A SURVEY ON VARIOUS MACHINE LEARNING APPROACHES USED FOR BREAST CANCER DETECTION

Neha Kumari and Khushboo Verma

Bansal Institute of Science and Technology,

Bhopal, India

Abstract: Now a day's cancer is one of the main decreases in all over the world. Several peoples have died in a day. According to the survey conducted by the US government, 40000 people died in 2012 only due to breast cancer. Cancer decease is classified into four types named type 1, type 2, type 3 and type 4. A previous survey that if cancer is detected in the early stage (i.e., type 1 and type 2) then only it can be procuring. But most of the time cancer is detected in the third and fourth stage. Due to this reason cancer detection in the early stage is one of the favorite areas of the researcher. In the past few decades, several machine learning approach has been used by various researchers. Cancer detection is a classification approach where the main aim is to find the cancer stage in the early stage. There are several classification approaches that can be used in cancer detection. This paper discusses the comparative analysis of some of the existing cancer detection approaches.

Keywords: machine, learning, breast, cancer, treatment

I. INTRODUCTION

Now a days Brest cancer is one of the common decease for the woman. Every year thousands of woman died due to the breast cancer. Estimated 1.38 and 1.67 million new cancer cases diagnosed in2008(23% of all cancers) and 2012 respectively (25% of all cancers) and ranks second overall (10.9% of all cancers) [1, 2]. According to the survey 2 million new cases have been identified in 2018 [3]. In Asia, 39 percentage of new breast cancer have been encountered and out of which 44 percentage patient have died. Table 1 shows countries wise breast cancer cases in 2008 and 2012.

It is observed breast cancer survival rate can be increased with the early cancer detection. The main cause of the low survival rate is the late detection of the cancer and the difficult diagnostic process. Mammography is the technique which is used to identified cancer tumour of the breast using x-rays. The Doctors use this X-ray to diagnosis the cancer. Some time it is very difficult to predict cancer stage due to the human fatigue and habituation [3]. Early detection of cancer boosts the increase of survival chance to 98% [3]. Figure 1. shows different types of cancers whereby breast cancer is leading with 24% as follows.
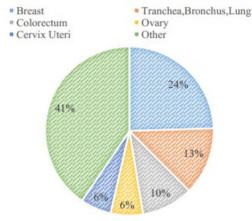

Figure 1: Types of Cancer

Table 1: Worldwide Breast Cancer Cases in 2008 and 2012

Estimated age standardized rates (World) (per 100,000) Estimated numbers (thousands)	Cases 2008	Deaths 2008	Cases 2012	Deaths 2012
World	1384	458	1677	522
More developed regions	692	189	794	198
Less developed regions	691	269	883	324
WHO Africa region (AFRO)	68	37	100	49
WHO Americas region (PAHO)	320	82	408	92
WHO East Mediterranean region (EMRO)	61	31	99	42
WHO Europe region	450	139	500	143
WHO South-East Asia region (SEARO)	203	93	240	110

WHO Western Pacific region (WPRO)	279	73	330	86
IARC membership (24 countries)	740	214	940	257
United States of America	182	40	233	44
China	169	44	187	48
India	115	53	145	70
European Union (EU-28)	332	89	367	91

Several machine learning approach have been used in the past few years in order to predict cancer stage. Machine learning is a sub domain of AI, where the main objective is to train machine in such a way that it can take decision just like a human being. Today's machine learning is heavily used in medical domain to predict cancer progress, planning for therapy and all assistant for patient. One of the main advantage of the machine learning approach is that it can predict decease based on the historical data so some time it gives better result compare to the physician [4]. Each machine approach has the following steps as shows in figure 2.

i.   Input
ii. Feature extraction
iii. Feature selection
iv. Apply machine learning algorithm
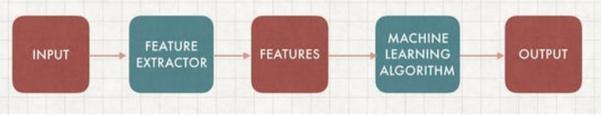v.   Output

Figure.2: Traditional Machine Learning Algorithm Flow

Some of the common approach used for the cancer diagnosis are Naive Bayes Classification, K Nearest Neighbor Classification, Multilayer Perceptron and support vector machine. A Naive Bayes classifier is a classifier which is based on Naïve hypothesis. It is a probabilistic model which assume two features are independents. Following equation is used for the Naïve classifier

$$p(c \mid v_1, v_2, ... v_n) = p(c \mid v_1, v_2 ... v_n) * \frac{p(v_1, v_2 ... v_n \mid c)}{p(v_1, v_2 ... v_n)}$$

Which can be written as

Posterior= prior* $\frac{liklihood}{evidence}$

One of the main advantage of the Naïve classifier is that it can be trained on small data set and gives result very fast. The main limitation of the this approach is that is gives less accuracy.

KNN is a supervised learning method which is used for diagnosing and classifying cancer [5]. In this method, the computer is trained in a specific field and new data is given to it. Additionally, similar data is used by the machine for detecting (K) hence, the machine starts finding KNN for the unknown data. It is recommended to choose a large dataset for training also K value must be an odd number.

Support vector machine (SVM) is a supervised pattern classification model which is used as a training algorithm for learning classification and regression rule from gathered data [6]. The purpose of this method is to separate data until a hyperplane with high minimum distance is found.

II. LITERATURE SURVEY

In the last few decades' lots of work have been done by the various researcher. This section will discuss various approach which is used in cancer detection. The uses of classification systems in medical diagnosis, including breast cancer diagnosis, are growing rapidly. Evaluation and decision making process from expert medical diagnosis is key important factor. The term of breast cancer is referred to a malignant tumor that has developed from cells in the breast. Mostly, it is found in women but men also can get breast cancer even it is rare. It is affecting about 10% of all women at some stages in their life. From American Cancer Society, breast cancer is the second reason of death for women in the United States. Factors that may cause the breast cancer are age, family history, race and many more. A tumor can be classified as benign or malignant. The Wisconsin Breast Cancer Dataset was developed by Dr.Wolberg in 1995. Thereafter, a lot of work is done by various medical and allied practitioners to simplify the breast cancers detection and prognosis.

Hiba et. al. [4], says that breast cancer is responsible for high number of deaths every year and

the percentage of death due to breast cancer is increasing every year. It is the most common type of all cancers and the main cause of women's deaths worldwide. Classification and data mining methods are an effective way to classify data. Especially in medical field, where those methods are widely used in diagnosis and analysis to make decisions. In this paper, a performance comparison between different machine learning algorithms: Support Vector Machine (SVM), Decision Tree (C4.5), Naive Bayes (NB) and k Nearest Neighbors (k-NN) on the Wisconsin Breast Cancer (original) datasets is conducted. The main objective is to assess the correctness in classifying data with respect to efficiency and effectiveness of each algorithm in terms of accuracy, precision, sensitivity and specificity. Experimental results show that SVM gives the highest accuracy (97.13%) with lowest error rate. This paper is considered as base paper for our research. However, intelligent classification algorithm may help doctor especially in minimizing error from unexperienced practitioners [5].

Comprehensive view of automated diagnostic systems implementation for breast cancer detection was provided by Ubeyli [6]. It compared the performances of multilayer perceptron neural network (MLPNN), combined neural network (CNN), probabilistic neural network (PNN), recurrent neural network (RNN) and support vector machine (SVM). The aim of that works was to be a guide for a reader who wants to develop this kind of systems.

According to kourou et. al. [7] has stated that early detection of cancer can be done by using machine learning algorithms and they have also shown how these algorithms work better in the area of classification between cancer and non cancer patients.

Zhou et. al. [8] has shown how cancer classification and prediction can be done using logistic regression with Bayesian gene selection approach.Several techniques have been deployed to predict and recognize meaningful pattern for breast cancer diagnosis.

Ryua [9] developed data classification method, called isotonic separation. The performances were compared against support vector machines, learning vector quantization, decision tree induction, and other methods based on two-breast cancer data set, sufficient and insufficient data. The experiment results demonstrated that isotonic separation was a practical tool for classification in the medical domain

Sweilam et. al. introduced least square support vector machine and active set strategy to show the classification on breast cancer dataset [10]. Khosravi et. al describes Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images. This work is mainly focused on several computational methods based on convolutional neural networks(CNN) and build a

stand-alone pipeline to effectively classify different histopathology images across different types of cancer [11] .

Hybrid machine learning method was applied by Sahan [12] in diagnosing breast cancer. The method hybridized a fuzzy artificial immune system with k-nearest neighbour algorithm. The hybrid method delivered good accuracy in Wisconsin Breast Cancer Dataset (WBCD). They believe it can also be tested in other breast cancer diagnosis problems.

In 1999 Xin Yao et al. has implemented artificial neural network [13] for breast cancer diagnosis using negative correlation training algorithm using two approaches viz. evolutionary and ensemble approach. In 2004 Tuba Kiyanet al.[14] has applied Neural Network on WBCD to estimate the diagnosis accuracy of various techniques. In 2007Sumathi et al.[14] have used genetic algorithms approach to WBDC and found that genetic algorithm not only improve the accuracy but also reduce the time taken to train the network. In 2009, Y. Iraneus et. al.[15] used SVM for early detection of breast cancer. In 2012, Muhammad Rafi et al. used SVM and RVM techniques for documentclassification without using minimum accuracy limit and find that predicting accuracy of RVM is much higher than SVM [16].

David B.fogel et al. [17] has discussed the evolving neural networks for detecting breast cancer and the related works used for breast cancer diagnosis using back propagation method with multilayer perceptron. In contrast to back propagation David B.fogel et al. found that evolution computational method and algorithms were used often, outperform more classic optimization techniques.

In 2012,Z.Qinli et al. [18] has presented an article on, a approach to SVM and its application to breast cancer diagnosis. In this article, the authors have proposed a method for improving the performance of SVM classifier by modifying kernel functions. This is based on the differential approximation of metric. The method is to enlarge margin around separating hyper plane by modifying the kernel functions using a positive scalar functions so that the seperability is increased. It is observed that it is competent to reduce the generalization error and computational cost.

Afzan Adam et al., [19] introduced a computerized breast cancer diagnosis by combining genetic algorithm and Back propagation neural network which was developed as faster classifier model to reduce the diagnose time as well as increasing the accuracy in classifying mass in breast to either benign or malignant. In these two different cleaning processes was carried out on the dataset. In Set A, it only eliminated records with missing values, while set B was trained with normal statistical cleaning process to identify any noisy or missing values. At last Set A gave 100% of highest accuracy percentage and set B gave 83.36% of accuracy. Hence the author has

concluded that medical data are best kept in its original value as it gives high accuracy percentage as compared to altered data.

III. CONCLUSION

This paper gives the introduction of some existing breast cancer algorithm which is used for breast cancer detection. Cancer detection is a classification problem where you need to classified cancer stage into two stages. There are several machine learning and AI based approach are used for the detecting cancer in early state. Based on the survey conducted in this paper suggest that support vector machine gives more accurate and fast result as compare to the others machine learning approaches. Maximum accuracy that can be achieved by SVM is 99.8%. Because of this reason SVM is one of the most popular approach for the classification. As the accuracy of the machine learning algorithm is depending upon the data set accuracy. There are several breast cancer datasets are available for testing, so accuracy of any ML algorithm is vary according to the dataset.

IV. REFERENCES

[1] http://www.breastcancerindia.net/

[2] GLOBOCAN: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in (2012).

[3] A. K. Dubey et al., "A Survey on Breast Cancer Scenario and Prediction Strategy", Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014 pp 367-375.

[4] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, ―Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis,‖ProcediaComput. Sci., vol. 83, no., pp. 1064–1069, 2016.

[5] M. F. Akay, ―Support vector machines combined with feature selection for breast cancer 50 diagnosis,‖ Expert Syst. Appl., 2009.

[6] E. D. Übeyli, ―Implementing automated diagnostic systems for breast cancer detection,‖ Expert Syst. Appl., 2007.

[7] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V Karamouzis, and D. I. Fotiadis, Machine learning applications in cancer prognosis and prediction,‖ CSBJ, vol. 13, pp. 8–17, 2015.

[8] X. Zhou, K. Liu, and S. T. C. Wong, ―Cancer classification and prediction using logistic regression with Bayesian gene selection,‖ vol. 37, pp. 249–259, 2004.

[9] Y. U. Ryu, R. Chandrasekaran, and V. S. Jacob, ―Breast cancer prediction using the isotonic separation technique,‖ Eur. J. Oper. Res., 2007.

[10] N. H. Sweilam, A. A. Tharwat, and N. K. A. Moniem, ―Support vector machine for diagnosis cancer disease : A comparative study,‖ Egypt. Informatics J., vol. 11, no. 2, pp. 81–92, 2010.

[11] P.Khosravi, E. Kazemi, and M. Imielinski, ―EBioMedicine Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images,‖ EBioMedicine, vol. 27, pp. 317–328, 2018.

[12] S.Şahan, K. Polat, H. Kodaz, and S. Güneş, ―A new hybrid method based on fuzzy artificial immune system and k-nn algorithm for breast cancer diagnosis,‖ Comput. Biol. Med., 2007.

[13] X. Y. X. Yao and Y. L. Y. Liu, ―Neural networks for breast cancer diagnosis,‖ Proc. 1999 Congr. Evol. Comput. Cat No 99TH8406, 1999.

[14] T.Kiyan and T. Yildirim, ―Breast cancer diagnosis using statistical neural networks,‖ IUJournalElectr. Electron. Eng., vol. 4, no. 2, pp. 1149–1153, 2011.

[15] Y. I. A. Rejani and S. T. Selvi, ―Early Detection of Breast Cancer Using SVM Classifier Technique,‖ Int. J. Comput. Sci. Eng., 2009.

[16] M. Rafi and M. S. Shaikh, ―A comparison of SVM and RVM for Document Classification,‖ Procedia Comput. Sci., 2013.

[17] K. I. Al-Sulaiti et al., ―Research Methods for Business Students,‖ Int. Mark. Rev., 2010.

[18] W.Shitong, ―A Novel SVM and Its Application to Breast Cancer Diagnosis,‖ 2007.

[19] A. Adam and K. Omar, ―Computerized Breast Cancer Diagnosis with Genetic Algorithms and Neural Network Keywords.