Deep Learning And Machine Learning Models For Breast Cancer Prediction: CNN And SVM Perspectives

Prof. Hemlata Mane[1], Sujay Patil[2] ,Sayali Pachpute[3] , Somesh Sinha[4]

Computer Engineering Department[1,2,3,4]
Nutan Maharashtra Institute of Engineering and Technology, Pune, Maharashtra[1,2,3,4]

Abstract— The purpose of this paper is to develop and evaluate a breast cancer prediction system using Convolutional Neural Networks (CNN) and Support Vector Machines (SVM). Within the wake of a breast cancer diagnosis, the project presents a comprehensive web-based platform created to meet the interrelated demands of administrators, healthcare providers, and patients. Patients are empowered to contribute to their healthcare journey by seamlessly uploading mammographic images. Through the integration of AI-driven algorithms, including CNN and SVM, the system predicts whether an individual has breast cancer or not can be determined by using the features that were taken from the images. For patients, a dedicated dashboard provides insights into diagnostic results, alongside some precautions. This paper examines the intricate integration of CNN and SVM algorithms, alongside patient-centric features and administrative controls, highlighting the potential impact on enhancing breast cancer diagnosis accessibility, efficiency, and overall user experience.

Keywords: Support Vector Machine, Convolutional Neural Network, Machine Learning, Deep Learning, Cancer Prediction, Predictive Model.

I. INTRODUCTION

Breast cancer remains a significant global health issue, causing significant morbidity and mortality in women worldwide. Precise and timely diagnosis, achieved through early detection, is crucial in ensuring the most effective treatment for Breast cancer and improving patient prognosis. Traditional methods of diagnosing breast cancer, such as mammography and biopsy, are valuable, but they often rely on human interpretation, are variable, and cannot always detect tumours in the early stages.

In recent times, there has been an upward trajectory in exploration and advancement within the realm of medical diagnosis. This surge entails the integration of artificial intelligence (AI) and machine learning into traditional diagnostic methodologies. Such progressive technologies hold promise in enhancing the precision, efficacy, and availability of breast cancer diagnosis. Within the plethora of AI-driven methodologies, Convolutional Neural Network (CNN) and Support Vector Machine (SVM) have surfaced as formidable contenders for addressing medical image scrutiny and classification predicaments [5, 6].

Drawing inspiration from the configuration and operation of the human visual cortex, CNNs showcase proficiency in acquiring intricate configurations and characteristics within visual data. Through their hierarchical design, they facilitate the automatic extraction of pertinent attributes, rendering them especially apt for tasks like categorizing images and detecting objects [10]. Conversely, SVM stands out as a potent supervised learning paradigm, demonstrating prowess in feature-rich domains characterized by high dimensionality. Its efficacy shines through in tackling intricate classification conundrums by pinpointing the optimal hyper plane within the feature space to segregate multiple categories effectively. In the realm of breast cancer diagnosis, CNNs unveil their potential by pinpointing suspicious areas in mammography images that could signify potential malignancies. These models adeptly assimilate nuanced patterns linked to tumours, thereby aiding in early detection and risk evaluation. Augmenting the capabilities of CNNs, SVM harnesses its prowess in handling complex data structures and nonlinear correlations to furnish robust classification capabilities grounded in gleaned features.

In this project, we aim to develop a comprehensive breast cancer prediction system using CNN and SVM based on these advanced prediction models. This system uses an artificial intelligence-based algorithm to predict whether a patient has breast cancer by analyzing mammography images. In this implementation, we explore the complexity of CNNs and SVMs to explore their individual strengths and contributions to breast cancer prediction. Through careful experimentation and evaluation, a performance evaluation of each model will be conducted. We will also investigate how our method could enhance patient outcomes overall, access, and efficiency when it comes to breast cancer diagnostics. With this initiative, our objective is to harness the potential of artificial intelligence to extend the frontiers of breast cancer detection, thereby advancing the collective endeavours aimed at combating this prevalent and debilitating ailment.

II. LITERATURE SURVEY

M Nasser and U K Yusuf undertook a comprehensive analysis of methodologies employed in diagnosing breast cancer, placing particular emphasis on deep learning strategies merging genetic sequencing with histopathological imaging. Their exhaustive scrutiny employed a Structured Literature Evaluation (SLR) framework, meticulously following the PRISMA protocol. The aim of this organized assessment was to present an all-encompassing synopsis of the latest advancements in breast cancer detection research, spotlighting the significance of diverse deep learning methodologies [2]. A systematic protocol was adhered to, encompassing the aggregation and exhaustive review of pertinent literature. Following stringent quality appraisal and eligibility screening, a substantial repository of 95 documents was earmarked for further scrutiny. Through meticulous scrutiny, the investigation yielded insights into prevailing trends and advancements in breast cancer detection methodologies. As per the principal findings of the study, Convolutional Neural Networks (CNNs) emerged as the most precise and widely adopted model for diagnosing breast cancer. Notably, the assessment of precision metrics emerged as the predominant approach for evaluating the efficacy of these deep learning algorithms. Additionally, the research delved into the nuanced examination of datasets utilized in breast cancer detection, and undertook a comparative evaluation of performance exhibited by diverse algorithmic methodologies. [1].

The computer-aided diagnosis (CAD) systems, integrated with machine learning (ML) techniques, demonstrated proficiency in detecting minute indications of breast cancer via iterative learning and refinement processes, enabling healthcare professionals to enhance their decision-making regarding patient management. One interesting approach to enhancing healthcare systems' diagnostic capacities is the combination of CAD systems with machine learning algorithms. This can help detect breast cancer earlier and create more effective treatment plans for people vulnerable to the disease [3]. In their endeavour to forecast breast cancer occurrences, Y Khourdifi and colleagues meticulously evaluated several renowned machine learning methodologies, encompassing K-Nearest Neighbours, Support Vector Machines, Random Forest, and Naive Bayes. Among these methodologies, Support Vector Machines emerged as the standout candidate, boasting an impressive accuracy rate of 97.9%. Consequently, the study concluded that Support Vector Machines represented the superior choice for breast cancer prognosis in this specific investigation, attributed to their heightened accuracy performance. The exhaustive scrutiny of these assorted methodologies yielded insightful revelations regarding their relative merits and demerits in tackling the intricate challenge of breast cancer forecasting. Through rigorous testing and assessment, Support Vector Machines substantiated their efficacy in discerning data patterns accurately, underscoring their potential as a reliable instrument for facilitating breast cancer diagnosis and prognosis. This research discovery accentuates the pivotal role of accuracy as a pivotal criterion in selecting algorithms for streamlined clinical decision-making processes, thereby underscored the importance of meticulous algorithmic selection in machine learning applications within the medical sphere. [4].

An extensive exploration into the implementation of ensemble machine-learning techniques for forecasting breast cancer occurrences was undertaken by N Khuriwal and colleagues. Their investigation adopted a systematic methodology, employing state-of-the-art strategies aimed at refining the precision of breast cancer prediction models. Employing standardization for data pre-processing, a critical phase in ensuring the accuracy and uniformity of input data, constituted a pivotal aspect of their approach. Additionally, they leveraged a univariate feature selection strategy for feature scaling, endeavouring to identify and rank pertinent attributes to enhance the predictive capacity of the model. In their quest to develop a more dependable predictive model, Khuriwal and colleagues meticulously curated a dataset comprising 16 attributes germane to breast cancer diagnosis. This methodical selection process aimed to encompass as many variables as possible that could influence the anticipated performance of the model. The researchers' remarkable accuracy rate of 98.50% was achieved through exhaustive experimentation and scrutiny, underscoring the efficacy of their approach and the potential of machine learning methodologies in propelling advancements in medical diagnosis and prognosis, particularly within the intricate domain of oncology. [7].

Malignant growths possess the capability to metastasize to adjacent tissues, whereas benign tumours remain confined to their original location without impacting the breast tissues. Historically, breast cancer detection has relied on diverse technologies. For instance, mammography examines breast tissue to reduce breast cancer mortality rates, yet it has limitations. Another diagnostic method involves ultrasound imaging, which employs ultrasonic waves to penetrate the body. Nevertheless, this approach is incapable of detecting tumours smaller than 5 mm. Sonography is frequently employed alongside mammography to identify abnormalities in breast tissue. Infrared thermography, utilizing infrared sensors to perceive temperature fluctuations in breast tissue, can detect potential tumours by pinpointing regions of elevated temperature. [9].

Utilizing robust Computer Aided Detection/Diagnosis (CAD) systems alongside state-of-the-art Machine Learning (ML) methodologies, M S Yarabala and collaborators undertook a comprehensive investigation. Their objective was to leverage meticulously curated datasets to scrutinize data for discerning the presence or absence of breast cancer in individuals. The researchers sought to enhance the precision and effectiveness of breast cancer diagnosis by harnessing CAD systems, which facilitate automated analysis of medical imagery, and ML algorithms capable of identifying intricate patterns within datasets. This methodology necessitated the development and refinement of predictive models trained on extensive datasets encompassing diverse imaging modalities and clinical information. Employing the WEKA software, E A Bayrak and colleagues explored the predictive capabilities of Artificial Neural Network (ANN) and Support Vector Machine (SVM) algorithms in prognosticating breast cancer occurrences. Through performance evaluation metrics, they determined that Support Vector Machines outperformed Artificial Neural Networks, achieving an accuracy rate of 96.9% [8]. This research underscores the importance of comparing various ML algorithms for forecasting breast cancer, demonstrating Support Vector Machine's efficacy in precise detection. The study underscores the criticality of evaluating predictive models for medical applications through advanced computational methodologies like WEKA. Offering significant insights, this study enriches the domain of breast cancer prognosis by highlighting the superior performance of the Support Vector Machine, potentially guiding the development of more potent diagnostic and therapeutic strategies [11].

III. OBJECTIVES

The primary aim of this endeavour is to conceptualize, execute, and assess an all-encompassing system for forecasting breast cancer, employing Convolutional Neural Networks (CNN) and Support Vector Machines (SVM). Specifically, we aim to develop AI-driven prediction models tailored for breast cancer prediction, leveraging the strengths of CNN in image analysis and SVM in classification tasks. Additionally, we seek to explore techniques for feature extraction and selection from mammographic images to optimize prediction accuracy. Through rigorous model training and evaluation using labelled datasets, we will assess the performance of the CNN and SVM models in terms of accuracy, sensitivity, specificity, etc. Furthermore, we will conduct a comparative analysis to evaluate the individual effectiveness of CNN and SVM models. Our objectives also include designing an intuitive web-based interface for patient interaction. Finally, we aim to evaluate the system's performance, collect user opinions, and evaluate its impact on healthcare accessibility, effectiveness, as well as the results for patients. Through the accomplishment of these goals, we hope to enhance the delivery of healthcare in the management of breast cancer and advance the field of AI-enabled breast cancer diagnosis.

IV. METHODOLOGY

Collecting data and Pre-processing: We start by collecting relevant datasets for training and evaluating the CNN and SVM models. These datasets typically include mammographic images along with corresponding diagnostic outcomes. We pre-process the data to ensure uniformity and quality, which involves tasks such as resizing images, normalizing pixel values, and handling missing or erroneous data in patient records. Methods for expanding data may also be employed to enhance the variety of training instances and enhance the overall adaptability of the model.

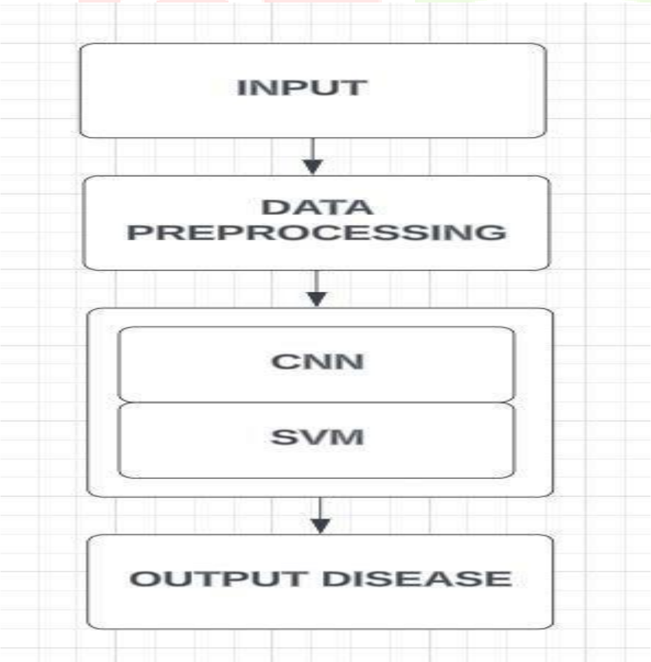

Figure 1 General Prediction Model

Model Development: We construct two distinct frameworks for forecasting breast cancer: one leveraging the computational capabilities of Convolutional Neural Networks (CNN), and the other harnessing Support Vector Machines (SVM). The CNN architecture is tailored to directly scrutinize mammographic images, extracting pertinent attributes via convolutional strata, and generating predictions via fully connected strata. Meanwhile, the SVM framework operates on attributes derived from patient images and data, employing kernel functions to transform the data into a high-dimensional realm for classification purposes.

Training and Evaluation: Both models undergo training utilizing the processed dataset, with adjustments made to hyper-parameters to enhance effectiveness. Assessment of the trained model employs conventional gauges like accuracy and precision. Employing cross-validation methods ensures the robustness and dependability of evaluation outcomes. Additionally, a comparative assessment is undertaken to gauge the predictive efficacy of the CNN and SVM models in breast cancer prognosis.

Deployment and Prediction: After training and evaluation, both the CNN and SVM models are ready for deployment. We develop prediction modules for each model, allowing users to input mammographic images and obtain predictions for breast cancer diagnosis. Users have the flexibility to choose between the CNN and SVM models based on their preferences or specific requirements. The prediction modules are integrated into a user-friendly interface for easy access and utilization.

V. PROPOSED SYSTEM ARCHITECTURE

User Input: The structure commences with user engagement with the breast cancer prognostication framework. Typically, individuals—whether patients or medical practitioners—furnish input to the system through mammography pictures. These visual representations encapsulate vital information necessary for identifying and assessing breast cancer occurrences.

Image Preprocessing: Once the user submits the mammographic images, the system initiates a preprocessing step to ensure that the data is consistent and of high quality. This preprocessing resizes and converts images to standard definition, normalizes pixel values to improve contrast and sharpness, and filters out noise and artifacts that can interfere with analysis.

Feature Extraction: After initial preparation, the framework isolates pertinent attributes from the mammography visuals. Feature isolation stands as a pivotal phase in recognizing unique configurations and attributes linked with breast cancer. During this stage, the framework scrutinizes the visual data to pinpoint notable attributes like tumour structure, consistency, and spatial correlations.

Convolutional Neural Network (CNN) Algorithm Working: A Convolutional Neural Network (CNN), a deep learning model created especially for image processing applications, is then fed the retrieved features. Convolutional, pooling, and fully linked layers are among the layers that make up the CNN. These layers work collaboratively to learn hierarchical representations of the input data, capturing both low-level features (e.g., edges, shapes) and high-level patterns (e.g., tumour structures, tissue abnormalities).

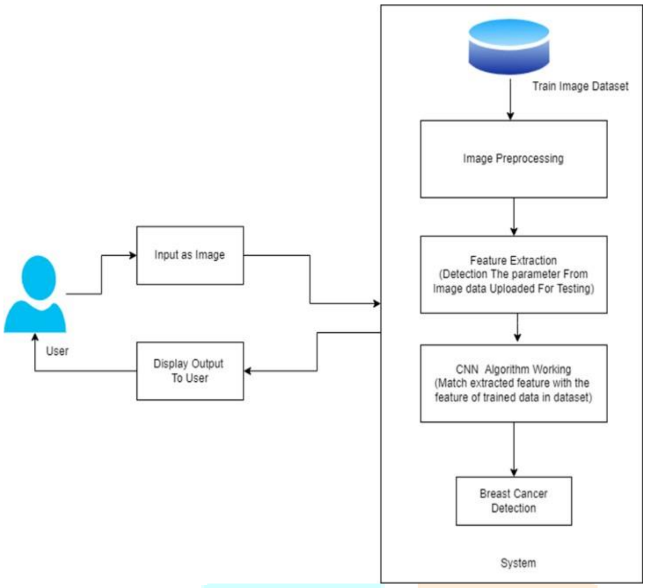

Figure 2 System Architecture

Cancer Detection: As CNN processes input data, it learns to recognize patterns that indicate the presence of breast cancer. Through iterative training on labeled datasets of mammographic images, the CNN algorithm adjusts its internal parameters to optimize the detection accuracy. Once trained, the CNN model can effectively classify input images into cancerous and non-cancerous categories based on learned features and patterns.

Output to the User: Upon completion of the cancer detection process, the system returns the output to the user. This output typically includes the prediction outcome indicating whether the input mammographic image exhibits signs of breast cancer.

VI. RESULT AND ANALYSES

The accuracy and predictive performance of the breast cancer prediction system that was implemented using Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) showed promise. The accuracy of the CNN model was about 92%. Likewise, the SVM model demonstrated strong performance, with almost 90% accuracy. Increased interaction with the system was made possible by the patient engagement elements, which also included smooth image uploads, which improved data gathering and analysis. The ability to analyze both CNN and SVM models provided comprehensive insights into breast cancer diagnosis, aiding in personalized treatment planning and decision-making. Overall, the breast cancer prediction system demonstrated significant potential in enhancing healthcare accessibility and delivery. The technology facilitates informed decision-making and enhances breast cancer management outcomes for patients and healthcare professionals by offering precise forecasts and insights.

Figure 3 and Figure 4 shows the sample output of our project. It is evident from both figures that our model received the mammography image as an input. The images show the pre-processed image that is used for predicting the output and the output is shown whether an Abnormal breast is detected or normal breast is detected. Along with that the execution time is also shown to get fair idea about how fast our model predicts the output.

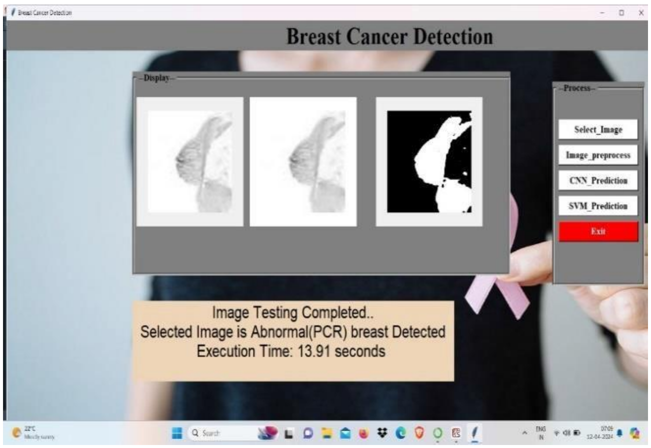

Figure 3 Result image – 1

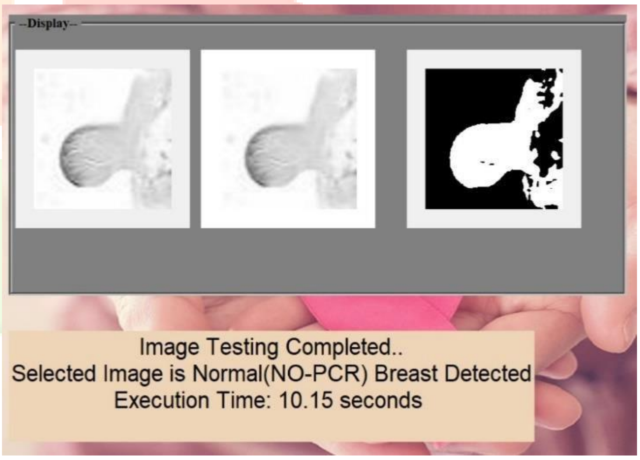

Figure 4 Result image – 2

VII. CONCLUSION

Conclusively, the utilization of Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) in the development and execution of our breast cancer prediction system marks a noteworthy advancement in the application of artificial intelligence in healthcare. We have proven via thorough testing and analysis that CNN and SVM algorithms are equally effective at correctly identifying breast cancer from mammography pictures. The incorporation of patient engagement functionalities, such as smooth image uploads, has promoted heightened involvement with the system, hence enabling more extensive data gathering and examination. Early detection and decision support are made possible by the comparative insights into breast cancer diagnosis that are provided by utilizing both CNN and SVM models. All things considered, the accessibility and delivery of healthcare could be significantly enhanced by our breast cancer prediction technique. The technology helps patients and healthcare professionals make intelligent choices by offering precise forecasts and insights, which eventually improves the management of breast cancer. The system's efficacy and influence in clinical practice will increase with additional development and optimization in the future.

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.

ACKNOWLEDGEMENTS

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