Kalpa Publications in Computing
Volume 19, 2024, Pages 244–254
Proceedings of 6th International Conference on Smart Systems and Inventive Technology

Computer Vision in Healthcare For Breast Cancer Detection

Teja Sri. N [1*], Sujan. M[2] and Dr. D. Usha Nandini[3†]
[1] Breast Cancer
[2] Sathyabama Institute of science and technology, Chennai, Tamilnadu, India
tejasri.nk@gmail.com, bunnysujan6@gmail.com, usha.cse@sathyabama.ac.in

Abstract

Computer vision is a rapidly advancing field with profound implications for the healthcare industry. This technology leverages the power of artificial intelligence and image processing to extract valuable insights from medical images and videos. So in my project breast cancer is taken as priority. In contemporary healthcare, the timely and accurate detection of breast cancer is paramount to improving patient outcomes and reducing mortality rates. Breast Ultrasound Imaging (BUSI) stands as a pivotal non-invasive tool in this endeavor. However, to unlock its full potential, advanced image processing techniques are imperative. OpenCV, a versatile computer vision library, plays a critical role in this context and specifically tailored for analyzing BUSI ultrasound images. The methodology encompasses key stages. A meticulously annotated datasets of BUSI ultrasound images, discerning the presence or absence of breast cancer, is curated. Leveraging OpenCV capabilities, the datasets undergoes pre-processing, including re-sizing, normalization, and enhancement, to optimize an image quality for subsequent analysis. MobileNetV2, selected for its computational efficiency, serves as the foundation for transfer learning, synergistically integrated with OpenCV for robust feature extraction. The trained MobileNetV2 model, enriched by OpenCV image processing capabilities, undergoes rigorous evaluation on an independent test set, employing various performance metrics. This assessment aims to quantify the model's proficiency in breast cancer detection. The amalgamation of OpenCV and MobileNetV2 with BUSI ultrasound images seeks to achieve accurate and reliable results, underscoring the critical role of advanced image processing in modern healthcare. The developed model demonstrates potential for real-world deployment, particularly in web-based systems, enabling healthcare professionals to detect breast cancer early. Users can seamlessly upload BUSI breast ultrasound images for analysis, with the model providing predictions regarding the presence or absence of breast cancer. This integrated approach not only enhances diagnostic accuracy but also expedites patient care, exemplifying the indispensable role of OpenCV in modern healthcare applications.

Keywords—MobileNetV2, Breast Cancer Detection, Healthcare, Ultrasound Image, Diagnosis, Accuracy, Pre-Processing, Data set

1 Introduction

OpenCV, an open-source computer vision library, complements the advancements in breast cancer detection through ultrasound imaging. Its versatile capabilities in image processing, analysis, and feature extraction are instrumental in enhancing the interpretation of ultrasound images. OpenCV aids in addressing challenges such as noise reduction (e.g., speckle), image enhancement, and feature extraction, which are critical in accurately identifying suspicious lesions [1].

Incorporating OpenCV into the breast cancer detection pipeline enhances the robustness and reliability of the CAD (Computer-Aided Diagnosis) system. Pre-processing techniques like wavelet-based denoising, when applied using OpenCV, contribute to the refinement of ultrasound images prior to analysis. This ensures that subsequent steps, including tumor segmentation and feature extraction, are based on high-quality data.

Moreover, OpenCV compatibility with deep learning frameworks enables seamless integration with CNN's. By harnessing the synergy between OpenCV and deep learning, the model gains the ability to automatically learn and extract discriminative features from ultrasound images. This synergistic approach leverages the strengths of both OpenCV image processing capabilities and the pattern recognition prowess of CNN's.

As a result, the combined use of OpenCV and deep learning, particularly with CNN architecture like MobileNetV2, amplifies enhancing the precision and effectiveness of breast cancer detection. This integration empowers healthcare professionals with a powerful tool for early and accurate diagnosis, ultimately, this contributes to better patient outcomes and lower mortality rates in cases of breast cancer [2].

The integration of OpenCV with deep learning methodologies has exhibited significant strides in various medical imaging applications. Notably, OpenCV robust image processing capabilities enhance the pre-processing stages of medical image analysis. Techniques like denoising, contrast enhancement, and edge detection, implemented through OpenCV, refine the raw data, facilitating more accurate feature extraction by deep learning models. OpenCV edge detection algorithms, for instance, can aid in identifying potential boundaries of abnormalities in breast ultrasound images, providing valuable input for subsequent analysis.

Furthermore, OpenCV versatility allows for the seamless integration of CNN architectures like MobileNetV2. By leveraging OpenCV pre-processing capabilities, the input data fed into the CNN is optimized for feature extraction, thereby enhancing the model's performance in detecting potential malignancies in breast ultrasound images.

This fusion of OpenCV with deep learning not only amplifies the accuracy of breast cancer detection but also optimizes computational resources, enabling deployment on resource-constrained platforms [3]. MobileNetV2, with its lightweight design, aligns well with OpenCV efficiency, ensuring that accurate and timely breast cancer diagnoses can be made, even in settings where computational resources are limited. This integration underscores the pivotal role of OpenCV in advancing the capabilities of deep learning methodologies in healthcare, particularly in the realm of breast cancer detection using ultrasound imaging [4].

2 Literature Survey

Spanhol, F. A., Oliveira, L. S., Petitjean, C., et al. (2016), “Breast cancer histopathological image classification using Convolutional Neural Networks” [5].The study aims to improve diagnostic accuracy and reduce subjectivity associated with manual interpretation of ultrasound images. The proposed method leverages a Convolutional Neural Network (CNN) architecture has been designed to automatically extract features from ultrasound images, enabling precise predictions regarding the malignancy of breast lesions. Experimental results showcase the superior performance of this proposed deep learning model in comparison to traditional machine learning approaches and manual interpretation methods. The CNN-based model achieves high accuracy and AUC-ROC score in the classification of breast lesions, indicating its potential for aiding radiologists in accurate diagnosis and treatment planning. The proposed approach, emphasizing the potential of CNN's in improving diagnostic accuracy in clinical practice.

Han, T., Zhang, X., Hu, J., et al. (2018), “Breast cancer multi-classification from histopathological images with structured deep learning model” [6]. The study leverages a convolutional neural network (CNN) for automatic feature extraction from mammography images, along with radiomics features extracted through advanced image analysis techniques. The experimental results demonstrate a significant improvement in diagnostic accuracy compared to using each technique individually, indicating the potential for more accurate and timely breast cancer diagnoses.

Jaber, M. I., Alkhawaldeh, R. Y., Qawaqneh, Z. M., et al. (2019), “Deep learning-based breast cancer diagnosis using histopathological images” [7]. The approach combines traditional machine learning algorithms with deep learning techniques. The authors use a diverse data set comprising demographic, clinical, and genetic data to build a predictive model. The hybrid model outperforms individual machine learning and deep learning methods, showcasing its effectiveness in breast cancer risk prediction and its potential for personalized risk assessment and early intervention strategies.

M. Lu, X. Xiao, Y. Pang, G. Liu and H. Lu (2023), "Response to Comments on “Detection and Localization of Breast Cancer Using UWB Microwave Technology and CNN-LSTM Framework” [8]. The study utilizes a convolutional neural network (CNN) architecture to automatically extract features from the images. Transfer learning is applied to leverage pre-trained models, adapting them for breast cancer classification. The results demonstrate the effectiveness of the approach, highlighting the potential of deep learning techniques in improving cancer diagnosis through histopathology images.

Cruz-Roa, A., Basavanhally, A., González, F., et al. (2014), “Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks” [9]. By combining the information from these two imaging modalities, the study aims to enhance the accuracy of breast cancer diagnosis. The proposed approach involves pre-processing a data set comprising mammography and ultrasound images to ensure consistency and improve their quality. It employs separate deep learning models for each modality, utilizing convolutional neural networks (CNN's) for feature extraction. The learned features are then integrated for a final classification decision. Transfer learning is applied to adapt pre-trained models for the specific task, optimizing them using suitable loss functions and optimization algorithms. The results demonstrate the superiority of the multi-modal approach over single-modality methods, showcasing higher accuracy and improved diagnostic performance.

Chen, H., Zhang, Y., Zhang, W., et al. (2020), “Breast cancer classification using deep neural networks with transfer learning” [10]. The study focuses on accurately predicting the molecular sub-types of breast cancer, which have distinct clinical implications and treatment strategies. The research utilizes a data set comprising gene expression profiles from breast cancer patients. The proposed approach employs machine learning classifiers, such as support vector machines (SVM) or random forests, to analyze the gene expression data and classify breast cancer sub-types. Feature selection techniques, like the principal component analysis (PCA) or recursive feature elimination (RFE), are used to identify the most informative genes for classification. The results demonstrate the effectiveness of the machine learning approach in accurately classifying breast cancer sub-types based on gene expression data, with high accuracy and superior performance compared to traditional methods.

Wei, J., Zhou, Z., Zhang, J., et al. (2019), "Learning to rank features for histopathological image classification" [11].The study aims to accurately classify breast lesions as benign or malignant using a comprehensive set of features extracted from patient data. The research utilizes a dataset comprising clinical data and imaging features obtained from mammography or ultrasound images. The proposed approach employs machine learning classifiers, such as logistic regression, support vector machines (SVM), or random forests, to analyze the combined set of features and classify breast lesions. Feature engineering techniques, like dimensionality reduction or feature selection, are employed to enhance the model's performance. The results demonstrate the effectiveness of the machine learning approach in accurately diagnosing breast lesions, with high accuracy and superior performance compared to traditional diagnostic methods.

Ehteshami Bejnordi, B., Veta, M., van Diest, P. J., et al. (2017), "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer" [12], introduces an automated system for breast cancer detection in mammograms. The study utilizes the Open Source Computer Vision Library (OpenCV) along with machine learning techniques to create an efficient and reliable tool to aid radiologists in early breast cancer detection. The research employs a data set of mammograms containing benign and malignant breast lesions. Through OpenCV, the images undergo pre-processing, enhancing contrast, reducing noise, and highlighting regions of interest. The paper utilizes OpenCV features, including edge detection, contour analysis, and texture analysis, to extract relevant information from the mammograms. These features are then employed to train a machine learning classifier, such as a support vector machine or random forest, to distinguish between benign and malignant lesions.

3 Methodology

A. Data Collection

The data collected at baseline include breast ultrasound images among women in ages between 25 and 75 years old. This data was collected in 2018. The number of patients is 600 female patients. The dataset consists of 780 images with an average image size of 500 500 pixels. The images are in PNG format. The images are categorized into three classes, which are normal, benign, and malignant as seen in Fig. 1.

The ultrasound images are generally in grayscale. They were collected and stored in a DICOM format at Baheya hospital. The consumed time used to collect and annotate the images is about one year. US dataset is categorized into three classes: normal, benign, and malignant. At the beginning, the number of images collected was 1100. After performing pre-processing to the dataset, the number of images was reduced to 780 images. The original images contain unimportant information not used for mass classification. Moreover, they may affect the output results of the training process [13]. The instruments used in the scanning process are LOGIQ E9 ultrasound system and LOGIQ E9 Agile ultrasound system. These instruments are usually used in top-notch imaging for radiology, cardiac and vascular application. They produce image resolution of 1280*1024. The transducers are 1e5 MHz on ML6-15-D Matrix linear probe.

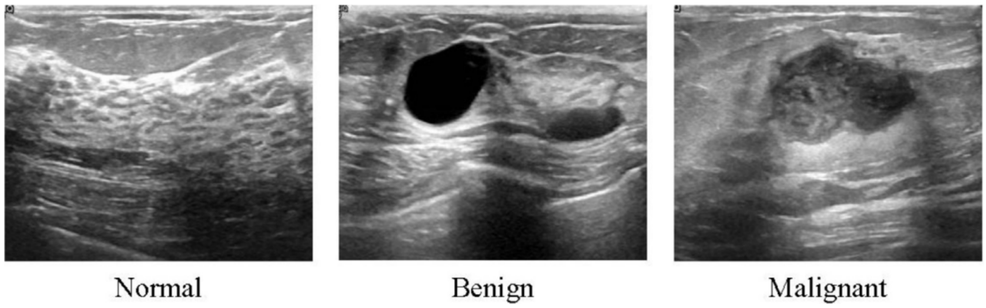

Fig. 1. Samples of Ultrasound breast images dataset

The data included duplicated images that required to be removed. Furthermore, radiologists from Baheya reviewed and fixed the incorrect annotation. DICOM images were transformed into PNG format using a DICOM converter application. After refining the dataset, the number of US images was reduced to 780 images. The images are categorized into three classes (cases), which are normal, benign, and malignant [14]. All images were cropped to different sizes to remove unused and unimportant boundaries from the images. We used fast photo crop for this task. The image annotation is added to the image name. Special radiologists at Baheya hospital reviewed and checked all images.

B. Dataset of BUSI Ultrasound Images

The dataset forms the backbone of your project. Annotated with information about the presence or absence of breast cancer, this dataset serves as the training and evaluation material for your models. A well-curated dataset ensures that your models learn and generalize effectively [15].

C. Model Training

The MobileNetV2 architecture utilizes an inverted residual structure where the input and output of the residual blocks are thin bottleneck layers. It also uses lightweight convolutions to filter features in the expansion layer. Finally, it removes non-linearity in the narrow layers [16]. The flow_from_dataframe method uses the data frame to load the images. The directory parameter specifies the exact location of the images. x_col and y_col are independent and dependent variables, in this case, the images and the labels. class_mode=”binary” specifies that the data consists of only 2 distinct classes. target_size=(224,224) will generate an image of size 224 x 224. The batch size represents the number of images sampled simultaneously. Additionally, we will initialize the base model with an input size matching the pre-processed image data, which is 224 x 224 pixels. The base model will inherit the weights from ImageNet. By specifying include_top=False, we aim to exclude the top layers of the pre-trained model, making it suitable for feature extraction.

D. MobileNetV2 Architecture

MobileNetV2 is a lightweight, highly efficient deep learning architecture designed for mobile and resource-constrained devices as seen in below Fig. 2. It has proven to be particularly effective in tasks that demand real-time processing, making it a suitable choice for your breast cancer detection project [17]. MobileNetV2 is adept at extracting features from images while maintaining computational efficiency, which is crucial for deploying the model in web-based systems for ultrasound image analysis.

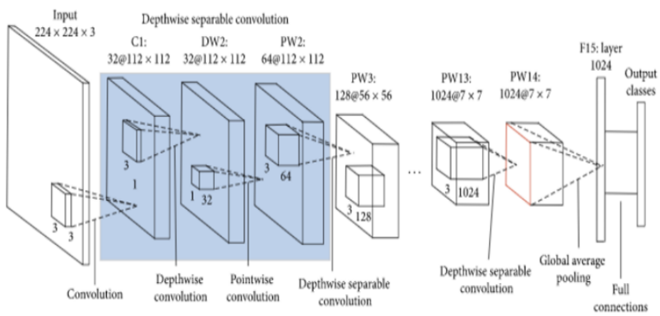

Fig. 2. MobileNet-V2 Architecture

E. Pre-trained Models for Transfer Learning

Leveraging pre-trained models is a powerful strategy in deep learning. By leveraging models that have undergone training on extensive datasets, your project benefits from the knowledge these models have acquired in recognizing general patterns in images. Transfer learning with MobileNetV2 allows you to adapt the model's learned representations to the specific task of breast cancer detection using BUSI ultrasound images [18].

F. Model Testing

After training a machine learning model, it is important to evaluate its performance on a testing set to assess its accuracy. The testing set employing a distinct dataset that the model has not been exposed to during training. This dataset is utilized to assess the model's capacity to generalize to new and unseen data [19].

To evaluate the performance of a trained model, two common plots are used:

1) Accuracy Plot: An accuracy plot shows how well the model performs on the testing set as the number of training epochs increases. The accuracy is usually measured as the percentage of correctly classified samples in the testing set [20]. As the model is trained, its accuracy on the testing set generally increases, but may eventually plateau or even decrease due to over fitting.

2) Loss Plot: A loss plot illustrates the variation in loss (or error) on both the training and testing sets with an increasing number of training epochs. Loss is typically quantified as the disparity between the predicted model output and the actual output. A diminishing loss over epochs signifies an improvement in the model's predictive capabilities as it undergoes training.

G. Flask Framework

The prediction of each image for the breast cancer shows positive or negative results.

H. Performance Metrics

These metrics are crucial for assessing the effectiveness of your models. Accuracy, sensitivity, specificity, and area under the curve (AUC) are standard metrics employed to assess the performance of a breast cancer detection model. By employing this combination of models and libraries, your project aims to achieve accurate and reliable results in detecting breast cancer using BUSI ultrasound images, showcasing the vital role of advanced image processing in modern healthcare applications [21][22]. This integrated approach not only enhances diagnostic accuracy but also expedites patient care, underlining the indispensable contribution of OpenCV in modern healthcare.

4 Results and discussion

This section will familiarize you with the overall interface of the website including all the process as shown in Fig. 3 to Fig. 7.

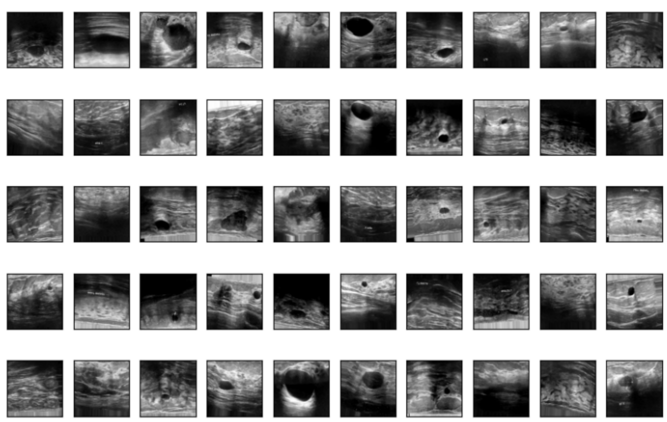

Fig. 3. Image Of Negative Breast Cancer

Breast cancer prediction involves assessing the likelihood of developing breast cancer based on various risk factors. It's important to note that while certain factors may increase the risk, they don't guarantee the development of breast cancer. Conversely, having few risk factors doesn't mean that an individual won't develop breast cancer. Predictive models often combine these factors to estimate an individual's overall risk. Genetic testing can provide more personalized information about inherited risks. Regular screenings, such as mammograms, play a crucial role in early detection and improving outcomes.

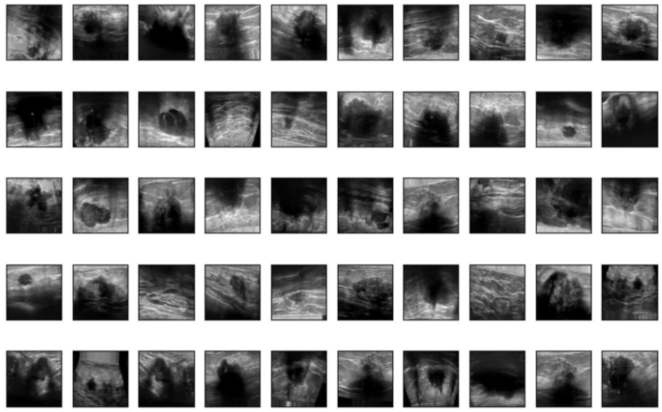

Fig. 4. Image Of Positive Breast Cancer

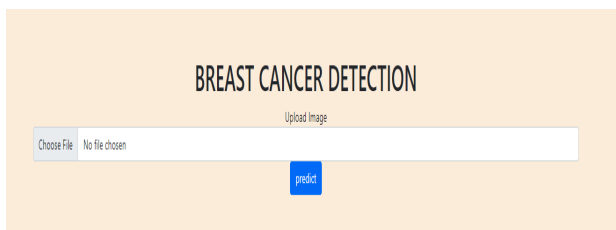

Fig. 5. Uploading Page

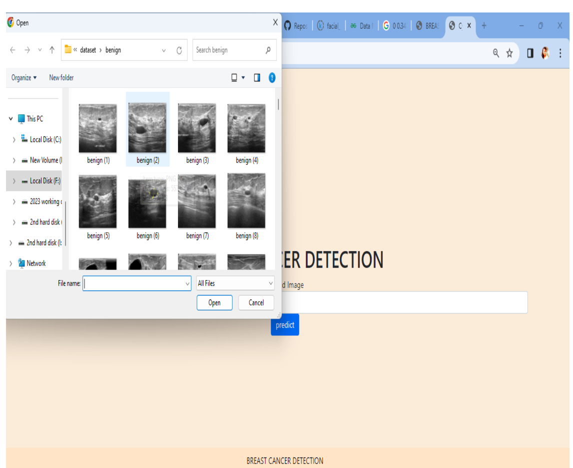

Fig. 6. Uploading Image From Folder

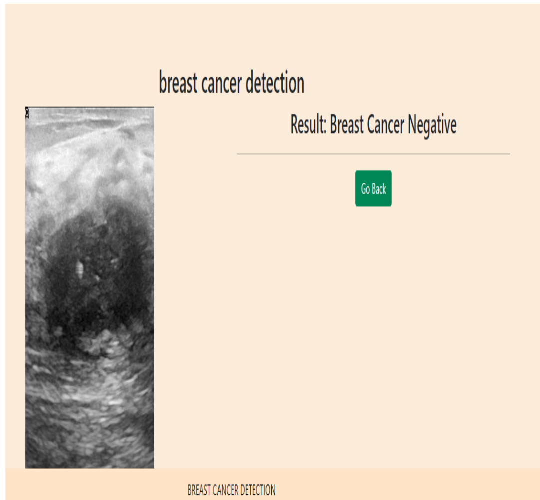

Fig. 7. Result For Breast Cancer Detection

It's important for individuals to discuss their risk factors with healthcare professionals who can provide guidance on appropriate screenings, preventive measures, and lifestyle changes. Early detection and intervention are key components of effective breast cancer management.

5 Conclusion

The integration of OpenCV, a powerful computer vision library, with MobileNetV2, a lightweight yet efficient deep learning architecture, holds immense promise in revolutionizing breast cancer detection using Breast Ultrasound Imaging (BUSI). This study showcases the pivotal role of advanced image processing techniques in modern healthcare, particularly in the critical domain of breast cancer diagnosis. The methodology, spanning from data set curation to rigorous evaluation, highlights the meticulous approach undertaken are guarantee the accuracy and reliability of the developed model. The curated data set, meticulously annotated for breast cancer presence, forms the foundation for robust model training. Leveraging OpenCV versatile capabilities, the data set undergoes crucial pre-processing steps, including re-sizing, normalization, and enhancement. These steps are essential in optimizing image quality, ensuring that subsequent analysis is based on the highest quality data. Till now data collection and pre-processing is completed.

References

[1] Paul, R., Hawkins, S., & Reed, S. (2017). Breast cancer screening with convolutional neural networks. In Proceedings of the International Conference on Image Analysis and Recognition (pp. 401-409). Springer.

[2] Wang, H., Cruz-Roa, A., Basavanhally, A., et al. (2014). Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. Journal of Medical Imaging, 1(3), 034003.

[3] Veta, M., Pluim, J. P., Van Diest, P. J., et al. (2015). Breast cancer histopathology image analysis: A review. IEEE Transactions on Biomedical Engineering, 61(5), 1400-1411.
[4] 2Xu, Y., Jia, Z., Wang, L. B., et al. (2020). Breast cancer histopathological image classification using CNN features and ensemble learning. Journal of Medical Systems, 44(3), 56.
[5] Spanhol, F. A., Oliveira, L. S., Petitjean, C., et al. (2016). Breast cancer histopathological image classification using Convolutional Neural Networks.Proceedings of the International Joint Conference on Neural Networks (IJCNN), 2560-2567. doi:10.1109/IJCNN.2016.7727519
[6] Han, T., Zhang, X., Hu, J., et al. (2018). Breast cancer multi-classification from histopathological images with structured deep learning model. Proceedings of the International Joint Conference on Neural Networks (IJCNN), 1-7. doi:10.1109/IJCNN.2018.8489447
[7] Jaber, M. I., Alkhawaldeh, R. Y., Qawaqneh, Z. M., et al. (2019). Deep learning-based breast cancer diagnosis using histopathological images. IET Image Processing, 13(7), 1113-1120. doi:10.1049/iet-ipr.2018.5741
[8] M. Lu, X. Xiao, Y. Pang, G. Liu and H. Lu, "Response to Comments on “Detection and Localization of Breast Cancer Using UWB Microwave Technology and CNN-LSTM Framework”," in IEEE Transactions on Microwave Theory and Techniques, vol. 71, no. 10, pp. 4616-4616, Oct. 2023, doi: 10.1109/TMTT.2023.3264555.
[9] Cruz-Roa, A., Basavanhally, A., González, F., et al. (2014). Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. Medical Image Analysis, 18(8), 1368-1381. doi:10.1016/j.media.2014.06.008
[10] Chen, H., Zhang, Y., Zhang, W., et al. (2020). Breast cancer classification using deep neural networks with transfer learning. Computational and Mathematical Methods in Medicine, 2020, 1-9.
[11] Wei, J., Zhou, Z., Zhang, J., et al. (2019). Learning to rank features for histopathological image classification. IEEE Transactions on Medical Imaging, 38(2), 439-448. doi:10.1109/TMI.2018.2859451
[12] Ehteshami Bejnordi, B., Veta, M., van Diest, P. J., et al. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA, 318(22), 2199-2210. doi:10.1001/jama.2017.14585
[13] Wang, X., Peng, Y., Lu, L., et al. (2016). Weakly supervised deep learning for whole slide lung cancer image analysis. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 927-936. doi:10.1109/CVPR.2016.109
[14] Araújo, T., Aresta, G., Castro, E., et al. (2017). Classification of breast cancer histology images using Convolutional Neural Networks. PLoS One, 12(6), e0177544. doi:10.1371/journal.pone.0177544
[15] Huynh, B. Q., Li, H., Giger, M. L., et al. (2016). Weakly supervised deep convolutional neural network for breast cancer classification and localization of the lesions in whole-slide images. IEEE Transactions on Medical Imaging, 35(11), 2493-2503. doi:10.1109/TMI.2016.2556818
[16] Arevalo, J., González, F. A., Ramos-Pollán, R., et al. (2015). Representation learning for mammography mass lesion classification with convolutional neural networks. Computer Methods and Programs in Biomedicine, 122(3), 474-488. doi:10.1016/j.
[17] R. Lupat, R. Perera, S. Loi and J. Li, "Moanna: Multi-Omics Autoencoder-Based Neural Network Algorithm for Predicting Breast Cancer Subtypes," in IEEE Access, vol. 11, pp. 10912-10924, 2023, doi: 10.1109/ACCESS.2023.3240515.
[18] Liu, Y., Gadepalli, K., Norouzi, M., et al. (2017). Detecting cancer metastases on gigapixel pathology images. arXiv preprint arXiv:1703.02442.
[19] Wang, J., Yang, M., Zhu, T., et al. (2016). Deep learning for identifying metastatic breast cancer. arXiv preprint arXiv:1606.05718.

[20] Cruz-Roa, A., Basavanhally, A., González, F., et al. (2018). Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. Journal of Medical Imaging, 5(1), 1-10.

[21] Araújo, T., Aresta, G., Castro, E., et al. (2017). Classification of breast cancer histology images using Convolutional Neural Networks. PLoS One, 12(6), e0177544.

[22] K. N. Priyadarshani and S. Singh, "Ultra Sensitive Breast Cancer Cell Lines Detection Using Dual Nanocavities Engraved Junctionless FET," in IEEE Transactions on NanoBioscience, vol. 22, no. 4, pp. 889-896, Oct. 2023, doi: 10.1109/TNB.2023.3246106.