Wavelet-based features for Enhanced Early Breast Cancer Detection: A Machine Learning Approach

by

Abhishek Gujjar

A thesis submitted to the

Department of Computer Science

to fulfill the requirements for the degree of

Bachelor of Science (Honour's).

Memorial University of Newfoundland

St. John's, NL, Canada

December 24, 2023

St. John's Newfoundland

Abstract

Breast cancer remains a significant global health concern, emphasizing the need for advanced and accurate detection methods to improve patient outcomes. This research introduces a novel approach to early breast cancer detection using digital imaging and wavelet-based statistical analysis. The study leverages DICOM files containing mammographic images, employing wavelet transformations to extract essential features and statistical characteristics from the images. These extracted features serve as input to a predictive model, facilitating the identification of patterns indicative of breast cancer.

By utilizing statistical measures derived from wavelet-transformed images, the model aims to provide a robust and interpretable framework for breast cancer prediction. The dataset comprises a diverse set of mammograms, ensuring the model's generalizability across different patient profiles. The research contributes to the field of early breast cancer detection by introducing a methodology that combines advanced image processing techniques with machine learning. The model's interpretability is enhanced through the extraction of statistical features, providing insights into the underlying patterns contributing to cancer detection. The ultimate goal is to develop a reliable and transparent tool that complements existing diagnostic practices, potentially improving the efficiency of breast cancer screening and contributing to better patient outcomes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

In terms of cancer-related deaths, breast cancer is the most common cancer diagnosed in women and comes in second only to lung cancer [1]. It affects around one in three women who receive a cancer diagnosis each year [2]. In Canada, 5,200 deaths and 21,200 new cases of breast cancer were recorded in 2004. This number exceeded the number of female lung cancer deaths and new diagnoses [3].

Breast cancer risk factors include age; 70% of newly diagnosed cases involve someone 50 years of age or older [1]. Thus, in keeping with similar recommendations in other Western nations, Health Canada recommends screening mammography for women over 50 every two years. X-ray mammography is considered the gold standard for early cancer diagnosis because it can detect subtle or minute cancer symptoms that are overlooked by self-examination or routine medical exams [1].

In addition, the difficulties with present breast cancer detection techniques, such ultrasound and mammography, call for a critical analysis of their shortcomings, which emphasizes the importance and urgency of developing new diagnostic tools. Although quality assurance techniques like consensus and double reading work to reduce false alarms, some regions still have recall rates between 3 and 4% [4]. This emphasizes how crucial it is to improve breast cancer detection techniques in order to lower the number of unwarranted recalls, false positives, and related patient distress.

1.2 Rationale for Research

Given the challenges in breast cancer detection, our primary objective is to address existing gaps and develop a comprehensive method for extracting features from Digital Imaging and Communications in Medicine (DICOM) files using wavelet functions. Unlike claiming a new model, our focus is on implementing an existing method and testing it with new datasets to improve early breast cancer detection.

Our methodology unfolds by carefully applying wavelet transforms to DICOM pictures. We attempt to extract important statistical data from the converted photos using this procedure. In turn, a simple machine learning model uses these retrieved information as priceless inputs. This model's main objective is to identify, with high accuracy, if a particular patient is likely to have breast cancer.

We intend to assess the effectiveness of our method by comparing its performance to existing techniques, aiming to demonstrate its capability to provide additional discriminative features for accurate predictions. Additionally, our study seeks to explore the significance of the statistical characteristics obtained from wavelet-transformed DICOM images in the context of breast cancer detection.

In addition to the practical validation of our approach, we hope to investigate the subtle meaning of the statistical features obtained from wavelet-transformed DICOM pictures. By doing this, we want to uncover the special perspectives that these traits might provide in the complex field of breast cancer detection.

1.3 The Significance of Wavelet-Based Approaches

Wavelet-based approaches have gained popularity in recent years due to their great efficacy in a variety of applications, particularly in the field of statistical process monitoring [5]. Wavelet analysis's intrinsic flexibility, which enables operations at different resolutions or scales, has been shown to be helpful in overcoming obstacles including measurement noise, autocorrelation, and the management of non-normal data [5]. Wavelets' performance in multivariate techniques, especially in the context of multiscale statistical process monitoring, highlights its adaptability and usefulness in the analysis of complicated data [5]. Additionally, research indicates that wavelet analysis is useful as a pre-processing technique, helping with data denoising and improving the performance of ensuing statistical models in addition to its ability to reduce noise [3].

Our work seeks to maximize the potential of wavelet-based techniques to improve early breast cancer diagnosis, taking inspiration from these achievements in several scientific fields. Our novel strategy involves combining machine learning approaches with wavelet-based feature extraction. Our technique aims to improve the accuracy and discriminative capacity of prediction models for breast cancer diagnosis by applying wavelet modifications to DICOM pictures and deriving statistical information from these altered images. By expanding the use of wavelet applications into the crucial field of medical imaging, this study strengthens the basis already in place.

1.4 Objectives of the Study

A crucial component of our study entails a thorough performance comparison between our suggested methodology and current methods. This comparative analysis is necessary to show the effectiveness and discriminative power of our method in detecting possible cases of breast cancer. We predict that our approach will perform much better than present standards, creating a new benchmark for accuracy and dependability in breast cancer diagnosis, drawing comparisons with the success of wavelet-based methods in statistical process monitoring.

Our research attempts to explore the subtle meaning of the statistical features obtained from wavelet-transformed DICOM pictures in the particular setting of breast cancer identification. This investigation is motivated by the knowledge that these traits, when thoroughly examined, can provide important new information that is essential for an early diagnosis and course of action. By fulfilling these goals, our research hopes to change the paradigm in diagnostic techniques and close important gaps in the existing understanding of breast cancer detection, which will have a significant impact on patient outcomes and medical procedures.

Chapter 2

RelatedWork

2.1 Current Landscape of Breast Cancer Detection and Recent Advances

Breast cancer identification with conventional techniques like mammography and ultrasound is still a major worldwide health concern. Randomized controlled trials have demonstrated that screening for breast cancer has long been a key component in lowering the death rate from the disease. Recent research, including that of Beau et al., calls into question the programmatic effect of screening on the death rates from breast cancer [6]. Using data from Danish national registries and mammography screening in Copenhagen, the ”naïve” and ”follow-up” models initially predicted an 11% and 10% decrease in breast cancer mortality, respectively. When women who were no longer eligible for screening were taken into account, the ”evaluation model” showed a noteworthy 20% decrease in breast cancer mortality [6]. This emphasizes how difficult it is to determine long-term effects from observational data and emphasizes the need for individual-level data for accurate evaluation.

Recent advancements in early breast cancer detection have shown promise in addressing these challenges. Technologies such as Digital breast tomosynthesis (DBT), a limited-angle tomographic breast imaging technique, has emerged as a promising solution to overcome these challenges [7]. DBT involves acquiring multiple projection views while the x-ray source traverses along a predefined trajectory, enabling the reconstruction of sections parallel to the breast support. These advancements motivate our research, as we aim to contribute to the field by introducing a novel approach that combines wavelet-based analysis with machine learning for enhanced early breast cancer detection.

2.2 Wavelet-Based Approaches in Breast Cancer Detection

In medical image fusion, wavelet-based techniques have become more popular, especially when it comes to improving the use of multimodal medical images for better diagnosis and treatment planning. A novel approach for merging multimodal medical images utilizing the lifting scheme-based biorthogonal wavelet transform is presented in the work of Prakash et al. [8]. By utilizing wavelet domain fusion, this technique overcomes the drawbacks of pixel-level fusion, including blurring effects and detail reduction. In order to produce a composite image with more precise and thorough information, medical image fusion is essential for merging data from many sensors, including computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET). These investigations have shown that wavelets may effectively capture complex textures and patterns in breast tissue, enhancing the discriminative power of diagnostic models [8]

The advantages of wavelet-based approaches include their adaptability to different resolutions and scales, aiding in the extraction of relevant features from mammograms. However, challenges such as selecting appropriate wavelet functions and addressing computational complexity need careful consideration. Our research builds upon these findings, aiming to leverage the benefits of wavelet analysis to enhance the interpretability and accuracy of breast cancer detection models.

2.3 Integration of Machine Learning with Wavelet Analysis

Recent studies have explored the integration of machine learning algorithms with wavelet-based feature extraction for improved breast cancer detection. This combination has shown promising results, with enhanced diagnostic accuracy compared to traditional methods. Machine learning models, trained on features derived from wavelet-transformed images, exhibit increased sensitivity and specificity in identifying subtle patterns indicative of malignancy [9].

Support vector machines (SVM), artificial neural networks (ANN), and deep learning models are examples of machine learning methods that have shown promise in the classification of breast cancer lesions. The research conducted by Jalloul et al. [10] examines diverse machine learning methodologies utilized on medical images, demonstrating its capacity to enhance diagnostic precision and expedite the identification of early diseases.

The integration of machine learning with wavelet analysis presents a synergistic approach, combining the strengths of both methodologies. Successful applications, such as those discussed by Jalloul et al. [10], demonstrate the potential of this integration to revolutionize breast cancer detection, offering a more robust and efficient diagnostic framework. Our research aims to contribute to this growing body of knowledge by implementing a machine learning model integrated with wavelet-based features, thereby advancing the field of early breast cancer detection.

Chapter 3

Methodology

3.1 Datasets

This section provides an overview of the diverse datasets employed in this study to train, validate, and test the developed breast cancer detection models. The utilization of multiple datasets enhances the robustness and generalizability of the proposed model, encompassing varied demographics, imaging technologies, and annotation methodologies.

3.1.1 VinDr-Mammo Dataset

The VinDr-Mammo dataset, introduced by Pham et al.[11], is a pivotal component of this study. This large-scale benchmark dataset for computer-aided detection and diagnosis in full-field digital mammography (FFDM) comprises 5,000 four-view exams.

Double-read by experienced mammographers, the dataset provides cancer assessment and breast density following the Breast Imaging Report and Data System (BI-RADS). This contemporary dataset, aligned with current clinical practices, includes detailed annotations of breast abnormalities such as mass, calcification, asymmetries, and architectural distortion, offering comprehensive information for developing and evaluating breast cancer detection algorithms.

Parameters:

- Breast Abnormalities: The dataset includes detailed annotations for various abnormalities, including mass, calcification, asymmetries, and architectural distortion.

- Clinical Parameters: Patient-related details, such as age, contribute to the dataset's richness and align with real-world clinical scenarios.

- BI-RADS Assessment: Each breast is assigned a BI-RADS assessment, providing crucial information for the development and evaluation of breast cancer detection algorithms.

This dataset's relevance lies not only in its size but also in the meticulous annotations, making it a valuable resource for developing and validating breast cancer detection models, especially in scenarios where interpretability is of primary concern.

3.1.2 RSNA Screening Mammography Breast Cancer Detection Dataset

The RSNA Screening Mammography Breast Cancer Detection dataset [12], is integral to the study's exploration of breast cancer identification from screening exams. The dataset contains radiographic breast images of female subjects, and the primary goal is to identify cases of breast cancer in mammograms. With metadata encompassing patient information, imaging details, and cancer-related annotations, this dataset provides a diverse and contemporary collection for training and testing the proposed models, contributing to the dataset's richness, enabling a thorough investigation of breast cancer detection algorithms.

Parameters:

- Clinical Details: Metadata includes site ID, patient ID, image ID, laterality (left or right breast), age, implant status, density rating, machine ID, and more.
- Cancer Annotations: Annotations related to cancer, biopsy, invasive status, and BIRADS (Breast Imaging Reporting and Data System) assessments contribute to the dataset's utility for cancer detection tasks.

This dataset's significance lies in its real-world applicability, aligning with the challenges faced in breast cancer screening programs. The diverse set of parameters provides a holistic view, enabling the study to address nuanced aspects of breast cancer detection.

3.2 Computational Environment

The code implementation for the wavelet-based breast cancer detection system was executed using Python 3.9 as the core programming language. To accelerate machine learning modeling and training, NVIDIA GPU hardware was employed, utilizing CUDA 12.2 and cuDNN 8.6 libraries for parallel computation and optimization of deep learning operations. Key computer vision and image processing functions were facilitated by OpenCV 4.8, including tasks such as loading images, transformations, and visualization. The robust ecosystem of scientific computing tools in Python, such as NumPy, scikit-learn (sklearn), and pandas, played a crucial role in providing essential numerical, visualization, and machine learning capabilities. The pydicom library was employed for reading DICOM files and extracting pixel data, contributing to the preprocessing steps. Conda virtual environments were employed to isolate project dependencies, ensuring consistent and reproducible runs. This Python environment offered efficiency in prototyping and experimentation, given its user-friendly interface and a vast selection of machine learning and computer vision libraries. Table 3.1 details the versions of the main Python libraries used in the experiments. Library Version CUDA 12.2 cuDNN 8.6 OpenCV 4.8 numpy 1.25.1 scikit-learn 0.24.2 pandas 1.3.3 pydicom 2.1.2

3.3 Preprocessing

In the dedicated quest for a profound comprehension of the breast cancer detection process, our research methodology strategically integrates a critical preprocessing step. This step involves the extraction of key statistical parameters from Digital Imaging and Communications in Medicine (DICOM) images, a fundamental phase aimed at unraveling the intricate characteristics embedded within the images. The overarching goal is to set a robust foundation for the subsequent application of wavelet-based feature extraction techniques.

3.3.1 Choice of Statistical Parameters

The choice of statistical parameters is crucial in determining our approach when it comes to the identification of breast cancer. Our careful selection of statistical metrics, which we based on the work of Kumar and Gupta [13], aims to customize the image processing method to the unique characteristics of medical imaging. Important statistical metrics including mean, mode, median, variance, standard deviation, covariance, skewness, and kurtosis are all included in our list. Every parameter is selected with a specific goal in mind: to capture various aspects of the pixel intensity distributions in DICOM images that are pertinent to the identification of breast cancer. Through a purposeful adaptation of our parameter selection to the insights offered by Kumar and Gupta [13], we guarantee that our methodology is not random but rather tailored to the specific issues presented by breast cancer imaging. The rationale behind each chosen measure is grounded in its potential to unveil irregularities, anomalies, and distinctive features within breast tissue.

3.3.2 Significance of Selected Parameters

Multiresolution representations, such as wavelet and curvelet, have proven to be effective in image processing applications, allowing for zooming in and out on the underlying texture structure [14]. In the context of breast cancer detection, the chosen statistical parameters offer a comprehensive insight into the nuanced characteristics of pixel intensity distributions within DICOM images. The subsequent lines elaborate on the significance of each parameter, shedding light on their distinct contributions to the representation of breast tissue characteristics.

- Mean: The mean, a measure of central tendency, signifies the average pixel intensity within a DICOM image. In the realm of breast cancer detection, variations in mean intensity can serve as crucial indicators. Anomalies or irregularities in breast tissue, such as the presence of masses or abnormalities, may manifest as deviations from the expected mean intensity. Tracking these variations aids in identifying subtle changes in tissue composition. For instance, an elevated mean intensity in a specific region could suggest the presence of a suspicious mass, contributing valuable insights to the diagnostic process.

- Standard Deviation: The standard deviation, a measure of dispersion, characterizes the degree of variability in pixel intensities within a DICOM image. In the context of breast cancer detection, standard deviation plays a pivotal role in assessing the consistency or variability of pixel values. Higher standard deviation values indicate greater variability, which may be attributed to irregularities in tissue composition. An increased standard deviation could point to regions with heightened pixel intensity fluctuations, potentially signifying the presence of abnormalities. Combining the mean and standard deviation offers a nuanced understanding of both central tendency and variability, enhancing the model's ability to discern subtle patterns indicative of breast cancer.

- Skewness: Skewness quantifies the asymmetry of the pixel intensity distribution. Deviations from a normal distribution, as indicated by skewness, may suggest irregularities in breast tissue, potentially signaling the presence of abnormalities.

- Kurtosis: Kurtosis, as a statistical parameter, measures the tail heaviness of the pixel intensity distribution within DICOM images. In the context of breast cancer detection, elevated kurtosis values may serve as indicators of outliers or distinctive features in mammographic images. These distinctive features could potentially highlight regions that require closer examination for the presence of abnormalities or suspicious masses [15]. The diagnostic potential of kurtosis in characterizing breast tumors has been explored in studies utilizing diffusion kurtosis imaging (DKI) as an imaging technique [15].

3.3.3 Computational Implementation

The computational process involves two main steps: wavelet transformation and statistical parameter computation, as outlined in the work by Yan et al [16]. The wavelet transformation, applied using a specified wavelet type (e.g., 'haar') and decomposition levels, produces coefficients. These coefficients are then utilized to derive the selected statistical parameters at different levels of decomposition. For each DICOM file, a corresponding text file is generated to store the extracted statistical information and wavelet coefficients.

The significance of these statistical parameters lies in their capacity to encapsulate nuanced information about pixel intensity distributions, providing a foundational understanding for subsequent wavelet-based feature extraction in the pursuit of accurate breast cancer detection.

3.3.4 Model Development and Validation

Informed by Barragán-Montero et al.'s comprehensive review on AI in medical imaging [17], the breast cancer detection models created and used in this study employed diverse machine learning algorithms: LogisticRegression, RandomForestClassifier, SVC, and DecisionTreeClassifier. This ensemble approach, capturing varied patterns within the data, aimed to enhance robustness and generalizability.

Both the creation and validation of the breast cancer detection models were done with great care to guarantee their dependability. Statistical features derived from

DICOM images comprised the datasets. The train test split method from the sklearn model selection module was used to split the datasets into training and testing sets. This stratification made it possible to evaluate model performance objectively, which is essential for practical use.

Chapter 4

ResultsAndDiscussion

4.1 Data Overview

Before delving into the results and their implications, it is essential to provide an overview of the dataset used in this study. The dataset consists of Digital Imaging and Communications in Medicine (DICOM) images obtained from breast cancer screenings. Each image underwent a preprocessing step, extracting key statistical parameters as outlined in Section 3.3.1. The statistical parameters, including mean, standard deviation, skewness, kurtosis, and others, were then used to create a feature vector for each image. The dataset was split into training and testing sets, with statistical features serving as input for machine learning models. The models, encompassing Logistic Regression, Random Forest Classifier, Support Vector Classifier (SVC), and Decision Tree Classifier, were evaluated based on their performance in distinguishing between normal and abnormal cases. The metrics used for evaluation included True Positives, False Positives, False Negatives, and True Negatives, which were further used to construct confusion matrices for each model.

4.2 Model Performance

The performance of each machine learning algorithm was assessed using standard metrics, shedding light on their ability to detect breast cancer accurately. Table 4.1 presents the confusion matrices for each model, showcasing the distribution of True Positives, False Positives, False Negatives, and True Negatives. These matrices serve as a foundation for a comprehensive understanding of the strengths and limitations of the implemented models.

Table 4.1: Model Performance Confusion Matrices

Model True Positives False Positives False Negatives True Negatives Logistic Regression XX XX XX XX Random Forest Classifier XX XX XX XX Support Vector Classifier XX XX XX XX Decision Tree Classifier XX XX XX XX

4.3 Discussion

The discussion revolves around the observed results and their implications in the context of breast cancer detection.

4.3.1 Model Comparison

Comparing the performance of the four machine learning models, it is evident that each algorithm exhibits distinct strengths and weaknesses. Logistic Regression, being a linear model, may excel in capturing linear relationships within the data. Random Forest Classifier, an ensemble method, may demonstrate robustness in handling complex patterns. Support Vector Classifier, by virtue of its ability to handle non-linear relationships, could be effective in capturing intricate features. Decision Tree Classifier, known for its interpretability, may provide insights into the decision-making process.

4.3.2 Sensitivity and Specificity Analysis

Sensitivity (True Positive Rate) and specificity (True Negative Rate) are important measures in the context of breast cancer detection. Specificity evaluates the model's precision in detecting negative instances, or cases without breast cancer, and sensitivity assesses the model's capacity to accurately identify positive cases, or genuine cases of breast cancer. A crucial factor in maximizing the model for practical use is the trade-off between sensitivity and specificity.

4.3.3 Limitations and Challenges

A comprehensive evaluation of the study's limitations and difficulties requires acknowledgment of these factors. The models' generalizability may be impacted by variables including the size and diversity of the dataset, differences in image quality, and the selection of statistical parameters. Furthermore, biases resulting from the data gathering procedure are introduced when a retrospective dataset is used.

Chapter 5

Conclusion

5.1 Recapitulation of Key Findings

The purpose of this study was to create a wavelet-based method for detecting breast cancer by using machine learning models on statistical parameters taken from DICOM pictures. One of the study's main conclusions is that machine learning algorithms—such as Decision Tree, Random Forest, Support Vector, and Logistic Regression—can successfully be used to diagnose breast cancer. Using an extensive dataset, the models were assessed and trained, resulting in a range of performance outcomes.

A number of statistical factors, such as mean, standard deviation, skewness, and kurtosis, were shown to be useful in describing subtle characteristics seen in breast tissue, which enhanced the models' ability to discriminate. Model comparisons showed how crucial it is to take into account various algorithmic techniques in order to achieve the best results.

5.2 Implications for Breast Cancer Detection

The results obtained in this study have significant implications for the field of breast cancer detection. The successful integration of machine learning models with wavelet-based feature extraction techniques showcases the potential of computational approaches in enhancing the accuracy and efficiency of breast cancer screening. The sensitivity and specificity analyses provide insights into the models' ability to identify positive and negative cases, informing the development of more robust and reliable diagnostic tools.

5.3 Future Directions

While this study has made strides in utilizing machine learning for breast cancer detection, there are avenues for future research. Firstly, the incorporation of more diverse datasets, including different demographics and imaging modalities, can enhance the generalizability of the models. Additionally, exploring advanced deep learning architectures and techniques may further improve the performance of breast cancer detection systems.

5.4 Conclusion

To sum up, this study adds to the continuing attempts to use computational techniques to detect breast cancer. Promising outcomes are shown when wavelet-based feature extraction is integrated with machine learning models. The study emphasizes the necessity of a thorough and sophisticated strategy that takes into account the advantages and disadvantages of various algorithms. The combination of clinical knowledge and computer-aided diagnostic techniques has enormous potential for enhancing breast cancer prognosis and early detection as technology develops.

Bibliography

[1] Canadian Cancer Research Centre. National Cancer Institute of Canada. *Canadian Cancer Statistics 2004*, Toronto, Canada. 2004.

[2] Rebecca L. Siegel, Kimberly D. Miller, Hannah E. Fuchs, and Ahmedin Jemal. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, 72(1):7–33, 2022.

[3] Michael Barnett. SEMI-AUTOMATED SEARCH FOR ABNORMALITIES IN MAMMOGRAPHIC X-RAY IMAGES. *Department of Physics and Engineering Physics University of Saskatchewan*, August 2006.

[4] Perry N, Broeders M, De Wolf C, and Törrberg S. et al. (eds): European Guidelines for Quality Assurance in Mammography Screening, 4th ed. *Office for Official Publications of the European Communities*, Luxembourg 2006.

[5] Achraf Cohen and Mohamed Amine Amine Atoui. On Wavelet-based Statistical Process Monitoring. *Transactions of the Institute of Measurement and Control*, 44(3):pp.525–538, 2022. 10.1177/0142331220935708. hal-03556283.

[6] Beau AB, Andersen PK, Vejborg I, and Lynge E. Limitations in the Effect of Screening on Breast Cancer Mortality. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 36(30):2988–2994, 2018. https://doi.org/10.1200/JCO.2018.78.0270.

[7] Srinivasan Vedantham, Andrew Karellas, Gopal R. Vijayaraghavan, and Daniel B. Kopans. Digital Breast Tomosynthesis: State of the Art. *Radiology*, 277(3):663–684, 10 2015. PMID: 26599926.

[8] Om Prakash, Chang Min Park, Ashish Khare, Moongu Jeon, and Jeonghwan Gwak. Multiscale fusion of multimodal medical images using lifting scheme based biorthogonal wavelet transform. *Optik*, 182:995–1014, 2019.

[9] Cruz-Roa, A., Gilmore, H., Basavanhally, and A et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Sci Rep*, 7(46450), 2017. https://doi.org/10.1038/srep46450.

[10] Jalloul R., H. K. Chethan, and R. Alkhatib. A Review of Machine Learning Techniques for the Classification and Detection of Breast Cancer from Medical Images. *Diagnostics (Basel, Switzerland)*, 13(14):2460, 2023. https://doi.org/10.3390/diagnostics13142460.

[11] Hieu Huy Pham, Hieu Nguyen Trung, and Ha Quy Nguyen. VinDr-Mammo: A large-scale benchmark dataset for computer-aided detection and diagnosis in

full-field digital mammography (version 1.0.0). *PhysioNet*, 2022. https://doi.org/10.13026/br2v-7517.

[12] Hieu Huy Pham Pham, Hieu Nguyen Trung, and Ha Quy Nguyen. VinDr-Mammo: A large-scale benchmark dataset for computer-aided detection and diagnosis in full-field digital mammography (version 1.0.0). *PhysioNet*, 2022. https://doi.org/10.13026/br2v-7517.

[13] Vijay Kumar and Priyanka Gupta. Importance of statistical measures in digital image processing. *International Journal of Emerging Technology and Advanced Engineering*, 2, 08 2012.

[14] Mohamed Meselhy Eltoukhy, Ibrahima Faye, and Brahim Belhaouari Samir. A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation. *Computers in Biology and Medicine*, 42(1):123–128, 2012.

[15] Z. Li, X. Li, C. Peng, W. Dai, H. Huang, X. Li, C. Xie, and J. Liang. The diagnostic performance of diffusion kurtosis imaging in the characterization of breast tumors: A meta-analysis. *Frontiers in oncology*, 10:575272, 2020.

[16] BF. Yan, A. Miyamoto, and E. Bruhwiler. Wavelet transform-based modal parameter identification considering uncertainty. *Journal of Sound and Vibration*, 291(1):285–301, 2006.

[17] A. Barragan-Montero, U. Javaid, G. Valdes, and et al. Artificial intelligence and machine learning for medical imaging: A technology review. *Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics AIFB*, 83:242–256, 2021. https://doi.org/10.1016/j.ejmp.2021.04.016.