states simultaneously.

V. EXPERIMENTAL ANALYSIS

TABLE I
EXPERIMENTAL ACCURACY[]

ACCURACY COMPARISON			
Algorithm	Cancer Type	Dataset Type	Accuracy
BN	Oral Cancer	tissue genomic	100%
SVM	Breast cancer	pathologic	89%
Graph SSL	Colon cancer	Gene expression	76.7%
SVM	Cervical cancer	pathologic	68%
ANN	Lung cancer	gene expression	83.5%
Graph SSL	Breast Cancer	SEER	71%
DT	Breast Cancer	SEER	93%
BC and Ada boost	Breast Cancer	SEER	95.75%
RCNN	Breast Cancer	SEER	91.3%
BiRNN	Breast Cancer	SEER	82.50%

VI. EVALUATION AND DISCUSSION

In the present review, the most recent works relevant to cancer diagnosis using a machine and deep learning techniques are presented. After a brief description of cancer disease and a small comparison between surveyed studies, we deduced that a large amount of cancer diagnosis studies using the
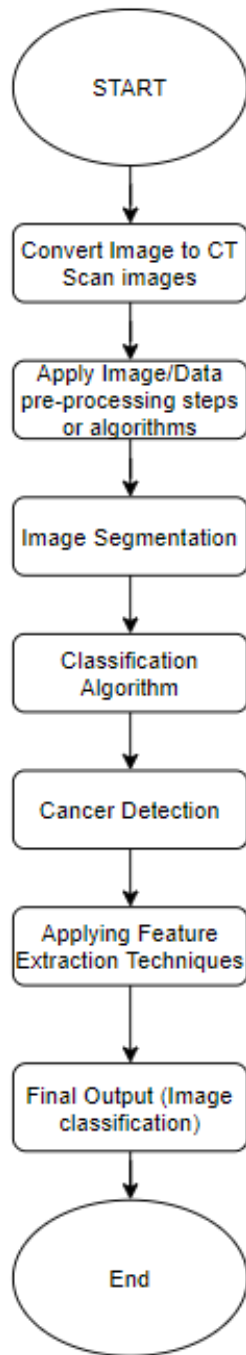
Fig. 1. CANCER DETECTION PROCESS

machine and deep learning technics in the last decade provide accurate results. Moreover, most of them combined multiple technics to optimize their system's performance and time response. Artificial Neural Networks (ANN) and Support Vector Machine (SVM) are most used technics in cancer diagnosis because it provides an accurate predictive performance.

However, selecting the appropriate method depends on many parameters, including the types and the size of data, the time limitations, and the type of prediction results. We also noticed three major limitations:1. Small dataset: To achieve preferment classification schemes, the size of the training datasets needs to be sufficiently large. However, the studies surveyed in this review used small amounts of data.2. Data quality: choosing the most significant data subset utilizing feature selection methods allows providing a robust system.3. Computational cost: most reviewed works are computationally expensive, which is not beneficial for both clinician and patient. Concerning the future of breast cancer diagnosis, new methods should be proposed to overcome the limitations quoted above, and more public datasets must be constructed.

Strengths of the method include:

- The improved accuracy of cancer nodule detection
- classifies the detected lung cancer as malignant or benign
- Removes the noises that create false detection of cancer

VII. CONTRIBUTION

This project was done in a team of two. Nishchay's contribution was identifying the recent research done in the computer science community for detecting cancer using computer-aided diagnosis. Nishchay collected the resources as much as possible, and he figured out which algorithm of machine learning was optimal for detecting cancer. Secondly, he did small implementation by taking The breast cancer dataset imported from sci-kit-learn, which contains 569 samples, including 30 real, positive features (including cancer mass attributes like mean radius, mean texture, mean perimeter, etcetera). Of the samples, 212 are labeled malignant, and 357 are labeled benign. He loads this data into a 569-by-30 feature matrix and a 569-dimensional target vector. Then he randomly shuffles the data and designates two thirds for training and one third for testing. He tried three different algorithms Logistic Regression, k-nearest neighbor, and Support Vector Machines for higher accuracy. He also figured out that the most advanced technique of LogitBoost gives much higher accuracy on this particular dataset.

Shiv's contribution was reading all the resources online and try to understand which machine learning technique was more suited for the corresponding dataset and why they gave that particular accuracy. He found that Bayesian Network, Support vector machine, Artificial Neural Network, Decision tree, Convolution neural network, and Bi-directional Convolution Neural Network were majorly used to solve the problem. Secondly, he tried to do a small implementation of Breast cancer detection using a Deep Convolution neural network with a Kaggle dataset. The original dataset consisted of 162 whole mounted slide images of Breast Cancer (BCa) specimens. They were scanned at 40x. Now from that, 277,524 patches of size 50 x 50 were extracted. Out of that, 198,738 IDC were negative, and 78,786 IDC were positive. Each of the patch's file name is of the format: uxXyYclassC.png(example 10253idx5x1351y1101class0.png), Where u is the patient ID (10253idx5), X is the x-coordinate of the place where this

patch was cropped from, and Y is the y-coordinate of where this patch was cropped from, and C indicates the class where 0 is non-IDC, and 1 is IDC. He used the Resnet-V50 Deep Convolution Neural network for experimentation purposes and achieved an accuracy of 84%.

VIII. Acknowledgment

As a project partner and student, we are thankful to Dr. Zubair Fadlullah, their teaching assistants Sadman Sakib and Tahrat Tazrin, for motivating us throughout the coursework and curriculum. Suppose we were not part of these subjects. In that case, we will not understand what computer science has contributed and how it is collaborated with medical science in solving the medical problems, which were extremely difficult for the doctors to solve without the machine. For instance, doctors can decide if there is cancer at the first stage or the initial level. This was new for both of us, and it was like exploring the new dimension of Healthcare and the computer science world. Moreover, most importantly, we would like to apologize if we have used any wrong terminologies as we have out knowledge limitations. Lastly, we are both grateful for your assistance and guidance, which would elevate our character for being part of the computer science society.

IX. Conclusion

Most of the studies that have been proposed in the last years focus on the development of predictive models using supervised ML methods and classification algorithms aiming to predict ultimate disease outcomes. Based on the analysis of their results, it is evident that the integration of multidimensional heterogeneous data, combined with the application of different feature selection techniques and classification, can provide promising tools for inference in the cancer domain. In the present review, the most recent works relevant to cancer diagnosis using the machine and deep learning techniques are presented. After a brief description of cancer disease and a small comparison between surveyed studies, we deduced that a large amount of cancer diagnosis studies using the machine and deep learning technics in the last decade provide accurate results. Moreover, most of them combined multiple technics in order to optimize their system's performance and time response. Artificial Neural Networks (ANN) and Support Vector Machine (SVM) are the most used techniques in breast cancer diagnosis because they provide accurate predictive performance. However, the selection of an appropriate method does depend on many parameters, including the types and the size of data, the time limitations, and the type of prediction results. We also noticed three significant limitations:-

- Dataset: In order to achieve preferment classification schemes, the size of the training datasets needs to be sufficiently large. However, the studies surveyed in this review used different datasets.
- Data quality: Choosing the most significant data subset utilizing feature selection methods allows providing a robust system.

- Computational cost: most reviewed works are computationally expensive, which is not beneficial for both clinician and patient
- The application of a Computer-Aided Diagnosis (CAD) system is made using Computed-Tomography (CT) scan images to help the early diagnosis of lung cancer (to distinguish between benign and malignant tumors). We report a framework for developing a model for early cancer detection using CADCT(Computer-Aided Diagnosis Computed-Tomography) image analysis.

Concerning the future of cancer diagnosis, new methods should be proposed to overcome the limitations quoted above, and more public datasets must be constructed.

References

[1] Jeyaprakash Vasanth Wason, Ayyappan Nagarajan, "Image processing techniques for analyzing CT scan images towards the early detection of lung cancer", 2019 Sep 12.
[2] Mesleh AM. Jordan Journal of Biological Sciences,2017, Google Scholar.
[3] Anjali Sevani, et al. International Journal of Advance Engineering and Research Development, Google Scholar
[4] Abdillah B, et al. The Asian Journal, 2016, Google Scholar
[5] Vaishaw K, et al. An Innovative Approach for Investigation and Diagnosis of Lung Cancer by Utilizing Average Information Parameters, Elsevier Procedia of Computer Science, 2018, Google Scholar
[6] Dimililer K, et al. Journal of Science and Technology, . 2017, Google Scholar
[7] Sakthivel A, Nagarajan A. International Journal of EngineeringSciences and Research Technology, . 2018;7:727. [Google Scholar]
[8] Djihane Houfani1, Sihem Slatnia1, Okba Kazar1, Noureddine Zerhouni2,Abdelhak Merizig1, and Hamza Saouli, Machine Learning Techniques for Breast Cancer Diagnosis, FEB 2020.
[9] Arpit, B., Aruna, T.: Breast Cancer Diagnosis Using Genetically Optimized NeuralNetwork Model. Expert Systems with Applications, Elsevier, pp. 1-15(2015)
[10] 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(2016)
[11] 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)
[12] 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)
[13] 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)
[14] Haifeng, W., Bichen, Z., Sang, W.Y., Hoo, S. K.: A Support Vector Machine-BasedEnsemble Algorithm for Breast Cancer Diagnosis. European Journal of OperationalResearch, Elsevier, pp. 1-33 (2017)
[15] 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 Multi-disciplinary Studies and Inno-vative Technologies (ISMSIT),IEEE, pp. 1-4 (2018)
[16] 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)
[17] 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)

[18] KonstantinaKourouaThemis V.KaramouziscDimitrios I.Fotiadisab, "Machine learning applications in cancer prognosis and prediction",15 November 2014, Science Direct.

[19] Anji ReddyVakaaBadalSoniaSudheer ReddyK.b, "Breast cancer detection by leveraging Machine Learning ", 7 May 2020, Science Direct.

[20] Joseph A. Cruz, David S. Wishart, "Applications of Machine Learning in Cancer Prediction and Prognosis ",2006, University of Alberta Edmonton.

[21] Li Shen, Laurie R. Margolies, Joseph H. Rothstein, Eugene Fluder, Russell McBride Weiva Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography",29 August 2019.

[22] Habib Dhahri, Eslam Al Maghayreh,Awais Mahmood, Wail Elkilani,and Mohammed Faisal Nagi,"Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms", 2019, Journal of Healthcare Engineering.

[23] Pier Paolo Ippolito, "Understanding Cancer using Machine Learning", Towards Data Science.

[24] 1Regina, "AI to Detect Cancer", Nov 22, 2019, Towards Data Science.

[25] Khushboo Munir,Hassan Elahi,Afsheen Ayub ,Fabrizio Frezza and Antonello Rizzi, " Cancer Diagnosis Using Deep Learning: A Bibliographic Review ", 3 June 2019, MPDI.