

A Study on Machine Learning bound Algorithms for Tumor Detection

M.Tahmooresi¹, V. Sivakumar²

^{1,2}School of Computing

Faculty of Computing, Engineering and Technology,
Asia Pacific University of Technology and Innovation (APU), Kuala Lumpur,
Malaysia – 57000

¹maryam.tahmooresi@yahoo.com, ²dr.sivakumar@apu.edu.my

Article Info

Volume 83

Page Number: 241 - 253

Publication Issue:

March - April 2020

Article History

Article Received: 24 July 2019

Revised: 12 September 2019

Accepted: 15 February 2020

Publication: 12 March 2020

Abstract

Cancer is being considered as the deadliest disease among all diseases across the globe. Since two decades, many academic researchers are contributing their work for automated detection of tumor from the data of screening modalities and the results of those researches are enhancing the probability of earlier detection of the disease which enable for proper timely treatment. This review article does a survey of several machine learning algorithms being involved in the earlier detection and classification of tumors and this review is expected to provide a detailed basis for the researchers who carry out their study in the development of automated diagnosis system for cancer detection.

Keywords: Cancer, Tumor Detection, Machine Learning.

1. Introduction

Cancer is defined as a tumor with unlimited and disordered spread of cells that are genetically changed. This proliferation damages the base fabric and leads to the tumor expansion. Therefore, the body is not able to control the situation. According to [1] the state in which the tumor cellules raise in one area is called cancer or tumor which is malignant, and when the cells of cancer spread from the main tumor to the other organs and parts of the patient's body, the tumor is called metastasis.

Different cancers which could be detected and diagnosed through machine learning bound automated detection systems from the respective screening modalities are given as follows: Lung carcinoma or lung cancer, happens due to the excessive growth of cells in the tissue of lung. The cancer progresses on the epithelial cells of our body. In order to diagnose lung cancer, radiologists

have to inspect through different CT scans to discover malignant tumors which might be the possibility of the disease.

Thyroid nodes are another type of really common endocrine tumors. Generally, many people have nodules, which are asymptomatic. The probable incidence rate that is detected by palpation is just 3%–7%; though, by implementing ultrasound in detecting thyroid nodules this rate raises from 19% to 67%. While many of the thyroid nodules are benign, 3% to 7% of cases are malignant [2, 3].

The beginning point of Pancreatic cancer is in the tissues of pancreas. The bad thing about pancreatic cancer is that in rare cases it is detected in early stages. Thereof, it is believed that those people who have pancreatic tumors or in their family background there is a history of pancreatic cancer, screening steps may be beneficial in detecting the problem in early stages [4].Pancreas cancer can be

classified as ductal adenocarcinoma and special types of cancer in pancreatic, are adenocarcinoma, mucinous neoplasm, signet- ring cell carcinoma, small cell cancer, based on its histopathology.

The prostate is an exocrine gland as big as a walnut. Therefore, its fluids and emissions are used outside the body. The prostate gland is responsible for producing the liquid by which the sperms are nourished and transported to join a female egg or ovum, and produce a baby. These fluids are contracted and forced out during orgasm by the prostate. Prostate-specific antigen (PSA) that is excreted by the prostate lets semen stay liquid. An initial symptoms of prostate cancer is this protein's increase in the blood. The urethra that is a pipe from which urine and sperm goes out of the body, goes through the prostate as well. As a result, another responsibility of the prostate is to control urine using hundreds of small muscle fibers, prostate tightens and restricts the urine flow through the urethra [5].

Breast cancer, is an unusual kind of cancer which mainly affects women. There are two main factors that lead to breast cancer; modifiable or non-modifiable factors. Some factors such life style, habits and environmental issues are considered Modifiable factors as they can be controlled. According to a study by Giftson, Hari and Christy [6], 1 in every 28 women in India is exposed to breast cancer. There are not enough detection techniques to diagnose breast cancer in early stages and they are still suffering from this shortage in the exact prediction of this cancer. Besides, the risks of survival increases due to not having enough information and awareness, practical measures and facilities of treatment.

When the large intestine or the final part of your digestive tract is infected by cancer, it is called Colon cancer. In many cases, the beginning of

colon cancer is tiny, noncancerous (benign) masses of cellules called adenomatous polyps. After a while, some of these polyps might change to colon cancers.

2. Machine Learning in Tumor Detection

The researchers [7]studied 114 benign nodules in a group of patients (106 people) and 89 malignant nodules in another group of patients (81 patients). They have investigated the effect of the amount of neurons that are hidden and the kinds of activation function on the ELM performance. Using feature selection method, they gained the feature subsets' relationship. Their findings suggest that meaningful differences exist between the benign and malignant thyroid nodules. The shape, echogenicity, margin, composition and classification are the most discriminative features. In comparison to other methods, not only this proposed method has attained a really hopeful classification accuracy through 10-fold cross-validation (CV) scheme, it even has critically decreased the computational costs in comparison to the rest of methods. A 94.55% specificity, 0.8672 AUC, 78.89% sensitivity, and 87.72% ACC, was achieved by this ELM-based approach.

A novel unsupervised learning model [8] which is on the basis of matrix tri-factorization via learning the resemblances from pairwise restraints to spot driver genes from the data of pan-cancer. Compared to approaches that are based on the current matrix factorization that do not pay attention to the pairwise similarities between cancers, their model presented a better performance in evaluating the identified benchmarking genes. Additionally, their model's performance in detecting is also greatly improved (area under the precision-recall curve = 9.1% for Vogelstein genes). Furthermore, some of driver genes being reported in current printed studies, are discovered by their model, thus it is

proved that this model is capable of recognizing driver gene candidates for future studies.

A surgical navigation system which is based on innovative virtual reality is presented by [9] to perform targeted prostate biopsies, without needing the painful TRUS. In this system, the needles for biopsy and anatomy of the patient are continuously traced via an electromagnetic tracking system and with respect to the surgical bed, their 3Dimensional position and orientation are provided. During MRI scanning Multiple fiducial markers are put over patient's skin at the pubic bone and iliac crest. Later, a virtual image of the tracked needle and a stereoscopic three dimensional volumetric versions and several orthogonal visions of the sick people's anatomy is presented to the surgeon in the operation room. After a process of registration between the tracker coordinate system and the MRI, the needle is inserted through the navigation system in the patient perineum via some anatomical layers toward the targets of biopsy.

Normal/Abnormal mammogram images are studied and classified by [10]. They implemented the mixture of automatic thresholding Based-Harris Corner Detection and SVM to classify those images. According to findings of their study, the accuracy of this method in classifying Normal/Abnormal images of mammogram is better than the previous ones. The experimental findings yield a more reliable result that is capable of being generated for early and further diagnosis of abnormalities.

The researchers [11] suggested a Fuzzy Multiple parameter SVM, that is implemented in CAD system. When the images are processed and features are extracted, the technique is utilized for classification through fuzzy association with improved several parameters. The findings support the idea that compared to the conventional SVM,

Fuzzy Multiple-parameter SVM gives out enhanced sensitivity, MCC, and accuracy. The findings mean that suggested technique contrary to the conventional classification tool, SVM, allows building a stronger CAD system which aids doctors by handing satisfying results for detecting breast cancer in early stage. The value of MCC found to be sensitivity, specificity, and accuracy are respectively 78.78%, 82%, 96% and 88% for existing SVM, and 86.16%, 96%, 90% and 93%.

The study [12] integrated deep learning methods and visual saliency model to a system of mass detection. By assigning higher saliency to the areas of interest from the previously processed mammograms, the visual saliency model, experimentally expands the outcome of the additional mass detection. Moreover, features of CNN is different from features of traditional handcraft which are served to a linear SVM for making final decision. Providing radiologists with suspicious masses, is one of the main aims of CAD system, so the marking method is adapted for more accurate notice of the suspicious masses. According to the experimental results, this proposed system for mammographic mass detection is proved to be effective. Nonetheless, the way specific saliency detection approach is implemented to reduce false positive rate is still an issue which has to be focused on.

The research [13] proposed techniques for identifying discussions along with examples of users from an online health community posts according to LSTM and CNN. Experimental outcomes prove that these kinds of techniques, can effectively locate debatable CAM disputes and help understand opinions about disputable health issues. Principally, LSTM can effectively take post-to-post connections, that are very important in recognizing discussions in discourse. The results of the study on alternative medicine for breast cancer, propose that

although CAM is broadly used in breast cancer management, its efficiency perception varies amongst patients. A lot of particular treatments cause discussions more than the rest, principally when considered a replacement of predictable cures.

An automated technique of detecting cancer of colon is presented by [14] that measures the differences between normal and malignant lumen 's shape with regard to several features like the slightest square lumen space, descriptors of elliptic Fourier, and characters of morphology. In order to classify, sets of individual features and the hybrid features were utilized. In order to classify, SVM classifiers ensemble were suggested. The suggested technique has been assessed on 174 colon biopsy images as the dataset, and through using hybrid feature set and ensemble classifier, success of classification was 96.65%. The influential outcomes of classification demonstrate capturing the fundamental geometrical differences by the proposed features in the form of malignant and normal tissues of colon. Moreover, examination show that compared to individual features, amalgam feature set that encompasses different information on the image tissue has more discriminating capability. The ensemble classifier that is based on SVM is going to be an encouraging method of classification for modeling the thought-provoking issue of classifying colon biopsy images.

The study [15] used automatic methods for detecting the texture, shape, and size of nucleus of the cell to find the rule out mistakes. In this study morphological operations are done by cell segmentation and classification of cancer is done implementing a support vector machine (SVM). In order to code, the image processing toolbox MATLAB has been used. The dataset consisted of 50 images, with JPEG format images. Among the images, fifteen were normal cellules and twenty were CIN1 and the rest were CIN2 or CIN3. Images

contain several cells and overlapped. Once feature was extracted, dataset was classified into 3 classes. The first class was related to normal cells; the second was the CIN1 cellules, and the 3rd one was CIN2 or CIN3 cellules. Since the used cell images are overlapped and several nucleuses, the dataset was classified into only 3 classes. In previous studies, just single cell was utilized in classification. The SVM classifier was implemented in cell classification, ten-cross fold was implemented in data training and testing. Findings of the studies show that the accuracy is 96%, specificity is 100% and sensitivity is 95.6%.

In their study [16] it showed that the false positive rates which outbreak existing techniques could be meaningfully reduced. As mentioned earlier, in some of the current methods of diagnosis the false positive rates are very high, even 95% in some circumstances. The average value of false positive rate in the present ensemble technique is 0:50%. It shows an approximately 99:47% reduction in the false positive rate, which is an important enhancement in the presently implemented techniques. The method called simple voting didn't improve the accuracy significantly on the single models that are unsmoothed and smoothed. Nevertheless, it can reduce the false positives' amount meaningfully and significantly enhances this model's accuracy. The false positives are targeted in voting method, while nothing is done about false negatives. Thus, the false negatives are not decreased in voting. As the rates of false negative of this particular model is higher than the ones that are observed clinically, we need an alternative technique to deal with them.

A CAD system is presented by [17] in order to help the radiologists in classification of breast lesions through images of MR. The design of a GA enhanced ANN permits achieving decent performance with 89.77% accuracy on average and

the best accuracy as 100%. This suggested automated methodology is applicable in clinical practice to distinguish benign breast lesions from malignant breast lesions by implementing CE-MRI, and assists the radiologist in classifying all noticed nodules, plus excluding areas that do not need any more examinations. These innovative results of diagnosing breast cancer and its stages reached to 100% accuracy, with 0:5822 threshold value; while the least accuracy was 73:08% with a 0:4309 threshold value. The consequential average indexes were 89:77% accurate (the standard deviation is 5:84), sensitivity of 0:8908 (the standard deviation is 0:1021) and specificity of 0:9046 (the standard deviation is 0:0875).

A study was conducted by [18] which helps physicians in diagnosing cancer of lung and improving the classification accuracy in malignant and benign pulmonary nodules. A novel intelligent diagnosis model is presented in this paper that aims at pulmonary nodules' CT imaging features. In this method, the convolutional neural network is used for extracting the pulmonary nodules' features. After that, principal component analysis was implemented to decrease the elicited features' dimension, besides lastly, the ultimate features are classified by particle swarm optimization optimized SVM. According to the pulmonary nodules resulted by the LIDC-IDRI database, four hundred pulmonary nodules have been utilized in the process of teaching and 310 pulmonary nodules for the analysis phase, the accuracy of classification found to be 91.94%. The proposed technique can present neutral, appropriate and effective assisting method for solving problems in classifying malignant and benign pulmonary nodules in medical images.

Neural Network Threshold Selection, Half Selection, and Mean Selection are the three features selection strategies that have been described with an RBF Network by [19] on the basis of fuzzy entropy

values classifier. They aimed at choosing a suitable group of features of classification for noisy datasets containing vague and/or redundant information. For evaluation, 5 standard data sets are used through the Machine Learning Repository of UCI. The accuracies of suggested feature selection strategies were satisfactory and even better than the accuracy of the whole feature set without being feature-selected. The fuzzy entropy with Neural Network Threshold Selection was the proponent that yielded the maximum accuracy.

By reducing the number of attributes in a study [20], it is proved that the prediction of cardiovascular disease presence is more accurate. Shilaskar and Ghatol investigated an intelligent system to produce feature subset by improving performance of diagnosis. In order to find subset which results in enhanced classification Features, forward selection, forward inclusion, and backward elimination search techniques are implemented to search through the ranked with distance measures. The selection technique of hybrid forward is proposed for diagnosing cardiovascular disease. This experimental study suggests that smaller subsets are found by this approach and in comparison to forward inclusion and back-elimination methods the diagnosis accuracy is increased.

According to [21] a random forest classifier (RFC) method is useful in lymph diseases diagnosis. In the first stage of the approach, the focus is on selection of features, the purpose is to construct different algorithms for feature selection such as Principal Component Analysis (PCA), genetic algorithm (GA), Sequential Forward Floating Search (SFFS), Relief-F, Sequential Backward Floating Search (SBFS) and Fisher, to reduce the dimension of lymph diseases dataset. In the second stage, the focused is switched from selection of features to construction of the model, and for an effective

classification the gained subsets of features are applied into the RFC. The result show that the highest classification accuracy belonged to GA-RFC which was 92.2%. By means of GA, the dimension of input feature space with eighteen feature decreases to six features.

For enhancing the mining process, [22] concentrated on techniques of selecting features as an approach for gaining high quality attributes. In Feature selection techniques all restraints which need discovery of information from large data is touched. In the above mentioned study, they compared benchmark feature selection techniques on the basis of three famous data sets with three well-known algorithms of machine learning. The result of the study suggest that techniques of feature selection can develop the learning algorithms' performance. Though, it is not possible to resort to a single method of feature selection that best satisfies all learning algorithms and datasets. Hence, to obtain better outcomes, researchers on machine learning are required to realize the characteristics of learning algorithms and the nature of datasets. All in all, consistency based subset evaluation (CB) and correlation based feature selection (CFS) outperformed Relief (RF), symmetrical uncertainty, and principle components analysis (PC) and information gain.

In one study [23] VAG signals are investigated by suggesting a decomposition based on wavelet. The signals of VAG are disintegrated to different frequency sub-band signals. Sample entropy (SampEn), approximate entropy (ApEn), and recurrence quantification analysis (RQA) which are all Non-linear features are taken out as VAG signal features to characterize a VAG signal, totally 24 features form a vector. Two techniques of feature selection (FS), genetic algorithm (GA) and a priori algorithm and, select 6 and 4 features are selected as the vital ones as classifiers for evaluating the

performance of FS methods random forest and Least Square Support Vector Machines (LS-SVM) are suggested. Findings suggest that features selected from FS algorithms have more noticeable classification accuracy. According to the findings of this study, implementing apriori algorithm, the best accuracy of LS-SVM was 94.31% with a 0.0892 false discovery rate (FDR). In the mentioned study, a better accuracy of classification is described in comparison to the earlier studies that yielded accuracy of 88%. Implementing the findings of this work, the performance of present technology for precise distinguish of normal and abnormal VAG signals is enhanced. It also can offer a useful non-invasive tool for diagnosing disorders of knee joint.

In one study, the researchers investigated the application of K nearest neighbor with feature subset selection in diagnosing heart disease [24]. According to the results, feature subset selection application into KNN enhances the accuracy of heart disease diagnose for the population of Andhra Pradesh. The KNN algorithm was employed with feature subset selection for determining which features contribute more to the prediction of the disease. They tested the proposed methods with focus on heart disease on Andhra Pradesh beside other machine learning datasets which have been taken from UCI sources. Implementing this method, greatly affects the number of tests that each patient has to take, and reduces them. This model of prediction even enables the doctors in making decisions efficiently with less attributes to detect the heart disease. It is obvious that there is an urgent need to fight with the effect of heart disease and minimize the number of people dying from this disease in Andhra Pradesh.

The tumor and normal tissues of breast cancer patients have gone under inspection for miRNA expression level classification through some machine-learning algorithms which have been run

and evaluated [25]. Wan claims that the supervised learning based approach LightGBM is capable of routinely recognizing miRNA related to cancer highly accurate. It is really influential in cancer classification, as it can recognize miRNA effectively, accurately, and quickly in breast cancer. Later the LightGBM method is going to be tested on the ability to analyze the data of protein expression and SNP in breast cancer. LightGBM is also really effective in examining omics dataset, as it can be used in new cancer transcriptomes to notice miRNA targets.

The study [26] used CNN in classifying breast cancer where an unidentified patient's image is given to be diagnosed accurately. They propose a novel network design is for extracting the most revealing features of a histopathology images collection by BreakHis data base of images of microscopic breast tumor. 1,995 histopathological images have been examined (with a $\times 40$ factor of magnifying), and the findings proved that its accuracy has improved in comparison to some previous works, except one study in which the performance is comparable. With an average increase of 2.67%, it could be concluded that the proposed technique (without data augmentation) was better than four out of five standard techniques, and it is comparable to CLBP with a 0.1% difference. However, the proposed technique (with data augmentation) was better than all of the standard methods with $\sim 2.7\%$ growth in rate of recognition.

The performance of 3 famous interpretable classifiers (Naïve Bayes, decision tree, probabilistic neural networks) is compared by [27] for classification of breast cancer patients to the category of recurrence risk (high or low recurrence risk in 5 / 10 years). In five-year prediction of recurrence risk, the probabilistic neural networks classifier could achieve the highest prediction

accuracy (Acc = $76.88\% \pm 1.09\%$, AUC=77.41%). The probabilistic neural networks and the decision tree classifier both gained comparable prediction accuracies in the prediction of recurrence risk in 10 years ($70.40\% \pm 1.36\%$ and $70.50\% \pm 1.13\%$, respectively). Although the PNN classifier attained the mentioned accuracy with just ten features with the maximum data achievement, the decision tree classifier required one hundred features for achieving a similar accuracy and had a significantly lower AUC (66.4% vs. 77.1%).

The research study [28] compared 5 models of classification (Logistic Regression, Decision Tree, k-Nearest Neighbor, Linear Support Vector Machine and Cubic Support Vector Machine) and studied the way selecting the features implemented in building such models can influence the related performance. The mentioned classifiers were taught with decreased feature subsets recognized by Principal Component Analysis (PCA), correlation selection, selection based on t-test significance, and Random Feature Selection. The most effective technique of feature reduction was Random Feature Selection that recognized subset of features, it generated the highest cross validation prediction accuracies when used to build the models. The logistic regression model that was trained with the six features that were recognized by random selection, was 97.77% accurate, enhancement of 1.5% over the logistic regression model with 3 features which have been observed. 10 features were used to train the linear SVM model, the features have been identified by random selection and the cross validated accuracy of which was 97.87%, a 0.4% enhancement was observed in the linear model with 3 features. A similar accuracy was reported for the cubic SVM model containing eleven features that is 97.98%. The classification accuracy is improved to 98.56% by an ensemble

learner of Stacking the logistic, linear SVM and cubic SVM models.

According to measures of variability a shape feature measurement [29] for classifying images of breast tumors into malignant and benign cases. In order to take out two features variance and range in this approach, the researchers used distance pixel histogram and SVM was used as tool for machine learning and classification. In order to prove the efficacy of the suggested method, 4 performance statistical measures of a binary classification test (MCC, Specificity, Sensitivity and Accuracy) were used. Findings suggest that, implementation of variance and range overtakes compactness and the quality of prediction (MCC), sensitivity, specificity and accuracy can be improved by it to respectively 13.2%, 9%, 2% and 6.5% and 12.1%, 8%, 4% and 5%. Thus, a proficient benign/malignant classification is obtained by the implementation of the proposed shape features. It can aid doctors in providing preliminary decision support information for more diagnosis.

The study [30] used CT of chests of patients and non-patients to diagnose lung cancer. In their study, they introduce an improved version of Deep Convolutional Neural Network that contains implementing multiple two dimensional convolutional neural networks on individual slices and combines the outcomes for diagnosing patients and differentiating them with non-patients. The data of each patient/non-patient's chest CT is firstly divided to the lung features and is kept in 3D arrays. The 3D arrays that are preprocessed are then fed into the CNN framework, and the network parameters trained. Many duplications of the process with adequate statistics, can modify the parameters of network so that patients' CT scans are diagnosed by the network with an accuracy between 70~80%.

In order to differentiate benign and malignant breast tumors [31] developed computer-aided diagnosis (CADs) system which uses diffusion-weighted Magnetic resonance image (DW-MRI). The data was collected from 61 cases, comprising 38 patients with malignant tumors, and 23 patients with benign tumors. Each lesion gave out 2 kinds of texture features, containing 6 histogram statistical features and 16 gray-level co-occurrence matrix (GLCM) features. The method for selecting the feature Random Forest-Recursive Feature Elimination (RF-RFE). In order to make a classification model, they used Random Forest, to evaluate the performance of the classifier, area under the receiver operating characteristic curve (AUC), and leave-one-out cross validation(LOOCV) were implemented. This approach was used to select 6 texture features (comprising 3 GLCM features and 3 histogram statistical features) and the obtained AUC was 0.76, also the classification accuracy, sensitivity, and specificity were respectively 77.05%, 84.21%, 65.21%. According to the finding of this study, for developing CADs of breast cancer we can use the texture features as it shows a high sensitivity.

Through embedding the high dimensional RNA-Seq data in two dimensional images [32] benefited from a convolutional Neural Network in classifying 33 types of tumors. The best resulted accuracy found to be 95.59%. Additionally, on the basis of Guided Grad Cam, in each class, they created the significance heat-map of each gene. Doing functional analysis with high intensities in the heat-maps on the genes, they could validate and conclude that these top genes are associated with pathways that are specific to tumors, some of which are used as biomarkers, so the effectiveness of their technique is proved. The researchers [32] claim that they are the first ones who have applied a convolutional neural network on Pan Cancer Atlas to classify types of tumors, they are the first ones

who implements the contribution of gene's in classification to the importance of genes in identifying candidate biomarkers. Their results suggest that the method performed well and is applicable to other genomics data.

3. Conclusions

In the present paper, different cancers and the machine learning algorithms being used for the detection of those were discussed with regard to an in-depth literature review of the Machine Learning techniques. The findings of most of these studies were acceptable and even in many cases, the scientists could achieve a 100% accuracy of using a machine learning in cancer diagnosis.

Among all the studies that have implemented different machine learning methods, SVM is the one which has been used more for classification of tumors. While most methods are highly appropriate in diagnosing disease, it could be argued that they still need more effort and study, to achieve a better performance in Cancer diagnosis in terms of detection and classification.

References

- [1] Ngadi. M, Amine. A and Nassih. B. (2016). Robust Approach for Mammographic Image Classification Using NSVC Algorithm. *Proc. Mediterr. Conf. Pattern Recognit. Artif. Intell. - MedPRAI-2016*, pp. 44–49.
- [2] Xia. J, Et al. (2017). Ultrasound-based differentiation of malignant and benign thyroid Nodules: An extreme learning machine approach. *Comput. Methods Programs Biomed.* vol. 147, pp. 37–49.
- [3] Thyroid Cancer—Patient Version - National Cancer Institute. [Online]. Available: <https://www.cancer.gov/types/thyroid>. [Accessed: 01-Oct-2018].
- [4] Pancreatic cancer - Symptoms and causes - Mayo Clinic. [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/pancreatic-cancer/symptoms-causes/syc-20355421>. [Accessed: 01-Oct-2018].
- [5] Smith. M. R., Et al. (2018). Apalutamide Treatment and Metastasis-free Survival in Prostate Cancer. *N. Engl. J. Med.* vol. 378, no. 15, pp. 1408–1418.
- [6] Giftson. S, Hari. D. S. and Christy. G. A. J. (2017). *Novel Linear Regressive Classifier for the Diagnosis of Breast Cancer*.
- [7] Tahmooresi. M, Et al. (2018). Early Detection of Breast Cancer Using Machine Learning Techniques. vol. 10, no. 3, pp. 21–27.
- [8] J. Xi, A. Li, and M. Wang, Et al. (2018). A novel unsupervised learning model for detecting driver genes from pan-cancer data through matrix tri-factorization framework with pairwise similarities constraints. *Neurocomputing*, vol. 296, pp. 64–73.
- [9] Rapetti . L Et al. (2017). *Virtual Reality Navigation System for Prostate Biopsy*.
- [10] Taheri .M, Hamer .G, Son S. H, and Shin S. Y., Et al. (2016). Enhanced breast cancer classification with automatic thresholding using SVM and Harris corner detection. *In Proc. of Res. Adapt. Converg. Syst. RACS 2016*. pp. 56–60.
- [11] Pack .C, Et al. (2015). Computer aided breast cancer diagnosis system with fuzzy multiple-parameter support vector machine. *In Proc. of Conf. Res. Adapt. Converg. Syst. – RACS*. pp. 172–176.
- [12] Hu .Y, Li .J and Jiao .Z., (2016). Mammographic Mass Detection Based on Saliency with Deep Features. *In Proc. of Int. Conf. Internet Multimed. Comput. Serv. - ICIMCS'16*. pp. 292–297.
- [13] Zhang .S, Et al. (2017). "We Make Choices We Think are Going to Save Us

- "; Debate and Stance Identification for Online Breast Cancer CAM Discussions. pp. 978–1.
- [14] Naiyar .M,Asim .Y andShahid .A, (2015). Automated Colon Cancer Detection Using Structural and Morphological Features.*In proc. of 13th Int. Conf. Front. Inf. Technol.* pp. 240–245.
- [15] Arya .M, Mittal .N, and Singh .G., (2016). Cervical Cancer Detection Using Segmentation on Pap smear Images. *In Proc. of Int. Conf. Informatics Anal. - ICIA-16.* pp. 1–5.
- [16] Rossetto . A. M and Zhou . W. (2017). Ensemble Convolution Neural Network with a Simple Voting Method for Lung Tumor Detection. *In Proc. of 8th ACM Int. Conf. Bioinformatics.* Comput. Biol. Heal. Informatics - ACM-BCB'17. pp. 729–734.
- [17] Bevilacqua .V,Et al. (2016). An Optimized Feed-forward Artificial Neural Network Topology to Support Radiologists in Breast Lesions Classification. *In Proc. of Genet. Evol. Comput. Conf. Companion - GECCO '16 Companion.* pp. 1385–1392.
- [18] Liu.L, Liu.Y and Zhao. H. (2018). Benign and Malignant Solitary Pulmonary Nodules Classification Based on CNN and SVM.
- [19]Jaganathan.P. and Kuppuchamy.R. (2013). A threshold fuzzy entropy based feature selection for medical database classification.*Comput. Biol. Med.* vol. 43, pp. 2222–2229.
- [20]Shilaskar.Sand Ghatol.A. (2013). Feature selection for medical diagnosis : Evaluation for cardiovascular diseases.*Expert Syst. Appl.* vol. 40, pp. 4146–4153.
- [21]Azar. A., Et al. (2014). A random forest classifier for lymph diseases. *Comput. Methods Programs Biomed.* vol. 113, no. 2, pp. 465–473.
- [22]Ashraf.D andChetty.M.G. (2013). Feature Selection Techniques on Thyroid, Hepatitis and Breast Cancer.*Int. J. Data Min. Intell. Inf. Technol. Appl.*, vol. 3, no. 1.
- [23] Nalband.S, Et al. (2016). Feature selection and classification methodology for the detection of knee-joint disorders.*Comput. Methods Programs Biomed.* vol. 127, pp. 94–104.
- [24] Jabbar. M. A, Deekshatulu. B. L and Chandra. P. (2013). Heart Disease Classification Using Nearest Neighbor Classifier With Feature Subset Selection.
- [25] Wang.D. (2017). LightGBM: An Effective miRNA Classification Method in Breast Cancer Patients. ICCBB-17, pp. 7–11.
- [26] Nejad. E. M, Et al. (2017). Classification of histopathology images of breast into benign and malignant using a single-layer convolutional neural network. *In proc. of Int. Conf. Imaging, Signal Process. Commun. (ICISPC 2017),* pp. 50–53.
- [27] Eyal. N, Last .M, and Rubin .E. (2015). Comparison of three classifiers for breast cancer outcome prediction. *In Proc. of Int. Conf. Eng. Appl. Neural Networks - EANN '15.* pp. 1–6.
- [28] Avramov. T. K. and Si .D. (2017). Comparison of Feature Reduction Methods and Machine Learning Models for Breast Cancer Diagnosis.
- [29] Gc.S, Kasaudhan.R, Et al. (2015). Variability measurement for breast cancer classification of mammographic masses. *In Proc. of Conf. Res. Adapt. Conver. Syst. – RACS-15.* pp. 177–182.
- [30] Kim.J, Lee. H and Yoon.T. (2017). Automated Diagnosis of Lung Cancer with the Use of Deep Convolutional Neural Networks on Chest CT.
- [31] Jiang.Z, Xu.W and Chen.S. (2017). Classification of benign and malignant breast cancer based on DWI texture features. *In Proc. of Int. Conf. Bioinforma. Comput. Intell. – ICBCI-2017,* pp. 11–15.
- [32]Lyu.B and Haque.L. (2018). Deep Learning Based Tumor Type Classification Using Gene

Expression Data. Comput. Biol. Heal.
Informatics.

Authors Profile

Dr. V. Sivakumar is currently working as Senior Faculty at School of Computing in Asia Pacific University of Technology and Innovation (APU), Kuala Lumpur, Malaysia. He had earlier served as Lecturer, Senior Lecturer, Assistant Professor in Gandhigram University (Central Government), India; VNR VignanaJyothi Institute of Engineering & Technology, Hyderabad, India and in the Public Universities of Ethiopia and Libya with a total 20 years of Experience. He was the Gold Medalist of his batch in his Masters. He has done Ph.D. in

Computer Science pertaining to the area of Fractals bound Breast Cancer Analysis. His areas of research include Medical Image Processing, Fractal Analysis and Big Data. He has more than 30 publications in scopus/peer-reviewed journals. He has participated and presented his research articles in several International Conferences. He has got around 100 citation indexes and 6 h-index to his credit.

Ms. M.Tahmooresi was Masters student earlier and currently a Ph.D student of Computing in Asia Pacific University of Technology and Innovation, Malaysia. Her research areas of interest include Machine Learning and Bio-Medical Research.

Table I. Performance analysis on different types of methodology, dataset, and references for Tumor detection

Reference	Type of Cancer	Performance	Machine Learning Technique
[2]	Thyroid	AUC:0.8672 Sensitivity: 78.89% Specificity: 94.55%	ELM
[9]	Breast	Precision: 96.8% Recall Rate: 92.5%	SVM
[10]	Breast	MCC: 86.16 Sensitivity: 96% Specificity: 90% Accuracy: 93%	Fuzzy Multiple Parameter
[13]	Colon	Accuracy: 96.55% Sensitivity: 0.98 Specificity: 0.94	SVM
[14]	Lung	Accuracy: 96% Specificity: 100% Sensitivity: 95.6%	SVM
[16]	Breast	Accuracy: 89.77% Sensitivity: 89.08% Specificity: 90.46%	CAD
[17]	Lung	Accuracy: 91.94%	CNN+ PSO+ SVM
[18]	Breast Diabetes Heart Statlog Hepatitis Heart Diseases	Accuracy: 97.28% Accuracy: 76.04% Accuracy: 85.19% Accuracy: %85.16 Accuracy: 84.46%	RBF
[19]	SPECTF Arrhythmia Heart Diseases	Accuracy: 78% Accuracy: 88% Accuracy: 85%	SVM
[20]	Lymph	Accuracy: 92.2%	Random Forest
[21]	Breast Hepatitis Thyroid	Accuracy: 96.28% Accuracy: 87.74% Accuracy: 97.22%	NB CFS DT
[22]	Knee Joint Disorders	Accuracy: 94.31%	Least Square SVM
[23]	Liver Diabetes Lymphography Primary Tumor Heart Statlog Breast Cancer Heart Diseases 1 Heart Diseases 2	Accuracy: 100% Accuracy: 100% Accuracy: 100% Accuracy: 61.35% Accuracy: 100% Accuracy: 96.85% Accuracy: 97.5% Accuracy: 100%	KNN
[24]	Breast	Best Accuracy: 97.56% Best Accuracy: 100%	XGBoost LightGBM
[25]	Breast	Accuracy (Without Data Augmentation): 77.3% Accuracy (With Data Augmentation): 77.5%	CNN

[26]	Breast	Accuracy (NN): 70.50% Accuracy (DT): 70.40%	NN, DT
[28]	Breast	Variance: (MCC:83.2%, Sensitivity: 95%, Specificity: 88%, Accuracy: 91.5%) Range: (MCC:82.1%, Sensitivity: 94%, Specificity: 88%, Accuracy: 90.5%) Compactness: (MCC:70%, Sensitivity: 86%, Specificity: 84%, Accuracy: 85%)	SVM
[29]	Lung	Accuracy (with 3 Hidden Layer): 72% Accuracy (with 5 Hidden Layer): 72% Accuracy (with 7 Hidden Layer): 74% Accuracy (with 10 Hidden Layer): 73%	Deep CNN
[31]	Breast	Histogram: (Accuracy: 68.85%, Sensitivity: 76.32%, Specificity: 56.52%, AUC: 0.73) GLCM: (Accuracy: 65.57%, Sensitivity: 71.05%, Specificity: 56.52%, AUC: 0.63) Histogram + GLCM: (Accuracy: 77.05%, Sensitivity: 84.21%, Specificity: 65.21%, AUC: 0.76)	CAD
[32]	Pan Cancer	Accuracy: 95.59% Precision: 95.54% Recall: 95.59% F1.Score: 95.43%	CNN