Utilizing Convolutional Neural Networks and Support Vector Machines for Breast Cancer Prediction: A Comprehensive Machine Learning Approach

Prof Hemalta Mane[1], Sayali Pachpute[2] ,Somesh Sinha[3] , Sujay Patil [4]
Computer Engineering Department[1,2,3,4]
Nutan Maharashtra Institute of Engineering and Technology, Pune, Maharashtra[1,2,3,4]

Abstract - Breast cancer represents a significant global health issue, with timely detection being paramount for treatment success. Over recent times, the fields of machine learning and deep learning have risen as crucial instruments, showcasing remarkable effectiveness in improving the accuracy and efficacy of breast cancer identification. This manuscript thoroughly investigates prior literature pertaining to the application of algorithms from machine learning (ML) and deep learning (DL) in the context of breast cancer detection. The examination encompasses diverse algorithms and measures of performance employed in these investigations. Furthermore, significant obstacles and prospective avenues for research advancement in this domain are deliberated upon.

Keywords: Machine Learning, Deep Learning, Detection, Diagnosis.

I. INTRODUCTION

Breast cancer stands as a prominent malignancy affecting women globally, contributing significantly to cancer-related fatalities annually. Timely identification of breast cancer is pivotal in enhancing treatment outcomes and reducing mortality rates. Throughout the years, various imaging techniques and screening protocols have emerged to detect breast cancer in its early stages [2]. Nonetheless, these approaches frequently encounter constraints in terms of accuracy, effectiveness, and accessibility.

With the fast headways in machine learning (ML) and profound learning (DL) advances, there has been a developing intrigued in leveraging these strategies for breast cancer location. ML and DL calculations have appeared promising comes about in moving forward the precision of breast cancer discovery, helping clinicians in making more educated choices, and decreasing the workload related with manual screening.

This writing overview points to give an outline of the existing inquire about on the application of ML and DL methods for breast cancer location. The overview will cover a extend of themes, counting the techniques utilized, the sorts of information and imaging modalities utilized, the execution measurements utilized to assess the calculations, and the challenges and future inquire about headings in this field.

Through meticulous examination and synthesis of extant literature, this research endeavours to furnish analysts, clinicians, and policymakers with an all-encompassing comprehension of the contemporary landscape in breast cancer detection employing ML and DL methodologies [1] . Such insights are poised to guide forthcoming research initiatives and foster the evolution of more efficient and effective strategies for breast cancer detection[5, 6].

II. LITERATURE SURVEY

A benign growth remains localized, refraining from spreading to surrounding tissues, whereas malignant growths possess the capability to extend into adjacent areas. Various methodologies have been utilized historically for breast cancer detection. For example, mammography examines breast tissue to decrease mortality rates associated with breast cancer; however, it exhibits limitations. Ultrasound imaging, employing the transmission of ultrasonic waves into the body, serves as an alternative diagnostic approach; nevertheless, it lacks the capacity to identify tumours smaller than 5 milli meters. Sonography, often conducted concurrently with mammography, aids in the detection of abnormalities within breast tissue. Infrared thermography utilizes sensors sensitive to infrared radiation to detect variations in heat across breast tissue, with areas of heightened temperature indicating potential tumour presence. [9].

N Khuriwal and colleagues conducted an extensive examination into the application of ensemble techniques within machine learning for the anticipation of breast cancer occurrences. Their research adopted a thorough methodology, integrating sophisticated approaches to heighten the precision of breast cancer prognostication models. A fundamental element of their procedural framework involved the utilization of normalization during data pre-processing, a critical measure in ensuring the dependability and uniformity of input data. Furthermore, they utilized a singular feature selection algorithm for scaling features, a tactic aimed at refining the prognostic potential of their model by pinpointing and prioritizing pertinent attributes. In their endeavour to forge a resilient predictive framework, Khuriwal and associates fastidiously compiled a dataset encompassing 16 distinct attributes relevant to the diagnosis of breast cancer. This meticulous curation aimed to encapsulate a comprehensive spectrum of variables potentially influencing the prognostic accuracy of the model. Through systematic experimentation and scrutiny, the researchers achieved an exceptional accuracy rating of 98.50%, underscoring the effectiveness of their approach and the promise inherent in machine learning methodologies for advancing medical prognosis and diagnosis, notably within the intricate realm of oncology. [7].

M S Yarabala and her collaborators conducted an extensive investigation utilizing advanced Computer-Assisted Detection/Diagnosis (CAD) frameworks alongside sophisticated Machine Learning (ML) methodologies. Their inquiry aimed to differentiate the presence or absence of breast malignancy in subjects via the examination of meticulously prepared datasets. By capitalizing on CAD frameworks' capabilities, which enable the automated scrutiny of medical imagery, and harnessing ML algorithms adept at discerning intricate data patterns, the team endeavoured to heighten the precision and efficiency of breast cancer diagnosis. This strategy encompassed the formulation and enhancement of prognostic models trained on expansive datasets incorporating varied clinical data and image data. Through iterative learning and optimization procedures, the CAD frameworks fused with ML methodologies exhibited proficiency in identifying nuanced indicators of breast malignancy, empowering healthcare practitioners to render more informed decisions regarding patient management. The fusion of CAD frameworks and ML algorithms presents a promising avenue for augmenting healthcare systems' diagnostic capacities, potentially culminating in earlier detection and more efficacious treatment modalities for individuals susceptible to breast cancer [3].

E A Bayrak and colleagues investigated the efficiency of (ANN) and (SVM) algorithms in forecasting breast cancer utilizing the WEKA software. Their examination concluded that, based on efficiency measures, Support Vector Machine surpassed Artificial Neural Networks, achieving an accuracy rate of 96%. This exploration underscores the importance of contrasting various machine learning algorithms in breast cancer prognosis, spotlighting Support Vector Machine as a promising avenue for precise diagnosis. The research highlights the significance of employing advanced computational utilities such as WEKA in appraising predictive frameworks for medical applications. By showcasing the superior performance of Support Vector Machine, this inquiry offers invaluable insights into breast cancer forecasting, potentially directing the formulation of more efficient diagnostic instruments and therapeutic strategies[11].

In a research undertaken by Y Khourdifi and associates, an extensive comparison was conducted among various prevalent machine learning methods, including K-Nearest Neighbours, Support Vector Machines, Random Forest, and Naive Bayes, for the purpose of breast cancer prognosis. Among these methods, Support Vector Machines emerged with a noteworthy accuracy rate of 98%. This discovery led to the inference that, owing to its exceptional precision performance, Support Vector Machines emerged as the favoured option for breast cancer prognosis in this specific study. The thorough examination of these varied methods yielded valuable insights into their distinct advantages and disadvantages in addressing the intricate task of breast cancer prognosis. Through meticulous experimentation and assessment, Support Vector Machines demonstrated their efficacy in precisely identifying patterns within the data, thereby showcasing their potential as a dependable tool for aiding in the prognosis and diagnosis of breast cancer. This outcome of the research emphasizes the significance of systematic method selection in machine learning applications within the healthcare field, accentuating the importance of accuracy as a crucial criterion in method selection for efficient clinical decision-making processes. [4].

M Nasser and U K Yusuf embarked on a thorough examination of breast cancer diagnostic techniques, with a particular focus on deep learning-based methodologies integrating genetic sequencing and histo-pathological imaging. Their exhaustive study employed a Rigorous Literature Scrutiny (RLS) framework, meticulously adhering to the PRISMA methodology. This methodical review aimed to offer a comprehensive comprehension of the present research landscape in breast cancer detection, accentuating the adaptability of diverse methodologies within deep learning paradigms. A detailed process was executed, encompassing a meticulous exploration and compilation of numerous studies relevant to the subject matter. Following stringent criteria for eligibility screening and quality assessment, a significant collection of 95 articles was identified for thorough scrutiny. Through scrupulous analysis, the study illuminated prevailing trends and advancements in breast cancer detection methodologies. Among the primary discoveries, the study emphasized the dominance of Convolutional Neural Networks (CNNs) as the most precise and widely employed model for breast cancer detection [10]. Particularly, the evaluation of precision metrics emerged as the prevailing technique for assessing the efficiency of these deep learning algorithms. Moreover, the study delved into the intricate examination of datasets utilized in breast cancer diagnosis, alongside a comparative assessment of the performance exhibited by various algorithmic approaches [1].

III. PROPOSED METHODOLOGY

Using Convolutional Neural Networks (CNNs) and Support Vector Machine (SVM) for breast cancer detection can be a powerful approach due to their ability to learn intricate patterns in images. Here's a proposed methodology for using CNNs for this purpose:

*A.U*Using Convulation Neural Network Algorithm and Support Vector Machine

1) Data Collection and Preprocessing:

Assemble a dataset comprising mammogram images paired with labels denoting the existence or nonexistence of breast malignancies. Preprocess the images to amplify distinctive characteristics and ready them for integration into the CNN-SVM framework. For CNN Feature Extraction, employ a pre-existing CNN model (such as VGG or ResNet) to abstract features from the mammogram images. This stage capitalizes on the CNN's aptitude for assimilating intricate features from images.

2) Feature Selection and Transformation:

Choose the most pertinent attributes derived from the CNN for categorization. Conduct any requisite modifications or standardizations on the attributes to prime them for inclusion into the SVM.

3) Model Training:

Split the dataset into preparing, approval, and test sets. Amid preparing, the demonstrate learns to distinguish designs and highlights that are demonstrative of breast cancer. Use the approval set to tune hyper-parameters and avoid over-fitting.

4) Model Integration:

Combine the CNN feature extraction and SVM classification stages into a single model. This can be achieved by using the output of the CNN as input to the SVM.

*5)*Model Evaluation:

Assess the performance of the trained model using the test dataset to gauge its effectiveness in detecting breast malignancies. Metrics such as precision, accuracy, recall, and F1 score are frequently employed to evaluate the model's efficiency.

6) Model Testing:

Evaluate the finalized model on an independent test dataset to analyze its proficiency in identifying breast cancer occurrences.

7) Deployment:

Deploy the trained CNN-SVM model in a production environment for real-world breast cancer detection.

8) Monitoring and Updating:

Continuously monitor the model's performance and update it with new data to maintain its effectiveness over time.

Algorithms Used:

1) Convolutional Neural Network:

Convolutional Neural Networks (CNNs) represent a subset of deep learning frameworks frequently utilized for processing visual information. They excel in tasks such as categorizing images, identifying objects, and segmenting visual content. CNNs are distinguished by their aptitude for autonomously acquiring hierarchical portrayals of features from unprocessed pixel data. These networks comprise numerous concealed tiers, encompassing convolutional tiers, pooling tiers, and fully linked tiers. Leveraging their capacity to capture hierarchical feature structures, CNNs have attained cutting-edge performance across a spectrum of computer vision assignments.

2) Support Vector Machines (SVMs) :

Support Vector Machines represent a class of supervised learning algorithms utilized for both classification and regression assessments. These models demonstrate efficiency in spaces with numerous dimensions and are notably adept for assignments with evident distinctions between categories. Their functionality involves identifying the hyper-plane that optimally divides distinct categories within the feature space, aiming to maximize the margin between these categories. SVMs exhibit versatility in managing data of both linear and non-linear nature, facilitated by diverse kernel functions. These models have garnered extensive utilization across diverse domains, encompassing but not limited to bioinformatics, textual categorization, and visual pattern identification.

IV. ADVANTAGES

A. Advantages :

1. Early Detection: Machine learning (ML) and deep learning (DL) methodologies possess the capacity to scrutinize extensive patient datasets, unveiling patterns suggestive of early-stage breast malignancy. This capability facilitates prompt intervention and enhances treatment results.

2. Accuracy: These algorithms can learn from diverse datasets, including genetic, imaging, and clinical data, to make highly accurate predictions, potentially surpassing the diagnostic accuracy of traditional methods.

3. Personalized Medicine: ML and DL models can tailor predictions and treatment recommendations based on individual patient characteristics, such as genetic predispositions, tumour characteristics, and response to previous therapies, leading to more personalized and effective treatment strategies.

4. Integration with Imaging Techniques: Machine learning techniques have the capability to scrutinize mammograms, ultrasounds, and MRI scans with remarkable accuracy, aiding radiologists in identifying subtle irregularities and mitigating the occurrence of erroneous positive and negative diagnoses.

5. Cost-Efficiency: By streamlining diagnostic processes and reducing unnecessary tests and procedures, ML and DL models can potentially lower healthcare costs associated with breast cancer diagnosis and treatment.

6. Continuous Learning: ML and DL algorithms can continuously learn from new data and updates, improving their predictive capabilities over time and keeping pace with advancements in medical knowledge and technology.

V. CONCLUSION

In conclusion, Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) stand out as highly efficient methods for detecting breast cancer, surpassing alternative algorithms and methodologies. CNNs demonstrate exceptional proficiency in deciphering intricate patterns within mammogram images, thereby discerning subtle cues indicative of cancerous tissues. Their layered structure enables the extraction of intricate features, leading to superior accuracy in classification endeavours. Conversely, SVMs excel in binary classification tasks and excel in processing multi-dimensional data, rendering them well-suited for handling the features extracted by CNNs. Their capacity to discern the optimal boundary for class segregation, particularly in scenarios with distinct boundaries, enhances their effectiveness in distinguishing between cancerous and non-cancerous tissues. Additionally, the amalgamation of CNNs for feature extraction and SVMs for classification exhibits notable potential, often surpassing alternative methodologies in terms of precision and effectiveness. In essence, the collaborative utilization of CNNs and SVMs offers a robust and dependable approach for breast cancer detection, showcasing their supremacy over alternative algorithms and methodologies in this critical medical domain.

VII. FUTURE SCOPE

The triumph of the endeavour unveils prospects for advancing initial breast cancer identification through the fusion of nascent AI and deep learning methodologies to enhance precision and responsiveness. Subsequent investigations may prioritize customization of therapeutic strategies according to the distinct attributes of identified tumours, transitioning towards individualized healthcare for individuals diagnosed with breast cancer.

ACKNOWLEDGMENTS

We wish to thank our parents and associates for their valuable support and encouragement throughout the development of the project work and we would also like to thank our guide Prof. Hemlata Mane for guiding us throughout the project work.

REFERENCES

[1] Maged Nasser and Umi Kalsom "Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction 2023" Jan; 13(1): 161.Published 2023 Jan 3.

[2] Arasi, Munya & Babu, Sangita. (2019). "Survey of Machine Learning Techniques in Medical Imaging", International Journal of Advanced Trends in Computer Science and Engineering.

[3] M. S. Yarabarla, L. K. Ravi and A. Sivasangari,"Breast Cancer Prediction via Machine Learning", 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019.

[4] Y. Khourdifi and M. Bahaj,"Applying Best Machine Learning Algorithms for Breast Cancer Prediction and Classification", 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, 2018.

[5] M. NEMISSI, H. SALAH and H. SERIDI, "Breast cancer diagnosis using an enhanced Extreme Learning Machine based-Neural Network", 2018 International Conference on Signal, Image, Vision and their Applications (SIVA), Guelma, Algeria, 2018

[6] M. Amrane, S. Oukid, I. Gagaoua and T. Ensarİ, "Breast cancer classification using machine learning", 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), Istanbul, 2018.

[7] N. Khuriwal and N. Mishra, "Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm", 2018 IEEMA Engineer Infinite Conference (eTechNxT), New Delhi, 2018.

[8] D. Bazazeh and R. Shubair,"Comparative study of machine learning algorithms for breast cancer detection and diagnosis", 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), Ras Al Khaimah, 2016.

[9] Institute of Medicine (US) and National Research Council (US) Committee on New Approaches to Early Detection and Diagnosis of Breast Cancer; Joy JE, Penhoet EE, Petitti DB, editors. Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis. Washington (DC): National Academies Press (US); 2005. Appendix A, Breast Cancer Technology.

[10] Akinsola, J E T & Adeagbo, Moruf & Awoseyi, Ayomikun. (2020). "Breast Cancer Predictive Analytics Using Supervised Machine Learning Techniques", International Journal of Advanced Trends in Computer Science and Engineering.

[11] E. A. Bayrak, P. Kırcı and T. Ensari, "Comparison of Machine Learning Methods for Breast Cancer Diagnosis", 2019 Scientific Meeting on ElectricalElectronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 2019.