International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| ISSN: 2395-7852 | www.ijarasem.com | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 6, November 2020 |

Machine Learning Approach in Prognosis of Cancer Using Support Vector Machine

Atharva Bansode

Academic Researcher, Department of Information Technology, B. K. Birla College of Arts , Science and Commerce (Autonomous), Kalyan, Maharashtra, India

ABSTRACT: Now-a-days the humans have prioritized the use of machine learning in almost every sectors. In recent years with the concepts of AI, ML and DL, more advancements can be noticed in this fields, as along with accuracy the use of machine learning algorithms are capable of taking decisions on the spot. Machine Learning (ML) has been developed over a past few years to develop prognostic classification models that can be used for predicting the outcomes of an individual cancer patient. ML involves the development of algorithms that, rather being explicitly programmed to assign specific outputs, it analyses the data and its properties by using statistical tools and gives output accordingly by determining the current scenario actions.[1] The support vector machine (SVM) is a rising machine learning approach that offers robust classification of high-dimensional big data into small numbers of data points i.e. (support vectors), achieving differentiation of subgroups in a shorter amount of time.[3] The SVM models are trained in such a way that they are used to forecast breast cancer subtypes using diverse system science data that includes transcriptomics, epigenetics, proteomics and radiomics along with biological pathway, clinical and biochemical data.[3]

KEYWORDS: Breast Cancer detection, Machine Learning, Support Vector Machine, Breast Cancer Risk Prediction

I. INTRODUCTION

In today's world diagnosis and treatment of cancer has become most formidable challenge in health care innovations.[3] Today breast cancer is one of the most widely spread disease and second most leading cause of cancer death in women.[4] Breast cancer has become one of the most common diseases worldwide. Approximately 2.09 million new cancer cases were diagnosed and examined.Over 627,000 related deaths occurred due to breast cancer globally.[7]Breast cancer majorly occurs in women aged 40 years and above and it occurs because of the cells in the glands that produce milk become abnormal and divide drastically.[4] Breast cancer starts when the malignant lumps which are also cancerous begin to grow from the breast cells in women. Detecting breast cancer efficiently has become a major concern and hence breast cancer risk predictions can be carried out efficiently through screening technique, SVM (support vector machines) and other preventative action. Since breast cancer can be detected in early stages with proper machine learning techniques there are high possibilities of survival of an individual as machine learning (ML) plays a vital role in a wide range of critical applications such as data mining, natural language processing, image recognition, expert systems and prediction, etc..[5]

Machine learning is a sub -field of computer science that has evolved from the study of pattern recognition and different computational learning theories.[10] ML uses pattern recognition systems to learn and predict the outcomes accordingly with the help or supervised or unsupervised learning. [5] Sometimes there is also the need of computer aided detection (CAD) systems which uses machine learning approach to diagnose breast cancer in the early stages.[4] Moreover ML is also considered as a field of study in artificial intelligence that gives computers the ability to learn from data.

II. RELATED WORK

Breast cancer is second most severe cancer among women if compared to all the different types of cancers.With rapid growth in population in medical research, it has become crucial in detecting the disease and hence the death incurred by breast cancer is rising rapidly. Various Machine learning techniques are introduced to diagnose and predict the outcomes of an individual cancer patients. The rate of breast cancer has decreased steadily over past 2 decades because of the development in management pathways , from early detection to treatment. Breast cancer still is the leading cause of cancer deaths among women from all over the world. A. Sidhu et al. stated that in Malaysia , breast cancer has found to the most common form of cancer in women. Breast cancer being the most common cancer in women is known to affect atleast 10% to all women at some stage in their life [4]. In a developed world , between 1 in 8 and 1 in 12 women is likely to have breast cancer during their life time [2]. Diagnosis and treatment of cancers become most formidable challenge in the health care innovations for reducing the rate of cancer [3].

The interest in machine learning is growing rapidly over a past few decades because of the cheaper computing power and low-cost memory [4]. Large amount data can be stored analysed and processed using various machine learning techniques. David A. et al. stated that doctors may sometime wrongly diagnose benign tumour (which is non-cancerous) as malignant tumour. Therefore there is a need of a proper Computer Aided Detection (CAD) system that uses machine learning approach in diagnosis of breast cancer .[6]

K. Arga et al. stated that the SVM models are designed to forecast breast cancer subtypes using diverse system science data as well as biological pathways and biochemical data. SVM and machine learning offers new solution and ways forward in biomedical ,clinical and bioengineering applications. G. Stark et al. stated that there are two types of primary breast cancer risk . First type represents the probability that individual will contract breast cancer at a specific age period and second type is that there is a probability that mutation will occur at high risk gene [2]. Machine learning algorithms in robotics are used to tackle difficult problems using autonomous control and sensing techniques, where large quantities of datasets are available enabling the robots to effectively teach themselves which helps in determining and effectively curing the cancer[12].

III. METHODOLOGY

BREAST CANCER RISK PREDICTION SURVEY TABLE (BCRAT)

For the below given table an experiment was conducted in which machine learning models were developed that used highly accessible personal health data that helped in predicting the five year breast cancer risk. The breast cancer risk prediction tool [BCRAT] was implemented using the R programming method of version 3.4.3 with BCRA packages. For every individual we set the values for the previous breast biopsies and for indicator in hyper plasia in a biopsy to 99 which showed that the values presented were unknown . The BCRAT predicted the table for the five years risk using the “BCRA” packages and “absolute risk” function to calculate BCART risk prediction.

Input	Breast Cancer	Non- breast Cancer
Age	62.6 (± 5.4)	62.5 (± 5.4)
Age at Menarche	12.6 (± 1.6)	12.7 (± 1.5)
Age at First Live Birth	23.7 (± 4.6)	23.0 (± 4.3)
Number of First-Degree Relatives Who previously had Breast Cancer	0.2 (± 0.6)	0.2 (± 0.4)
Race / Ethnicity		
White	94.8%	92.8%
Black	3.6%	5.6%
Hispanic (Born in US)	1.4%	1.3%
Hispanic (Born Abroad)	0.2%	0.3%
Age at Menopause	48.6 (± 5.0)	48.0 (± 5.1)
Current Hormone Usage	58.8%	50.6%
Years of Hormone Usage	4.7 (± 4.1)	4.1 (± 4.1)
BMI	I 27.1 (± 5.3)	27.3 (± 5.6)
Pack Years of Cigarettes Smoke	14.6 (± 23.9)	13.3 (± 22.4)

Birth Control Usage	2.9 (± 3.6)	2.7 (± 3.6)
Number of Live Births	2.8 (± 1.4)	2.9 (± 1.5)
Personal History of Prior Cancer	4.2%	3.4%

https://doi.org/10.1371/journal.pone.0226765.t0

After comparing the machine learning models with border set of inputs it suggested that the model that incorporate both BCRAT inputs and additional easy to obtain personal health inputs can effectively predict breast cancer risk.

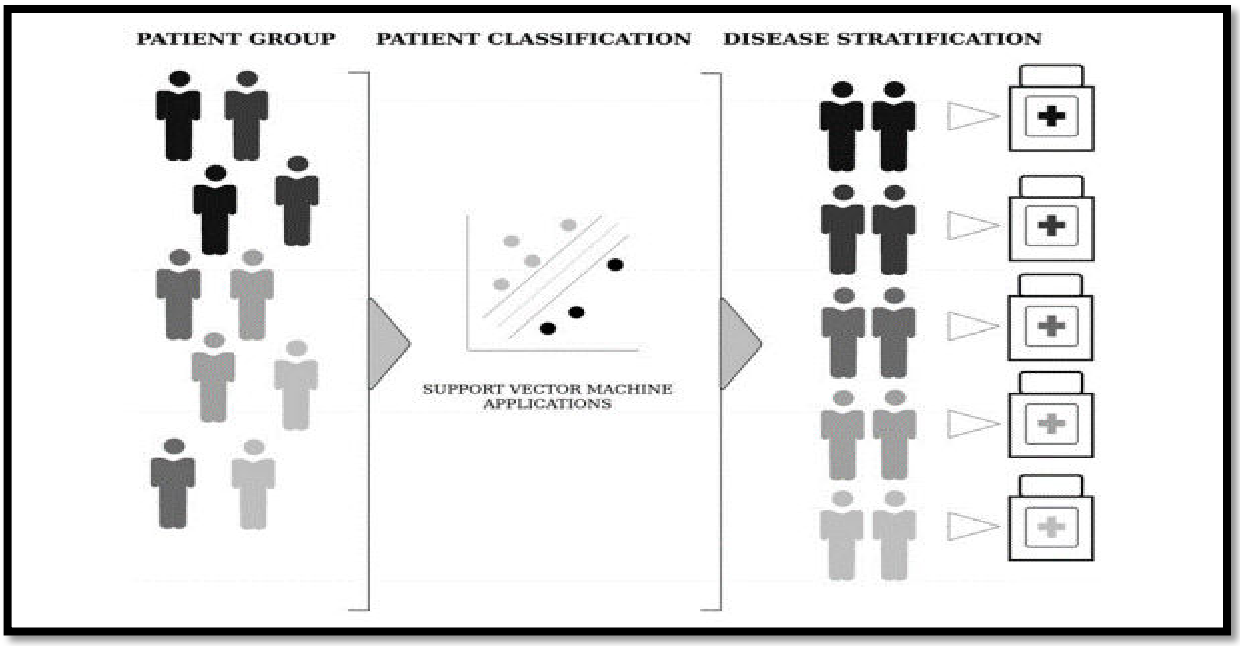
Support Vector Machine (SVM)

SVM is a rising machine learning approach that offers robust classification of high dimension big data into small number of data points called support vectors. In SVM method, hyperplane is an imaginary boundary line that is used to separate the data points forming two sides of different classes.[3] During classification , data points that belong to either sides of the hyperplane are optimized in such a way that it provides highest distance margin between two sides of data. These data points are called support vectors.

Breast cancer is the most malignant disease among women worldwide and also the leading cause of death after lung cancer.(Kazim et al.,2020) .Important features in determining the breast cancer are , lymphovascular invasion of cells and axillary lymph nodes (ALN). Besides this there are also other features such as estrogen receptor (ER) , human epidermal growth factor receptor (HER2), progesterone receptor (PR) on surface of tumour cells. Breast cancer is also known for having heterogeneity under both temporal and spatial dimensions, which eventually causes problems during the dissection process.

SVM technique is important for its application in classifying and selecting informative features such as genes , proteins and pathways.[3] Since the classification of big data is computationally expensive, in many applications a hybrid methodology of SVM is observed which also includes data filtration components. In SVM , due to the high dimensionality in applied data sets, groups having small data set and high number of dimensions can be properly classified. (Gao et al., 2017).In this methodology a brief review of Breast Conserving Surveys (BCS) classification is presented using the SVM models. it also discusses the applicability of various data types in these models and compares its accuracy to other prediction models. it should be noted that once the problem of lack of data integration is solved , the prediction accuracy and performance of the models will highly improve.

Fig 1: CONCEPTUAL DESCRIPTION OF DISEASE STRATIFICATION USING SUPPORT VECTOR MACHINE APPLICATION.



Src: https://www.liebertpub.com/doi/abs/10.1089/omi.2020.0001

Omics-Based Prediction Models :

Fig 2: SVM SUBTYPES FOR SIGNIFICANT RELEVANCE TO STRATIFICATION OF BREAST CANCER.

Approach	Technology / Data	SVM type
Transcriptome	RNA -seq	v-SVM
		SVDD
	miRNA	SVM with 5-fold CV
	Microarray	SVM
Methylome	DNA methylation	SVM with 5-fold CV
Proteome	SILAC	SVM with ONE-vs- REST
	2D-DIGE	SVM-LOOCV
Pathway	Pathway enrichment analysis	SVM
		SVM with 20-fold CV
Radiome	MRI	SVM with LOOCV
	Ultrasound	SVM with 3-fold CV
	DCE-MRI	SVM
	DWI	SVM
Clinical–pathological	Patient clinical data and tumor characteristics	SVM
Biochemical	Raman spectra of lipids, nucleic acids, and proteins	SVM

Src: https://www.liebertpub.com/doi/abs/10.1089/omi.2020.0001

(2D-DIGE : 2-dimensional, differential in-gel electrophoresis; LOOCV: leave-one-out cross-validation; miRNA: microRNA; DCE-MRI: dynamic contrast-enhanced magnetic resonance imaging; RNA-seq: RNA sequencing; SVDD: support vector data description; SVM: support vector machine; SILAC: stable isotope labelling with amino acids in cell culture ; SVM: variant of support vector machine.)

SVM models are also developed using the x-comics data sets which also includes high-throughput data from different biological levels. High prediction accuracies can be obtained depending on the quality of data these models present. Predictive models are constructed on basis of the inter-condition expressions such as differences of RNA's and ncRNA's (non coding RNA). similarly, mRNA and microRNA (miRNA) transcriptome produces the data sets that can be used in different technologies such as RNA sequencing and microarrays which are among the frequently used omics datasets employed with ML techniques which also includes SVM for differentiation of BCS's. SVM methods are applied using two or more class methods . On the contradiction to the two class methods, negative and positive classes are given together that does not affect the accuracy of the model performance in one-class SVM.

Various machine learning methods which are used in employing transcriptome data sets are used in stratification of BCS's. However, its accuracy depends on the applied data sets or the sample data sets that are used. Furthermore in the case of employing the RNA-seq-based miRNA expression data in distinguishing BCS's , it was found that SVM with fivefold cross-validation indicated higher performance in terms of AUC metrics when compared with fisher score method (Gu et al.,2012). SVM also shows high accuracy in the identification of potential miRNA biomarkers using sub-type microarray data.[3] The SVM with fivefold cross validation data yielded 100% accuracy in prediction of intrinsic subtypes and 90% in prediction of BRCA1 (breast cancer 1) whereas the predictions of BRCA2 and BRCAx failed. The predictive performance of the omics- based approaches can be improved if the quality of data is enhanced.

Besides Omics based prediction models other prediction models were also tested namely Pathway-Based Prediction Models and Image-Based Prediction models. Harnessing the pathway information and relationships between the cellular molecules have helped in reducing the noise in omics data and has paved a way for pathway based higher performance prediction models. The SVM based classification of BCS's using pathway-based biomarkers showed the accuracy of more than 90% in the prediction of luminal and basal-like patients , but it was inaccurate in prediction of other BCS's. In addition to omics-based prediction models and pathway-based models, imaging-based prediction tools are frequently used. Magnetic Resonance Imaging (MRI) and ultrasonography are the widespread techniques in detecting breast cancer. Several studies employed the SVM process to improve the interpretation of images and check whether the characteristics of images could provide differentiation among BCS's. SVM used LOOCV data process for analysing MRI data. The experts examined the employment of SVM prediction models in classification of BCS's and it showed that the SVM methods involving radiomic data had higher accuracy in predicting breast cancer without failure in discriminating BCS.

IV. CONCLUSION

SVM is one of the promising approach for development and application of best medical practices and will be used more frequently in the upcoming decades. Along with this, there are few drawbacks of machine learning at the moment , one of them is that the inner working of these self-learning machines is a black box , which makes it difficult to understand the reasoning process and justify the recommendations made.[5] Although machine learning helps in creating brilliant and useful tools most of the times the inner working is not clearly explained . The Black Box inside its algorithms makes recommendations more mysterious to understand even to the creators which eventually leads to information asymmetries among AI experts , users and other stakeholders.[6]Sometimes trusting the machines becomes a challenge because machine learning using he big data can be susceptible to human biases that are inherent in the data. People can completely trust the machine learning only when the reasoning process will be known and will be interpretable.[6] In the upcoming time period with the advancement in machine learning approaches cancer prediction will become more accurate and efficient.

REFERENCES

[1] Goldenberg, S., Nir, G. &Salcudean, S.E. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol* 16, 391–403 (2019).https://doi.org/10.1038/s41585-019-0193-

[2] G.F.S.G.R.H.B.J.N. (2019, December 27). Predicting breast cancer risk using personal health data and machine learning models. Https://Doi.Org/10.1371/Journal.Pone.0226765. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0226765e/article?id=10.1371/journal.pone.0226765.

[3] New Machine Learning Applications to Accelerate Personalized Medicine in Breast Cancer: Rise of the Support Vector Machines. (2020, May 7). Https://Doi.Org/10.1089/Omi.2020.0001. https://www.liebertpub.com/doi/abs/10.1089/omi.2020.0001

[4] D.A.O. (n.d.). Machine Learning Classification Techniques for Breast Cancer Diagnosis. Https://Iopscience.Iop.Org. Retrieved November 21, 2020, from https://iopscience.iop.org/article/10.1088/1757-899X/495/1/012033/meta

[5] Rimmer, L., Howard, C., Picca, L. et al. The automaton as a surgeon: the future of artificial intelligence in emergency and general surgery. Eur J Trauma EmergSurg (2020). https://doi.org/10.1007/s00068-020-01444-8

[6] Wang, Weiyu and KengSiau. "Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda." JDM 30.1 (2019): 61-79. Web. 21 Nov. 2020. doi:10.4018/JDM.2019010104

[7] Murata, T., Yanagisawa, T., Kurihara, T. et al. Salivary metabolomics with alternative decision tree-based machine learning methods for breast cancer discrimination. Breast Cancer Res Treat 177, 591–601 (2019). https://doi.org/10.1007/s10549-019-05330-9

[8] P.F.M.R. (2018, December 18). Breast Cancer Prognosis Using a Machine Learning Approach. Https://Doi.Org/10.3390/Cancers11030328. https://www.mdpi.com/2072-6694/11/3/328#cite

[9] smail Fawaz, H., Forestier, G., Weber, J. et al. Accurate and interpretable evaluation of surgical skills from kinematic data using fully convolutional neural networks. Int J CARS 14, 1611–1617 (2019). https://doi.org/10.1007/s11548-019-02039-4

[10] Zhao, B., Waterman, R.S., Urman, R.D. et al. A Machine Learning Approach to Predicting Case Duration for Robot-Assisted Surgery. J Med Syst 43, 32 (2019). https://doi.org/10.1007/s10916-018-1151-y

[11] D.A.O. (n.d.). Machine Learning Classification Techniques for Breast Cancer Diagnosis. Https://Iopscience.Iop.Org. Retrieved November 21, 2020, from https://iopscience.iop.org/article/10.1088/1757-899X/495/1/012033/meta

[12] A.M.A.R.V.-K. (2016, September 23). Integration of Machine Learning and Optimization for Robot Learning. Https://Link.Springer.Com. https://link.springer.com/chapter/10.1007/978-3-319-46490-9_47

[13] K.S.W.W. (n.d.-b). Building Trust in Artificial Intelligence, Machine Learning, and Robotics. Https://Www.Researchgate.Net. Retrieved November 21, 2020, from https://www.researchgate.net/profile/Keng_Siau/publication/324006061_Building_Trust_in_Artificial_Intelligence_Machine_Learning_and_Robotics/links/5ab8744baca2722b97cf9d33/Building-Trust-in-Artificial-Intelligence-Machine-Learning-and-Robotics.pdf

[14] A.M. .a.n.d. .A.R.V.-K. (n.d.). Integration of Machine Learning and Optimization for Robot Learning. Https://Eprints.Qut.Edu. Retrieved August 25, 2020 from https://link.springer.com/chapter/10.1007/978-3-319-46490-9_47

[15] V.S.Y.-H.C.J.E.A.S.Z.Z.M.I. .o.f. .T.C.M.A. (2017, October 17). Hardware for Machine Learning: Challenges and Opportunities. Https://Ieeexplore.Ieee.Org. https://arxiv.org/pdf/1612.07625.pdf

[16] S.O.N.N. (2018, November 5). Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. DOI.ORG. https://onlinelibrary.wiley.com/doi/abs/10.1002/rcs.1968

[17] P.S. (2019, August). Artificial Intelligence and the Future of Surgical Robotics. Annalsofsurgery.Com. https://journals.lww.com/annalsofsurgery/Fulltext/2019/08000/Artificial_Intelligence_and_the_Future_of_Surgical.7.aspx

[18] James Johnson (2019) The end of military-techno Pax Americana? Washington’s strategic responses to Chinese AI-enabled military technology, The Pacific Review, DOI: 10.1080/09512748.2019.1676299. https://www.tandfonline.com

[19] J. (2020a, June 6). Dentronics: Towards robotics and artificialintelligence in dentistry. Sciencedirect.Com. https://www.sciencedirect.com/science/article/abs/pii/S0109564120300762

[20] Benjamin M Jensen, Christopher Whyte, Scott Cuomo, Algorithms at War: The Promise, Peril, and Limits of Artificial Intelligence, International Studies Review, Volume 22, Issue 3, September 2020, Pages 526–550, https://doi.org/10.1093/isr/viz025

[21] K.E.I.T.H. (n.d.-b). Space War and AI. Emrald.Com. Retrieved September 12, 2020, from https://www.emerald.com/insight/content/doi/10.1108/978-1-78973-811-720201004/full/html

[22] Islam, M.M., Haque, M.R., Iqbal, H. et al. Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. SN COMPUT. SCI. 1, 290 (2020). https://doi.org/10.1007/s42979-020-00305-w

[23] Lee S, Liang X, Woods M, Reiner AS, Concannon P, Bernstein L, et al. (2020) Machine learning on genome-wide association studies to predict the risk of radiation-associated contralateral breast cancer in the WECARE Study. PLoS ONE 15(2): e0226157. https://doi.org/10.1371/journal.pone.0226157

[24] G.F.S.G.R.H.B.J.N. (2019, December 27). Predicting breast cancer risk using personal health data and machine learning models. Https://Doi.Org/10.1371/Journal.Pone.0226765. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0226765e/article?id=10.1371/journal.pone.0226765.