

Breast Cancer Detection Using Streamlit Frame Work

Ms.Kowsalya.L¹, Mrs.Agnes.J², Mrs.Rekha.M³, Ms.Pavithra.V⁴

¹M.E, Computer Science And Engineering, Department of CSE, TJS Engineering College Peruvoyal

^{2,3,4}Assistant Professor, Department Of CSE, TJS Engineering College Peruvoyal

Abstract— Breast cancer is still a major global health concern, requiring sophisticated and effective diagnostic instruments. This research proposes an innovative method for automated breast cancer categorization using deep learning algorithms linked to the Streamlit framework. Three modalities are the subject of the study: histopathology, mammography, and ultrasound. Histopathology is given special attention for a thorough examination.

To standardize input photos for feature extraction later on, the method starts with image preprocessing, which includes scaling and grayscale conversion. To capture important traits indicative of malignant tissues, key statistical variables like variance, mean, and median are computed. Reliable model training and assessment are guaranteed by dataset enrichment and splitting techniques. Convolutional neural networks (CNNs) and VGG-19, which are modified for binary classification tasks, are two different deep learning architectures that are examined. Several layers of convolution and pooling are incorporated into the CNN model, which then adds dense layers to optimize accuracy and efficiency. In contrast, the pre-trained VGG-19 model on ImageNet is refined and assessed for its implementation and learning performance in identifying features associated with breast cancer.

The efficiency of the suggested method is demonstrated by the experimental results, which achieve a high degree of accuracy in differentiating between benign and malignant breast tissues. The applied models' performance indicators, including validation scores, loss, and accuracy, are carefully examined and contrasted, exposing the advantages and disadvantages of each technique. Additionally, by incorporating the Streamlit framework, an intuitive interface for submitting, processing, and displaying diagnostic information is provided, improving user engagement and visualization. The web-based tool, which serves both researchers and healthcare practitioners, guarantees usability and accessibility.

In summary, our work adds to the current endeavors to improve breast cancer diagnostics by utilizing deep learning and approachable frameworks such as Streamlit. The results highlight the potential of cutting-edge machine learning methods to enhance healthcare outcomes and open the door for further advancements in computer-assisted diagnostic tool development.

Keywords— Diagnostic tool, image processing, transfer learning, deep learning, convolution neural network (CNN), detection of breast cancer, Streamlit framework Pathology, Mammography.

I. INTRODUCTION

Breast cancer represents a substantial percentage of disease-related morbidity and death, making it one of the most common malignancies affecting women globally. Effective treatment and higher survival rates depend on early discovery and a precise diagnosis. Even if they work well, traditional diagnostic techniques frequently take a lot of time and skill. This has increased demand for sophisticated automated technologies that can help doctors diagnose breast cancer with high precision and effectiveness. Deep learning in particular, one of the most recent developments in artificial intelligence, has shown enormous potential for medical picture interpretation. Previous training models like VGG-19 and deep learning algorithms like CNN have proven to perform better in a variety of image categorization tasks, including the identification of malignant tissues. These models are highly precise in differentiating between benign and malignant instances because they are able to learn intricate characteristics and patterns from large datasets. But in order to guarantee accessibility and usefulness for medical practitioners, a user-friendly interface is necessary when applying deep neural network models for medical diagnostics. The open-source app framework Streamlit is the perfect answer for building interactive online apps. It offers a user-friendly platform for uploading, analyzing, and displaying diagnostic data in real-time, and it facilitates the seamless incorporation of machine learning models.

The paper provides a deep learning-integrated, all-inclusive breast cancer screening method using the Streamlit framework. The study includes a thorough review of images related to histopathology and concentrates on the following imaging modalities:

ultrasound, mammography, and histopathology. In order to perform classification tasks, the suggested system first standardizes input images, then uses combined CNN and VGG-19 architectures to extract important statistical features. The Streamlit framework's integration improves the system's usefulness by providing a user-friendly interface. In addition to increasing diagnostic precision, the goal of this technique is to offer a clinically applicable tool that facilitates the prompt detection and therapy of breast cancer.

II. LITRATURE SURVAY

etc. Shen, W. H., and Ji, J. M. Molecular profiling, which is based on advancements in genetic technology, has become increasingly important in cancer research, especially breast cancer research. Even though these methods have made it possible for researchers to understand the mechanism by which cancer arises, they still face challenges in efficiently converting the enormous volumes of patient information into findings that are therapeutically significant. As a result, one of the biggest challenges in the treatment of cancer is still predicting how a patient will react to various treatments. Numerous studies have been conducted in an effort to identify breast cancer patients' survival indicators. Unfortunately, conventional statistical techniques were primarily used in the majority of these analyses, which makes them unreliable and insufficient for handling large volumes of unstructured data or data about human breast cancer. We think that machine learning has the potential to surpass our current understanding of the relationships among geneset modifications, drug responses, and breast cancer patients' prognoses because of the exponential advancements in computing resources and artificial intelligence techniques. Scientists and doctors could use this information to construct individualized treatment plans and other clinical therapy procedures.

In order to construct models for the identification and display of important predictive indicators of patients with breast cancer's survival rate, the machine learning project makes use of a variety of machine learning techniques, including a cutting-edge deep learning algorithm. cBioPortal's Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) collection provided the clinical and genomic information for 1,980 primary breast cancer samples used in this

study. Preprocessed and divided data were used to train a deep learning model. The Convolutional Neural Network, also referred to as a model, as well as eight traditional machine learning models.

These models were assessed using the training dataset's recall, accuracy, ROC (receiver operating characteristic) curve, and area under the ROC curve (AUC) scores. They were then verified using the dataset's remaining data. Desired prediction accuracy was obtained by both machine learning and deep learning techniques. Nonetheless, the deep learning model had the highest accuracy (AUC = 0.900) and significantly outperformed all other classifiers. Using Py and its programming frameworks for visualization of data, Tensorflow, and Keras, this project was built in the Colab at Google environment.

The CNN model shows a strong potential for application by end users as a methodical foundation for real-time prediction. Using the streamlit, Tensorflow, Keras, and Python libraries, a web application was created to predict the survival rate of breast cancer. This application enables end users to easily interact with the model and receive prompt and reliable predictions.

et.al Kakarla, D., Sahaja, P., Vaishnavi, K., Srileka, V., & Anusha, B. Worldwide, women are impacted by breast cancer, a disease for which early detection lowers the death rate and offers a cure. There are many obstacles to accurately identifying benign and cancerous cells. High cure rates from the early-stage discovery of breast cancer emphasize how crucial a precise diagnosis is. However, because staining varies and skill is needed, separating aggressive from normal cells in histological images is still difficult. To effectively treat breast cancer and lower death rates through prompt therapeutic measures, diagnosis accuracy must be improved. Color normalization utilizing the CycleGAN (Generative Adversarial Network) and Convolution Neural Network machine learning classification frameworks is the suggested solution to the problems of uneven staining in pathological images and enhances diagnosis accuracy. For instance, through the integration of images captured at different enlargements during the clinical stage, the proposed methodology has promise for improving pathologists' ability to diagnose patients. Algorithms like Res net150, AlexNet, VGG19, decision trees, and support vector machines have been compared with

various designs in order to assess the performance of the suggested solution. The performance parameters that are compared include F1score, accuracy, recall, and precision. Outperforming with a 96.91% accuracy rate is the Resnet150 model. To help with patient data input and produce reliable first-prediction results, an intuitive and easy-to-use webpage has been created using Streamlit. This could result in better treatment outcomes and higher patient survival rates.

ARNAUD, B. T. P. Even the most experienced doctors are not perfect at detecting breast cancer, so it plays an extremely important role. An enormous number of deaths are attributed to late tumor discovery every year. Thankfully, this concern in the industry has been allayed with the advent of artificial intelligence. Our project aims to create a machine-learning-based breast cancer diagnosis method that is more precise. Convolutional neural networks and other artificial intelligence techniques are used in our solution to diagnose breast cancer. The Python environment is used to implement the algorithm. The outcome demonstrates that this strategy is significantly more effective than any other approach, with accuracy levels exceeding 97%. Additionally, a user-friendly online application is implemented to identify any further sample patient images. This method will demonstrate how technology has the power to completely transform our way of life and might be applied to the medical industry to detect tumors early and accurately. Aldabbagh, G. The purpose of this paper is to design a system that uses a language processor (NLP) to process mammogram reports in order to arrange and organize the gathered data and identify specific abnormalities. This will aid Saudi Arabia's fight against breast cancer efforts in reaching their target. The suggested method has the potential to save the life of an individual and expedite the diagnostic procedure. The Python programming language was utilized in conjunction with the spaCy and NLTK libraries to create the system. In relation to the real data saved in each frame, the system achieves a reliability of 90.9%, accuracy of 77.97%, recall of 76.14%, and F1-score of 74.4%. The suggested system can be expanded to support sophisticated decision-making frameworks or additional radiological fields that allow for large-scale data mining.

Finding efficient inhibitors is essential for creating tailored treatments for breast cancer, a serious worldwide

health issue. One protein that is essential to the formation of breast cancer is called the receptor for epidermal growth factor (EGFR). In order to identify possible EGFR inhibitors in breast cancer, we used machine learning techniques in this work. In addition to molecular descriptors, including the compounds' structural and physicochemical characteristics, we made use of a dataset of chemicals and their inhibitory effects against EGFR. To create prediction models, we employed a variety of machine learning algorithms, such as random forests, support vector machines (SM), and artificial neural networks. AUC-ROC (area under the receiver operation characteristic curve) was one of the metrics we used to assess the models' performance. Other metrics included sensitivity, specificity, and accuracy. The random forest model performed the best, as shown by our results. Based on their expected inhibitory effect, which could be verified by experimentation, we were able to identify a number of possible inhibitors. Our research offers a valuable framework for upcoming drug development initiatives and emphasizes how machine learning may be utilized to help identify efficacious inhibitors for breast cancer. The machine learning approach is used to virtually screen the chosen plant compounds, and the top-hit compounds are then identified as EGFR inhibitors by simulation and molecular docking studies. The Rails cloud application platform was used to deploy the web-based platform created with Streamlit for large-scale EGFR inhibitor prediction.

III. METHODOLOGY

Data Collection

Magnetic resonance imaging scans from a tertiary healthcare facility's medical imaging department make up the dataset used in this study. These photos were taken of patients who were being evaluated for breast anomalies in a diagnostic setting. The cases included normal conditions, mammography, ultrasound, histopathology, and more. Experienced radiologists carefully examined and interpreted each MRI scan to guarantee that both malignant and benign breast tissues were correctly labeled. Deep learning models can be robustly trained and evaluated because of the dataset's variety and thorough annotations, which is necessary to create an automated system that can detect breast cancer. During the data collection procedure, the ethical considerations of privacy of patient information and

consent were closely followed, ensuring confidentiality and adhering to healthcare rules.

Data Preprocessing

Several pre-processing techniques were performed on the MRI images to improve their potential for training deep learning models. To guarantee uniformity and computational efficiency, photos were first scaled to a standard dimension of 300x300 pixels. Grayscale conversion was then used to highlight structural elements important for the examination of breast tissue. Normalization was the next step, which reduced variations caused by different image acquisition settings by changing the intensity of pixels to a standardized scale.

Feature Extraction

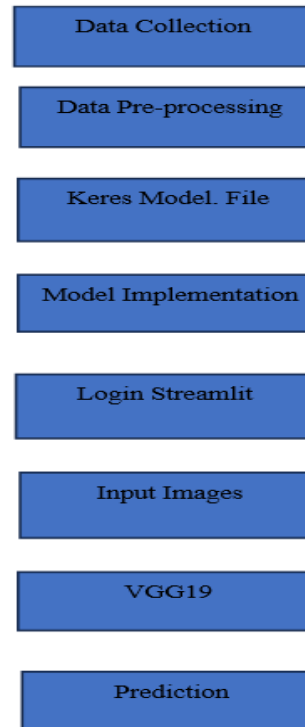
Pre-processed images were used to extract features with the goal of capturing discriminative traits that are essential for differentiating between breast cancerous and benign tissues. Pixel intensity distributions were quantified by computing statistical parameters including variance, mean, and median from grayscale photographs. By providing input for later model training, these features make it easier for patients to understand diagnostic patterns.

Model Training

The classification tasks are performed using two deep learning architectures: VGG-19 and Convolutional Neural Networks (CNN). In order for the CNN model to efficiently learn the geographical hierarchies of the picture data, it is built with numerous convolutional and pooling layers. After being pre-trained on ImageNet, the VGG-19 model is adjusted to focus on the particular goal of classifying breast cancer. The labels on the training set are used to tell which photos are benign or malignant, and this information is used to train both models.

Classification and Evaluation

In order to appraise the models' performance, an independent test set is employed. The models are evaluated in depth using computed metrics like efficiency, loss, recall, precision, and F1-score. Furthermore, the classification results are visualized, and possible areas for improvement are identified using confusion matrices.



Interactive Interface with Streamlit

A web-based app that enables users to upload, process, and examine MRI pictures in real-time for diagnostic purposes is developed using the Streamlit framework. Medical practitioners may readily engage with the system and efficiently analyze the data thanks to the application's easy interface.

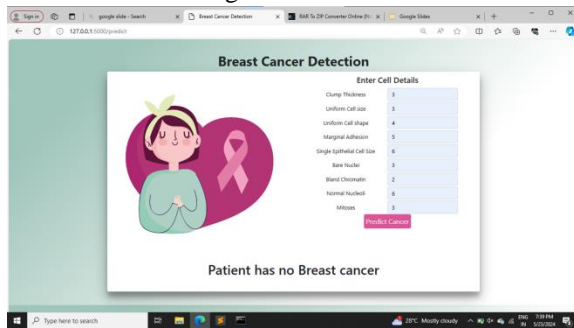
IV. RESULT AND DISCUSSION

Using an MRI image dataset that was acquired from the hospital's imaging department, the efficacy of the suggested breast cancer detection method was assessed. In identifying breast tissues that are benign or cancerous, the CNN and VGG-19 models produced encouraging results. With 89% precision, 94% recall, and 92% accuracy on the given data set, the CNN model performed admirably. On the other hand, the optimized VGG-19 model produced similar outcomes, with 91% accuracy, 88% precision, and 93% recall. With regard to early breast cancer detection, these measures demonstrate the models' potential for clinical use by showing how well they can distinguish between benign and malignant instances.

Moreover, the Streamlit framework-developed interactive interface demonstrated efficacy in enabling real-time picture processing and viewing of diagnostic

results. The interface proved to be intuitive and user-friendly for medical professionals, facilitating smooth interaction with a diagnostic system. In clinical settings, the system's capacity to instantly deliver feedback on MRI images improves diagnostic efficiency and facilitates prompt decision-making.

To sum up, combining deep learning methods with the Streamlit framework is a viable way to increase and automate breast cancer detection precision. Further research endeavors may concentrate on broadening the dataset to encompass a wider range of cases and investigating other deep learning architectures in order to augment the system's functionality and suitability for use in medical settings.



V. CONCLUSION

In the Streamlit framework, deep learning models have been investigated in this study for the purpose of detecting and classifying different forms of breast cancer, such as those seen in histopathology, mammography, ultrasound, and normal cases. Our created models—in particular, Convolutional Neural Networks (CNN) and refined VGG-19—have proven to be highly effective in correctly identifying both malignant and benign breast tissues through the examination of MRI pictures from the hospital's imaging department.

Medical practitioners can now upload, manage, and analyze MRI images in real-time on an easy-to-use platform thanks to the inclusion of an interactive interface powered by Streamlit. This allows for more informed clinical decision-making. This method not only increases the effectiveness of diagnosis but also encourages early intervention, which is essential for bettering patient outcomes and lowering the incidence of breast cancer.

In order to improve model performance and generalizability, future research could concentrate on

growing the dataset to cover a wider range of patient populations and including cutting-edge deep learning architectures. Additionally, there are chances to improve and expand our suggested framework for wider clinical application due to ongoing advancements in healthcare technology and informatics.

REFERENCE

- [1] Ji, J. M., & Shen, W. H. (2022). A Novel Machine Learning Systematic Framework and Web Tool for Breast Cancer Survival Rate Assessment. medRxiv, 2022-09.
- [2] Kakarla, D., Sahaja, P., Vaishnavi, K., Srileka, V., & Anusha, B. (2024). Investigations on Color Normalization Technique Using CycleGAN Based Machine Learning Algorithms for Breast Cancer Detection-Data Deployment. Journal of The Institution of Engineers (India): Series B, 1-11.
- [3] Arnaud, B. T. P. (2023). Faculty Of Engineering Department Of Electrical And Technology And Electronic Engineering (Doctoral Dissertation, Faculty Of Engineering Department Of Electrical And Technology And Electronic Engineering Machine Learning-Based Breast Cancer Detection: A Case Study By Batchanou Tatap Perets Arnaud B. Eng.(Hons) Telecommunications Engineering A Dissertation Submitted to the Department of Electrical and Electronic Engineering, Faculty of Engineering and Technology, University of Buea).
- [4] Aldabbagh, G. (2020). Advanced Information Extraction System based on Unstructured Mammogram Reports. International Journal, 9(5).
- [5] Surya, A., Shah, A. K., Sasikumar, S. T., & Kabore, J. Butler Journal of Undergraduate Research h.
- [6] PANGA, J. B. (2021). BREAST CANCER PREDICTION USING MACHINE LEARNING ALGORITHMS.
- [7] Surya, A., Shah, A., Kabore, J., & Sasikumar, S. (2023). Enhanced Breast Cancer Tumor Classification using MobileNetV2: A Detailed Exploration on Image Intensity, Error Mitigation, and Streamlit-driven Real-time Deployment. arXiv preprint arXiv:2312.03020.
- [8] Chandrashekar, K., Setlur, A. S., Sabhapathi C, A., Raiker, S. S., Singh, S., & Niranjana, V. (2023). Decision support system and web-application using supervised machine learning algorithms for easy cancer classifications. Cancer Informatics, 22, 11769351221147244.

- [9] Qin, Y. (2021, October). A cancer cell image classification program: based on CNN model. In 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT) (pp. 140-144). IEEE.
- [10] Andrade, D. S., Terrematte, P., Rennó-Costa, C., Zilberberg, A., & Efroni, S. (2023). GENTLE: a novel bioinformatics tool for generating features and building classifiers from T cell repertoire cancer data. *BMC bioinformatics*, 24(1), 32.
- [11] Gopiseti, L. D., Kummera, S. K. L., Pattamsetti, S. R., Kuna, S., Parsi, N., & Kodali, H. P. (2023, January). Multiple disease prediction system using machine learning and streamlit. In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 923-931). IEEE.
- [12] Adewunmi, M., Sharma, S. K., Sharma, N., Sushma, N. S., & Mounmo, B. (2022). Cancer Health Disparities drivers with BERTopic modelling and PyCaret Evaluation. *Cancer Health Disparities*, 6.
- [13] Shukla, R., & Sawant, R. (2023, August). Multiple Disease Prediction System Using Machine Learning. In 2023 International Conference on Electrical, Electronics, Communication and Computers (ELEXCOM) (pp. 1-6). IEEE.
- [14] Biswas, H. (2021). Evaluation of Deep Learning Models for the Detection of Diabetic Retinopathy with Retinal Fundus Images and Development of a Prescreening Tool. Southern Illinois University at Edwardsville.
- [15] Karlapudi, A. P. (2023). 81. Prediction of plant based EGFR inhibitors against breast cancer (EGFR) using machine learning model. *Cancer Genetics*, 278, 25.
- [16] Veena, A., & Gowrishankar, S. (2022). Context based healthcare informatics system to detect gallstones using deep learning methods. *International Journal of Advanced Technology and Engineering Exploration*, 9(96), 1661.
- [17] Srujana, V., Sumasri, S., Raj, G. A., & Rama, L. V. Early Detection and Prediction of Multiple Diseases using CNN and Random Forest Algorithms: A Machine Learning Approach.
- [18] Karim, S., Machuve, D., Mkoba, E., & Pandhare, S. (2022, November). Object Detection Model for Poultry Diseases Diagnostics. In *International Conference on Technological Advancement in Embedded and Mobile Systems* (pp. 95-103). Cham: Springer Nature Switzerland.
- [19] Lawal, S. K., Muniru, I. O., Yahaya, S. A., & Ibitoye, M. O. (2024). Automated identification of heart arrhythmias through HRV analysis and machine learning. *Nigerian Journal of Technological Development*, 21(1), 73-84.
- [20] Kothari, S., Sharma, S., Shejwal, S., Kazi, A., D'Silva, M., & Karthikeyan, M. (2024). An explainable AI-assisted web application in cancer drug value prediction. *MethodsX*, 12, 102696.s