INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 08, August 2024, pp : 947-949
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 5.725

REVIEW OF VARIOUS MACHINE LEARNING TECHNIQUES APPROACHED BY EARLIER RESEARCHERS IN BREAST CANCER PREDICTION

Rajiv Kumar[1], Dr. Nishant Kumar Pathak[2]

[1]Research Scholar, Department of Computer Science & Engineering, Shobhit Institute of Engineering & Technology, Deemed to be University, Meerut

[2]Associate Professor, Department of Computer Science & Engineering, Ajay Kumar Garg Engineering College, Ghaziabad

ABSTRACT

Cancer diagnostic categorization systems are built using machine learning (ML) approaches. These machine learning techniques can assist professionals and novices alike in mitigating potential errors and enable accurate and timely inspection of healthcare databases. However, useful machine learning categorization is limited by the amount of the data. Numerous cancer forecast models have been developed, and they use machine learning approaches to anticipate the onset of the disease and identify its distinguishing features. Experts in health management are concentrating on creating new algorithms and modifying machine learning to make it easier to categorise patient data and analyse it to produce accurate, dependable, and error-free predictions. For this, the ML platform makes use of numerous parameters.

Keywords: Cancer diagnostic; Machine learning; Prediction.

1. INTRODUCTION

By offering the appropriate guidance in a timely manner, machine learning (ML) techniques can assist healthcare professionals in safeguarding patient lives. Machine learning technologies are highly beneficial for accurately classifying a variety of records. As such, it provides cancer sufferers with an abundance of bravery. Effective prediction algorithms generate a dataset that surgeons can use to track a drug's effectiveness in treating breast cancer as well as side effects, treatments, and patient responses to specific drugs. The ML techniques aid in the creation of practical inquiry tools that uncover hidden links in large volumes of BC information. Cancer diagnostic categorization systems are built using machine learning (ML) approaches. These machine learning techniques can assist professionals and novices alike in mitigating potential errors and enable accurate and timely inspection of healthcare databases. However, useful machine learning categorization is limited by the amount of the data; **Gupta and Kaushik (2018).**

ML approaches are used to forecast classifier output based on input from past data. The machine learning workflow, which outlines the stages that are carried out, provides a clear explanation of a machine learning project. **Korenberg (2007)** has been observed that issues in computational biology and bioinformatics are gradually being addressed with machine learning techniques. Protein expressions, pictures, pathways, and high throughput data in the form of sequences are all examined computationally. This provides a general grasp of illnesses and the strategic creation of drugs. Because machine learning (ML) techniques including support vector machines (SVMs), Markov models, neural networks (NNs), and graphical models are adept at managing data noise, generalisation, and uncertainty, they prove to be comparatively beneficial when analysing biological data.

According to **Sharma et al. (2017),** cancer poses a significant challenge to medical professionals and specialists in the field. Using the LR technique, an analysis of the cancer dataset for benign and malignant breast cancer revealed an accuracy rate of 59.1% on WPBC and 95.24% on WDBC. It has been reported that a logistic regression-based analysis with cluster adjustments was used for the investigation and prediction of impatient mortality of hospitalised cancer patients. Breast surgery is typically the first step towards managing early-stage signs of breast cancer, which include difficulty urinating, behaviours, skin and mouth ulcers, persistent cough, nausea, and stomach pain. The final stage of BC's prognostic procedure depends entirely on how well the corrective care is administered. Prognostic forecasts are an additional benefit over casual problem-solving examination.

Review of Various Earlier Studies related to Machine Learning Techniques in Breast Cancer Prediction

ML algorithms have been extensively employed in BC diagnosis and prognosis over the years to extract various insights from data sets. Machine learning (ML) is a type of artificial intelligence (AI) that learns and improves performance automatically using a range of statistical, probabilistic, and optimization methods. Designs derived from historical data and fresh data, devoid of specifically stated programming instructions. Data analysis is done with both

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 08, August 2024, pp : 947-949
e-ISSN : 2583-1062
Impact Factor : 5.725
www.ijprems.com
editor@ijprems.com

statistics and machine learning. ML techniques are often able to extract important features and possible rules from complicated, large-scale, high-dimensional data that would be challenging to find with standard statistics. These days, artificial intelligence techniques including supervised and unsupervised machine learning classifiers are used for BC forecasting, clustering, and the classification of benign and cancerous data. By using classification approaches, supervised machine learning methods are being deployed and receiving a lot of interest in the identification and therapy of BC. When machine learning is combined with categorization techniques, the prognosis for BC becomes significantly simpler.

Table 1: Techniques described in previous research studies

Author (s)	Technique	Advantages	Limitations /Research Gap	Remarks
Shin and Nam (2014)	Hybrid approach with a predictor and descriptor	Forecasting the survivability of patients with chronic cancer cell growth.	The approach has not been tried with different types of tumours.	Breast cancer survival dataset(SEER)
M U Khan etal. (2008)	Weighted Fuzzy Decision Trees	Breast cancer prognosis, namely survivability with enhanced accuracy.	The method's optimisation using rule weights and genetic algorithms is still ongoing.	SEER breast cancer dataset
Thongkam etal. (2008)	AdaBoost algorithms	Improved accuracy and computational efficiency for predicting breast cancer survival.	AdaBoost requires integration with other algorithms.	Breast cancersurvival databases in Thailand
Maglogianniset al. (2009)	SVM basedclassifiers	Automatic breast cancer diagnosis and prediction.	The evaluation of their method using the SEER dataset is still ongoing.	Wisconsin Diagnostic Breast Cancer (WDBC) and Wisconsin Prognostic BreastCancer (WPBC) Datasets
Kim and Hyunjung Shin. (2013)	Semi- Supervised Learning (SSL)using tagged, unlabelled and the pseudo- labeled Structure	Survivability forecast for patients with chronic cancer cell proliferation.	Member models with co-training still need to be completed.	SEER breastcancer survivabilitydataset

Shin and Nam developed a hybrid technique combining a predictor and descriptor to forecast patient survivability amidst chronic cancer cell growth. Their approach, though promising for predicting survival in breast cancer cases using SEER dataset, lacks validation across various tumor types. This method's advantage lies in its potential to provide crucial insights into long-term prognosis, particularly for chronic cancer patients. However, its applicability beyond breast cancer remains uncertain, necessitating further validation and adaptation to different tumor contexts to enhance its clinical utility and broaden its impact in oncology research. Khan et al. introduced Weighted Fuzzy Decision Trees as a technique to improve breast cancer prognosis accuracy using SEER dataset. Their method enhances survivability predictions by incorporating fuzzy logic and decision tree structures, promising significant advancements in clinical outcomes assessment. However, ongoing optimization efforts involving rule weights and genetic algorithms indicate that further refinement is necessary for maximizing its predictive power. Despite these challenges, the approach represents a robust framework for leveraging complex data to support clinical decision-making in breast cancer treatment, highlighting its potential impact on enhancing patient care through more precise prognostic assessments. Thongkam et al. employed AdaBoost algorithms to enhance accuracy and computational efficiency in predicting breast cancer survival using databases from Thailand. Their approach leverages the strengths of ensemble learning to achieve improved prognostic outcomes. However, the method's effectiveness relies on integrating AdaBoost with complementary algorithms to maximize predictive performance. Despite this requirement, Thongkam et al.'s work signifies a significant step forward in utilizing machine learning for precise survival predictions in breast cancer, demonstrating its potential to contribute valuable insights for clinical decision-making and patient management strategies in oncology. Maglogiannis et al. utilized SVM-based classifiers for automatic breast cancer diagnosis and prediction, focusing on the Wisconsin Diagnostic Breast Cancer (WDBC) and Wisconsin Prognostic Breast Cancer (WPBC) datasets. Their approach aims to streamline diagnostic processes and enhance predictive accuracy through machine learning. However, ongoing evaluation using the SEER dataset indicates ongoing refinement and validation are necessary. Despite these challenges, their work underscores the potential of SVM classifiers in advancing clinical applications of machine learning, offering valuable tools for early detection and prognosis assessment in breast cancer, thus potentially improving patient outcomes and treatment strategies in oncology.

Kim and Hyunjung Shin implemented Semi-Supervised Learning (SSL) utilizing tagged, unlabelled, and pseudo-labeled structures to forecast survivability in patients with chronic cancer cell proliferation, focusing on the SEER breast cancer survivability dataset. Their approach aims to leverage unlabeled data effectively to improve prediction accuracy in challenging clinical scenarios. However, the method's full potential, particularly involving member models with co-training, remains under development. Despite ongoing refinements, their work represents a promising advancement in using machine learning to enhance prognostic models for chronic cancer, potentially offering clinicians valuable insights for personalized treatment strategies and patient management in oncology.

2. CONCLUSION

Machine learning techniques have revolutionized breast cancer diagnosis and prognosis by leveraging vast datasets to enhance accuracy and efficiency. These methods, including supervised and unsupervised learning classifiers, extract critical insights that aid in predicting outcomes, tracking treatment effectiveness, and identifying hidden correlations within complex biological data. While challenges such as data quantity and model generalization persist, the integration of machine learning into healthcare offers promising advancements in personalized medicine and clinical decision-making. By mitigating errors and enabling precise analysis of healthcare databases, machine learning continues to play a pivotal role in improving patient care and outcomes in breast cancer treatment and management.

Recent studies by Shin and Nam, Khan et al., Thongkam et al., Maglogiannis et al., and Kim and Hyunjung Shin highlight diverse approaches leveraging machine learning for breast cancer prognosis and treatment prediction. Shin and Nam's hybrid technique shows promise in predicting survivability in breast cancer patients, yet it requires further validation across different tumor types to enhance its applicability in clinical settings. Khan et al.'s Weighted Fuzzy Decision Trees offer significant advancements in prognosis accuracy but require ongoing optimization efforts for maximizing predictive power. Thongkam et al.'s use of AdaBoost demonstrates improved prognostic outcomes, contingent upon effective integration with other algorithms. Maglogiannis et al.'s SVM-based classifiers streamline diagnostic processes, yet ongoing evaluation is crucial for refining their accuracy and clinical utility. Kim and Hyunjung Shin's Semi-Supervised Learning approach shows potential in leveraging unlabeled data for improved predictions but requires further development in member model co-training. Collectively, these studies underscore the evolving role of machine learning in advancing personalized breast cancer care, aiming to enhance clinical decision-making and ultimately improve patient outcomes.

3. REFERENCES

[1] Gupta, A. & Kaushik, B.N. (2018). Feature selection from biological database for breast cancer prediction and detection using machine learning classifier, *Journal of Artificial Intelligence*, 11, 55-64.

[2] Khan, M.U., Choi, J.P., Shin, H., Kim, M. (2008). Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare, Annual International Conference of IEEE in Engineering & Medicine and Biology Society, 48-51.

[3] Korenberg, M. J. (2007). Microarray data analysis: methods and applications, *Springer Science & Business Media*, 377.

[4] Maglogiannis, I., Zafiropoulos, E. & Anagnostopoulos, I. (2009). An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers, *Applied Intelligence*, 30(1), 24-36.

[5] Sharma, A., Kulshrestha, S., & Daniel, S. (2017, December). Machine Learning approaches for breast cancer diagnosis and prognosis. In Soft Computing and its Engineering Applications (icSoftComp), 2017 International Conference on (1-5). IEEE.

[6] Shin, H. & Nam, Y. (2014). A coupling approach of a predictor and a descriptor for breast cancer prognosis, BMC Med Genomics, 7(Suppl 1): S4.

[7] Thongkom, J., Xu, G., Zhang, Y. & Huang, F. (2008). Breast Cancer Survivability via AdaBoost Algorithms, Conference in Research and Practice in Information Technology, Vol 80.