Implementing Artificial Intelligence Algorithms for Predicting Survival Rates in Breast Cancer Patients

Kalluru Madhavi

Department of Computer Science Sri Venkateswara University, Tirupati

Abstract— *Breast cancer ranks as the most prevalent cancer type among women worldwide, being the second highest cause of female mortality among all cancer types. Accurately predicting the survival rate of breast cancer patients is a critical concern for cancer researchers. Machine Learning (ML) has garnered considerable attention for its potential to provide precise results, yet its methodologies and predictive performance remain debatable. This paper focuses on employing ML algorithms to predict Haberman's Breast Cancer Survival study. Specifically, two different ML approaches, namely Multilayer Perceptron (MLP) and Support Vector Machine (SVM) models, are explored for Breast Cancer Survival prognosis. The classification performance of abnormal and normal Breast Cancer Survival patients is assessed across various metrics including training and testing accuracy, precision, and recall. The objective of this systematic review is to identify and critically evaluate current studies regarding the application of ML in predicting the 5-year survival rate of breast cancer. Test results on Haberman's Breast Cancer Survival dataset demonstrate the effectiveness of the proposed MLP approach, achieving an accuracy of 97.54%.*

Keywords: Prediction, SVM, MLP and ML

I. INTRODUCTION

Breast cancer ranks as the second most lethal disease after Lung Cancer, accounting for the highest mortality rate among all cancers. Approximately 12% of new cancer cases worldwide are attributed to breast cancer, with women constituting nearly 25% of these cases [5]. Individuals typically consult an oncologist upon noticing any signs or symptoms of the disease, who may conduct various diagnostic tests such as mammograms, MRI scans, ultrasound, X-rays, and tissue biopsies to detect breast cancer. Sentinel lymph node biopsy is commonly performed to identify cancerous cells in lymph nodes. AI techniques are also employed to classify cancers as benign or malignant, aiding in early detection to improve patient prognosis and survival rates [1].

Survival, defined as the period a patient survives after cancer diagnosis, is crucial for standardizing reporting and assessing survivability. The 5-year threshold is significant for monitoring survival rates, with previous studies using this timeframe to determine survival outcomes [7]. Breast cancer exhibits considerable variability in survival rates between individuals, making accurate prediction of survival essential for guiding clinical treatment decisions, avoiding unnecessary interventions, reducing financial costs, and facilitating palliative and hospice care. Consequently, predicting survival has become a major focus of breast cancer research, enabling patients to receive timely and appropriate interventions while avoiding unnecessary treatments for benign cancers."

II. MACHINE LEARNING (ML)

Artificial intelligence, a subset of AI, is a specialized field concerned with designing and developing algorithms that enable computers to perform tasks based on input data, such as sensor data or datasets. The primary focus of AI research is to devise efficient methods for understanding complex patterns and making intelligent decisions based on data [4]. Machine learning (ML) has diverse applications, including search engines, medical diagnosis, text and handwriting recognition, image classification, load forecasting, fraud detection, and sales prediction, among others.

AI approaches enable the discovery and acquisition of information through models that may not be discernible by human understanding alone. These models, known as classifiers, analyze input data to determine whether an activity is an attack or a regular occurrence. The training process of a machine learning model involves providing training data to a machine learning algorithm [2][3]. The term "machine learning model" refers to the predictive model generated by the training process.

III. DIRECTED LEARNING ALGORITHMS

In this evaluation work, Supervised ML Algorithms like SVM and MLP are discussed.

3.1 Support Vector Machine (SVM)

The SVM is one more kind of AI systems subject to veritable learning hypothesis. Considering incredible movement and a higher precision, SVM has turned into the examination point of combination of the AI social class. SVMs are set of related supervised learning techniques utilized for social affair and apostatize [6]. Several late assessments have unequivocal that the SVM by and large are ready for passing on preferred in basically the same manner as solicitation accuracy over different information strategy calculations. SVM relies upon genuine learning hypothesis by Vapnik [9] et al proposed another learning procedure, which relies upon a set number of tests in the data contained in the ongoing arranging text to get the best assembling results.

A phenomenal property of SVM can't abstain from being, SVM at the same time limit the preliminary social occasion stumble and extend the mathematical edge. So SVM called Maximum Margin Classifiers. SVM depends upon the Structural danger Minimization. SVM map input vector to a higher layered space where a maximal limiting hyperplane is gathered. Two identical hyperplanes are made on each side of the hyperplane that different the information. The disengaging hyperplane is the hyperplane that help the distance between the two identical hyperplanes. A thought that is made that the more prominent the edge or distance between these comparable hyperplanes the better the hypothesis stir up of the classifier [10].

3.2 Multilayer Perceptron (MLP)

MLP is a supervisor among the most thoroughly saw Neural Network Design that has been used for various applications. The MLP organize is ordinarily made from different center concentrations or managing units, and it is figured out into an advancement of in any event layers [4]. The crucial layer is named as an information layer where it gets the external information while the last layer is a yield layer where the response for the issue is gotten. The covered layer is the extensively captivating layer in the data layer and the yield layer, and may shape with in any event one layer. The objective of Multilayer Perceptron learning is to find the best loads that limit the partition between the information and the yield. Most of arrangement computation are used in Neural Network is Back Propagation and it has been used in overseeing various issues in model affirmation and portrayal [6]. This estimation depends a few cut-off points like Learning Rate, Momentum Rate and Activation work, and so on.

IV. EXPERIMENTAL RESULTS

This part gives results and related conversation on information driven analysis of Haberman's Breast Cancer Survival dataset was gathered from UCI repository [8]. This exploration work was executed utilizing Weka. WEKA is made by analysts at the University of Waikato in New Zealand. The product is written in the Java language and contains a GUI for communicating with information documents. WEKA additionally gives the graphical UI of the client and gives numerous offices. WEKA is a cutting edge office for creating AI (ML) methods and their application to true information mining issues. The information record typically utilized by WEKA is in ARFF document design. ARFF represents Attribute Relation File Format, which comprises of extraordinary labels to demonstrate separating in the information document. WEKA implements algorithms for data pre-processing, classification. We tracked down that the dataset had 4 credits for every one of the 306 patients. These records were arranged into two classes, the patient survived 5 years or longer contains 225 instances and the patient died within

5 year has 81 instances. The analyses were performed considering 306 examples which implies 70% of the complete examples were preparing information and 30% were trying information. The statistical summary of the dataset as shown in the figure-1 and figure-2.

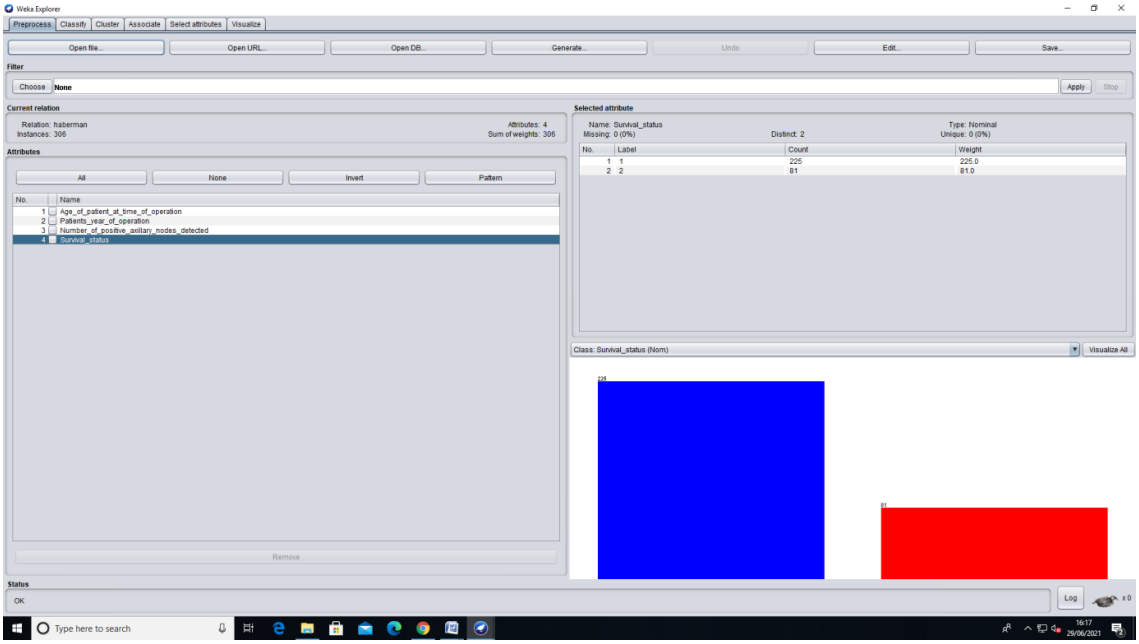

Figure-1: Statistical summary of the dataset

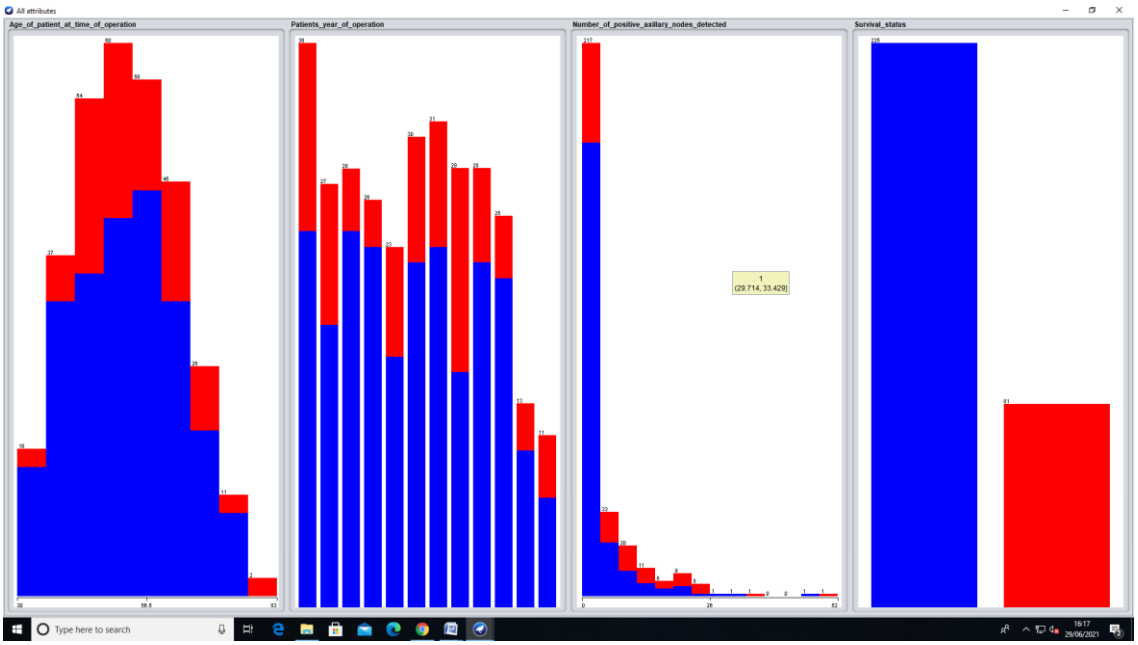

Figure-2: Statistical summary of the dataset

The Experimental outcomes are displayed in the table-1 and furthermore same displayed in the figure-3.

Table-1
Performance of classifiers

Algorithm	Accuracy	Precision	Recall
SVM	95.64	95	95
MLP	**97.54**	96.68	96.72

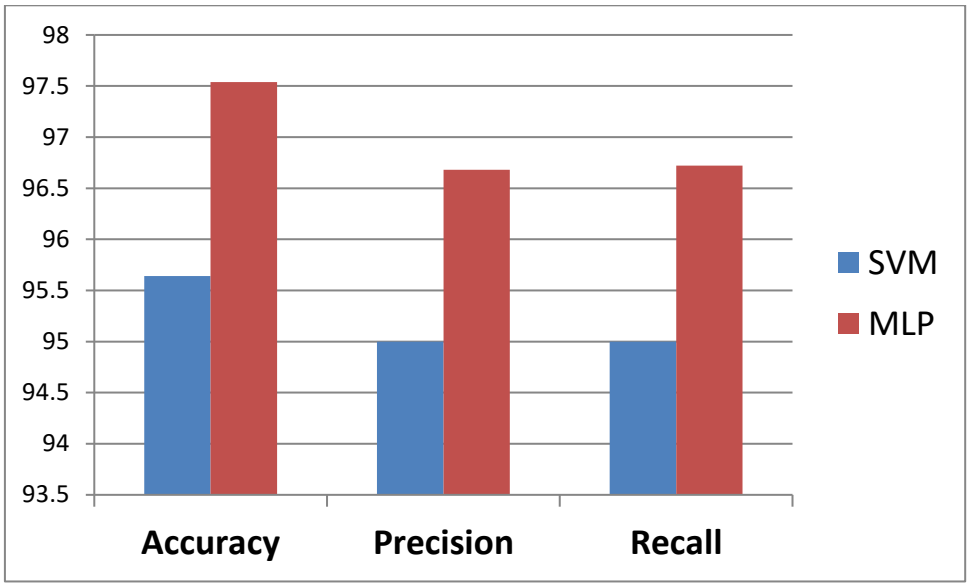

Figure-3: Performance of ML algorithms

We notice the exhibition of two ML calculations as displayed in the figure-3 dependent on precision of MLP classifier calculation gives huge improvement in the exactness (97.54%) when contrasted with a SVM classifier.

V. CONCLUSION

This paper investigates anomalies using two AI algorithms. Our initial findings indicate that the MLP algorithm achieves higher classification accuracy in detecting various abnormalities compared to SVM models. The results demonstrate that MLP is the most suitable approach for data-driven anomaly detection compared to other methods such as SVM.

REFERENCES

[1] Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications. 2009 Mar 1;36(2):3240-7

[2] G. Ravi Kumar, K. Nagamani and G. Anjan Babu, "A Framework of Dimensionality Reduction Utilizing PCA for Neural Network Prediction", Lecture Notes on Data Engineering and Communications Technologies, ISBN 978-981-15-0977-3, Volume 37, PP:173-180, Springer Nature Singapore Pte Ltd. 2020

[3] G. Ravi Kumar, Venkata Sheshanna Kongara & Dr. G. A. Ramachandra, "An Efficient Ensemble Based Classification Techniques for Medical Diagnosis", International Journal of Latest Technology in Engineering, Management and Applied Sciences, Volume II, Issue VIII, Pages: 5-9, ISSN-2278-2540,August-2013

[4] H. Witten and E. Frank, "Data mining: practical machine learning tools and techniques with Java implementations", San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., (2000)

[5] Jemal A, et al.(2005).. Cancer statistics, 2005. CA: a cancer journal for clinicians. 2005 Jan 1;55(1):10-30.

[6] J. Han and M. Kamber," Data Mining concepts and Techniques", the Morgan Kaufmann

[7] Polat K, Güneş S. Breast cancer diagnosis using least square support vector machine. Digital Signal Processing. 2007 Jul 1;17(4):694-701.

[8] UCI Machine Learning repository (https://archive.ics.uci.edu/ml/datasets.html)

[9] Vapnik V.N, "Statistical learning Theory", John Wiley and Sons, New York, USA, 1998.

[10] Vapnik V.N,"The Natural of Statistical Learning Theory, Springer-Verlag, New York, USA, 1995.