NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF INFORMATION SYSTEM ENGINEERING

DETECTION OF BREAST CANCER USING MACHINE LEARNING

MSc. THESIS

TAMARAEBILAYEFA PIBOWEI

Nicosia

July, 2023

NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF INFORMATION SYSTEM ENGINEERING

DETECTION OF BREAST CANCER USING MACHINE LEARNING

M.Sc. THESIS

TAMARAEBILAYEFA PIBOWEI

Supervisor
Assist Prof. Dr. Kaan UYAR

Nicosia
July, 2023

APPROVAL

We certify that we have read the thesis submitted by Tamaraebilayefa PIBOWEI titled **"DETECTION OF BREAST CANCER USING MACHINE LEARNING"** and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Educational Sciences.

Examining Committee	Name-Surname	Signature
Head of the Committee:	Assist. Prof. Umit Ilhan	…………………………
Committee Member:	Assist. Prof. John Bush Idoko	…………………………
Supervisor:	Assist. Prof. Kaan Uyar	…………………………

Approved by the Head of the Department

…06/…12../20..23

……………………………….

Assoc. Prof. DR, Boran Sekeroglu

Head of Department

Approved by the Institute of Graduate Studies

(u.)

……/……/20…

Prof. Dr. Kemal Hüsnü Can Başer

Head of the Institute

DECLARATION

I hereby declare that all information, documents, analysis and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of Institute of Graduate Studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Tamaraebilayefa Pibowei

19/07/2023

ACKNOWLEDGMENTS

During the entire process of writing this thesis, I have been given a significant amount of help and assistance. As a starting point, I'd like to express my gratitude to Assist. Prof. Dr. Kaan Uyar, my supervisor, who gave so generously of his time and expertise in helping me complete this research project. In addition, I'd like to express my gratitude to my parents for the sound advice and understanding they've always provided. You are always there for me when I need you. In conclusion, I cannot thank my Wife, Mrs Esther Pibowei enough for her support, without which I would not have been able to finish this thesis. She helped me by providing engaging conversations as well as cheerful distractions, which allowed me to fulfil my commitments to my thesis.

Tamaraebilayefa Pibowei

ABSTRACT

Detection of breast cancer using machine learning

Pibowei Tamaraebilayefa
MA, Department of Information System Engineering
July 2023 84 pages

One of the common forms of female malignancy tumors is influenced by a number of clinical, lifestyle, social, and economic variables. Based on hidden patterns in data, machine learning has the ability to predict breast cancer (BC). A BC screening program cannot be successful without accurate risk assessment. More than 1 million women died from cancer globally 2022, with breast cancer accounting for 22.6 % of those deaths. In India, 14.7% of all cancer cases in women are BC, making it the most prevalent malignancy in women. A number of research have been done on early BC identification to aid in timely treatment initiation and lower mortality. Only 86% of patients who are diagnosed worldwide receive an accurate diagnosis. Cell biopsy images run the risk of raising false alarms that could put lives in jeopardy. Finding new, alternative techniques that are simple to use on various datasets, affordable, dependable, secure, and capable of making accurate predictions is urgently needed. Within the article, we provide a machine learning algorithm-based model.

Key Words: machine learning, random forest, k- nearest neighbour.

ÖZ

Detection Of Breast Cancer Using Machine Learning

Pibowei, Tamaraebilayefa

MA, Department of Information System Engineering

July 2023 84 pages

Kadın malignitelerinin en az yaygın biçimlerinden biri olan tümörler, bir dizi klinik, yaşam tarzı, sosyal ve ekonomik değişkenden etkilenir. Verilerdeki gizli kalıplara dayanan makine öğrenimi, meme kanserini tahmin etme yeteneğine sahiptir. Bir meme kanseri tarama programı, doğru risk değerlendirmesi yapılmadan başarılı olamaz. 2022 yılında dünya çapında bir milyondan fazla kadın kanserden dolayı hayatını kaybetti ve bu ölümlerin yüzde 22,6'sını meme kanseri oluşturdu. Hindistan'da, kadınlardaki tüm kanser vakalarının %14,7'si meme kanseri olup, bu da onu kadınlarda en sık görülen malignite yapmaktadır. Tedavinin zamanında başlatılmasına ve ölüm oranının düşürülmesine yardımcı olmak için erken meme kanseri tanımlaması üzerine bir dizi araştırma yapılmıştır. Dünya çapında teşhis konan hastaların sadece %86'sına doğru teşhis konuluyor. Hücre biyopsi görüntüleri, hayatları tehlikeye atabilecek yanlış alarmlar verme riski taşır. Çeşitli veri kümelerinde kullanımı basit, uygun maliyetli, güvenilir, güvenli ve doğru tahminler yapabilen yeni, alternatif tekniklerin bulunmasına acilen ihtiyaç vardır. Bu çalışmada, makine öğrenimi algoritmasına dayalı bir model sunulmuştur.

Anahtar kelimeler: machine learning, random forest, k- nearest neighbor,

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial intelligence
CNN	Convolutional Neural Network
DBT	Digital Breast Tomosynthessis
DL	Deep Learning
DT	Decision Tree
KNN	K Nearest Neighbor
ML	Machine Learning
MLP	Multi-Layer Perceptron
RF	Random Forest
SVM	Support Vector Machines
WHO	World Health Organization

CHAPTER I

Introduction

1.1 Background

The most frequent cancer is still BC among women. Modern medical imaging techniques and technology have dramatically lowered patient mortality and helped in detection of tumors in its initial stages. Since the great variety BC apart from that fibro glandular muscle, it is difficult to read and interpret breast pictures, which results in reduced sensitivity and specificity for cancer diagnosis as well as high inter-reader variability. Researchers have worked hard to create software to detect and evaluate breast imaging in order to solve these clinical issues, and to offer radiologists judgments. Recent quick developments in radiomics and deep learning as well as high-throughout data processing methodologies just as importantly Machine learning (ML) approaches have accelerated the creation of new breast imaging several versions with different uses. In present times, A World Health Organization (WHO) study claims that tumour is one of by far common diseases between women globally. BC is among one of the main reasons why women die. In India, prostate in women is greater than virtually every other group of people experience cancer in the world, and its about 14% death rate, which is incredibly high. When specific symptoms occur, it is typically simpler to identify BC. However, some BC patients experience no indications at all. Cancer can only be found, diagnosed, or treated if it is found early, early diagnosis lowers the risk of mortality that is crucial to the patient's survival. Cancer may spread more widely and be more challenging to cure if it receives a delayed diagnosis or late detection. According to the American Cancer Statistics Working Group, (2012), mammary tumor is one of the most frequent reason for lost of life for females specifically, following respiratory system diseases.

Conclusions drawn from the classification of cancer according to medical and histological evidence may be incorrect or incomplete. The rise and advancement of technology has created many challenges for the healthcare industry. The processing of enormous volumes a large amount (big data), complexity not only that a range various from healthcare services, the length of time required for diagnosis, and rising expenses in almost all areas will have an influence on the majority of healthcare professionals and patients. Numerous applications in this discipline, particularly in the field of medical diagnostics, heavily rely on ML techniques, enhances symptom recognition for more accurate disease diagnosis. Additionally, new diagnostic hypotheses can be proposed and more individualized treatment recommendations can be made using analysis of data (such as medical imaging). Many ML approaches, including KNN, distinctive perceptrons, a helper stroke machines, Neuromorphic methods, among others, are used in the classification of cancer. Early cancer research found a substantial correlation between the impact of a delayed cancer diagnosis and the disease progressing to an advanced stage and having a decreased chance of being cured. Several research have already shown that. This report's key focus is the evaluation and examination of the role played by several scanning for prostate cancer using techniques ML.

BC in British Columbia is anticipated to be the main reason for cancer-related deaths among women, according to Cancer Facts & Figures 2022 (June 2022). ML is a critical component of healthcare systems according to Javaid. et al (2022). ML techniques can more precisely detect early breast cancer according to (Amethiya, et al 2022). The diagnosis of diseases can benefit from ML classification algorithms. A successful diagnosis depends on knowledge, and patient data (Mirbabaie, Stieglitz, & Frick, N, 2021). ML-enabled technologies can help the healthcare industry address its physician shortfall. These tools can speed up the analysis of medical data and decrease the possibility of mistakes being made by unskilled staff. In supervised ML, there are various classification models. Some are non-linear, whereas others are linear. The steps of creating a system for machine learning known as test data are crucial. To hone and evaluate the system a dataset is necessary. Many scientists worldwide are working to lower the fatalities related to breast cancer, early detection and treatment.

The deadliest illnesses harming women globally is BC. It happens as a result of the breast tissue's unchecked cell growth. Histopathological data-based BC diagnosis can produce unreliable results. Machine learning (ML) methods have become widely employed throughout the decade before to diagnose BC, assisting pathologists and medical professionals with early identification, choice-making, and efficient treatment planning. The most recent US cancer data predict done at 2022, breast cancer would account for 31% of all malignancies reported in women, killing 43,250 people. 15% of all cancer-related deaths are attributable to it (Siegel, Miller, Fuchs & Jemal, 2022). As a result mammary cancer continues to also be the disease in which women are most frequently diagnosed and has the second- greatest chance of fatality. Population-based breast cancer veiling has contributed significantly to the early identification and decline in breast cancer mortality during the past 30 years. Deaths from breast cancer decreased by 40% between 1989 and 2017. This translates into 375,900 fewer deaths from breast cancer (DeSantis, Ma, Gaudet, Newman, & Miller, 2019). The rate of deterioration reduced from 1.9% per year between 1998 and 2011 to 1.3% per year between 2011 and 2017, yet mortality is still falling (DeSantis, Ma, Gaudet, Newman, & Miller 2019). However, because of the low cancer incidence (0.3%), low detection rates, and high false-positive rates in annual breast screening, society melanoma screening's efficacy is still debatable (Berlin, 2010). This high rate of falsified tests reveals a increase number of pointless tests, which are performed elsewhere a financial strain on healthcare systems but also causes unwarranted patient concern, which in turn encourages women to undertake routine breast cancer screening. are frequently less likely to go through screening (McCann, Stockton, & Godward, 2002). Its effectiveness in lowering mortality is currently the subject of heated dispute (Gøtzsche, 2015). For instance, ductal carcinoma in situ (DCIS), an early-stage intrusive cancer that does not spread or endanger the patient, is detected more frequently than incurable cancer. This is known as overdiagnosis, and it frequently results in needless treatments that could actually worsen the malignancy (Brennan & Houssami). Therefore, enhancing the effectiveness of breast cancer screenings/or diagnosis continues to be a very pressing global public health concern (Wilkinson & Gathani, 2022). Combining ML and AI can increase consequently identification of BC while minimizing overtreatment. However, precise predictions and choices can be made by combining AI and ML techniques. For instance, to recognize BC and decide whether a patient needs surgery based on the results of a biopsy. The most widely used diagnostic today, mammograms is a great risk of erroneous good results that might result in unnecessary biopsies and treatments. When cancerous cells are surgically removed, it's possible that the removed cells be benign but not cancerous. This implies that people have expensive, unneeded, and inconvenient operations.

With regard to health-related datasets, such as pictures, X-rays, and blood samples, ML algorithms have various benefits. While some tactics work best with small datasets, others work best with large datasets. Some approaches may have problems with noise.

ML has demonstrated encouraging results very soon diagnosis of breast cancer, particularly when combined using cancer screening other diagnostic imaging methods. Large datasets of mammograms can be applied to teach machines to learn algorithms to find patterns and characteristics linked to breast cancer. The analysis of fresh mammograms by algorithms can then be used to determine a person's risk for breast cancer. Medical personnel to create better notified judgments regarding patient safety thanks to the method' high accuracy. ML technologies are fundamentally altering the health sector. The goal of ML subclass about computer vision, is to grow better the effectiveness not only that accuracy of medical interventions.

Many quantitative tools that enable machines to acquire knowledge from data without becoming pattern recognition are referred to as “machine learning” (ML). The health industry is one of the biggest global sectors that potentially benefit from this technology (Abdelaziz, Elhoseny, Salama, & Riad, 2018), (Char, Abra’moff, & Feudtner, 2020), (Ahmad, Eckert & Teredesai, 2018). Over the past century, technological advancements have greatly extended life expectancy. Over the past century, technological advancements have greatly extended life expectancy. Although technology has come a long way since the past, modern breakthroughs like cognitive computing and computer vision has capacity to usher in a new era of healthcare of fact, also the simplest and the least important aspects of the operation can be virtually eliminated with the use of computers. Training a neurons on a dataset of mammograms is a typical method for detecting BC using ML. Convolutional Neural Network may be trained to recognize characteristics like complexes and microcalcifications, which are typical breast cancer red flags. This network may be applied to correctly identify regular and unusual new mammography. Another option is to utilize ML to analyze the biopsy image and determine the breast cancers' malignancy. This helps medical practitioners in choosing the most effective course of action for every patient. There is capacity for machine learning to greatly improve tumor detection accuracy and efficacy. It's indeed clear ML will greatly improve the accuracy and profitability of cancer detection, despite the need for greater study and improvement. With this technology, the possibilities are virtually limitless. The use of cutting-edge ML applications is improving the healthcare sector. Even more value is added to this process by ML tools in public health systems and primary/tertiary patient care, they raise the standard of automation and intelligent decision-making. There are several ways that ML technology can be used to improve clinical trial research. Medical experts can examine a wider number of records and cut the cost and time of medical testing by using advanced prediction modeling for potential clinical research participants. The effectiveness of clinical trials is further increased by a variety of ML applications. Making use of electronic health records to assist in choosing the ideal sample size in order to improve efficiency and lower the risk of data related problems. This strategy resolves a key problem in the medical field brought on by the doctors with the requisite training are in scarce supply worldwide. Combining individual health with predictive analytics can result in more dynamic and effective tailored care (Chen, Joshi, Ghassemi, Ranganath, 2020) (Siddique & Chow, 2021), which is one more advantage of ML in medicine (Waring, Lindvall & Umeton, 2020).

In both medical and academic research, there are numerous alternate uses for ML. several graphs, including those from press, past hospital procedures platforms, etc., are inputs used by researchers to act utilizing ML-based prediction studies to find possible participants in clinical trials. Additionally, it supports ideal sample sizes for investigations, makes use of the power of electronics, and lowers database mistakes so that experimental staff members may recognize and identify patterns and anomalies (Ahmad, Patel, Eckert, Kumar) (Manogaran & Lopez, 2018). Similar to how 1 very qualified physician spot suspicious body grafts, tumors, and brain hemorrhages, ML algorithms can assess image data. Bhardwaj, et al. (2017), anticipated a significant increase in the use of these platforms to assist radiologists (Roth, et al, 2018).

Another area where ML is anticipated to have a significant impact is research. Clinical studies can take years and are quite expensive. By applying ML-based predictive analytics to discover individuals from different data sources like social media, previous doctor visits, etc., researchers can decrease the pool of possible clinical trial participants. Observing study subjects in real-time is another method for applying ML in this situation. Using electronic records instead of paper records, these technologies may aid scientists in identifying the best sample size for testing (Kushwaha et al., 2020). (Rudin & Ustun, 2018).

The significant decrease in breast cancer fatality has been largely attributed to improvements in pharmaceutical imaging technology and research into one deeper comprehension of the complicated organic and synthetic nature of breast cancer. However, BC is an intricate and charismatic processes, making the battle against cancer a challenging path with many obstacles to overcome. The monitoring and control of cancer involves a number of steps, such as the identification the dubious tumors, the assessment even those tumors as cancerous, the stage of a cancer's classification, histology, and prognosis, the development of a good therapeutic intervention, the identification of tumor therapeutic excision boundaries, meaning analysis and the forecasting of chemotherapeutic or radiotherapy response treatments, or the forecasting of long term risk of event, reoccurrence. The selection procedure for all of them clinical steps is greatly influenced by medical imaging in this pipeline. In this therapeutic pipeline, the method used to make decisions for each of these jobs is greatly influenced by medical imaging. To identify suspect growths, gauge the possibility of cancer, especially assess the assessment for cancer, radiologists traditionally use Visual representation of subjective or quasi data gleaned due to clinical pictures.

The information that is clinically significant the frequency and extent of suspected tumors, enhancing patterns, the presence or absence of necrosis or bleeding, tumor perimeter margin are sometimes, or an area of suspected cancer. Although, it is not an easy process to evaluate and combine visually acquired data from curative photos to arrive at final diagnostic conclusions. However, it is not an easy process to evaluate and combine visually acquired data from curative photos to arrive at final diagnostic conclusions. Improvements in pharmaceutical imaging technologies and research into a greater comprehension of the intricate organic and synthetic basis of BC have been largely credited for the significant drop in breast cancer fatalities. However, because breast cancer is a complicated and dramatic disease, managing it is a challenging path with many obstacles along the way. The process of identifying suspicious tumors, determining whether they are cancerous or benign, diagnosing those tumors, creating an ideal treatment strategy, determining the surgery's tumor borders during surgical treatment, analyzing and foretelling the response to chemotherapy or radioactivity treatments, or forecasting the likelihood of an upcoming event or recurrence are all steps in rapid disease detection and therapy. Medical imaging that is pipeline has a significant impact on the judgments made for each of these clinical phases.

The rest of this work is organized as follows: Section II gives a selection of earlier research work carried out by others. Section III describes the methodology. Section IV shows the results. Findings, Discussions, and Limitations are included in Section V and conclusion is presented in Section VI.

1.1.1 Purpose of the study

The purpose of this study is to develop and validate effective algorithms and models that can assist in the early and accurate identification of BC. The overaraching goal is to improve medical outcomes, increase survival rates, and enhance patient care through the application of ML. Here are some specific purpose and objective of the study:

I. **Early Detection:** Detecting breast cancer at an early stage greatly improves treatment success and patient prognosis. ML can help identify subtle patterns that might not be easily discernible by human experts, enabling earlier intervention.

II. **Accuracy Improvement:** Traditional diagnostic methods can sometimes yield false positives or false negatives. ML algorithms aim to achieve higher accuracy by learning from a diverse and comprehensive dataset.

III. **Risk Assessment:** Develop models that assess an individual's risk of developing breast cancer based on various factors, such as medical history, genetic markers, and lifestyle.

IV. **Personalized Medicine:** Tailoring treatment plans based on a patient's specific characteristics and the predicted behavior of their cancer. ML can assist in recommending optimal treatments for individual patients.

V. **Reducing Biopsy Rates:** By accurately identifying cases that are more likely benign, ML can help reduce unnecessary biopsies, thereby minimizing patient discomfort and healthcare costs.

VI. **Automated Analysis:** Automating the analysis of mammograms and other medical images can increase efficiency and consistency in diagnosis, freeing up radiologists' time for more complex cases.

VII. **Handling Large Data:** Modern medical databases are vast and complex. ML techniques are capable of handling and analyzing these large datasets to extract meaningful information.

VIII. **Feature Extraction:** Discovering subtle features and patterns in medical images or data that may not be immediately apparent to human observers.

IX. **Continuous Learning:** ML models can continuously improve as new data becomes available, adapting to changes in the disease's presentation and characteristics over time.

X. **Assisting Medical Professionals:** ML algorithms can serve as decision support tools for healthcare providers, providing additional insights and information to aid in clinical decision-making.

XI. **Interpretable AI:** Developing models that provide explanations for their predictions, allowing medical professionals to understand and trust the decisions made by the algorithm.

XII. **Validation and Benchmarking:** Rigorously evaluating the developed models against established benchmarks and clinical standards to ensure their reliability and effectiveness.

XIII. **Ethical Considerations:** Exploring the ethical implications of using ML in healthcare, including patient privacy, data security, and potential biases in algorithmic decisions.

The ultimate purpose of this study is to bridge the gap between cutting-edge technology and medical practice, enabling more accurate, timely, and efficient detection and treatment for BC. Collaboration between data scientists, machine learning experts, and medical professionals is crucial to ensure that the developed solutions are practical, ethical, and aligned with the needs of the medical field.

1.2 Problem Statament

Breast cancer is a significant global health concern, and early detection plays a crucial role in improving patient outcomes. However, the current diagnostic methods, while effective, can benefit from enhanced accuracy and efficiency. The objective of this study is to develop and validate a machine learning-based model that can accurately classify mammogram images as malignant or benign with improved accuracy, thereby aiding medical professionals in timely and reliable breast cancer detection. The ML method that is being suggested can forecast BC. Through effective care at the correct time, early discovery of this condition will aid in slowing the disease's course and reducing death. One of the most prevalent and fatal illnesses impacting women globally is breast cancer. For patients' survival and outcomes to be improved, early Prostate cancer identification and therapy are crucial. Manual tumor screening, but is inaccurate, moment, and susceptible to mistakes. By automating the screening process and lowering reliance on individual analysis, machine learning algorithms having the capacity to increase the efficacy as well as that diagnosis of breast cancer with precision. However, several challenges when using ML techniques to identify breast cancer, such as:

I. **Data variability:**

Breast tissue looks very different on mammograms, making it difficult to develop an algorithm that can accurately detect cancer in all cases

II. **Data imbalance:**

The A training dataset's proportion of affirmative to negative instances can significantly affect how well machine learning methods perform. The amount of incorrect data in the identification of breast cancer is significantly larger than the number of samples tested, which can result in biased systems that are not particularly good at identifying cancer.

III. **Data privacy:**

The sensitivity of medical imaging data raises significant privacy and security concerns when using methods for machine learning to find breast cancer.

IV. **Algorithm Interpretability:**

Because many machine learning algorithms, including deep learning models, are difficult to solve, it can be difficult to accept how decisions are made as well as to spot and correct possible bias. The accuracy, efficacy, and interpretability of techniques for machine learning are used to identify cancerous tumors in breasts being improved by researchers and practitioners by experimenting with novel methodologies like transfer learning, data augmentation, and explainable AI.

1.3 Risk factors

General Risk Factors:

I. **Age:** Risk increases as aging.

The greatest major vulnerability component is a woman's gender.

II. **Race:** On average, much more probably to be white ladies than black ladies to acquire tumor. Ladies of Asian, Hispanic, and Native American descent are less likely to acquire breast cancer and pass away from it.

Genetic Risk Factors

I. **Family history:** The risk may be increased for when there is a familial connection to the illness, breast cancer may arise in a mom, sister, or child.

II. **Factors that are inherited:** Genetic variations (alterations) within some. The BRCA1 as well as BRCA2 mutations may raise the danger.

III. **Weight:** Adipose epithelium could assist in enhance estradiol concentrations after menopause, just as importantly higher estrogen levels may increase how often tumor occurs. A part seems to be also played through adult body growth then there is excess body fat around the waist.

IV. **Childlessness:** Breast cancer risk may be higher for women who never have children or who get pregnant later in life. Breastfeeding lowers the chance of developing breast cancer.

V. Breast tissue that is more glandular and less fatty is said to be dense. Breast cancer risk is increased females having thick breast cells.

VI. **Monthly flow:** Tumor danger is marginally elevated for females whose menstruate early (prior age 12) and/or experience a late menopause (beyond age 55). The elevated risk could be brought on by a lifetime of estrogen and progesterone exposure.

Lifestyle Risk Elements

I. sedentary kind the lifetime.

II. The likelihood of cervical tumor is reduced by physical activity.

III. **Alcohol:** Drinking consuming liquor may a higher danger of breast cancer. The danger rises as alcohol consumption does.

Previous therapy

I. **Administration of DES:** Pregnant ladies whom use medicine DES (diethylstilbestrol) have a marginally higher danger of tumor.

II. **Utilization of oestrogen treatment after menopause:** After menopause, using Glycine and oestrogen raises the danger of breast tumor.

III. **Radiation exposure:** Breast cancer risk is significantly increased in women (Toddlers or grownups) whom have undergone radiation counseling for another type of cancer.

1.4 Objective Of The Study

The objectives of this study focused on breast cancer detection using machine learning. Outlines the specific goals and outcomes the aim to achieve. These objectives help guide the study's design, methodology, and evaluation criteria. Cancer analysis aims to eradicate the disease so that you can lead a normal life. Depending on the specific circumstances, this might or might not be achievable. If a cure is not possible, treatment can be offered to reduce the cancer's size or limit its progress, allowing patients to experience symptom-free living for as long as possible. Here's a sample set of objectives for a study on breast cancer detection using machine learning:

I. **Develop Accurate Classification Models:** Design and implement ML algorithms that can accurately classify mammogram images as either malignant or benign, surpassing the performance of existing diagnostic methods.

II. **Enhance Early Detection:** Create models that can identify subtle signs of breast cancer at an early stage, allowing for timely interventions and improved patient outcomes.

III. **Improve Sensitivity and Specificity:** Strive for models that achieve high sensitivity (detecting true positives) and specificity (minimizing false positives) to ensure accurate and reliable diagnoses.

IV. **Utilize Diverse Dataset:** Curate and preprocess a comprehensive dataset of mammogram images representing diverse patient.

V. **Feature Extraction and Selection:** Explore advanced image processing techniques to extract relevant features from mammograms, ensuring that the chosen features are both discriminative and interpretable.

VI. **Compare and Benchmark:** Compare the performance of different machine learning algorithms, providing insights into which models are most effective for breast cancer detection.

VII. **Address Imbalanced Data:** Investigate methods to handle class imbalance in the dataset, ensuring that the model's performance is not biased towards the majority class.

VIII. **Provide Interpretability:** Implement techniques that offer insights into the model's decision-making process, making the predictions more understandable and trustworthy for medical professionals.

IX. **Validate Clinical Applicability:** Evaluate the developed models using clinically relevant metrics and standards, demonstrating their potential real-world impact in a medical setting.

X. **Ethical Considerations:** Address ethical concerns related to patient privacy, data security, and potential biases in the machine learning algorithms.

XI. **Optimize Model Parameters:** Perform hyperparameter tuning to fine-tune model parameters, aiming to maximize accuracy while preventing overfitting.

XII. **External Validation:** Validate the trained models on an external dataset to assess their generalization capability and robustness.

XIII. **Contribution to Medical Field:** Contribute insights and knowledge to the medical community by demonstrating the capabilities of machine learning in breast cancer detection, potentially inspiring further research and collaboration.

By setting clear and measurable objectives, it will help guide their study's progress, ensure that the efforts are aligned with the goals, and provide a basis for evaluating the success and impact of the work in the field of breast cancer detection using ML.

1.5 Research Question

- How can medical pictures like mammograms and ultrasound utilizing scans and algorithms for machine learning, prostate cancer can be accurately detected?
- In comparison to employing a single algorithm, may mixing different ML techniques improve BC predictive performance disclosure?
- How is the quantity and variety of the datasets used to train machine learning models has an impact on the accuracy of breast cancer diagnosis?
- Can machine learning algorithms effectively enhance the accuracy of breast cancer detection through the analysis of mammogram images, and if so, which specific algorithms and techniques yield the best results?

Approach:

Data Collection and Preprocessing:

Gather a diverse and comprehensive dataset of mammogram images, ensuring representation of both malignant and benign cases. Preprocess the images to standardize size, format, and quality, while maintaining relevant clinical information.

Feature Extraction and Selection:

Explore image processing techniques to extract meaningful features from the mammograms, such as texture, shape, and intensity information. Investigate methods for selecting the most relevant features to improve model efficiency and performance.

Algorithm Selection:

Experiment with a range of ML algorithms suitable for classification tasks, including support vector machines (SVM), random forests, multilayer perceptron (MLP), and k-nearest neighbor. Consider ensembling techniques to combine the strengths of multiple algorithms.

Handling Class Imbalance:

Implement strategies to address class imbalance, such as oversampling, undersampling, or generating synthetic samples, to prevent bias in favor of the majority class.

Model Training and Validation:

Split the dataset into training, validation, and test sets.

Train different machine learning models on the training set and fine-tune their hyperparameters using the validation set.

Model Evaluation:

Evaluate the trained models using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). Compare the performance of different models and techniques to identify the most effective approaches.

Interpretability and Visualization:

Utilize techniques such as feature importance analysis and saliency maps to provide insights into the models' decision-making processes, enhancing their interpretability.

External Validation:

Validate the best-performing models on an external dataset to assess their generalization capability and real-world applicability.

Ethical Considerations:

Address ethical concerns related to patient privacy, data security, and potential biases in the models' predictions.

Clinical Validation:

Collaborate with medical professionals to validate the models' results and assess their clinical relevance and impact.

Communication and Dissemination:

Present the findings through research papers, presentations, and discussions with the medical community to contribute valuable insights to BC detection practices. By following this approach, systematically explore various machine learning algorithms and techniques to address the research question and provide valuable insights into enhancing breast cancer detection using ML.

1.6 Significant For Research

For medication and timely identification of an affliction, finding mammary tumor is a significant problem. The identification of mammary tumor and prevention have both benefited greatly from machine learning. Using a lot of data, machine learning algorithms can figure out patterns that are challenging for human specialists to spot. Transfer learning for screening mammography has been investigated in a number of studies. These findings show that, in some cases, automated analysis techniques can categorize breast tumors as benign or cancerous with greater accuracy than human specialists.

The capacity to integrate a large number of characteristics in the analysis Among the important benefits of utilizing machine learning to diagnose breast cancer. To establish a more accurate diagnosis, for instance, machine learning algorithms can examine the size, shape, and texture of breast lesions as well as patient characteristics and medical history.

The fact that ML algorithms get better over time as more data is fed into them is another benefit of employing them to find breast cancer. This enhances the algorithms' accuracy over time and gives patients more trustworthy outcomes. In conclusion, employing ML to identify BC offers enormous potential to enhance the disease's early identification and care. Early results are encouraging, and the technology could have a considerable impact on the medical treatment of breast cancer patients. nevertheless, additional investigation is essential to properly comprehend what possibilities and limitations of machine learning in this context.

This Research focus on BC detection using machine learning is significant due to its potential to revolutionize early diagnosis and treatment strategies. Here are some reasons why this research is significant:

I. **Improved Early Detection:** Early detection of BC is crucial for successful treatment and improved patient outcomes. ML algorithms can potentially identify subtle patterns and anomalies in medical images that might not be easily noticeable to human observers, leading to earlier interventions.

II. **Enhanced Accuracy:** ML models have the capability to analyze complex and large datasets, leading to potentially higher accuracy rates in distinguishing between malignant and benign cases. This can reduce false positives and false negatives in diagnosis, resulting in more reliable results.

III. **Efficiency in Diagnosis:** Automated analysis of mammogram images through ML can speed up the diagnostic process, allowing radiologists and medical professionals to focus on more complex cases and patient care.

IV. **Personalized Medicine:** ML can help in the development of personalized treatment plans based on a patient's unique characteristics, medical history, and genetic factors, leading to more tailored and effective therapies.

V. **Reduced Biopsy Rates:** By accurately classifying cases as benign, ML can help reduce unnecessary biopsies, minimizing patient discomfort and healthcare costs.

VI. **Handling Data Complexity:** Mammogram images are complex and can contain subtle patterns that are challenging to interpret manually. ML algorithms can handle this complexity and learn intricate relationships within the data.

VII. **Exploring New Features:** ML can identify novel features and markers in medical images that may not have been previously considered by medical professionals, leading to potentially new insights in breast cancer diagnosis.

VIII. **Translational Research:** Bridging the gap between technology and medical practice, this research has the potential to directly impact patient care by providing tools that can be integrated into clinical workflows.

IX. **Interpretability and Explainability:** Research in this area can also focus on making machine learning models more interpretable, allowing medical professionals to understand the rationale behind the models' decisions.

X. **Leveraging Big Data:** Modern healthcare generates massive amounts of data. Machine learning can leverage this data to extract meaningful insights, contributing to advancements in medical research and patient care.

XI. Collaboration between Fields: BC detection using ML requires collaboration between medical professionals, radiologists, oncologists, and data scientists, fostering interdisciplinary communication and learning.

XII. Ethical and Social Impact: This research addresses ethical concerns related to patient privacy, data security, and algorithmic biases, encouraging the responsible use of technology in healthcare.

XIII. **Potential for Scalability:** Successful ML models can potentially be scaled to different healthcare settings, making BC detection more accessible and effective globally.

In summary, this research in breast cancer detection using machine learning has the potential to significantly impact the medical field by improving early detection, accuracy, and patient outcomes. It also demonstrates the synergy between advanced technology and medical expertise, setting the stage for future advancements in medical diagnosis and treatment.

1.7 Limitations

There are some restrictions on utilizing ML to detect tumor:

I. **Data bias apart from that imbalance**: The nature and volume of information utilized in education heavily influence exactly precise of machine learning models. Unbalanced or biased training data can produce a biased model that is unable to properly distinguish among various populations' occurrences of breast tumor.

II. **Restricted Illustratable**: Many machine learning algorithms are opaque and can be challenging to understand, despite the fact that they can generate precise predictions. This might make them less useful in medical settings where knowing the basis for a prognosis is vital.

III. **Rates of false positives and false negatives**

Breast tumour using algorithms that learn is constrained by frequencies of false positives and false negatives, just as other diagnostic methods. False positive results can result in pointless biopsies and patient anguish, whilst false negative results can cause a delayed diagnosis and poorer outcomes.

IV. **Oversampling:**

When a model is too specific to the training data to generalize to fresh, untried data, the issue of overfitting arises. This can reduce the model's usefulness in real-world settings and result in subpar performance on independent test sets. Difficulty of medical picture data:

Medical imaging data from mammograms is frequently used in ML the detection of breast cancer. To effectively prepare this data for analysis, which can be complicated and challenging to read, requires specialist knowledge and expertise. It is significant to emphasize that notwithstanding these drawbacks, machine learning continues to hold promise as just a means of the prompt breast cancer detection, and research to overcome these issues is underway.

1.8 Definition Of Terms

breast cancer

A specific type of tumor which appears in mammary gland.

machine learning

A Machine learning variant which enables a machine to gain knowledge from data and improve its efficiency before having explicitly designed.

CHAPTER II

Literature Review

2.1 Introduction

A literature review's primary goals are to summarize, assess, and synthesize previous findings and data on the subject at hand. It necessitates painstakingly reading, evaluating, and interpreting the pertinent literature in order to get a deeper understanding of the subject, spot research gaps, and present the background and theoretical underpinnings of a new enquiry.

2.2 Theoretical Framework

Information regarding prior relevant research that has been done is provided in this section. ML is a topic of extensive investigation, however, DL address some issues that machine learning methods cannot. Information on the intersection of ML and DL methods is provided in this section. A prove to show using a machine learning and layered system was developed by Megha & Vikas, (2016). Four classifiers and compared SVM, Naive Bays, Function Tree, and End Meta. SVM proves to be a useful Information regarding prior relevant research that has been done is provided in this section. In essence, he employs two methods to find breast cancer. Deep learning is the second, and machine learning is the first. Machine learning is a topic of extensive investigation.

A simulation built using a machine learning and mixed algorithm was developed by Megha & Vikas, (2016).

Other machine learning-based mixed model was proposed by Tahmooresi, et al, (2018) for better outcomes. As a result, SVM is a good classifier and offers the highest accuracy. SVM, ANN, and decision trees were contrasted. used on blood and image sets of data. The ML model outcome was proposed by (Muhammet, et al, (2018), but it made use of a different classifier. The author employed the Extreme Learning Machine (ELMA), SVM, KNN, and ANN as classifiers. The classifier had a minor adjustment to get improved outcomes. Anusha, Pooja, and Anishka, (2018) proposed a machine learning-based model. He employed SVM, Decision Tree (CART), ANN, and Naive Bayes as the four classifiers. The author claimed that ANN was more precise. CNN used Mobile Net and Inception V3 among the numerous models that were modified for it. Inception V3 was discovered to be more accurate when the authors compared the two models. Although the datasets were vast and machine learning approaches did not produce superior outcomes, the author also discussed them.

Deep learning was therefore applied to this problem.SVM has some shortcomings. SVM produced fantastic outcomes for binary variables. Multi-SVM was thus employed. Machine learning techniques were compared by Ebru, Pinar, TolgaEnsari, (2019). According to the author, SVM generated matrix findings with improved performance. To solve machine learning issues, the Deep His Learning approach was created. Shwetha, et al, (2018) suggested a convolutional neural network using a deep learning framework paradigm. A number of models were corrected by CNN using Mobile Net and Inception V3. The author's two models were put side by side, and Inception V3 came out on top. However, there was still room for machine learning in breast cancer treatment. A supervised machine learning model was proposed by Shravya, Pravalika, & ShaikSubhani, (2019). ANN, SVM, and logistic regression classifiers have all used this work as a basis. This UCI repository's information was downloaded, even better performance test was used to evaluate the findings. They claim that SVM, which operates on the Python platform, is a good classifier, providing an accuracy of 92.7%. Using several classifiers, Sivapriya, Aravind, Siddarth & Sriram, (2019) created a machine learning model. SVM, Logistic Regression, Random Forest, and Naive Bay were employed by the authors. It was incorporated into his Python Anaconda platform. With a reported accuracy of 99.76%, the authors determined that Random Forest was a decent predictor in regards to efficiency. To increase the accuracy of some bit changes in the network, a classifier could be utilized. An ANN-based model was suggested in this article Kalyani, Prashant, & Nikhil, (2019), article and the figure's effectiveness was assessed via an SVM classifier. The ANN's accuracy was 97.0%, and the SVM's was 91%, according to the author. The accuracy increased even without SVM, according to the authors. He put out his model-based SVM and grid search by Vishal & Mukta, (2019). The author used grid search after first utilizing SVM to conduct his study. The author compared them all and chose the best one. The comparison led to the creation of a new model. A model based on k-means GMM and CNN was proposed by Shamy and Dheeba (2019) for search algorithm to help optimize. The method of texture feature extraction is used after the author first calculates the ROI. The author's accuracy score for him was 95.8%.. Sansya & Lekshmy, (2019) proposed a system that utilizes deeper learning. They focused on Lloyd's CNN for classification and his clustering approach. The suggested procedure produced a precision of 96%. For diagnosis, histopathological imaging was employed. Deep learning and image processing were also discussed. A model for deep learning-based image improvement of histopathology images was proposed by Puspanjali, Baldev, and Samikshya, (2019). This paper made use of a variety of feature extraction algorithms, including PCA and LDA. Although the datasets were vast and machine learning approaches did not produce superior outcomes, the author also discussed them. Deep learning was therefore applied to this problem. On CNN, its accuracy was 81.0%. However, the accuracy increased by up to 89% when the images were trained on GPUs. a deep recursive neural network-based approach to IDC forecasting networks was proposed by (Chandra & Gopal, 2019). The authors used histopathology pictures as their dataset. With an AUC score of 0.9996, accuracy was 99.29% their approach. Canh, Anh, & BaoThien, (2019) put forth a deep learning-based model that improves accuracy by adding to the dataset. Dhivya & Dharani, (2019) used neural networks to carry out a survey on breast cancer diagnosis. They investigated a variety of methods during their research and discovered that machine learning algorithms increase the system's accuracy. On the basis of analyses, Ayush, Bhawna, & Kaushik, (2019) suggested a model. This model was built using deep learning methods with computer assistance. This work offered a concise summary of all current deep learning trends. Sornam, Kavita, & Vanitha, (2017) carried out a deep learning-based image identification experiment. The salient characteristics of deep learning applications were emphasized. It presented crucial details on all fronts and demonstrated why deep learning her program produced superior outcomes. Andrik, Bryan, and Hui (2018) worked on classification of breast masses using diagnostic techniques like mammography. For classification, a cnn was employed. Modifying AlexNet is effective, and based on their findings, a straightforward change might produce superior outcomes. They made use of the PReLu activation function, which produced superior outcomes to Relu. Using a BC decision support and information management system, the authors worked with previously published data. Prof. Dr. Cemil, Seyma, and Ahmet, (2019), used the CBISDDSM dataset. A study was conducted to classify breast cancer. They compared every CNN model and discovered that they were incredibly accurate. The Inception

Recurrent Residual Convolutional Neural Network hence produced better outcomes. Programming in R was used for the work. A study based on ad hoc data acquisition techniques deep learning-based was published by Yawen, Jun, Zongli, & Xiaodong, (2018). Only feature extraction was the goal of their strategy. A piled autoencoder was yet another tool the author employed. This basically made the data's dimensions smaller and gave it a more compact shape. The author's SVM served as the classifier. The University of California provided the studied data. Genes were studied by J, V, Yasar, Qaiser & Dietrich, (2018) to forecast the return of breast cancer, they used biological indicators. The author employed the GCNN model (Graph Convolution Neural Network). Therefore, when compared to other algorithms, GCNN produced the value obtained.

A model simulation SVM classifier was proposed by Srirambabu, Santhosh K, and Senthil M, (2010). This study involved numerous phases, including picture enhancement, segmentation, feature extraction, and SVM classifier application. In identification, de - noising and edge detection, a median filtering strategy was adopted. A design using the k-nearest neighbor method was proposed by Seyyid, Tamazouzt A, and Abdelkader, (2013). The information for Wisconsin's melanoma emerged the one the researchers used to assess performance in relation to distance. The author calculated accuracy using two distance algorithms and obtained results. They achieved a 98.70% accuracy on the Euclidean distance and a 98.48% accuracy on the Manhattan distance. Basavaraj & Prasannakuma, (2015) saw the proposal of a model based on the SVM classifier. The model used Gaussian and Gaussian filter differences for detection and was based on mammography images. From the mini-mias dataset, only 75 mammography images were used. This approach had an accuracy rate of 89.33%. Varsha, (2015) suggested a mammography-based model utilizing an SVM classifier. A variety of stages were taken in doing this investigation. Quantization, ROI extraction, segmentation, semantic segmentation, and categorization were covered first. The VIES data set, which was employed for this, demonstrated that SVM offered 83% accuracy. A model based on the breast cancer decision tree classifier was proposed by Tina, (2019). Decision trees were used to incorporate the Wisconsin breast cancer data set. It was also covered basic Bayesian trees, sometimes known as rotation forests, in this study. The study was carried out within his WEKA environment. To demonstrate their accuracy, also achieving the goal, dynamic enhancing, elevating, and REPtree. A model based on an expansion of the classifier was proposed by Deepa et al (2019). A melanoma information utilized by the authors was subjected to the categorization strategy. 5 categories, both include and excluding image segmentation processes, were employed by the authors. These feature selection techniques mostly rely on correlation and data. Finally, they demonstrated how these 5 categories performed both include and excluding image segmentation. A model based on categorization approaches was proposed by Badal, et al, (2019). Using information from the Wisconsin melanoma collection, we applied random forests and support vector machines. The pull test ratio was used to base the results, and an accuracy of 99.714% was attained. Amandeep & Prabhjeet, (2019) proposed a paradigm for gene backpropagation based on neural network techniques. They also used KPCA for feature extraction. The information will be uploaded. Key performance measures, including as sensitivity and specificity, were used to assess performance.

The literature is replete with studies examining cancer detection and detection-related difficulties. Bilgin & Albayrak, (2016) created a deep learning-based technique to histopathologically detect breast mitosis by extracting features from a CNN model and training a SVM with those features. Based on histological scans, benign and malignant tumors were classified using AlexNet Krizhevsky, Sutskever & Hinton, (2012). Chen et al (2020). reported an extensive spiral system proposed. For the purpose of detecting meiosis 1 in breast tissue samples Jia, et al., (2014). They modified his CaffeNet model after first training their fully convolutional network (FCN) prototype can identify potential mitotic sites. Xu, Xiang, Hang, and Wu (2014) suggested using the Stacked Sparse Auto-Encoder (SSAE) technique to categorize cancerous cell shots with mitochondria. Vateekul & Wichakam, (2016) recommending Sharp CNN and SVM system combinations for bulk computerized ultrasound monitoring. There are various a version of CNN has been created for categorization in addition to pre-trained models. Demyanov and others they suggested finding two different types of patterns using deep CNN. A stochastic gradient descent approach was used to train the CNN. Introduced a CNN-based boundary detection system. Kim, Kim, & Ro, (2016) used factorization decomposition and statistical personality to enhance the training data. The same has been applied to mammography pictures using multiple deep learning approaches in addition to convolutional neural networks. For learning prospective bilateral feature representations of volumes for digital breast tomosynthesis (DBT)., we suggested a 3D multi-view deep ConvoNet model. High-level characteristics were independently retrieved utilizing two CNNs from two VOIs. For the purpose of melanoma detection, we suggested a semi-supervised learning approach. According to Mahbod, Schaefer, Wang, Ecker, and Ellinge's suggested system from 2019, tagged data and data with three different labels are combined with a deep network with two radial basis functions (RBFs) and a polynomial kernel SVM. trained on location, made from two unrelated pieces of information. Using a a system with entire multilayer regression, we suggested a deep residual network for melanoma identification in dermoscopy pictures (FCRN). The manually retrieved features used by traditional CAD programs have various disadvantages. For instance, custom features are frequently domain-specific, and designing custom features can be time-consuming, challenging, and non-generalizable Yi, (2017). Another technique for feature extraction is to use using a 3D CNN-based deep learning algorithm for identifying directly learn features from the full image Lo, Chan, Lin, Li, and T, (1995). CNNs have excelled at a variety of image categorization tasks Sze, Chen, Yang, Emer, (2017). For instance, 2012 saw AlexNet, a traditional CNN example won the ImageNet competition with 1000 class color photos Krizhevsky, Sutskever & Hinton, (2012). The 1.2 million annotated pictures used to train AlexNet by Krizhevsky, Sutskever & Hinton, (2012). These specifications relate to certain types of health imaging statistics, including: B. malignant pictures from radiography were frequently impracticable due to their difficulty in obtaining, fewer accurate advantages than expected in the sample set, and high cost of expert labeling by Shin, (2016). Making use of test set CNN styles as feature extractors on very large picture datasets from different domains offers a promising option, as does doing such studies on a small sample of labeled medical images. Model retraining (fine-tuning) Tadschbachsch, (2016) also known as transfer learning, it applied to solve a number of computer vision issues with success by Sharif R et al, (2014), Azizpour, et al, (2015), and Penatti, Nogueira & Dos S., (2015).

The diagnosis of pulmonary embolism and melanocyte lesions in earlier work by Tajbakhsh, (2016) and Esteva, (2017) shown that characteristics (connection weights in CNN) learnt with organic photos may be used even to ct scans when the aim photos are substantially different. demonstrates the transfer's potential. from source photos with prior training. Currently, CNN uses three techniques to classify medical images:

1) Building a CNN from the ground up to extract features from medical images (Wolterink, Leiner, & Viergever, 2015; Pan, 2015; Shen, Zhou, Yang, Yang, & Tian, 2015; Langs, 2014; Carneiro & Nascimento, 2013; Li, 2014). We contrasted three

main approaches. Prior studies have employed a variety of automated learning techniques for mammogram-based mammary melanoma /tumor identification (Ganesan, et al., 2013). The most widely used public mammography database with 10 folds for testing trained models is the MIAS and the Screening Mammography Heath, Bowyer, Kopans, Moore, and Kegelmeyer Digital Database (DDSM) 2000). Commonly employed is cross-validation. Some research, including Hussain, Aboalsamh & Bebis (2017), Raghavendra, et al. (2016), Khan, et al. (2016), Zhang, Wang, Liu, & Yang (2016), and Narváez, Alvarez, Garcia-Arteaga, Tarquino & Romero (2017), used conventional automatic feature extraction techniques (rather than manual extraction). As classifiers, neural networks have also been employed (S, et al., 2017). (R&B, 2011). In various studies (Zhu, Q, Vang, S, & X, 2016; Z, X, Y, & J, 2016; N, G, & P, 2016), CNNs have been used to extract features from mammography pictures (S, D, S, A, and M, p. 2014). Pretrained CNNs were employed in some of these research as a transfer learning application. However, only a small amount of prior analyses has given the outcomes of mammary melanoma diagnosis in mammograms employing CNNs alone for feature generation and classification. In our study, we just used one CNN. The bottom was entirely joined (FC) layer serves as the categorizer, while the front convolutional layer creates the features. Therefore, the mammography picture serves as the source and the (anticipated) description serves as the result for CNN. Inside this study, I utilized mammography pictures utilized from the VIES and DDSM databases. We first put three training techniques to the test on VIES.

2) Build an initial CNN,

3) Utilize a previously VGG-16 has been programmed to retrieve images via data photos, and then use essential qualities for brain connection classification changed the weights and made a kid train. By using backpropagation (fine-tuning) to find anomalous areas in the VGG-16 prototype final layers. In contrast, the second approach used in their research. Next, folks applied technique 2 to categorize the territories.

DDSM compares aberrant, aggressive, and malignant conditions to realistic conditions. To analyze the categorization outcomes, we used merge 10 times. The curves for convergence of recognition rate without any obvious collinearity. The pre-trained model employed in this investigation was different, the classification architecture and classifier were simpler, and a greater amount of training images than earlier studies in the field. The outcomes disagree with earlier research, with an Area

Under Curve of 0.96, the average accuracy for classifying anything as abnormal or normal is roughly 0.905. For both the aberrant and normal situations, our top model was accurate to 0.950.

In nations where breast cancer mortality has been successfully decreased, it is falling by 2-4% annually. If world mortality dropped by 2.5% annually 2.5 million cases of breast cancer could be diagnosed between 2020 and 2040 prevented. Mammary gland tissue cells can alter and divide uncontrollably in a disease group called breast cancer, which results in tumors. The lobules that connect the lobules and nipples are where the majority of breast cancers start. Breast cancer symptoms include breast soreness, breast skin color changes, breast lump development, and changes in breast size and shape. For the detection X-rays, magnetic imaging, and ultrasound are useful in the diagnosis of mammary frequently employed (Bower, 2008). The most effective treatment for carcinoma of the breast screening is mammography, which creates images using low doses of X-rays (Shen, et al., 2019). Deep learning models based on medical pictures are being developed by researchers all around the world for breast cancer detection. A thorough visual inspection may be necessary for breast cancer screening in order to find lumps or other anomalies that could be signs of disease. When these nodules are discovered, the accompanying measures can be made to assist clinicians in determining whether or not malignant tissue is present. architectural deformation, left-right asymmetry, and additional pronounced anomalies evident ones, like masses and calcifications, can be found with mammography. During a mammography, nodules, lumps, or density are all potential problems. Not all anomalies, nevertheless, are cancerous. For instance, well defined bulges are frequently benign. However, stellate, irregularly shaped tumors may a probe is required since it can be cancerous to determine this (Rodriguez-Ruiz, et al., 2019). Breast cancer cells have the ability to spread to the lymph nodes and harm the lungs and other organs. The dysfunction of the tubes that breast feed, also called "intrusive tubes", is the most frequent reason for BC. They also develop from some types of glandular tissue, breast lobules, or other cells or tissues. Researchers have discovered a link between environmental, hormonal, and lifestyle factors and a higher risk of BC. Breast tumor formation and demise are caused by unequal function and aberrant cell proliferation (Rodrguez-Ruiz, et al., 2018). A radiologist examines these mammography images to identify breast cancer. However, because to disparities in prior experience and knowledge, The results of specialists' examinations for melanoma may differ. Deep CNN-based breast cancer detection as a consequence methods can be used to enhance radiologists' assurance and act as a feedback in the evaluation of breast cancer. Many studies on a variety of deep CNN styles for mammography image-based breast cancer detection are now being conducted. In order to achieve 98.1% accuracy and diagnose BC with a 0.01% mistake rate, Naji, et al., (2021) used Decision Trees (DT), Naive Bayes (NB), Simple, sophisticated logic ensemble innovations including Random Forest (RF)and Voting Majority. DDSM, MI-AS, and INbreast collections, correspondingly developed by Chakravarthy & Rajaguru, (2022) an enhanced ICSELM Using the (Crow Search Efficient Extreme Learning Machine) method with 98.26%, 97.193%, and 98.137%. obtained a precision of Faisal, et al (2018) For comparison, they employed individual classifiers such the DT, Gradient Boosted Tree (GBT), MLP, NB, SVM, and Neural Network. Additionally thought of are MV-based ensembles and RF. In the GBT Ensemble, the author obtained a 90% accuracy rating. A classification model based was performed using a back propagation neural network (BPNN) utilized by Mughal, (2019). The approach was 99 times more accurate at detecting cancers in early-stage MIAS and DDSM statistics. A BiCNN model suggested by Wei, Han, He, and Yin, (2017) has been demonstrated to be 97.97% accurate. Khuriwal & Mishra, (2018) employed voting algorithm approaches, using survival analysis and artificial neural networks (ANNs) to identify breast cancer with 98% accuracy. Thuy & Hoang, (2019) a hybrid deep learning model's enhanced predictive accuracy employing a VGG16 model with a VGG19 model, a the efficiency of the generative adversarial network (GAN) of 98.1%. Bhowal, Sen, Velasquez & Sarkar, (2022) presented a CNN with Choquet integrals model for a four-class issue nanoparticles in melanoma histopathology a 95 validity using Information Theory and the Alliance Match. Khan, et al(2019) suggested using an original CNN prototype in conjunction with several methods for domain adaptation and attained an accuracy of 97.67. A new deep CNN model with an accuracy of 96.55%. In addition, a number of studies Nguyen et al., (2019), Ezzat, Hassanien, & Ella, (2021), Rajaraman & Antani, (2020), Lakhani & Sundaram, (2017), Hernández, Panizo, & Camacho, (2019), Wang, et al., (2019), Zheng, et al., (2020).

Breast malignancy is a leading force of death among women globally and with fast detection plays a pivotal role in increasing survival rates. A proposal for computer vision as a promising weapon for detecting breast cancer, and several surveys were carried out to assess its performance. One common approach in these studies is the use of digital mammography images to train machine learning algorithms. The use of convolutional neural networks (CNNs) is very common in these studies and have been demonstrated to reach high precision in detecting breast cancer in mammography images. In a study by Wang et al. (2018), a robust training model based on a CNN was trained on a dataset of over 85,000 mammography images and achieved an accuracy of over 90% in detecting breast cancer. Another popular approach is via means of ultrasound images to melanoma detection. In a study by Petrick et al. (2018), a computer learning procedure based on a random forest classifier was trained on a predictor variables of more than 500 ultrasound pictures and achieved a precision of over 80% in detecting breast cancer.

Some studies have also combined multiple modalities, such as ultrasonography and mammogram scans, to improve the precision of breast cancer detection. In a study by Rajpurkar et al. (2017), a deep learning model based on a CNN was trained on an integrated collection of mammography and ultrasound pictures and achieved an accuracy of over 90% in detecting breast cancer. Tumor is a pervasive type of tumor in females globally, apart from that early detection are crucial for improving patient's chances of survival. Convolutional neural networks (CNNs) has produced encouraging outcomes for the identification of tumors in medical images, such as mammograms, ultrasound, and magnetic resonance imaging (MRI). In this literature review, consultants shall discuss some of the recent studies that use CNNs for breast tumor diagnosis. Among these recent studies was conducted by Wang et al. (2021), who suggested deep learning framework for breast cancer detection. They used a Similar format which was before system, which was refined using a sample of mammogram images. They also image enhancement was utilized techniques, like rotational, flipping, apart from that scaling, in order to enlarge the information. Their proposed framework obtained a precision of 91.4% in the binary separating normal from cancerous tumours.

Another study was conducted by Huynh et al. (2020), who posited a CNN predicated framework for breast tumor detection in ultrasound images. They used a pre-trained DenseNet-121 network, which was fine-tuned on a dataset of ultrasound images. They also used a method of acquisition algorithms whereby you features learned from the which was before network were used as inputs to a small fully connected layer for classification. its suggested framework attained precision of 92.3% in the multimodal categorization task of benign and malignant breast tumors.

Chougrad et al. (2019) postulated a CNN-based strategy for identifying breast disease in mammograms. They used a conditioned ResNet-50 network, that's been fine-tuned on a dataset of mammogram images. Their proposed framework obtained a precision of 95.56% in the Dividing both non-malignant and aggressive breast cancers into two categories. Rahman et al. (2020) suggested a CNN-based method for breast tumor detection in MRI scan. They utilized a learned Hypothesis that was suggested v3 network, This was improved upon using a collection of MRI images. They also used a method of learning algorithms wherein one features learned from the pre-trained network were used as inputs to a small fully connected layer for classification. The suggested framework obtained a precision of 95.8% in the binary classification task of benign and malignant breast tumors.

In conclusion, CNN-based frameworks have produced encouraging outcomes for the identification of tumors in various medical images, such as mammograms, ultrasound, and MRI. These frameworks have the potential to increase the precision of breast cancer detection and reduce the number of false negatives and false positives, which can lead to improved patient outcomes. However, further research is needed to validate the results on larger datasets and to compare the performance of different CNN architectures for breast tumor recognition. A sort of synthetic neural network is the CNN that are created especially for image processing jobs.. It have the ability to gain knowledge complex features from the input picture through multiple layers of convolution, pooling, and activation functions. In the context of breast cancer detection, CNNs can learn to detect various patterns and abnormalities in mammography images, such as masses, microcalcifications, and architectural distortion. Several studies have demonstrated the effectiveness of CNNs in detecting breast cancer from mammography images. The model achieved an dimensionless region underneath the line of the receiver's operation characteristic (AUC) of 0.902, which was higher than that of the radiologists in the study.

The model achieved an AUC of 0.930 and a sensitivity of 88.5%, which was higher than that of the radiologists in the study. Overall, these studies suggest that CNNs have the opportunity to enhance accuracy apart from that breast disease's effectiveness detection from images from digital breast computed tomography and mammogram. However, more such investigation is required to confirm the performance of these models in larger and more diverse patient populations. Computer-aided detection (CADe) and computer-aided diagnosis are common categories for CAD in literature (CADx). By drawing the radiologist's attention to dubious portions of the image, the CADe method seeks to minimize observational oversight. Since 1998, a commercial mammography CADe system has been used in clinical settings (Freer & Ulissey, 2001). According to one study, CADe was utilized in over 92% of screening mammography conducted in the United States in 2016 (Freer & Ulissey, 2001). (Keen & Keen JM, Keen JE, 2018). The efficacy of CADe regimens in breast cancer screening is frequently contested despite their extensive clinical usage (Rodrguez-Ruiz, et al., 2018; Fenton, et al., 2007). (Henriksen, Carlsen, Vejborg, & Nielsen, 2019). The aim of computer-aided diagnostic systems (CADx), on the other hand, is to classify suspicious areas and characterize them. In 2017, the US FDA authorized Qlarity Imaging's QuantX, the company's first CADx program for chest MR imaging (Jiang, Edwards, & Newsstead, p. 2021). The purpose of QuantX is to assist radiologists in classifying lesions as benign or malignant. There are still not many people using this program, thus further clinical trials are required. Numerous research possess employed ML methods to diagnose BC in order to increase the precision and timeliness of categorization. By adding new intersection and mutation operators, they improved the neural network architecture. The ROC curve, AUC under the ROC, results in terms of accuracy, sensitivities, particular, non - linear data curve from GONN to the conventional model and the conventional backpropagation model were calculated using WBCD to assess their work. As opposed to this technique offers incredibly precise classification. However, he could enhance by making GONN more effective in the true breast tumor detection by employing a larger dataset than WBCD and component retrieval. A system-based process for automatically classifying tumor was proposed by Ashraf & Siti, (2018). The accuracy and network structure were optimized in this method using a Enhanced non-dominant multilayer perceptron (MLP) neural network genetic Selecting formula (NSGA-II). Comparing our study to other methods, the classification accuracy is improved. The MLP, though, might stay at a local minimum. Na, Qi, Xu, Bo, & Gui-Qiu, (2019) proposed utilizing a mixed feature selection method intelligent classification paradigm for identifying breast cancer. To create a paradigm for identifying breast cancer method, eliminate redundant and unnecessary features from the feature space using the Directed Simulated Annealing Genetic Algorithm Wrapper (IGSAGAW). This method lowers the cost of computing while increasing classification accuracy. To demonstrate the effectiveness of the suggested technique, it is used to Wisconsin primitive breast cancer (WBC) and WBCD. The suggested work performs well and lessens computing complexity.

A concept and execution of computer-aided detection (CAD) for the classification of mammography pictures were reported by Nawel, Nabiha, Nilanjan, & Mokhtar, (2016). The system employs a semi-supervised support vector machine (S3VM) for classification and a GA-based feature selection technique to minimize the site's number of features vectors. a computerized screening mammography database data sets (DDSM) was used to validate experiments. The suggested method increased accuracy. An automated system was created by Abdulkader, John, & Rahib, (2017) to categorize breast tissue. This system makes use of two of her ML methods: the Radial Basis Function Network (RBFN) and the Back propagate learning method in a multilayer perceptron (BPNN). Six distinct tissues, including connective, adipose, glandular, mammary, and carcinoma, were identified in breast cancer samples. Electrical impedance spectroscopy (EIS) was used to gather the data. When classifying six different breast tissues, the system with zonal fundamental functions performed better than the backpropagation network in terms of precision, least greatest eras, variance, and learning duration. The suggested system increased precision and cut down on preparation time. Neural network learning can become locked in local optimums and has poor generalization capabilities. For the purpose of diagnosing breast cancer, Haifeng, Bichen, Sang, & Hoo, (2017) developed an SVM-based ensemble learning model. Two different SVM structure types can be found in the proposed ensemble model. H. 6 kernel functions, -SVM, and C-SVM. To incorporate the knowledge of various base classifiers into the diagnostic job, a model hybridization weighted area under receiver operating characteristic curve ensemble (WAUCE) A process is shown. The Incidence, Epidemiology, and End Results (SEER) dataset, the Wisconsin Breast Cancer (WBC) dataset, the Wisconsin Diagnostic Breast Cancer (WDBC) raw information, and one big piece of data, were used to evaluate the model. When compared to other works that just used one SVM, the suggested model increases detection ability. Nevertheless, this method requires more processing and takes longer to train. A hybrid technique based on crazy normalization, KMC-based feature weighting, and the AdaBoostM1 classifier was proposed by Kemal & Mit, (2018). They outlines three steps for determining the presence of breast cancer. First, the information was standardized using her technique of MAD correction in the first phase. In the following phase, they weighted the normalized data using feature weighting based on K-Means Clustering (KMC). The weighted data set was finally classified using the AdaBoostM1 classifier. The UCI machine learning database's Breast Cancer Coimbra (BCC) dataset was utilised. In terms of accuracy, this strategy produces good results.

CNN were suggested by Teresa, et al., (2017) as a technique for technique of MAD correction in the images stained the use of the two substances. They offer four categories of medical applicability. Benign lesions, invasive cancer, carcinoma in situ, and normal tissue. Multiple histology scales will be integrated into the proposed CNN architecture. High-stained, high-resolution, unedited pictures based on the 2015 Bioimaging mammary tissue categorization endeavor make up the image dataset to which this model is applied. (Fabio, et al (2016) classified mammary melanoma histopathology images from the public dataset BreaKHis using a deep learning approach. A technique based on extracting CNN image extensions training and integrating the ultimate categorization of such locations was suggested. By skipping this stage, model adjustment, which can result in more intricate and computationally demanding designs, is avoided. Nevertheless, the investigation is costly due to the lack of data. A performance comparison of four classifiers was reported in Hiba, et al (2016). Using data from the Wisconsin Breast Cancer collection, supports Naive Bayes (NB), k Nearest Neighbors (k-NN), Decision Trees (C4.5), Vector Machines (SVM), and Naive Bayes.

The investigation initially focused on those who claimed to have very precise machine learning algorithms for mammary prognosis. The authors of Saxena & Gyanchandani, (2018) employs a back-propagation neural network (BPNN). We applied ML models, principal component analysis, and dimensionality reduction with the original data set (WBC) of Wisconsin breast cancer to improve performance. The researcher found that a multi-layer perceptron (MLP) classifier and a naive Bayes (NB) classifier both performed well on the WBC data set and achieved 95.99% of his desired accuracy, according to Bohacik, (2017). The Breast Cancer Diagnosis Record has been tested by a number of researchers (WDBC). Five machine learning algorithms were examined by the author of Filippakis & Kaklamanis, (2019). These algorithms were trained on three different sized data sets using the data mining program WEKA. Principal component analysis and correlation matrices were employed to look at the relationships between various variables. The efficiency score of a K-Nearest Neighbors (KNN) ML model was 96.73%. Houfani, (2020) evaluated decision trees

(DT), multilayer perceptrons (MLP), linear SVMs, random forests (RF), and kernel support vector machines (SVM) in a single experiment. Rodrigues, (2015) examined Bayesian networks and J48 decision trees as two different machine learning techniques for identifying breast cancer. He attained 97.80% accuracy using a Bayesian network. Salama, Abdelhalim, and Zeid, (2012) compared various classification algorithms utilizing three different breast cancer data sets and the fusion approach. Researchers created a method to differentiate between cancerous and healthy breasts tumors using genetic programming and machine learning techniques. The main shortcomings of ML-based computer-aided diagnostic systems are false negative (FN) and false positive (Fp) predictions. In order to lessen prognosis, both false positive and false negative prognosis, various researchers have suggested methods. A few of these have been looked at. According to Burnside, Sickles, (2002); Hayward, (2016) and Roelofs, (2007), taking past physical examination results into account during mammography screening can dramatically lower the false-positive rate.

The interpretation of mammography pictures is also being improved by radiologists thanks to recently created ML algorithms (Rodrguez-Ruiz, 2019). For instance, the author shows how his evident BI-RADS 3 and BI-RADS 4 lesions on high-density breast magnetic resonance imaging (MRI) are anticipated described in Burnside, et al (2002). This factor lessens the likelihood of inaccurately favorable predictions. Furthermore, to decrease false positive predictions in breast ultrasound pictures, researchers in Kim, (2021) which developed a deep learning-based CAD system. Related studies have demonstrated that ML classifiers have been utilized by researchers to enhance accuracy and lessen false-positive or false-negative BC predictions. However, a non-linear link exists between the remaining qualities and the WBC diagnostic features. As a result, it is impossible for ML linear classifiers to distinguish between cancers, both malignant as well as benign with accuracy. The average-weighted perceptron classifier is nonlinear, whereas the perceptron classifier is linear.

2.3 RELATED RESEARCH

There have been several surveys of recognize melanoma that there have written up in literature. To extract knowledge from databases, a number of Systems for computer vision, such as supervised learning methods, there have been created. The most typical applications for these algorithms are the recognition and categorization melanoma of the mammary. The different current studies on this topic are compiled in this section. Nilash et al. (2017) used a regression and classification tree CART on adopting WBC information, generate vague guidelines for diagnosing in melanoma an experience and understanding approach. They succeeded in 93.20 accuracy. Chaurasia et al. (2018) utilized his three algorithms and the WBC dataset to create a predictive model (J48, RBF Network, and Naive Bayes).The findings indicate believing Nave Bayes to be the most reliable predictor, outperforming J48 (accuracy of 93.41%), RBF Network (accuracy of 96.77%), and holdout sample (accuracy of 97.36%). In 2020, created the ELM-RBF, a business structure using the extreme learning machine (ELM) using a radial basis function kernel (RBF) that had a 95.39% accuracy rate. The Wisconsin's categorization of melanoma data sets (WBC and WDBC) is summarized in Table 1, along with studies on the use of WPBC for cancer relapse prediction.

Dataset	Ref	Year	Classifiers	ACC
WBC	Nilashi	2017	CART	93.20%
	Chaurasia et al.	2018	NB	97.36%
	Wang et al.	2020	RIPPER	95.39%
	Mojrian et al.	2020	ELM-RBF	95.69%
	Bayrak et al.	2019	SMO	96.90%
WDBC	Ramos et al.	2019	LDA	98.82%
	Najmu et al.	2020	DT	97.29%
	Sharma et al.	2018	KNN	94.00%
	Rufai et al.	2020	SVM	94.30%
	Salama et al.	2013	SMO	97.7%.
WPBC	Ojha &	2017	SVM	68.00%

	Goel			
	Pritom et al.	2017	SVM	75.70%
	Salama et al.	2013	fusion of MLP, J48, SMO and IBK	77.00%
	Kiage	2015	NB, KNN, RT	73.00%
	Chi et al.	2007	ANN	64.90%

Table 1. Using Various Datasets to Compare Past Research For the

In conclusion, The capacity of computer vision is enormous to identify breast cancer in a variety of imaging modalities. It is significant to highlight that this study frequently makes use of moderate sample sizes. To confirm the effectiveness of these algorithms in truly huge populations, much more research is required. Additionally, it takes a lot of work to achieve explainability and regulatory approval for machine learning models and the algorithms that integrate them into clinical practice.

CHAPTER III

Methodology

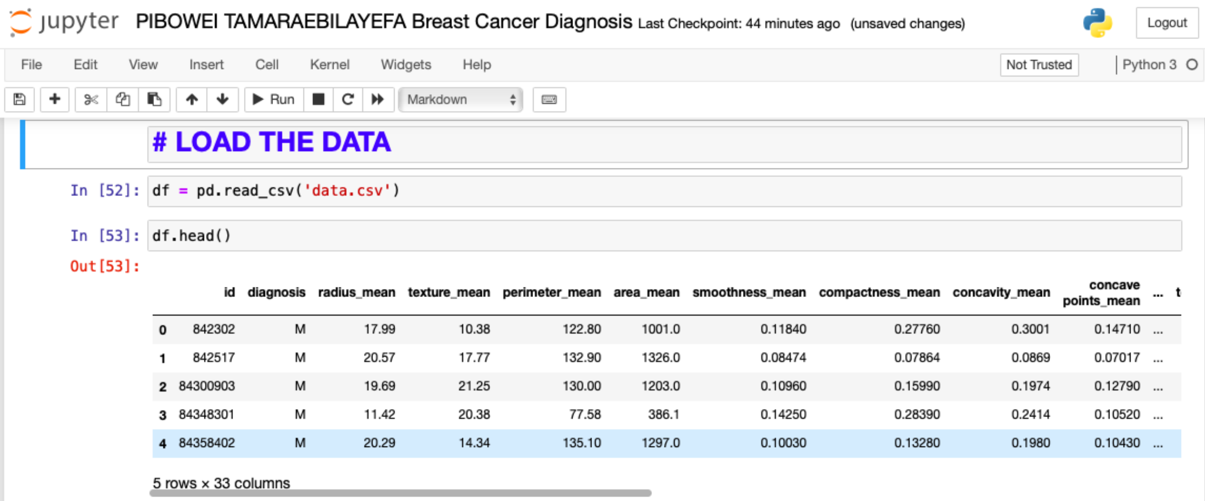
3.1 Introduction

This portion of the research project focuses on a thorough examination of the numerous techniques, steps, and tactics used to collect vital research data. The numerous statistical methods used to analyze the secondary data acquired for this study are also thoroughly analyzed and explained in this part.

3.2 DATA PRE-PROCESSING

3.2.1 Data Collection

The dataset used in this study is obtained from the UCI at the university provided by Wisconsin Breast Cancer Dataset (WBC). It is utilized by numerous researchers studying breast cancer at Wisconsin hospitals. These datasets are little spikes of enormous volumes of scanned picture feature-based data. Each feature matches a distinct visible cell nucleus in the image. A description of that same several datasets is provided. Instances of classification systems or characteristics with numbers are present in each record shown in Figure 1.



	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	...	t
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	...	
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	...	
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	...	
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	...	
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	...	

Figure 1. Data collection

3.2.2 Data Cleaning:

Eliminate or lessen clamor, and deal containing blank values. Eliminate empty values (ID & Unnamed: 32).

This work uses this collection after analysis. The WBC records have lacking yet superfluous. In order to tidy up the data, we replace any missing values with appropriate ones shown in figure 2.

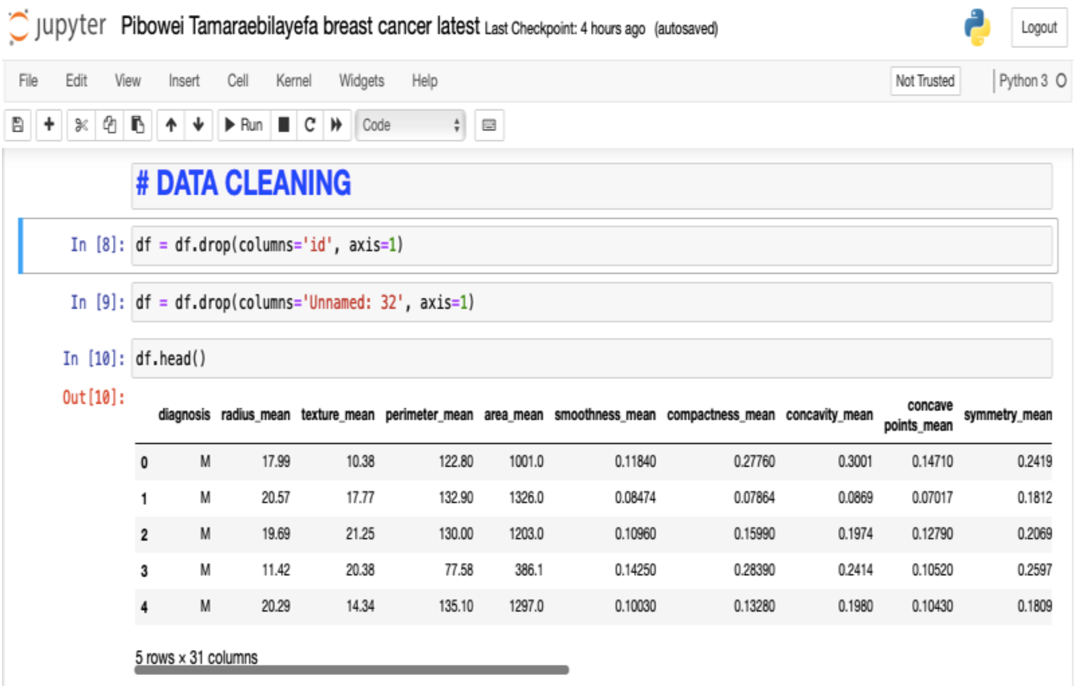


Figure 2. Data Cleaning

3.2.3 Wisconsin breast cancer dataset (WBC)

Some research made use of the Squamous cell carcinoma in Wisconsin (Original) (WBC) dataset. 699 examples of both benign and malignant breast cancer are represented. The record also includes eleven properties that have integer values.

Dataset name	WBC
Instances	568
Attribute	31
Classes	Benign (B) & Malignant (M)
Classes distribution	B= 356 and M= 212
Missing Values	No missing value

Table 2. Wisconsin Dataset Description

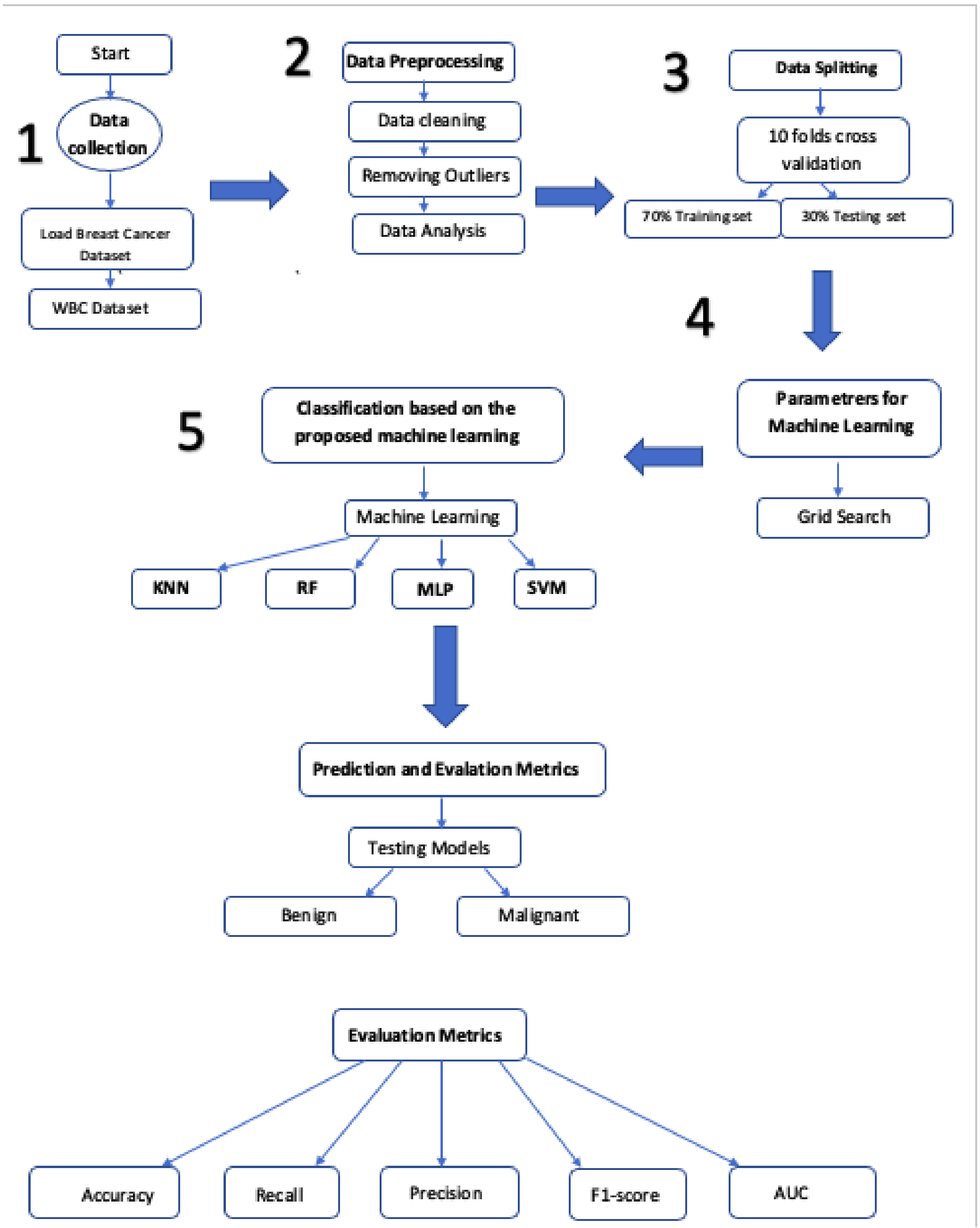


Figure 3. The proposed system of breast cancer Model

3.2.4 Wisconsin breast cancer diagnostic dataset (WDBC)

A Wisconsin diagnostic breast cancer (WDBC) data set is a minute piece of data made up of details extracted from digitally enhanced images. This database contains 568 sample entries, each with 33 characteristics (ID, diagnosis, 30 real-valued characteristics). All 31 input features can be linearly separated from the data collection. Each feature matches the characteristics of the cell nucleus seen in the photographs. The data on the attributes is summarized in Table 4. A unique identifier for each patient is represented by the first characteristic, and the malignant or benign class categorization is represented by the second attribute. The estimated parameters of each cell nucleus are represented by the attribute range from 3 to 32.

<table>
<tr><th>NO</th><th>Features</th></tr>
<tr><td>1</td><td>Id</td></tr>
<tr><td>2</td><td>Diagnosis</td></tr>
<tr><td>3-32</td><td>3-32). Eachnucleusisdescribed by ten computed characteristics.</td></tr>
<tr><td colspan="2">Radius</td></tr>
<tr><td colspan="2">Texture</td></tr>
<tr><td colspan="2">Perimeter</td></tr>
<tr><td colspan="2">Area</td></tr>
<tr><td colspan="2">Smoothness</td></tr>
<tr><td colspan="2">Compactness</td></tr>
<tr><td colspan="2">Concavity</td></tr>
<tr><td colspan="2">Concave points</td></tr>
<tr><td colspan="2">Symmetry</td></tr>
<tr><td colspan="2">Fractal dimension</td></tr>
</table>

Table 3. WDBC Dataset Features Information

The radius is the average distance from the entry to every other point in its immediate vicinity. The grayscale values' standard deviation is used to define texture settings. The degree of local variation in the radius length is known as smoothness. The formula for the compression ratio is (perimeter intensity 2/area - 1.0). The contour lines' concavity is measured in terms of their degree, which is the asymmetrical element (the coastal extrapolation) -1. Calculations are made for the average, standard deviation, and worst 30 traits. An illustration, box 3 shows the typical diameter, box 13 the average standard deviation of the perimeter, and box 23 the worst radiuses. Characteristics of WDBC dataset reveal that these properties contain the median, standard deviation, and worst three values, and 3 angles.

• The Mean is determined by Eq. (1) as follows:

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (1)$$

The Standard Error is determined by Eq. (2) as follows:

$$\mathbf{SE} = \frac{\sigma}{\sqrt{n}}$$

3.3 Data Pre-processing

Every categorization system needs data pre-processing it does this by converting picture data into a format that neural nets can use comprehend. To guarantee that accurate information is provided instead of mistakes, we process the data set using data preprocessing. Data should be clear, accurate, and complete because data quality affects categorization performance. By filling in the collection has incomplete data, data preprocessing removes disparities from the data set. To improve The accuracy level of an information and produce clean data appropriate for modeling, pre-processing is utilized (William H, et al., 1994). The data set included redundant and pointless material that was assembled from several sources. For the purpose of removing inconsistencies of this kind from the data gathering, we employ data cleaning procedures. Several Previously, periodically re techniques were used on the Melanoma information. completing using machine learning techniques for categorization jobs. Throughout preparation, having cleansed the material, non - key attribute remained eliminated, as well as tiers eliminated. The dataset is prepared for machine learning models using these pre-processing stages, which also include the cleaning phase. It improves performance results by removing unnecessary features from the data. The pre-processing process is broken down into a number of sub-phases, which are described beneath.

3.3.1 Filtering information:

Eliminate or lessen clamor, and deal containing blank virtues. Eliminate empty virtues. This work uses the collection after analysis. The WBC and WPBC records have blank and superfluous data, while the WDBC record is blunder. In order to tidy up the data, we replace any missing virtues with appropriate ones. One attribute value is missing in 16 WBC instances and 4 WPBC instances, and is denoted by the character "?". The term "attribute" refers to the process of adding missing values to all instances that belong to the same tier.

3.3.2 Eliminating oddity

Oddity can be quite damaging. These significantly affect how well machine learning models perform. Outliers are often assessed by researchers to ascertain if each data set is the outcome of errors in information gathering or of underlying phenomena that are taken into consideration when analyzing information. Statistics points known as outliers appear to lag behind the remainder of the data. While maintaining the initial information's dignity, removing outliers can lead to a lesser sample than the initial.

3.3.3 Evaluation of Authenticity:

Analysis of residuals correlations removes unnecessary elements of future study. There is an additional feature known as "Sample Code Number" that is shared by WBC, WPBC, and WDBC but is unimportant for classification operations. This characteristic is therefore disregarded.

3.3.4 Data normalization:

Start the training process using features that are similar in size to shorten training time. To make the range of feature values more understandable, normalization is used.

3.3.5 Information Division

Its essential principle of the Split a 10-Fold CV the data set along with 10 parts or creases, of whose 9 will be utilized for evaluation and instruction. With holdout (80% for training, 30% for The dataset is split into training and testing datasets for validation), development and certification collections are separated from the information. The data splitting procedure is performed k times (k=10).

3.3.6 Modification of the extracted features

Lattice look is a method as an adjusting hyperparameters which could be used to choose the best values for ML algorithms. The optimum model hyperparameter response should be provided after a machine learning model has been evaluated on every set of algorithmic conditions set in a grid. In this step, the best range for each parameter in the ML model is determined using a hyperparameter optimization approach (vector scan with filtered 10-fold intra- and inter**)**

3.4 Employing Machine Learning Models For Fragmentation

In this step, we put 3 standard common machine learning methods, such as discriminant analysis. —into practice. Multilayer perception (MLP), K-Nearest Neighbor (KNN), andrandom forest (RF).

3.4.1 K-Nearest Neighbor (KNN)

The K-Nearest Neighbor (KNN) classifier, is perhaps the most widely utilized well-liked designation using computer vision techniques (Guo, Ma, Cukic & Singh, 2004). Nonparametric trick learning employs the Neighbor KNearest taxonomy. This categorization technique arranges objects whose close neighbors refer to them as "k." Not intrinsic data mapping, but object neighborhood, is the focus here.

3.4.2 Random forest (RF)

The array of plants is produced using the Random Forest (RF) feature selection algorithm from the results of various plant choices. A decision made tree's potential to provide either particular models or quite simple models supports this (Xiong & Yao, 2021).

3.4.3 The Multilayer Perceptron (MLP)

A majority popular method for example identification in artificial neural networks (ANNs) is the multilayer perceptron (MLP), a feedforward backpropagation network (Witten, Frank & Geller, 2002). Three parts make up the supervised learning method known as MLP. A layer or shades, including input, output, and gather important data throughout the training process and provide the input layer's constituent parts adjustable weights (Riedmiller, 1994).

3.5 Performance Metrics

There are five common performance metrics in Table 5. The following formulas are used to determine precision (PREC), recall (REC), accuracy (ACC), the function of the receiver area, and (AUC) (AUC): The ratio of examples which be properly categorized (TP+TN) to all occurrences is used to calculate classifier accuracy.
(TP+TN+FP+FN).

Calculated using the Eq. (3)

Accuracy = $\frac{\mathrm{TP+TN}}{TP+TN+FP+FN}$ (3)

The accuracy is the ratio of all projected patients (TP + FP) to all illness samples (TP + TP) accurately categorized. formula. figured using (4).

Precision = $\frac{\mathrm{TP}}{TP+FP}$ (4)

The percentage of correctly categorized specimens (TP) relative from the overall count of ill patients is known as the callback score metric. The number of people who have been diagnosed with the disease influences how memories are perceived. Recall is also known as sensitivity.

The percentage of incidents that were appropriately categorized as malignant or, in our situation, as having the disease, is known as could be recalling accurate malignancy (favorable) Score (TM) or (TP). calculated out using the formula. (Five):

Recall = $\frac{\mathrm{TP}}{TP+FN}$ (5)

In addition to the F dimension, F1 Precision and recall are balanced, according to the F1 score. In real-world situations, improving model accuracy reduces recalling and the opposite. A single value, the F1 score, fits both trends in the equation (1). (6):

F1 score = $\frac{\mathrm{2TP}}{2TP+FP+FN}$ (6)

In order to evaluate several computer vision algorithms, the area under (AUC), or the receiver's operational efficiency, frequently were applied.

AUC = $\frac{\mathrm{s0-n0(n0+1)/2}}{n0n1}$ (7)

3.6 Correlation Matrix

A correlation matrix is a statistical tool used to analyze the relationships between variables in a dataset. It shows how strongly pairs of variables are related. In the context of breast cancer detection using machine learning, you might have a dataset with various features (attributes) that describe characteristics of tumors, and you want to understand how these features correlate with each other. This can help you identify which features are most informative for the detection of breast cancer. Here's a general process for creating a correlation matrix for breast cancer detection using machine learning:

I. **Data Preparation:** Obtain a dataset with features (e.g., tumor size, shape, texture, etc.) and labels (e.g., malignant or benign) for breast cancer cases. Make sure the dataset is clean and properly formatted.

II. **Calculate Correlation Coefficients:** Calculate the correlation coefficients between pairs of features in your dataset. Common correlation coefficients such as Pearson correlation, Spearman rank correlation, or Kendall tau correlation can be used.

III. **Pearson Correlation:** Measures the linear relationship between two continuous variables. It ranges from -1 to 1, where -1 indicates a strong negative correlation, 1 indicates a strong positive correlation, and 0 indicates no correlation.

IV. **Spearman Rank Correlation:** Measures the strength and direction of monotonic association between two continuous or ordinal variables. It's useful when the relationship is not strictly linear.

V. Kendall Tau Correlation: Similar to Spearman's correlation, it measures the strength and direction of a monotonic association between two variables. It's also robust to outliers.

VI. **Create the Correlation Matrix:** Construct a square matrix where each cell represents the correlation between two features. Rows and columns of the matrix correspond to the features showing benign (positive and negative) and maliganant (positive and negative) in your dataset.

VII. **Visualize the Correlation Matrix:** Visualization can provide a clearer understanding of the correlations. You can create a heatmap of the correlation matrix using libraries like Seaborn in Python.

<table>
<tr><td colspan="2"></td><td colspan="2">ACTUAL VALUES</td></tr>
<tr><td colspan="2"></td><td>Malignant (Positive)</td><td>Benign (Negative)</td></tr>
<tr><td rowspan="2">PREDICTED VALUES</td><td>Malignant (Positive)</td><td>TP</td><td>FN</td></tr>
<tr><td>Benign (Negative)</td><td>FP</td><td>TN</td></tr>
</table>

Table 4. Correlation Matrix

CHAPTER IV

Findings and Discussion

Wisconsin examined dataset in order to assess the effectiveness using ML methods for the early diagnosis of breast cancer. Using the holdout validation (70% for training, 30 percent for testing) and tenfold cross-validation learning procedures, produce WBC recordings. The usage is to get higher reliable analysis for each dataset, machine learning algorithms (KNN, RF, MLP & SVM). In order to assess the model about its precision, accuracy, recall used.

4.1 Maintain Verification

On the basis of three machine learning models, 70 training sessions and 30 tests to investigate common impacts on the detection of breast cancer conducted. About its precision, recall, f1 score, area under the curve and accuracy, Random Forest (RF) surpassed its rivals with achievements of 86%, 94%, 90%, 92% and 92%. Due to the robustness of the kernel and the distinctive capacity of his RF to handle binary issues, this bias exists. Additionally, on the WBC dataset, the ANN classifier performs the worst. The outcomes of the MLP classifier in fundamental machine learning are shown in Table 5.

ALGORITHM	**70 TRAINING & 30 TESTING METRICS**				
	ACC	**AUC**	**PREC**	**REC**	**F1-SCORE**
KNN	**83%**	**82%**	**76%**	**79%**	**78%**
RF	**92%**	**92%**	**86%**	**94%**	**90%**
MLP	**91%**	**90%**	**88%**	**88%**	**88%**
SVC	**86%**	**84%**	**86%**	**76%**	**81%**

Table 5. The performance results of Machine Learning for WBC dataset using HOLD OUT VALIDATION

CHAPTER V

Discussion

5.1 Introduction

The results of the study are outlined in each of the four sections that make up this chapter

The discussion portion of a research paper is where the author interprets and assesses the study's findings in light of the body of knowledge and hypotheses that have already been established. It is a critical section of the article that enables the author to show that they have a solid grasp of the research issue, the techniques employed, and the conclusions drawn. The research question and an overview of the key findings are often restated at the start of the discussion section. The results are then interpreted and explained by the author, who also discusses how they add to the body of knowledge already known in the subject.

5.2 For case study

Table 5 displays the overall actual outcomes of the cross-validation performance or test results based on the various machine learnings from the trials on the WBC dataset. They show the potency of the best models for each feature extraction method. We calculated the mean crossover of each model to provide a summary of the performance of the contrasted models employing various underlying The Severe Asymmetric Increase in Computer Vision (XGboost), Neural Network Layers of Sensation (MLP), and Naive Bayes (NB) algorithms. Consider testing and validation findings such as Random Forest, instance-based K-Nearest Neighbor (KNN), and Random Forest. Comparing the MLP model to other conventional machine learning techniques, the average cross-validation is the greatest. The RF technique achieved precision 86%, 94% recall, 92% accuracy, area under the curve 92% and 90% F1 score, according to cross-validation, and 98.3% accuracy, 99.3% AUC.

5.3 Comparative Analysis

The flowchart for the whole random forest model is shown in Figure 4. The ML system Random Forest is well-organized. It constructs "forests" from a collection of decision trees that are primarily prepared to employ the "sacking" strategy. The main justification for the packing technique is that combining a few learning models provides with a clear result. To achieve a more precise and reliable result, a random forest selects various decision trees and combines them. It benefits from attention to the game planning and backsliding problems that characterize most modern ML structures.

Finding the overall importance of each element on the gauge is a crucial component of the random forest technique, which is still another fascinating feature. Sklearn offers a remarkable mechanical assembly for evaluating a component's relevance by examining how much the tree centers that use it lessen tainting across the entire backwoods. After preparing, it records this score for each brand name and modifies the reveals with the aim of enhancing its overall significance.

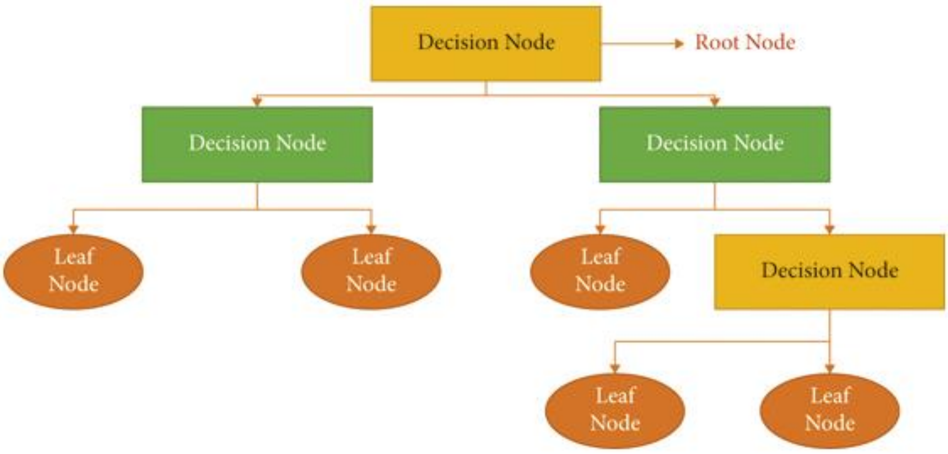


Figure 4. RANDOM FOREST FLOWCHART.

The adaptability of random forest is one of its most appealing qualities. It is applicable to both grouping and relapsing procedures, and it is evident how important it considers the data qualities as a whole. It is also a useful strategy since it frequently uses default hyperparameters that provide expectations that are obvious. Although there aren't many of them, understanding the hyperparameters is crucial. Even though overfitting is one of the most well-known problems in ML, the arbitrary random forest classifier seldom experiences it. The classifier won't overfit the model if there are enough trees in the forest.

The decision trees used in the random forest approach are each composed of a bootstrap sample taken from a training set. One-third of the training sample is kept as test data and is known as the out-of-bag (OOB) sample, which will be covered later. The dataset is then given a second dose of randomization via feature bagging, increasing its variety while decreasing the connection between decision trees. The method for making the forecast varies depending on the scenario..

Figure 5 displays the flowchart for the whole K-nearest neighbor model. The K-nearest neighbor computation, which depends on the supervised learning system, is one of the most fundamental machine learning calculations. The new example is placed in the class that is closest to the prior classifications by the KNN technique, which assumes that the new case and previous cases are equivalent. Every available data point is kept up to date by the KNN computation, which also prioritizes new information sources based on how well they match up with earlier data. This suggests that fresh information may be swiftly categorized into an obvious categorization using the KNN technique. The KNN technique may be used for grouping and relapsing, although it is mostly employed for arrangement.

The KNN technique is nonparametric, meaning it has no assumptions about the data. It is also known as a "languid student technique" since the student does not instantly benefit from the preparation; rather, the student saves the material and organizes it later. The KNN technique only stores data during the preparation phase, and when it receives fresh data, it classifies it into a set that is almost comparable to the fresh data.

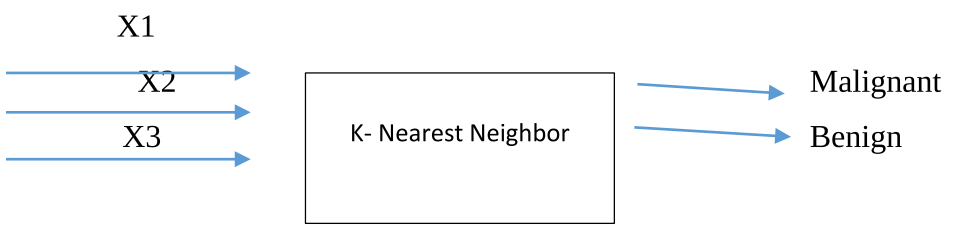


Figure 5 .

Information can be organized using the nonparametric slow learning technique known as K-nearest neighbor. With this classifier, objects are arranged according to their distance from "k" closest neighbors. Instead of the necessary information dispersion, it takes into account the item's immediate ambient factors.

5.4 Data Visualization

It has square bases with bases at the distances between class boundaries and districts in relation to the frequency of the various classes. It is a location framework. Every square structure in such representations is related because the basis fills up the spaces between class boundaries. The heights of square structures will correspond to the comparing repeat densities for distinct classes and are related to the frequencies of surrounding classes. Figures 6 show the plot of the whole dataset.

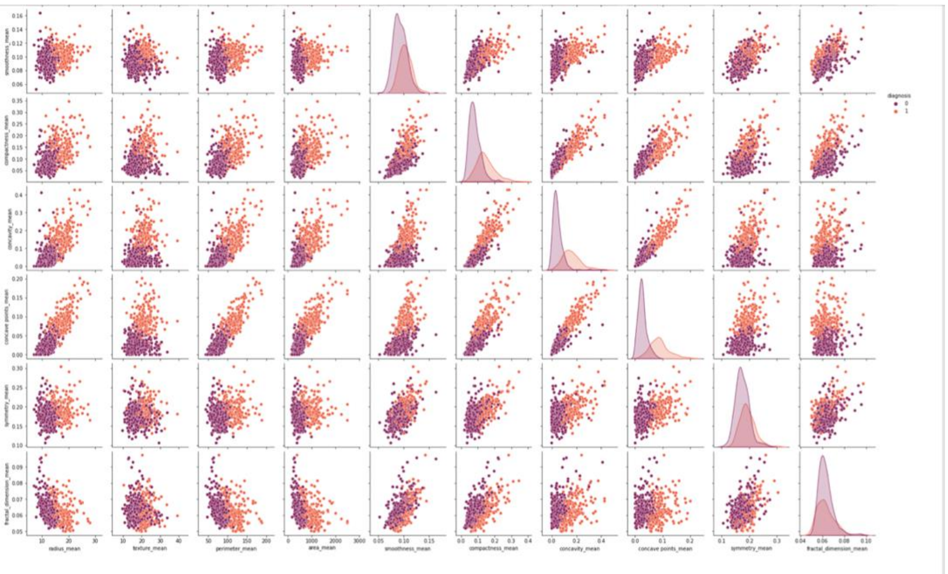


Figure 6

Model Comparision

Proposed approach	Accuracy(%)	Reference paper (sharma)	Accuracy(%)
Random forest	92	**Random forest**	97
k-nearest neighbour	83	**k-nearest neighbour**	94

Table 6

CHAPTER VI

Conclusion and Recommendations

An effective paradigm for diagnosing breast cancer and predicting recurrence was suggested in this research. The six main steps in our suggested framework for finding breast cancer are as follows.

Data collection, pre-processing, partitioning, hyperparameter tweaking for machine learning models, proposed ML-based categorization, and evaluation metrics are all included. Four methods of machine learning for identifying malignancy have been compared, such as Multi-Layer Perception, K Nearest Neighbors, Random Forest, and Support Vector Machines. To evaluate which classifier performed best on the Wisconsin dataset, we examined the performance of the frame works. This study made a significant contribution by developing an optimal framework for breast cancer detection and recurrence prediction that outperformed existing algorithms. In comparison to multiple prior art approaches, various feature combinations enhanced precision, recall, f1 score, accuracy and Area under curve measures.

The Random Forest performed with precision (86%), recall (94%), AUC (92%), and accuracy (92%) and the highest F1 score (90%) in the WBC dataset for breast cancer classification. Ranked due to a cross validation of 10 times. Area Under Curve (55,2%), sharpness (52,1%), recollection (51,7%), another thing is the F1 score (51,9%), on the other hand, placed ANN last. In conclusion, uneven information has a big impact negative influence on the effectiveness various systems for computer vision.

The K Nearest Neighbors model had the maximum power precision (76%), F1 score (78%), accuracy (83%), recall (79%), AUC (82%), using the WBC breast cancer detection dataset.

The multi-layer perceptron (MLP) achieved the best F1 score (88%), accuracy (91%), recall (88%), AUC (90%), and precision (88%) in the WBC dataset for breast cancer classification.

The Support Vector Machines model had precision (86%), F1 score (81%), accuracy (86%), recall (76%), AUC (84%), using the WBC breast cancer detection dataset

In conclusion, uneven data has a significant impact on how well machine learning algorithms function.

The WBC breast cancer detection dataset produced the highest power precision Classification performance Random Fores tmodel. 10 times of cross-validation KNN is the final model, followed by MLP.

Because melanoma is particularly common prevalent finding it is a difficult and risky sickness challenging task. Every year, breast cancer occurrence rates rises, and the likelihood of survival is declining. Breast cancer is detected using computational intelligence techniques. According to earlier research, computer vision approaches perform better in their particular fields. Several different computer vision techniques were indeed used in previous studies, with some enhancements and extensions to the dataset made to enhance performance. Yet, it is found that linear data is superior for machine learning. Previous studies have also revealed if data in the form of pictures of malfunctioning machines are accessible.

This work demonstrates that mammography may be used to detect breast cancer by using transfer learning to train CNN, and that developing her retrieval of qualities using NN heuristic is a quicker technique of adaptive learning.

REFERENCES

Abdelaziz A., Elhoseny M., Salama A, S., Riad, A, M., A machine learning model for improving health care services on cloud computing environment, Measurement 119 (2018) 117– 128.

Abdulkader, H., John, B. I., Rahib, H.A.: Machine learning techniques for classifi- cation of breast tissue. In: 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW, pp. 402-410. Procedia Computer Science, Elsevier. Budapest, Hungary (2017)

Ahmad, M, A., Eckert, C., A. Teredesai, Interpretable machine learning in healthcare, in: Proceedings of the 2018 ACM I*nternational Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2018, August, pp. 559–560.

Ahmad, M, A., Patel, A., Eckert, C., Kumar, V., Teredesai, A., Fairness in machine learning for healthcare, in: Proceedings of the 26th ACM S*IGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, August, pp. 3529–3530.

Ahmet K, A., Res. Assist., Seyma Yasar, Res. Assist., Cemil Colak, Prof. Dr., “ Breast cancer classification using a constructed convolutional neural network on the basis of the histopathlogical images by an interactive web-based interface,” 2019 3rd International Symposium on pp. 1-5, 2019.

Ajay, K, Sushil, R., Tiwari, A, K., “ Comparative study of classification techniques for breast cancer diagnosis,” International Journal of Computer Science and Engineering, vol. 7, no. 1, pp. 234-240, 2019.

Alokkumar, J., Ghanshyam, V., Yasar, K., Qaiser, M., Dietrich, R, S., “Deep convolution neural network model to predict relapse in breast cancer,” 17th IEEE International Conference on Machine Learning and Applications, pp. 351-358, 2018.

Amandeep, K., Prabhjeet, K., “Breast cancer detection and classification using analysis and gene-back propagation neural network algorithm,” International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 8, pp. 2798-2803, 2019.

Amethiya, Y., Pipariya, Y., Patel, S., Shah, M., Comparative analysis of breast cancer detection using machine learning and biosensors, Intelligent Medicine 2 (2) (2022) 69–81.

Andrik, R., Bryan W., Scotney, H, W., "Breast mass classification in mammogram using ensemble convolutional neural networks," IEEE 20th International Conference on e-Health Networking, Application and Services (Healthcom), pp. 1-6, 2018.

Anusha, B., Pooja, N., Anishka, R, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," IEEE 3rd International Conference on Circuits, Control, Communication and Computing, pp. 1-4, 2018.

Arpit, B., Aruna, T.: Breast Cancer Diagnosis Using Genetically Optimized Neural Network Model. Expert Systems with Applications, Elsevier, pp. 1-15(2015)

Ashraf, O, I., and Siti, M, S., : Intelligent breast cancer diagnosis based on enhanced Pareto optimal and multilayer perceptron neural network. International Journal of Computer Aided Engineering and Technology, Inderscience, Vol. 10, No. 5, pp. 543-556(2018)

Ayush D., Bhawna, G., Kaushik K., "A brief review of breast cancer detection via computer aided deep learning methods," International Journal of Engineering Research & Technology, vol. 8, no. 12, pp. 326-331, 2019.

Azizpour, H., Sharif, R, A., Sullivan, J., Maki, A., and Carlsson, S.,"From generic to specific deep representations for visual recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 36–45.

Badal, S., Angshuman, B., Arpita, G., Anji, R, "A novel classification technique for breast cancer diagnosis," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 12, pp. 3295-3305, 2019.

Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., and Greenspan, H., "Chest pathology detection using deep learning with non-medical training," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, 2015, pp. 294–297.

Basavaraj H., Prasannakumar, S, C., "Automated evalution of breast cancer detection using svm classifier," International Journal of Computer Science Engineering, vol. 5, no. 1, pp. 7-16, 2015.

Berlin, L., Hall, F, M., More mammography muddle: emotions, politics, science, costs, and polarization. *Radiology* (2010) 255(2):311–6. doi: 10.1148/radiol.10100056

Bhardwaj, R., Nambiar, A, R., Dutta, D., A study of machine learning in healthcare, in: 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC) vol. 2, *IEE*E, 2017, July, pp. 236–241.

Bhowal, P., Sen, S., Velasquez, J, D., Sarkar, R., Fuzzy ensemble of deep learning models using choquet fuzzy integral, coalition game and information theory for breast cancer histology classification. *Expert Syst. Appl.* **2022**, *190*, 116167.

Bohacik, J., Data mining applied to Breast Cancer Wisconsin data,, Journal of Information Technologies 10 (1) (2017) 1–5.

Borges,W., Sampaio, Moraes, E, D., Corrêa, S, A., Cardoso de Paiva, A. A., and Gattass, M., "Detection of masses in mammogram images using CNN, geostatistic functions and SVM," *Comput. Biol. Med.*, vol. 41, no. 8, pp. 653–664, Aug. 2011.

Bower, J, E., Behavioral symptoms in breast cancer patients and survivors: Fatigue, insomnia, depression, and cognitive disturbance. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 2008, *26*, 768.

Breast Cancer. 2021. Available online: https://www.who.int/news-room/fact-sheets/detail/breast- cancer (accessed on 1 February 2022).

Breast FNA (June 2022). URL https://icytology.wordpress.com/breast-fna/

Brennan, M., Houssami, N., Discussing the benefits and harms of screening mammography. *Maturitas* (2016) 92:150–3. doi: 10.1016/j.maturitas.2016.08.003

Burnside, E, S., Sickles, E, A., Sohlich, R, E., Dee, K, E., Differential Value of Comparison with Previous Examinations in Diagnostic Versus Screening Mammography, American Journal of Roentgenology 179 (5) (2002) 1173–1177.

Cancer Facts & Figures 2022 (June 2022). URL https://www.cancer.org/latest-news/facts-and-figures- 2022.html

Canh, P, N., Anh, H, Vo., Bao, T, N., "Breast cancer histology image classification using deep learning," 19th International Symposium on Communication and Information Technologies(ISCIT), pp. 366-370 , 2019.

Carneiro, G., and Nascimento, J, C., "Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2592–2607, 2013.

Chakravarthy, S, S., Rajaguru, H., Automatic Detection and Classification of Mammograms Using Improved Extreme Learning Machine with Deep Learning. *IRBM* **2022**, *43*, 49–61..

Chandra C, C., Gopal, K., "A noval method for IDC prediction in breast cancer histopathology images using deep residual neural networks," 2nd International Conference on Intelligent Communication and Computational techniques(ICCT), pp. 95-100, 2019.

Char, D, S., Abra`moff, M, D., Feudtner, C., Identifying ethical considerations for machine learning healthcare applications, Am. J. Bioeth. 20 (11) (2020) 717.

Chen, I, Y., Joshi, S., Ghassemi, M., Ranganath, R., Probabilistic machine learning for healthcare, Annual Review of Biomedical Data Science 4 (2020).

Chen, Y, H., Yang, T, J., and Emer, J., "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," *ArXiv170309039 Cs*, Mar. 2017.Technology, vol. 8, no. 07, pp.148-152, 2019.

Chtihrakkannan, R., Kavitha, P., Mangayarkarasi, T., Karthikeyan, R., "Breast cancer detection using machine learning," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 11, pp. 3123-3126, 2019.

Ciompi, F., *et al.*, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," *Med. Image Anal.*, vol. 26, no. 1, pp. 195– 202, 2015.

Control Conference July 25-27, Wuhan , China, pp. 9428-9433 , 2018.

Cruz, B, F., De Assis, J, T., Estrela, V, ., Khelassi, A., A compact SIFT-based strategy for visual information retrieval in large image databases, Medical Technologies Journal 3 (2), 402-412, 2019 https://doi.org/10.26415/2572-004X-vol3iss2p402-412

Deepa, B, G., Senthil, S., Gupta, R, M., Shah, V, R., "Augmentation of classifier accuracy through implication of feature selection for breast cancer prediction," International Journal of Recent Technology and Engineering, vol. 8, no. 2, pp. 6396-6399, 2019.

DeSantis, C, E., Ma, J., Gaudet, M, M., Newman, L, A., Miller KD, Goding Sauer A, et al.. Breast cancer statistics, 2019. *CA: A Cancer J Clin* (2019) 69(6):438–51. doi: 10.3322/caac.21583

Devi I., Karpagam G, R, and Vinoth Kumar, B., (2017), A survey of machine learning techniques. International Journal of Computational Systems Engineering. 3 (4): 203-212.https://doi.org/10.1504/IJCSYS E.2017.10010099

Dhivya, R., Dharani, R., "Survay on breast cancer detection using neural networks," International Research journal of Engineering and Technology, vol. 06, no. 03, pp.3943-

Dhungel, N., Carneiro, G., and Bradley, A. P., "The Automated Learning of Deep Features for Breast Mass Classification from Mammograms," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, 2016, 3945, 2019.

Ebru, A, B., Pinar, K., Tolga, E.,"Comparison of machine learning methods for breast cancer diagnosis.2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), pp. 1-3,2019.

Erhan, D., Manzagol, P, A., Bengio, Y., Bengio, S., and Vincent, P., "The Difficulty of Training Deep Architectures and the Effect of Unsupervised Pre-Training," 2009.

Esteva, A., *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017. pp. 106–114.

Ezzat, D., Hassanien, A, E., Ella, H, A., An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization. *Appl. Soft Comput.* **2021**, *98*, 106742.

Faisal, M, I., Bashir, S., Khan, Z, S., Khan, F, H., An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer. In Proceedings of the 2018 3rd International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST), Karachi, Pakistan, 21–22 December 2018; pp. 1–4.

Fenton, J, J., Taplin, S, H., Carney, P, A., Abraham, L., Sickles, E, A., D'Orsi, C., et al.. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* (2007) 356(14):1399–409. doi: 10.1056/

Freer, T, W., Ulissey, M, J., Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiol* (2001) 220(3):781–6. doi: 10.1148/radiol.2203001282

Ganesan, K., Acharya, U, R., Chua, C, K., Min, L, C., Abraham, K, T., and Ng, K, H., "Computer-Aided Breast Cancer Detection Using Mammograms: A Review," *IEEE Rev. Biomed. Eng.*, vol. 6, pp. 77–98, 2013.

Gøtzsche, P, C., Mammography screening is harmful and should be abandoned. *J R Soc Med* (2015) 108(9):341–5. doi: 10.1177/0141076815602452

Haifeng, W., Bichen, Z., Sang, W.Y., Hoo, S. K.: A Support Vector Machine-Based Ensemble Algorithm for Breast Cancer Diagnosis. European Journal of Operational Research, Elsevier, pp. 1-33 (2017)

Hayward, J, H., Improving Screening Mammography Outcomes Through Comparison With Multiple Prior Mammograms, American Journal of Roentgenology AJR 207 (4) (2016) 918–924. doi:10.2214/AJR.15.15917.

Heath, M., Bowyer, K., Kopans, D., Moore, R., and Kegelmeyer, W, P., "The digital database for screening mammography," in *Proceedings of the 5th international workshop on digital mammography*, 2000, pp. 212–218.

Henriksen, E, L., Carlsen, J, F., Vejborg, I, M, Nielsen, M, B, Lauridsen, C, A., The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review. *Acta Radiol* (2019) 60(1):13–8. doi: 10.1177/0284185118770917.

Hernández, A., Panizo, Á., Camacho, D., An ensemble algorithm based on deep learning for tuberculosis classification. In *Intelligent Data Engineering and Automated Learning—IDEAL 2019, Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, Manchester, UK, 14–16 November 2019*; Springer: Cham, Switzerland, 2019; pp. 145– 154.

Houfani, D., Breast cancer classification using machine learning techniques: a comparative study, Mèd Technologies J 4 (2) (2020) 535– 544.

Hungle, L, M., Manh, M, V., Tran, V, L., "Deep feature fusion for breast cancer diagnosis on histopathology images," 2019 11th International Conference on Knowledge and Systems Engineering (KSE), Da Nang, pp. 1-6, 2019.

Jamieson, A, R., Drukker, K., and Giger, M, L., "Breast image feature learning with adaptive deconvolutional networks," presented at the Medical Imaging 2012: Computer- Aided Diagnosis, 2012, vol. 8315, p. 831506.

Javaid, M., Haleem, A., Singh, R, P., R. Suman, Rab, S., Significance of machine learning in healthcare: Features, pillars and applications, International Journal of Intelligent Networks 3 (2022) 58–73.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadar- rama, S. and Darrell, T., 2014, November. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia pp. 675-678. ACM.

Jiang, Y., Edwards, A, V, Newstead, G, M. Artificial intelligence applied to breast MRI for improved diagnosis. *Radiology* (2021) 298(1):38–46. doi: 10.1148/radiol.2020200292

Jiao, Z., Gao, X., Wang, Y., and Li, J., "A deep feature based framework for breast masses classification," *Neurocomputing*, vol. 197, pp. 221–231, Jul. 2016.

Kaklamanis, M, M., Filippakis, M, E., A comparative survey of machine learning classification algorithms for breast cancer detection, in: Proceedings of the 23rd Pan-Hellenic Conference on Informatics, 2019, pp. 97–103. doi:https://doi.org/10.1145/3368640.3368642.

Kalyani, W., Prashant, P., Nikhil, W., " Breast cancer detection using ANN network and performance analysis with SVM," International Journal of Computer Engineering and Technology , vol. 10, no. 3, pp. 75-86, 2019.

Keen, J, D., Keen, J,M., Keen, J, E.. Utilization of computer- aided detection for digital screening mammography in the united states, 2008 to 2016. *J Am Coll* Demyanov, *Radiology* (2018) 15(1):44–8. doi: 10.1016/j.jacr.2017.08.033

Kemal, P., and mit, .: A Novel ML Approach to Prediction of Breast Cancer: Combining of mad normalization, KMC based feature weighting and AdaBoostM1 classifier. In: 2nd International Symposium on Multidisciplinary Studies and Inno- vative Technologies (ISMSIT),IEEE, pp. 1-4 (2018)

Khan, S, M., Hussain, H., Aboalsamh, H., Mathkour, G., Bebis, and Zakariah, M., "Optimized Gabor features for mass classification in mammography," *Appl. Soft Comput.*, vol. 44, pp. 267–280, Jul. 2016.

Khan, S., Hussain, M., H. Aboalsamh, and G. Bebis, “A comparison of different Gabor feature extraction approaches for mass classification in mammography,” *Multimed. Tools Appl.*, vol. 76, no. 1, pp. 33–57, Jan. 2017.

Khan, S., Islam, N., Jan, Z., Din, I, U., Rodrigues, J, .J, C., A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognit. Lett.* **2019**, *125*, 16.

Khuriwal, N., Mishra, N., Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm. In Proceedings of the 2018 IEEMA Engineer Infinite Conference (eTechNxT), New Delhi, India, 13–14 March 2018; pp. 1–5.

Kim, D, H., Kim, S, T., and Ro, Y, M., Latent feature representation with 3-D multi-view deep convolutional neural network for bilateral analysis in digital breast to-mosynthesis, 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, 2016, pp. 927-931.

Kim, S, Y., Deep learning-based computer-aided diagnosis in screening breast ultrasound to reduce false-positive diagnoses," Sci Rep 11 (1) (2021) 395–395.

Krizhevsky, A., Sutskever, I., and Hinton, E, G., “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

Krizhevsky, A., Sutskever, I., and Hinton, G, E., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems pp. 1097- 1105.

Kushwaha, S., Bahl, S., Bagha, A, K., Parmar, K, S., Javaid, M., Haleem, A., Singh, R, P., Significant applications of machine learning for COVID-19 pandemic, Journal of Industrial Integration and Management 5 (4) (2020) 453– 479.

Lakhani, P., Sundaram, B., Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* **2017**, *284*, 574–582.

Li, R., *et al.*, “Deep learning based imaging data completion for improved brain disease diagnosis,” in *International Conference on Medical Image Computing and Computer- Assisted Intervention*, 2014, pp. 305–312.

Lo, S, C, B., Chan, H, P., Lin, J, S., Li, H., Freedman, M, T., and Mun, S, K., "Artificial convolution neural network for medical image pattern recognition," *Neural Netw.*, vol. 8, no. 7–8, pp. 1201–1214, 1995.

M, D., W, N, S, PH. D., D, M, H, PH. D., O, L, M, PH. WİLLİAM, D., WOLBERG, H., "COMPUTERIZED BREAST CANCER DIAGNOSIS AND PROGNOSIS FROM FINE NEEDLE ASPIRATES," HUMAN ONCOLOGY, AND COMPUTER SCİENCES UNİVERSİTY OF WİSCONSİN, MADİSON, WI , NOV. 1994.

Mahbod, A., Schaefer, G., Wang, C., Ecker, R. and Ellinge, I., 2019, April. Skin lesion classification using hybrid deep neural networks. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp. 1229-1233. IEEE.

Mahmoud, K., Nedislko. S, N., "Probabilistic graphical models and deep belief networks for prognosis of breast cancer," IEEE 14th International Conference on Machine Learning and Applications, pp. 727-732, 2015.

Majid, N., Adel, A., Sewissy, T., Hassan, A, S., "Multi-class breast cancer classification using deep learning convolution neural network," International Journal of Advanced Computer Science and Application, vol. 9, no. 6, pp. 316-322, 2018.

Manogaran, G., Lopez, D., A survey of big data architectures and machine learning algorithms in healthcare, Int. J. Biomed. Eng. Technol. 25 (2–4) (2017) 182–211.

McCann, J., Stockton, D., Godward, S., Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer Res*(2002) 4(5):1–9. doi: 10.1186/bcr455

Megha, R., Vikas, P., "Hybrid approach to predict breast cancer using machine learning techniques," International Journal of Computer Science Engineering, vol. 5, no. 3, pp. 125- 136, 2016.

Memon, Q (2019), On assisted living of paralyzed persons through real-time eye features tracking and classification using Support Vector Machines, Medical Technologies Journal 3 (1), 316-333https://doi.org/10.26415/2572- 004X-vol3iss1p316-333

Mirbabaie, M., Stieglitz, S., Frick, N, R, J., Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction, Health and Technology 11 (4) (2021) 693–731.

Muduli, D., Dash, R., Majhi, B., Automated diagnosis of breast cancer using multi-modal datasets: A deep convolution neural network based approach. *Biomed. Signal Processing Control.* **2022**, *71*, 102825.

Mughal, B., Early Detection and Classification of Breast Tumor from Mammography. Doctoral Dissertation, COMSATS Institute of Information Technology, Islamabad, Pakistan, 2019.

Muhammet, F, A., Yunus C., Kadir S., Akif D., "Breast cancer diagnosis by different machine learning method using blood analysis data," International Journal of Intelligent System and Applications in Engineering, vol. 6, no. 4, pp. 289- 293, 2018.

Na, L., Qi, E., Xu, M., Bo, G., Gui-Qiu, L.,: A novel intelligent classification model for breast cancer diagnosis. Information Processing and Management, Elsevier, pp. 609-623(2019)

Naji, M, A., El Filali, S., Bouhlal, M., Benlahmar, E, H., Abdelouhahid, R, A., Debauche, O., Breast Cancer Prediction and Diagnosis through a New Approach based on Majority Voting Ensemble Classifier. *Procedia Comput. Sci.* **2021**, *191*, 481–486.

Naresh, K,, Nidhi, Mishra.,"Breast cancer detection from histopathological images using deep learning," 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering, 22-25 November 2018.

Narváez, F., Alvarez, J., Garcia-Arteaga, J, D., Tarquino, J., and Romero, E., "Characterizing Architectural Distortion in Mammograms by Linear Saliency," *J. Med. Syst.*, vol. 41, no. 2, p. 26, Feb. 2017.

Nawel, Z., Nabiha, A., Nilanjan, D., and Mokhtar, S.: Adaptive Semi Supervised Support Vector Machine Semi Supervised Learning with Features Cooperation for Breast Cancer Classification. Journal of Medical Imaging and Health Informatics, American scientific publisher, pp. 53- 62(2016)

Nguyen, Q, H., Nguyen, B, P., Dao, S, D., Unnikrishnan, B., Dhingra, R., Ravichandran, S, R., Chua, M,C., Deep learning models for tuberculosis detection from chest X- ray images. In Proceedings of the 2019 26th International Conference on Telecommunications (ICT), Hanoi, Vietnam, 8–10 April 2019; pp. 381–385.

Nikita Rane, Jean Sunny, RuchaKanade, Prof. Sulochana Devi, "Breast cancer classification and prediction using machine learning," International Journal of Engineering Research & Technology, vol. 9, no. 02, pp. 576-580, 2020.

Nithya, R., and Santhi, B., "Classification of normal and abnormal patterns in digital mammograms for diagnosis of breast cancer," *Int. J. Comput. Appl.,* vol. 28, no. 6, pp. 21–25, 2011.

Osman, A, H., Aljahdali, H. M. A., An Effective of Ensemble Boosting Learning Method for Breast Cancer Virtual Screening Using Neural Network Model, IEEE Access 8 (2020) 39165–39174.

Pan, Y., *et al.*, "Brain tumor grading based on neural networks and convolutional neural networks," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 699–702.

Panuwat M., Nutnicha T.,"Deep learning algorithms for predicting breast cancer based on tumor cells," 4th International Conference on Digital Arts, Media and Technology and 2nd ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, pp: 343-346, 2019.

Penatti, O, A., Nogueira, K., and dos Santos, J, A., "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 44–51.

Preetha, R., Vinila, S, J., "A research on breast cancer prediction using data mining techniques," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 11S2, pp. 362-370, 2019.

Puspanjali, M., Baldev, P., Samikshya, S., "Enhancing histopathological breast cancer image classification using deep learning," International Journal of Innovative technology and Exploring Engineering, vol. 8, no. 7, pp. 2024-2032, 2019.

Raghavendra, U., Rajendra A, H., Fujita, A., Gudigar, J,Tan, H., and Chokkadi, S., "Application of Gabor wavelet and Locality Sensitive Discriminant Analysis for automated identification of breast cancer using digitized mammogram images," *Appl. Soft Comput.*, vol. 46, pp. 151–161, Sep. 2016.

Rajaraman, S., Antani, S, K., Modality-specific deep learning model ensembles toward improving TB detection in chest radiographs. *IEEE Access* **2020**, *8*, 27318–27326.

Rajpurkar, P., (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. ArXiv, 1711.05225.

Rodrigues, L., Analysis of the Wisconsin Breast Cancer Dataset and Machine Learning for Breast Cancer Detection, in: and others (Ed.), Proceedings of XI Workshop de Visão Computacional, 2015, pp. 15–19.

Rodríguez-Ruiz, A., Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System, Radiology 290 (2) (2019) 305–314.

Rodríguez-Ruiz, A., Krupinski, E., Mordang, J, J., Schilling, K., Heywang-Köbrunner, S, H., Sechopoulos, I.; Mann, R, M. Detection of breast cancer with mammography: Effect of an support system. *Radiology* **2019**, *290*, 305–314.

Rodríguez-Ruiz, A., Krupinski, E., Mordang, J-J., Schilling, K., Heywang-Köbrunner, S, H., Sechopoulos, I., et al.. Detection of breast cancer with mammography: Effect of an artificial intelligence support system. *Radiology*(2018) 290(2):305–14. doi: 10.1148/radiol.2018181371

Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Broeders, M., Gennaro, G., Clauser, P., Sechopoulos, I., Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists. *JNCI J. Natl. Cancer Inst.* **2019**, *111*, 916–922.

Roelofs, A, A, J., et al., Importance of Comparison of Current and Prior Mammograms in Breast Cancer Screening, Radiology 242 (1) (2007) 70–77, doi:https://doi.org/10.1148/radiol.2421050684.

Roth, J, A., Battegay, M., Juchler, F., Vogt, J, E., Widmer, A, F., Introduction to machine learning in digital healthcare epidemiology, Infect. Control Hosp. Epidemiol. 39 (12) (2018) 1457–1462.

Rudin, C., Ustun, B., Optimized scoring systems: toward trust in machine learning for healthcare and criminal justice, Interfaces 48 (5) (2018) 449–466.

S., Chakravorty, R., Abedini, M., Halpern, A., and Garnavi, R., Classifica- tion of dermoscopy patterns using deep convolutional neural networks, 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, 2016, pp. 364-368.

Salama, G, I., Abdelhalim, M, B., Zeid, M, A., Breast Cancer Diagnosis on Three Different Datasets Using Multi- Classifiers, International Journal of Computer and Information Technology 01 (1) (2012) 36–43. URL https://www.ijcit.com/index.php/ijcit

Sansya, V, V., Lekshmy, P, L., "Deep learning based prediction of breast cancer in histopathological images," International Journal of Engineering Research & Sze, V.,

Sarah, M., : Cancers du sein et Immunologie anti-tumorale. UNIVERSITE DE REIMS CHAMPAGNE- ARDENNE, ECOLE DOCTORALE SCIENDES TECH- NOLOGIE SANTE (547), 2014, Doctoral thesis.

Saxena, S., Gyanchandani, M., A Model for Classification of Wisconsin Breast Cancer Datasets using Principal Component Analysis and Back- Propagation Neural Network, International Journal of Science and Researc h (IJSR) 8 (2018) 1324–1327. doi:10.21275/ART20199863. URL www.ijsr.net

Schlegl, T., Ofner, J., and Langs, G., "Unsupervised pre- training a image domains improves lung tissue classification," in *International MICCAI Workshop on Medical Computer Vision*, 2014, pp. 82–93.

Seyyid, A., Medjahed, T., Ait S., Abdelkader B., "Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules," International Journal of Computer Application, vol. 62, no. 1, pp. 1-5, 2013.

Shamy, S., Dheeba, J., "A research on detection and classification of breast cancer using k- means GMM & CNN algorithms," International Journal of Engineering and Advanced Technology, vol. 8, no. 6S, pp. 501-505, 2019.

Sharif, A, R., Azizpour, H., Sullivan, J., and Carlsson, S., "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806–813.

Shen, L., Margolies, L, R., Rothstein, J, H., Fluder, E., McBride, R., Sieh, W., Deep learning to improve breast cancer detection on screening mammography. *Sci. Rep.* **2019**, *9*, 12495.

Shen, W., Zhou, M., Yang, F., Yang, C., and J. Tian, "Multi- scale convolutional neural networks for lung nodule classification," in *International Conference on Information Processing in Medical Imaging*, 2015, pp. 588–599.

Shin, H, C., *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285– 1298, May 2016.

Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K.I., Matsui, M., Fujita, H., Kodera, Y. and Doi, K., 2000. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection o

Shravya, C, H., Pravalika, K., Shaik S., "Prediction of breast cancer using supervised machine learning techniques," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 6, pp. 1106-1110, 2019.

Shubham, S., Archit, A., Tanupriya, C., "Breast cancer detection using machinelearningalgorithms," InternationalConferenceonComutati nalTechniques,Electronics and Mechanical System(CTEMS), pp. 114-118, 2018.

Shwetha, K., Spoorthi, M., Sindhu, S, S., Chaithra, D., "Breast cancer detection using deep learning technique," International Journal of Engineering Research & Technology, vol. 6, no. 13, pp. 1-4, 2018.

Siddique, S., Chow, J, C., Machine learning in healthcare communication, Encyclopedia 1 (1) (2021) 220–239.

Siegel, R, L., Miller, K, D., Fuchs, H, E., Jemal, A., Cancer statistics, 2022. *CA: A Cancer J Clin* (2022) 72(1):7–33. doi: 10.3322/caac.21708

Sivapriya, J., Aravind, K,V., Siddarth, S, S., Sriram, S., "Breast cancer prediction using machine learning," International Journal of Recent Technology and Engineering, vol. 8, no. 4, pp. 4879- 4881, 2019.

Sornam, M., Kavita M., Vanitha, V., "A survey on images classification and activity recognition using deep convolution neural network architecture," Ninth

International Conference on Advanced Computing (ICoAC), pp.121-126, 2017.

Sri, H, N., Pragnyaban, M., Suvarna, V, K., "Breast cancer detection using machine learning way," International Journal of Recent Technology and Engineering, vol. 8,

Srirambabu, S., Santhosh, K., Senthil, V, B, M., "Classification of breast cancer using svm classifier technique," International Journal of Advanced Research in Computer Science, vol. 1, no. 4, pp. 261-263, 2010.

Suckling, J., *et al.*, "The Mammographic Image Analysis Society Digital Mammogram Database," *Exerpta Medica*, vol. 1069, pp. 375–378, 1994. no. 2S3, pp. 1402-1405, 2019.

Tahmooresi, M., Afshar, A., Bashari, B., Rad, K, B., Nowshath, M. A. Bamiah, "Early detection of breast cancer using machine learning techniques," Journal of Telecommunication, Electronic and Computer Engineering, vol. 10, no. 3-2, pp. 21- 27, 2018.

Tajbakhsh, N., *et al.*, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1299– 1312, May 2016.

Tan, Y.J., Sim, K, S., Ting, F, F., Breast cancer detection using convolutional neural networks for mammogram imaging system. In Proceedings of the 2017 International Conference on Robotics, Automation and Sciences (ICORAS), Melaka, Malaysia, 27–29 November 2017; pp. 1–5.

Thuy, M, B, H., Hoang, V, T, Fusing of deep learning, transfer learning and gan for breast cancer histopathological image classifica- tion. In *Advances in Intelligent Systems and Computing, Proceedings of the 6th International Conference on Computer Science, Applied Mathematics and Applications, ICCSAMA 2019, Hanoi, Vietnam, 19–20 December 2019*; Springer: Cham, Switzerland; pp. 255–266.

Tina E, M., "Simple and ensemble decision tree classifier based detection of breast cancer," International Journal of Scientific & Technology Research, vol. 8, no. 11, pp. 1628- 1637, 2019.

U.S. Cancer Statistics Working Group. United States Cancer Statistics: 19992008 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and

Human Services, Centers for Disease Control and Prevention, and National Cancer Institute; 2012.

Van, B., Ginneken, A, A, A., Setio, C., Jacobs, and Ciompi, F., "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 286–289.

Varsha, J., Gaikwad, "Detection of breast cancer in mammogram using support vector machine," International Journal of Scientific Engineering and Research, vol. 3, no. 2, pp. 26- 30, 2015.

Vishal, D., Mukta, S, "Breast cancer detection using SVM classifier with grid search techniques," International Journal of Computer Application, vol. 178, no. 31, pp. 18- 23, 2019.

Vikas, C., Saurabh, P., : A Novel Approach for Breast Cancer Detection using Data Mining Techniques. International Journal of Innovative Research in Computer and Communication Engineering. Vol. 2, Issue 1, pp. 2456-2465(2014) http://www.springer.com/lncs. Last accessed1 February 2019

Wang, S., Rao, R, V., Chen, P., Zhang, Y., Liu, A., and Wei, L., "Abnormal Breast Detection in Mammogram Images by Feed-forward Neural Network Trained by Jaya Algorithm," *Fundam. Informaticae*, vol. 151, no. 1–4, pp. 191–211, Jan. 2017.

Wang, Z., Li, M., Wang, H., Jiang, H., Yao, Y., Zhang, H., Xin, J., Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE Access* **2019**, *7*, 105146–105158.

Waring, J., Lindvall, C., Umeton, R., Automated machine learning: review of the state-of-the-art and opportunities for healthcare, Artif. Intell. Med. 104 (2020), 101822.

Weal, E. F., Amr, S., G., "A deep learning approach for breast cancer mass detection," International Journal of Advanced Computer Science and Applications, vol. 10, no. 1, pp. 175-182, 2019.

Wei, B., Han, Z., He, X., Yin, Y., Deep learning model based breast cancer histopathological image classification. In Proceedings of the 2017 IEEE 2nd

International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 28– 30 April 2017; pp. 348–353.

Wichakam, I., and Vateekul, P., Combining deep convolutional networks and SVMs for mass detection on digital mammograms, 2016 8th International Conference on Knowledge and Smart Technology (KST), Chiangmai, 2016, pp. 239-244.

Wilkinson, L., Gathani, T., Understanding breast cancer as a global health concern. *Br J Radiol* (2022) 95(1130):20211033–. doi: 10.1259/bjr.20211033

Wolterink, J, M., Leiner, T., Viergever, M, A., and Išgum, I. “Automatic coronary calcium scoring in cardiac CT angiography using convolutional neural networks,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 589–596.

Xu, J., Xiang, L., Hang, R., and Wu, J., Stacked Sparse Autoencoder (SSAE) based framework for nuclei patch classification on breast cancer histopathology, 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), Beijing, 2014, pp. 999-1002.

Yawen X., Jun W., Zongli, L., Xiaodong Z., “ Breast cancer diagnosis using an unsupervised feature extraction algorithm based on deep learning,” Proceedings of the 37th Chinese

Yi, D., *et al*., “Optimizing and Visualizing Deep Learning for Benign/Malignant Classification in Breast Tumors,” *ArXiv170506362 Cs*, May 2017.

Zhang, Y, D., Wang, S, H., Liu, G., and Yang, J., “Computer- aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform,” *Adv. Mech. Eng.*, vol. 8, no. 2, p. 1687814016634243, Feb. 2016.

Zheng, J., Lin, D., Gao, Z., Wang, S., He, M., Fan, J., Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis. *IEEE Access* **2020**, *8*, 96946–96954.

Zhu, W., Lou, Q., Vang, Y, S., and Xie, X., “Deep Multi- instance Networks with Sparse Label Assignment for Whole Mammogram Classification,” *ArXiv161205968 Cs*, Dec. 2016.

APPENDICES

Appendix A

Table 1: Using Various Datasets to Compare Past Research For the Identification of Breast Cancer

Dataset	Ref	Year	Classifiers	ACC
WBC	Nilashi [8]	2017	CART	93.20%
	Chaurasia et al. [9]	2018	NB	97.36%
	Wang et al. [10]	2020	RIPPER	95.39%
	Mojrian et al. [11]	2020	ELM-RBF	95.69%
	Bayrak et al. [12]	2019	SMO	96.90%
WDBC	Ramos et al. [13]	2019	LDA	98.82%
	Najmu et al. [14]	2020	DT	97.29%
	Sharma et al. [15]	2018	KNN	94.00%
	Rufai et al. [16]	2020	SVM	94.30%
	Salama et al. [17]	2013	SMO	97.7%.
WPBC	Ojha & Goel[18]	2017	SVM	68.00%
	Pritom et al. [19]	2017	SVM	75.70%
	Salama et al. [17]	2013	fusion of MLP, J48, SMO and IBK	77.00%
	Kiage [20]	2015	NB, KNN, RT	73.00%
	Chi et al. [21]	2007	ANN	64.90%

Table 2: Wisconsin Dataset Description

Dataset name	WDBC
Instances	569
Attribute	31
Attribute Type	Real
Classes	Benign (B) & Malignant (M)
Classes distribution	B= 357 and M= 212
Missing Values	No missing value

Table 3: Wisconsin Breast Cancer Dataset (WBC) Features

NO.	Feature	Domain
1	instance code number	id number
2	Thickness of Clump	From one to ten
3	Size Uniformity	From one to ten
4	Shape Uniformity	From one to ten
5	Marginal Adhesion	From one to ten
6	Single Epithelial Cell Size	From one to ten
7	Bare Nuclei	From one to ten
8	Bland Chromatin	From one to ten
9	Normal Nucleoli	From one to ten
10	Mitoses	From one to ten
11	Class	2 for benign, and 4 for malignant

Table 4: WDBC Dataset Features Information

NO	Features
1	Id
2	Diagnosis
3-32	3-32). Eachnucleusisdescribed by ten computed characteristics.
Texture	
Radius	
Area	
Perimeter	

Compactness
Smoothness
Concavity
Concave points
Symmetry
Fractal dimension

Table 5: Correlation Matrix

		actual values	
		Malignant (Positive)	Benign (Negative)
predicted values	Malignant (Positive)	TP	FN
	Benign (Negative)	FP	TN

Table 6: Quality of block evaluations for WBC dataset using machine learning

Algorithm	70 Training and 30 Testing

ML	ACC	AUC	PREC	REC	F1 SCORE
RF	92	92	86	94	90
MLP	91	90	88	88	88
KNN	83	82	76	79	78
SVM	86	84	86	76	81

.

.

Appendix X

Turnitin Similarity Report

(A plagiarism report is included at the end of the thesis immediately before the CV of the author.)

CV

Note: Please refer to the Institute's Guidelines for Thesis Writing to better understand this template!