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

March 3, 2024

St. John's Newfoundland

Abstract

Breast cancer is a major global health issue, that needs advanced and valid health-care methods and diagnosis to increase the patient's chances of survival. To allow researchers to facilitate in this space, we are going to conduct our research on early breast cancer detection using wavelet methods and the statistical composition of digital images. We employ DICOM-encapsulated mammography images and apply a wavelet modification to extract important features and statistical attributes from the pictures. By using these characteristics, a prediction algorithm can identify trends that point to the possibility of breast cancer.

The objective here is to create an understandable model for the diagnosis of breast cancer, which will be tested using different factors in many classes of classifiers using the statistics picked from the pictures of breast tissues. The study sample sets called the mammograms, detect specific patient attributes and aid in the examination of model accuracy across different clinical conditions and a wider range of patients. This novel method combines the advancement of image processing tools and machine learning with structured feature extraction to capture such breast tissue patterns that can be used to detect cancer. The model aims to be accurate, robust and straightforward, this is a step aimed at replacing current tools and techniques on breast cancer diagnosis, therefore early detection of breast cancer will be greatly facilitated.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Breast cancer is the most commonly diagnosed cancer when it comes to cancer deaths, and in women, second only to lung cancer [1]. It costs about one in three women diagnosed with cancer each year [2]. There are approximately 5500 people and 28,600 new deaths from breast cancer by 2022 in Canada. It is estimated that one in eight women will develop breast cancer in her lifetime, and one in thirty-four will die of disease [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 as ultrasound and mammography, call for a critical analysis of their shortcomings, which emphasizes the importance and urgency of developing new diagnostic tools. For the European guidelines on quality assurance in breast cancer diagnoses, there has been a need to review and strive for better screening and diagnosing techniques. Raising the detection techniques is an unavoidable need, so the patient's anxiety level can be reduced, and meaningless tests and false positive rates decrease, leading to improved lives and constant success of screening programs for breast cancer. [4]. This emphasizes how crucial it is to improve breast cancer detection techniques to lower the number of unwarranted recalls, false positives, and related patient distress.

1.2 Rationale for Research

For breast cancer detection, our initial goal is to narrow down the existing gaps and form a reliable data-transforming algorithm using Digital Imaging and Communications in Medicine (DICOM) files which can be achieved with the help of wavelet functions. Unlike claiming a new model, we utilize the existing technique by Michel Bernett [5] and apply it with different datasets to assist in better early breast cancer detection.

Our approach is implemented in two steps, the first step is to apply wavelet transfor-

mations to DICOM files, and the second is to use these wavelet features and extract statistical parameters from these images. This procedure will help us parse out important statistical features from the images. To sum up, in the data model that has been created, the retrieved statistical information is assigned as indispensable inputs to a simple machine learning model. The core purpose of this model is to find out with high accuracy if the patient has breast cancer or not.

We are going to evaluate the quality of our proposed methods by comparing its performance to existing techniques, aiming to demonstrate its capability to provide additional discriminative features for accurate predictions. Besides, we are also researching to investigate the discriminative features of DICOM images after wavelet transformation in connection with the context of breast cancer detection.

Moreover, data validation will not be the only competitive advantage of our approach. We would also like to find out the sensibility of the statistical attributes of wavelet-transformed DICOM images. The purpose of this is to find out the information these features can provide in this complex field of breast cancer detection.

1.3 The Significance of Wavelet-Based Approaches

Wavelet-based techniques are increasingly becoming the methods of choice in the field of research because they do a great job in several applications with one major use in statistical process monitoring being the fact that they are very effective [6]. Wavelet analysis is proven to be an effective tool because of its ability for data processing in-

volving different resolutions of scales. This helps solve problems like noise processing, autocorrelation, and the treatment of anomalous data [6]. The wavelet's performance in multivariate techniques, especially in the area of multiscale statistical process monitoring, adds another dimension of flexibility and effectiveness as these complicated data analysis processes are handled [6]. Furthermore, researchers have proved that wavelet analysis also serves well as a pre-treatment method for data cleansing as well as boosting the accuracy of subsequent statistical models [5].

Our research is aimed at discovering innovative methods based on wavelet techniques which will result in more accurate and fast detection of breast cancer, with applications of this achievement both in medicine and other fields. The approach we will use will consist of a combination of machine learning techniques, which will be enhanced by wavelet-based feature extraction. Our method intends to enhance the precision and capability of the model by applying wavelet transformations to DICOM images and getting statistical information about these changes.

1.4 Objectives of the Study

The most important part of our research is to prove that our suggested method is better than the ways people apply the methodology now. We therefore use the comparative technique which will show that the wavelet-based approach has an edge over others in terms of its effectiveness and power to distinguish those people who are most likely to have breast cancer.

The goal of our study is precisely to understand the importance of these statistical parameters obtained from the wavelet-transformed DICOM images and then apply such information to the breast cancer identification methods. The information that we are trying to achieve through our research is vital for the diagnosis and treatment of breast cancer. Thus, our research will achieve this with the ultimate goal of changing the method of cancer diagnosis and addressing some of the existing gaps in the understanding of how to detect breast cancers which results in improving the outcomes of the treatment and also the whole process of medical care.

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 [7]. 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 [7]. 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.

The recent breakthroughs in breast cancer detection in its early stages have been effective in the pursuit of these objectives. Technologies like DBT, which is digital breast tomosynthesis, an imaging technique based on limited-angle tomography, has been found to be the best solution for overcoming these problems [8]. DBT is based on collecting multiple projections views while the x-ray source is traversing along a predefined line, which make it capable of re-constructing sections that look parallel to the breast support.

2.2 Wavelet-Based Approaches in Breast Cancer Detection

The wavelet-based methods, for the good cause in diagnosing the breast cancer, are rapidly improving to provide a better interpretation of the mammogram images. According to a new study, it is possible to develop a new way to group mammograms of both benign and malignant breast cancer. Being especially observed with a research performed and published in Applied Sciences, the method suggested is multi-fractal dimension-oriented which is also feature fused, allowing a significant detection accuracy on INbreast, MIAS, DDSM and BCDR. [9]

The usefulness of wavelet-based methods is that they allow for the use of different resolutions and scales, thus enabling the selection of the most pertinent features from mammograms. However, these challenges include selecting the best wavelet functions and dealing with the computational complexity have to be resolved with care. The baseline of our investigation lies on these discoveries and uses wavelet analysis aim to enhance breast cancer detection models accuracy and interpretability.

2.3 Integration of Machine Learning with Wavelet Analysis

Recent investigations have considered the incorporation of machine learning algorithms with wavelet feature extraction for breast cancer diagnosis that is more exact. This mix is demonstrating its positive effect, as its diagnostic accuracy comes out to be higher than traditional screening methods. Feature-based machine learning models yielded improved sensitivity and specificity in detecting the faint signs which point towards malignancy in the wavelet-transformed images [10].

Support Vector Machine (SVM), Neural Networks, and deep learning models are some of the classifications of breast cancer lesions with machine learning methods that have shown some promise. The research by Jalloul et al. [11] illustrates the potential of machine learning to improve the diagnostic accuracy and shorten the time for detection of early diseases through the use of diverse machine learning methods on

medical images.

A synergistic approach is realized in the combination of the machine learning with wavelet analysis in that the two methodologies are integrated to complement each other. Applications that were shown to be successful, including those discussed by Jalloul et al. [11], present the integration as a way to revolutionize early breast cancer detection, by offering a more reliable and efficient diagnosis setup. The main goal of our research is to contribute to the existing large body of science by implementing a machine learning model along with wavelet-based features, thereby helping significantly in early breast cancer detection.

Chapter 3

Methodology

3.1 Datasets

In this part, the various datasets, which are used for training, testing, and validating the developed breast cancer detection models, are summarized. The employment of multiple datasets raises the resilience and portability of the proposed model, covering diverse demographics, imaging modalities, and annotation types.

3.1.1 VinDr-Mammo Dataset

The VinDr-Mammo dataset proposed by Pham et al. [12], crucial for advancing breast cancer detection research, is structured within a project directory that comprises DICOM files, annotations, and metadata across 5,000 mammography exams. Specifically, the dataset organization includes:

- images: A subfolder containing 5,000 subdirectories for each exam, with each subdirectory named after a hashed study identifier. Each contains four DICOM files representing two standard views (CC and MLO) of each breast.

- breast-level-annotations.csv: Provides BI-RADS assessment and image metadata for each image, including study-id, series-id, laterality (L or R), view-position (CC/MLO), image dimensions (height and width), breast-birads (BI-RADS assessment), breast-density, and the dataset split (training/test).

- finding-annotations.csv: Contains detailed annotations for breast abnormalities per image, including the metadata from the image and specific annotations like finding-categories, finding-birads (BI-RADS assessment of the finding), and bounding box coordinates (xmin, ymin, xmax, ymax).

- metadata.csv: Offers additional details relevant for research, such as the patient's age, and imaging device's model and manufacturer, provided by DICOM tags.

This dataset's relevance is not just because of its large size but the annotations which make it an excellent resource for building and validating breast cancer detection models especially in cases where interpretability is crucial.

3.1.2 RSNA Screening Mammography Breast Cancer Detection Dataset

The RSNA Screening Mammography Breast Cancer Detection dataset [13], is integral to the study's exploration of breast cancer identification from screening exams. The dataset contains radiographic breast images, and the primary goal is to identify cases of breast cancer in mammograms. This dataset provides a detailed framework for breast cancer detection through screening mammography, incorporating DICOM images and extensive metadata. The organization allows for nuanced analysis and model training:

- DICOM Images: Featuring mammograms, potentially in jpeg 2000 format, across roughly 8,000 patients. Each patient typically has four images.
- site-id: Identifier for the source hospital.
- patient-id: Unique code for the patient.
- image-id: Unique identifier for the image.
- laterality: Indicates if the image is of the left or right breast.
- view: Image orientation, with two views per breast being standard.
- age: Patient's age in years.
- implant: Indicates the presence of breast implants.

- density: Breast tissue density rating from A (least dense) to D (most dense).
- machine-id: Identifier for the imaging device.
- cancer: Malignancy status of the breast, with follow-up details like biopsy, invasive status, and BIRADS assessments provided for training data.

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.

3.1.3 INbreast Dataset

The INbreast dataset, introduced by Moreira et al. [14] in their technical report, is a valuable resource for breast cancer research and model development. This dataset contains 410 photos, that is 115 cases total in the INbreast archive which were used for the research in this essay. These ninety incidences involved women were with images (MLO and CC) of the two breasts being photographed that give us a total of four photographs for every individual case. These cases have 49 of them who had breast amputation, in which they are their only document of one of their breasts. DICOM format is the encoding used for the images in the collection that has characters regarding the information of the images, for instance, info related to equipment of capturing, size, voxel, color mode, quantity of bits, and so forth. The material on the

planning of mammogram by the database will be tested on different type of lesions, for instance masses, calcification, asymmetries, and distortions. As well, the database includes the segmentations of defects which are associated to medical subspecialty, such as imbalance, multi-directional growth, calcification etc.

The INbreast dataset's strengths lie in its focus on full-field digital mammograms, wide variability of cases, and the provision of precise annotations, making it a valuable asset for researchers aiming to develop and evaluate computer-aided detection and diagnosis systems for breast cancer.

3.2 Computational Environment

Python 3.9 served as the kernel language for the high-level wavelet-based breast cancer detection system. Tremendous speedup was achieved in the process of machine learning modeling and training by using NVIDIA GPU hardware, CUDA 12.2 and cuDNN 8.6 being the libraries for parallel computing and deep learning operations optimization. The strength of Python in scientific computing which is available in the form of NumPy, scikit-learn (sklearn), and pandas was critical in ensuring the numerical, visualization, and machine learning capabilities were in existence. Pydicom was utilized for reading DICOM files and extracting pixel data that were further used in the preprocessing phase. Conda virtual environments has been used to isolate project dependencies, achieved an identical and consistent results. This Python environment made it a quick, easy and agile prototyping and experimentation process

due to its user-friendly interface and many machine learning and computer vision libraries available.

Table 3.1: Versions of Python Main Libraries that Were Used

Library	Version
numpy	1.26.3
scipy	1.10.1
pylibjpeg	2.0.0
numpy	1.25.1
scikit-learn	1.1.0
pandas	2.2.0
pydicom	2.1.2

3.3 Preprocessing

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 parameter of statistical significance plays a significant role in the approach we take during the breast cancer identification. Our carefully chosen statistical indicators are based on the work of Kumar and Gupta [15], which is the purpose of designing a medical imaging method that fits the peculiarities of medical imaging. You can find in our list mean, mode, median, variance, standard deviation, covariance, skewness, and kurtosis among, which are important statistical metrics. Every parameter is selected with a specific goal in mind: to account for different ideas of DICOM picture intensity distributions which are focus of detecting breast cancer. By an attentive adjustment of our parameters to the information from Kumar and Gupta [15], we are able to have a guarantee that our methodology is not at random but rather fixed at the particulars of the given breast cancer imaging problems. The logical basis of the selected technique is diversified into its efficiency in detecting abnormalities, outlines, and specific features within breast tissue.

3.3.2 Significance of Selected Parameters

Wavelet and curvelet, which are the multiresolution representations, have been successfully used in image processing applications to zoom up and down on their underlying texture structure [16]. The statistical parameters give a rich picture of the fine details of the distributions intensities of pixel within DICOM images especially in the context of breast cancer detection. From the subsequent stanza all essential compo-

nents are highlighted, with each one outlined in detail to demonstrate its purpose in the manifestation of the appearance of breast tissue.

- Mean: The term, "mean" (measure of central tendency) implies the average pixel intensity in a DICOM image. In the aspect of breast cancer detection, the variations in the mean intensity provide vital information. Any abnormalities or deviations from the normal tissue pattern such as presence of irregularities or mass in the breast may be detected through the measurement of mean intensity whose value is different than the expected value. These changes track the tissue variations and help detect small, yet important tissue composition changes. An example of that would be, the mean intensity being elevated in some regions could be an indication of the mass being present which would be additional information for the diagnostics process to use.

- Standard Deviation: The standard deviation, a measure of dispersion of which the pixel intensities in a DICOM image reflect the level of variability, characterizes the spread of values in a DICOM image. In the context of mammography and breast cancer determination, standard deviation is very important in showing the pixel values consistency or variation. A larger standard deviation signifies a higher variability in response to frequency and is indicative of a possible microstructure deterioration. A greater standard deviation is likely to mean the existence of identified regions with high fluctuations in intensity level, which could very well signify the presence of abnormalities. Using the mean

and standard deviation provides a more complex knowledge of not only the central tendency and variability, but also the strength of the model in showing the subtle patterns specific to breast cancer.

- Skewness: Symmetry or Skewness describes which side of the pixel intensity distribution has more pixels. Skewness of data from a normal distribution suggests that there may be some irregularities in breast tissue, which gradually leads to a more thorough subsequent investigation.

- Kurtosis: The kurtosis as a statistical feature indicator detects the tail heaviness of the pixel intensity of the DICOM images. When searching for breast cancer on mammograms, the high research values represent outliers or distinct features. The contrast characteristics could help to identify the zones for which the closer inspection may be needed for exclusion of the abnormalities or the masses [17]. The diagnostic ability of kurtosis in mapping breast cancer cells has been investigated in the studies that used the diffusion kurtosis imaging technique (DKI) as the imaging technique [17].

3.3.3 Computational Implementation

The computational process involves two main steps: wavelet transformation and statistical parameter computation, described in the paper by Yan et al [18] . The wavelet transform, using a particular wavelet type (for example, 'haar') and decomposition levels, produce coefficients. These coefficients are then applied to derive the parame-

ters of interest at various stages of decomposition. Every DICOM file is followed by a text file where the statistical data and wavelet coefficients are stored.

The fact that these statistical parameters are able to encapsulate complex pixel intensity distributions makes them useful foundational information used in wavelet-based feature extraction, thereby allowing for 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 [19], the breast cancer detection models created and used in this study employed diverse machine learning algorithms: LogisticRegression, RandomForestClassifier, Support Vector Classifier (SVC) and DecisionTreeClassifier.

The development as well the validation of the breast cancer detection models was accomplished by being highly conscious to keep the utilization of the models fully reliable. The datasets had DICOM images of statistical features as its components. The train test split method from the sklearn Modular model selection module was used to to partition the data into training and test sets. This division allowed us to assess predictive power of models on the one hand and objective on the other: it is a crucial issue from the view of 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 4.3 provides different performance metrics for each model on both datasets, showcasing the distribution of True Positives, False Positives, False Negatives, and True Negatives.

Table 4.1: Performance Metrics - VinDr Dataset

Model	Accuracy	F1 Score	Sensitivity	Precision
Logistic Regression	0.9595	0.0	0.0	0.0
Random Forest Classifier	0.9605	0.1023	0.0556	0.6429
Support Vector Classifier	0.9595	0.0	0.0	0.0
Decision Tree Classifier	0.918	0.0939	0.1049	0.085

Table 4.2: Performance Metrics - INbreast Dataset

Model	Accuracy	F1 Score	Sensitivity	Precision
Logistic Regression	0.7111	0.8312	0.9143	0.7619
Random Forest Classifier	0.6444	0.7838	0.8286	0.7436
Support Vector Classifier	0.6889	0.8158	0.8857	0.7561
Decision Tree Classifier	0.5778	0.7246	0.7143	0.7353

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

In assessing machine learning models across both the Inbreast and Vindir datasets, a comprehensive understanding of their performances reveals limitations of using the wavelet features for breast cancer detection.

Logistic Regression demonstrates commendable accuracy, yet a closer inspection of its metrics reveals challenges in effectively identifying positive cases. The F1 Score, Sensitivity, and Precision all point to limitations in the model's ability to discern abnormal cases accurately. The Random Forest Classifier, as an ensemble method, exhibits resilience in capturing intricate patterns. However, the heightened number of false positives suggests a compromise in precision, emphasizing the inherent trade-offs

involved in its application.

Similarly, the Support Vector Classifier (SVC) shares similarities with Logistic Regression in grappling with sensitivity and precision. The model's struggle to distinguish between normal and abnormal cases reflects in its overall performance, characterized by a low F1 Score.

The Decision Tree Classifier, known for interpretability, falls short in overall performance. Despite providing insights into the decision-making process, the model's lower accuracy and F1 Score underscore challenges in comprehensively capturing the complexity of breast cancer patterns

4.3.2 Sensitivity and Specificity Analysis

Sensitivity (True Positive Rate) and specificity (True Negative Rate) play pivotal roles in evaluating the effectiveness of breast cancer detection models [20] across both the Inbreast and Vindir datasets.

Logistic Regression, despite its high specificity, exhibits a notable challenge in identifying positive cases, as evidenced by a sensitivity of 0.0. The model's proficiency in recognizing cases without breast cancer is contrasted by a struggle to capture instances of the disease.

The Random Forest Classifier strikes a balance between sensitivity and specificity, outperforming Logistic Regression in terms of identifying positive cases. However, the compromise in precision, indicated by an increased false positive rate, underscores the

model's inherent challenges.

Similar to Logistic Regression, the Support Vector Classifier (SVC) grapples with issues of both sensitivity and specificity, contributing to an overall suboptimal performance. The model's struggle to distinguish between normal and abnormal cases persists, echoing the challenges seen in Logistic Regression.

The Decision Tree Classifier, while showing a slightly improved sensitivity, continues to face challenges in accurately identifying positive cases. The model's interpretability comes at the cost of overall effectiveness in breast cancer detection

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 [21], 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 combined with wavelet based statistical parameters are not successfully 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] Cancer. *Public Health Agency of Canada*, March 2024. Accessed: 2024-03-02.

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

[6] 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.

[7] 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.

[8] 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.

[9] Diyar Akay Zebari, Dalia Abdulrahman Ibrahim, Diyar Qader Zeebaree, Mazin Abed Mohammed, Habibollah Haron, Nechirvan Asaad Zebari, Robertas Damaševičius, and Rytis Maskeliūnas. Breast Cancer Detection Using Mammogram Images with Improved Multi-Fractal Dimension Approach and Feature Fusion. *Applied Sciences*, 11(24):12122, 2021.

[10] 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.

[11] 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.

[12] 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.

[13] Chris Carr, Felipe Kitamura, George Partridge, inversion, Jayashree Kalpathy-Cramer, John Mongan, Katherine Andriole, Lavender, Maryam Vazirabad, Michelle Riopel, Robyn Ball, Sohier Dane, and Yan Chen. RSNA Screening Mammography Breast Cancer Detection, 2022. https://kaggle.com/competitions/rsna-breast-cancer-detection.

[14] Ines C. Moreira, Igor Amaral, and Ines et al. Domingues. INbreast: Toward a Full-field Digital Mammographic Database. *Academic Radiology*, 2011. https://doi:10.1016/j.acra.2011.09.014.

[15] Vijay Kumar and Priyanka Gupta. Importance of Statistical Measures in Digital Image Processing. *International Journal of Emerging Technology and Advanced Engineering*, 2, 08 2012.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] Noriaki Ohuchi, Akihiko Suzuki, Tomotaka Sobue, Masaaki Kawai, Seiichiro Yamamoto, Ying-Fang Zheng, and et al. Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): a randomised controlled trial. *The Lancet*, 387(10016):341–348, 2016.

[21] Samantha Forde, Robert Beardmore, Ivana Gudelj, et al. Understanding the limits to generalizability of experimental evolutionary models. *Nature*, 455:220–223, 2008.