MACHINE LEARNING ALGORITHM USING CLASSIFICATION AND REGRESSION TREES, LINEAR GAUSSIAN NAIVE BAYES, SUPPORT VECTOR MACHINES, CART AND K-NEAREST NEIGHBORS TO AID CANCER RESEARCH FIRMS IN DATA GROUPING AND BETTER ANALYSIS.

21206051022

MASTER OF SCIENCE

in Informatics engineering.

THESIS

by

Rebeccah Ndungi

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-277/Un.02/DST/PP.00.9/01/2023

Tugas Akhir dengan judul : MACHINE LEARNING ALGORITHM USING CLASSIFICATION AND REGRESSION TREES, LINEAR GAUSSIAN NAIVE BAYES, SUPPORT VECTOR MACHINES, CART AND K-NEARST NEIGHBORS TO AID CANCER RESEARCH FIRMS IN DATA GROUPING AND BETTER ANALYSIS

yang dipersiapkan dan disusun oleh:

Nama : REBECCAH NDUNGI
Nomor Induk Mahasiswa : 21206051022
Telah diujikan pada : Jumat, 20 Januari 2023
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR

Ketua Sidang

Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.
SIGNED

Valid ID: 63d3775c6af01

Penguji I

Dr. Agung Fatwanto, S.Si., M.Kom.
SIGNED

Valid ID: 63d339b1b687d

Penguji II

Ir. Muhammad Taufiq Nuruzzaman, S.T. M.Eng., Ph.D.
SIGNED

Valid ID: 63d0b8abe2cd5

Yogyakarta, 20 Januari 2023
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 63d5cd91e1137

DECLARATION OF AUTHENTICITY

Assalammualikum Warrahmatullahi Wabarokatuh

I here with:

Name	: Rebeccah Ndungi
Student ID	: 21206051022
Grade	: Post Graduate
Department	: Informatics

declare that I wrote this thesis on my own and did not use any unnamed sources or aid. Thus, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person except where due reference is made by correct citation The work contained in this thesis has not been previously submitted for examination.

Wassalammualikum Warrahmatullahi Wabarokatuh.

Kenya, 18 January 2023

Rebeccah Ndungi

21206051022

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
MAGISTER INFORMATIKA
Jl. Marsda Adisucipto Telp. (0274) 515856 Yogyakarta 55281

SURAT PERSETUJUAN TUGAS AKHIR

Hal : Persetujuan Tugas Akhir

Kepada:
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
Di Yogyakarta

Assalamualaikum wr.wb.
Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka saya selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Rebbecah Ndugi
NIM : 21206051022
Judul Tesis : *MACHINE LEARNING ALGORITHM USING CLASSIFICATION AND REGRESSION TREES, LINEAR GAUSSIAN NAIVE BAYES, SUPPORT VECTOR MACHINES, CART AND K-NEAREST NEIGHBORS TO AID CANCER RESEARCH FIRMS IN DATA GROUPING AND BETTER ANALYSIS.*

Sudah dapat diajukan kepada Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Magister Informatika.

Dengan ini saya mengharap agar tugas akhir tersebut di atas dapat segera dimunaqosyahkan. Atas perhatiannya saya ucapkan terimakasih.

Yogyakarta, 16 Januari 2023
Pembimbing

Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.
NIP. 19780106 200212 2 001

ABSTRACT

World Health Organization publications disclose that breast cancer is one of the most common diseases amongst the women, and that it has a high death rate. Its prevalence is growing in developing nations, where the vast majority of cases are discovered late in the disease's development. Preventing this malignancy may be accomplished with safeguards and frequent examinations. Aside from that, early detection of the illness may be beneficial in terms of battling the condition. This deep learning algorithm is a strong approach for identifying features about the breast mass that are often not visible to the human eye through deep learning analysis of the mammogram. The algorithm will then be evaluated through various metrics to validate on its accuracy. Deep learning is a technology that will be used in this project. The algorithm intends to help cancer firm research centres in Kenya and globally in quicker grouping, identification and aid doctors in quicker diagnosis and research.

Keywords: Breast cancer,Deep Learning,WHO,Health,Machine learning, Africa.

Contents

List of Figures

List of Tables

Acronyms

BCE:	Breast Self Examination
Ke:	Kenya
WHO:	World Health Organization
ML:	Machine Learning
KNH:	Kenyatta National Hospital
CC:	Cranio-caudal
MLO:	Medial-Lateral Oblique
SaaS:	Software as a Service
RT:	Radio Therapy
DSS:	Decision Support System
GBCI:	Global Breast Cancer Initiative
FNA:	Fine Needle Aspirate
CART:	Classification and Regression Tree
svm:	Linear Support Vector Machines
NB:	Gaussian Naive Bayes
KNN:	K-Nearest Neighbors

Acknowledgement

I'd want to say a big "thank you" to my Supervisors for providing me with the opportunity to complete this fantastic project on Mammography analysis using ML algorithms and thereby allowing me to learn so much via my research. Second, I'd like to express my gratitude to my family and friends, without whom I would not have been able to finish this project in time.

1 INTRODUCTION

1.1 Background

Cancer of the breast is one of the primary reasons why women lose their lives. The treatment of malignant cancer in tissue has led to a significant increase in the number of visual examinations for Cancer of the breast In the field of cancer research, the categorization of The diagnosis of cancer requires the examination of tissue samples[1]. This can be accomplished through the utilization of quantitative inference and extraction models to prevent further degradation of the tissue growth. The anatomical classification of the tissues that surround the malignant cancer cells can differentiate into benign and malignant cells predicting categories is an extremely difficult and challenging task using data mining algorithms. A fictional story is presented herein quantitative inferences can be made about breast cancer tissues explored by using a mechanism with only partial supervision. The records in the dataset are examined using a semi-supervised method grouped using the method of starting with the most distant members. The grouped together The use of an artificial neural network is employed to classify the records and classifiers based on Naive Bayesian-ism. The strategy that was suggested is assessed through the utilization of k-fold cross validation [1].

According to the Ministry of health (MOH) in Kenya, After infectious diseases and cardiovascular conditions, cancer is the third biggest cause of mortality worldwide. In 2018, it was anticipated that there would be 47,887 new cases of cancer diagnosed each year, while the number of deaths caused by cancer would total 32,987.Breast, cervical, and oesophageal cancers are the most prevalent forms of the disease in females, whereas colorectal, prostate, and esophageal cancers are the most prevalent forms of the disease in males.On Figure 1: Figure 1.1, Oesophageal cancer is the primary cause of death from cancer in Kenya, accounting for 13.2% (4,351 deaths) of the country's overall cancer mortality rate wheres Figure 2: Figure 1.2 shows that Breast cancer is the third largest cause of death from cancer, accounting for 7.7% (2,553 fatalities), whereas cervical cancer is the second greatest cause of death from cancer, accounting for 10% (3,266 deaths) (GLOBOCAN, 2018).

According to [26] Cancer of the breast is the most prevalent form of the disease in females. It is estimated that 2.3 million people around the world are newly diagnosed with BC every single year. On the basis of the levels of mRNA gene expression, breast cancer can be broken down into a number of different molecular subtypes. These molecular subtypes provide insights into new treatment strategies and patient stratifications, both of which have an effect on the management of breast cancer patients. This review provides

an overview of the epidemiology of breast cancer, including risk factors, classification, and molecular types, with an emphasis on prognostic biomarkers and possible treatment modalities.

Carcinogenesis, which can take place in any cell, tissue, or organ, is characterized by six major hallmarks and can lead to the pathological alterations that cause a wide variety of cancers. These alterations can be fatal [26]. The ability to evade apoptosis, limitless capacity to divide, enhanced angiogenesis, resistance to anti-growth signals and induction of own growth signals, and the capacity to metastasize are the primary mechanisms that enable its progression . Carcinogenesis is a multifactoral process that is primarily stimulated by both genetic predispositions and environmental causes. Carcinogenesis is a term that refers to the development of cancer. The alarmingly high number of deaths that can be attributed to cancer continues to rise each year, solidifying their position as one of the leading causes of death across the globe. Even though a significant number of cancers do not always have to result in death, they still significantly reduce the quality of life and require significantly higher costs in general[26].

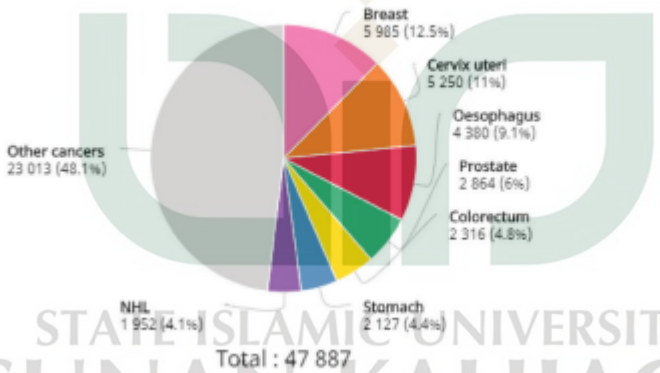

Figure 1: Figure 1.1

Late-stage presentation, when it is difficult to find a cure, is a common problem in Kenya, as it is in numerous LMICs (WHO, 2018b), when treatment and diagnostic services are poor or non-existent. This is the case in Kenya as well (WHO, 2017). According to research conducted at Kenyatta National Hospital between 2014 and 2016, around 64% of cancer patients were diagnosed at stage III or stage IV, which is when it is extremely challenging to find a treatment that would cure the disease. Registration and surveillance of cancer in Kenya have not been doing as well as they should.

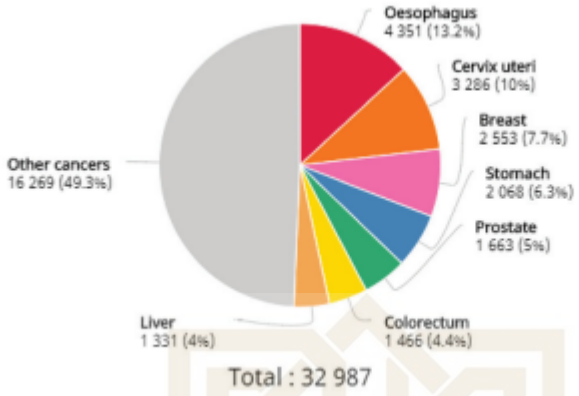

Figure 2: Figure 1.2

Eldoret and Nairobi each have their own regional population-based cancer registries, and together they cover approximately 10% of Kenya's total population. These registries are currently in operation. When it comes to malignancies that affect children, a lack of information and stigma among parents, guardians, and caregivers leads to patients presenting themselves to cancer treatment centers later than they should[3]. Diagnostic facilities with found to be ineffective, personnel, and consumables are one of the many factors that contribute to poor outcomes. Other contributing factors include other aspects.

The cancer load can be greatly decreased with early detection and care of precancerous diseases. While around one-third to one-half of all malignancies can be avoided, this percentage can be increased. On the other hand, relatively few people in Kenya take advantage of screening services. For instance, just 16 percent of women aged 30 to 49 get screened for cervical cancer, which is a far lower rate than the 47 percent of women who are aware that screening programs are available. It is regrettable that this is the case given that screening can discover early stages of several of the most common malignancies[4].

Prevention, early detection, and screening for cancer are the three primary priorities of the Kenya National Cancer Control Strategy (NCCS) 2017-2022. An early diagnosis can lead to improved treatment outcomes, less morbidity, and even reduced overall expenses associated with therapy. It is possible to accomplish this through the use of screening and early diagnosis. Screening can enhance the likelihood of finding cancer in its early stages in certain types of the disease. When appropriate tests are used in a competent manner and quality assurance is implemented, screening programs have the potential to be helpful for specific malignancies. This is followed by linkage to diagnostic

and treatment services [5].

Cancer that is detected in its early stages has a far better chance of responding favorably to treatment and of being cured entirely. An intervention known as secondary prevention is referred to as early detection when it entails the discovery of a disease at an early, pre-symptomatic stage when a patient would not have any reason to seek medical attention[5].

In Kenya, where the majority of patients present at more late stages of disease, this is an essential public health initiative that has the potential to have a significant impact. Its primary objective is to find symptomatic patients as early as possible by recognizing possible warning indications of cancer in order to take quick action. This can be accomplished by recognizing possible warning indicators of cancer. It is feasible to accomplish this goal through education, which can raise people's awareness of the probable warning symptoms of cancer among health care practitioners as well as among the general population. The provision of care at the earliest possible time is intended to achieve the goal of improving treatment outcomes [5].

The table Figure 1 Table 1 that follows provides a list of signs and symptoms that are connected with certain malignancies and that can assist in the early detection of these tumors. These should be recognized as probable warning symptoms of cancer by health care providers, and urgent action should be taken to diagnosis these malignancies at an earlier stage[5].

CANCER	SIGNS & SYMPTOMS
Cervix	Post-coital bleeding, excessive vaginal discharge
Breast	Lump in the breast, asymmetry, skin retraction, recent nipple retraction, blood-stained nipple discharge, eczematous changes in areola
Colorectal	Change in bowel habits, unexplained weight loss, anaemia, blood in the stool
Stomach	Upper abdominal pain, recent onset of indigestion, weight loss
Urinary bladder	Pain, frequent and uneasy urination, blood in urine
Prostate	Difficulty in urination, frequent nocturnal urination
Head & neck cancers	Lump in nose, throat, or neck (with or without pain), persistent sore throat, difficulty swallowing (dysphagia), persistent cough, hoarseness or change in voice, ear pain or hearing loss, persistent headaches, persistent bad breath not explained by hygiene, nasal obstruction or persistent congestion, difficulty breathing, frequent nose bleeding or unusual discharge,
Oral cavity	White lesions (leukoplakia) or red lesions (erythroplakia), growth or ulceration in mouth
Retinoblastoma	White spot in the pupil, convergent strabismus (in a child)

Figure 3: Table 1

1.2 Mammography

Mammography is a diagnostic method that creates an image of the breast by subjecting it to a low dosage of radiation in order to produce an image of it in a manner that is very similar to an x-ray. The result of this procedure is a mammogram. Mammograms are recognized as an essential tool for the early detection of abnormalities in women's breasts because they can show minuscule changes that are typically undetected by simple palpation or breast self-examination. This makes mammograms an essential instrument for the early detection of abnormalities in the breasts of women[6].

Mammogram can be used to examine symptomatic individuals, which is called diagnostic mammography, as well as to screen asymptomatic women for breast cancer (selected age groups)[7].One or more usually two views of each breast are obtained during a standard mammographic screening examination. These views are referred to as CC and MLO [7]. Figure 4: Figure b shows am MLO with a view of the subscapularis muscle, and there are some distinct benign nocs present.

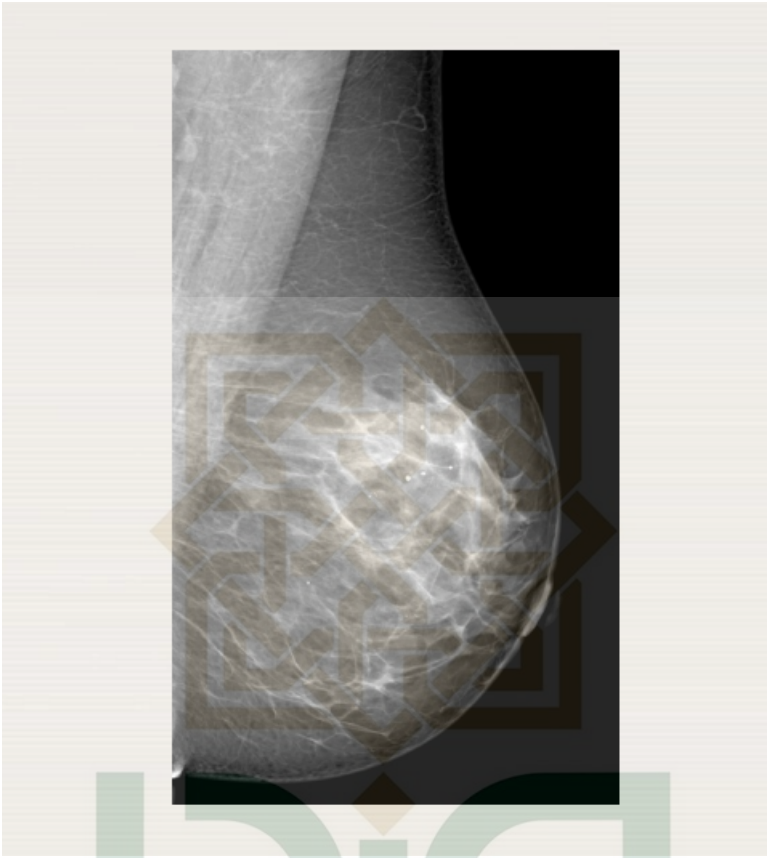

Figure 4: Figure b

Outpatient procedures such as mammography are typically carried out in a hospital or clinic laboratory setting. The steps to complete the process are as follows:In the room where the mammogram is being performed, the technician will inquire about the patient's personal and medical history. Patients are required to provide pertinent medical history when asked whether they have felt any lumps in their breasts, have recently undergone breast surgery, or have breast implants. In order to differentiate the breasts in the photographs, the technician will place adhesives on the genitals and other markings (such as birth marks) on the breasts[8].

After that, the patient will be instructed on how to place her breasts on the mammography machine. To alleviate the patient's suffering, there is a specialized platform that can be moved to different positions and is adjustable according to the patient's height. After providing support by placing the breast on top of this, the breast will next be squeezed using two flat, transparent panels.After that, a series of photographs of each breast from

a variety of angles will be shot. As soon as the photos have been obtained, the device will release the breast, and then it will begin the process all over again with the other breast. The amount of views that are taken will be determined by whether or not there are any abnormalities that need to be investigated, as well as whether or not the doctor has ordered an angle that is routinely available[8].

After the technician has obtained the necessary number of photos, the procedure is considered complete. After that, the patient is discharged and given an appointment to return for a follow-up appointment at which the results of the examination will be explained. Patients are urged to do the following things in order to ensure that the surgery is carried out correctly:Since mammograms might cause some discomfort, it is best to schedule the procedure during a time when the patient's breasts are less sensitive, such as at least one week after or prior to the start of their period.On the day of the mammography, refrain from applying deodorant, lotion, or body powder because these items have the potential to appear as calcium spots on the scan. Ask them whether they would rather have a female technician perform the procedure or whether they would be okay with a man technician performing it.Please bring any past mammograms for comparison. Mention to the technician whether or not the patient has breast augmentation so that they can keep track of it[6].

Using machine learning and other approaches, the writers of this study article [13] gathered data from ten separate breast cancer studies. When a breast tumor spreads to other parts of the body, the condition is known as breast cancer. Breast cancer is the most prevalent cancer in women worldwide, and its prevalence is increasing in developing countries, where the vast majority of cases are diagnosed at a late stage. Cancer can develop in either one or both breasts and can begin in either one or both of them if abnormal cell division occurs. According to the graph, breast cancer affects virtually exclusively women and men [13].Breast cancer can begin in a variety of breast sites. The breast is a female reproductive organ that rests on top of the upper ribs and chest muscles. Each booby has ducts, ducts, and fatty tissue. There are two boobies, one on each side. Each side possesses two boobs [13]. The mammary gland produces and secretes milk which is used to nurture neonates and infants. The degree of fatty tissue in each breast dictates the unique size of each breast. The global cancer burden is increasing, putting a pressure on the health-care systems at every socio - economic level[13]. After infectious diseases and cardiovascular disease, cancer is Kenya's third largest cause of death. According to the 2018 GLOBOCAN report by the International World Cancer Research fund, there are around 47,887 new instances of cancer recorded year, with 32,987 fatalities (IARC). Compared to the previous study, which expected 37,000 new cases of cancer every year in 2012, with a death rate of 28,500 cases every year in 2012, this is an increase of almost 45 percent in frequency [13].

On the other hand, types of cancer of the breast, cervix uteri (womb), oropharynx (food pipe), bladder (prostate cancer), and colon and rectum (colorectum) are the most ubiquitous both in males and women of all ages, with testicular cancer being its leading reason for cancer-related deaths, supplemented by bony cancer and breast cancer. It's vital to be aware that the majority of hyperplastic mammary are not malignant but rather benign growths (malignant) (malignant) [25].Cancer cells clinical features are malignant benign tumors that do not spread outside the chest and are therefore categorized as noncancerous. Although benign emphysematous breasts are not fatal, they may raise a woman's chance of developing breast cancer in the future. If you discover a tumor or change in your breast, consult a health care provider to determine whether it is malignant or benign (cancer), or whether it may raise your chance of having cancer in the future.

When classifying cancers, neoplasia are typically sorted into subtypes according to their degree of metastasis or lack thereof. One or more variables predicted by a classification system are corroborated by the presence of characteristics with the opposite distribution. Data processing software is required to put the classification strategies into action. Categorization is used

to determine which treatment option will be the least complicated. Classification is vital because it provides a framework for scientists to recognize, arrange, and name species according to agreed-upon criteria. Classification methods and clustering methods are two widely used approaches in the world of data processing [16].

Clustering techniques must first extract information from a data set and characterize the data set before they can generate knowledge clusters from it. Classification, also known as supervised learning in computer vision, is a method that aims to classify novel events by learning patterns and classes from the input set, and then making predictions based on the classification. The terms "training set" and "test set" are frequently heard while discussing classification jobs. The test set is used to assess the classifier's efficacy once it has been developed using the training set. Categorization can be an arduous optimization problem to resolve in some scenarios. Researchers are employing numerous machine learning strategies in an effort to classify these datasets [17].

Notable algorithms for classifying and predicting breast cancer include: neural networks, randomized forests, support vector machines, and others. The scientific community is constantly on the lookout for the simplest approach that will yield the most precise classification result; yet, data of varied reliability will also have an effect on the classification result. The number of times an algorithm is used will also be affected by the lack of data. With early detection, not only do patients have a better chance of survival, but they also have a wider range of treatment options to choose from. When detected within the first 5 years of the cancer's development, the survival percentage for women with breast cancer is 93% or higher. Getting regular checkups could provide you peace of mind about your health. Deaths could be prevented with early cancer detection[18].

1.3 Problem Formulation

World Health Organization publications disclose that breast cancer is one of the most common diseases amongst the women, and that it has a high death rate. Its prevalence is growing in developing nations, where the vast majority of cases are discovered late in the disease's development. Preventing this malignancy may be accomplished with safeguards and frequent examinations. Aside from that, early detection of the illness may be beneficial in terms of battling the condition. This deep learning algorithm is a strong approach for identifying features about the breast mass that are often not visible to the human eye through deep learning analysis of the mammogram. The algorithm will then be evaluated through various metrics to validate on its

accuracy. Deep learning is a technology that was used in this project. The algorithm intends to help cancer firm research centres in Kenya and globally in quicker grouping, identification and aid doctors in quicker diagnosis and research.

1.4 Problem Limitations

The major limitations in the development and deployment of the project will be; People adjusting to a computerised systems and Lack of smart devices among various research firms in Kenya.

1.5 Research Objectives

- Increase the level of collaboration and interdisciplinary care provided to breast cancer patients in Breast Units.

  providing a unified, multidisciplinary, multi-scale holistic view of the patient (digital patient) and its current needs; allowing for the handling of a vast amount of information that is heterogeneous, multi-scale, dynamic, and timely; and allowing for the handling of the information that is generated during the course of treatment. through the provision of a web-based Software as a Service (SaaS) collaboration environment that includes enhanced intuitive visual exploration interfaces for the purpose of analyzing, contrasting, and presenting the patient case based on a complicated digital breast cancer patient model.

- Utilizing unusual or previously undiscovered sources of data

  imaging, genetic and biological data, data on the administration of therapeutics, risk factors, or environmental or social aspects that are not used or under exploited in clinical practice but may have important prognostic or diagnostic value and may impact decision making may be examples of such information. By creating a complex digital breast cancer patient (DBCP) model that represents the case, incorporating all of this data in a structured way for agile exploration and case portrayal, as well as instinctive data mining and visualization tools that are capable of retrieving and comparing similar cases and testing the impact of some of these parameters on retrospective data, evaluating its potential decision- making value as well as influence.

- Utilizing the vast amounts of information that are available from standard imaging examinations

  both diagnostic and prognostic value through the provision of quantitative imaging biomarkers that are both helpful and objective. These can be computed and compared not just over the course of treatment but also between different cases, making use of the data that has been obtained in retrospective cases that have previously been resolved. Through the creation of highly automated and advanced medical image-analysis techniques, which will be integrated into web-based medical image patch analysis tools or as virtualized batch systems. These tools and processes will be accessible over the internet.

- The development of methods for the visual evaluation of the potential aesthetic results of Female Conservative Treatment

enhancing the connection between both the patients and the surgeons, and having the potential to have repercussions in the reduction of subsequent interventions or in the prognostic effects of therapeutic compounds, such as radiation or systematic treatments By incorporating, adapting, and tweaking an existing patient-specific multi-scale pharmacological model of BCT, which couples a genetic model of wound healing to a modeling method of the glandular tissue, and which may also combine the effects of systemic and RT treatments. This model couples a biological model of healing process to a mechanical model of the breast tissues.

- Offer judgment help

  Because to the wide variety of possible treatments for PBC (including surgical alternatives, radiation therapy (RT) treatments, and (neo) adjutant systemic therapies), By constructing a decision support system (DSS) that is based on a complex understand model that evolves by trying to incorporate the experience of the clinical team on past instances, decisions, and outcomes (defined by the DBCP data model), forecasts about treatment outcome (i.e. from the models), and the opinion and context knowledge of the patient.

References

[1] Chinniyan, K., & Subramani, R. (2021). Breast Cancer Prediction Using Machine Learning Algorithms.

[2] Kenya National Cancer Screening Guidelines. Nairobi, November 2018.

[3] Njuguna 2016, NATIONAL CANCER CONTROL STRATEGY Kenyatta National Hospital, Cancer treatment centre.

[4] STEPS survey, 2015; Ng'ang'a et al., 2018.KENYA STEPwise SURVEY FOR NON COMMUNICABLE DISEASES RISK FACTORS 2015 REPORT.

[5] Njuguna 2016, National-Cancer-Screening-Guidelines-2018.Kenyatta National Hospital, Cancer treatment centre.

[6] American Cancer Society. American Cancer Society Recommendations for Early Breast Cancer Detection in Women Without Breast Symptoms. Last revised October 24, 2013. Available at: www.cancer.org/cancer/breastcancer/moreinformation/breastcancerearlydetection/breast-cancer-early-detection-acs-recs. Accessed January 29, 2015.

[7] https://humanhealth.iaea.org/HHW/MedicalPhysics/TheMedicalPhysicist/Studentscorner/

[8] American College of Obstetricians and Gynecologists (ACOG). ACOG practice bulletin no.122. Breast cancer screening. Obstet Gynecol. 2011;118:372-382. PMID 21778569. Available at: www.ncbi.nlm.nih.gov/pubmed/21775869.

[9] Fehniger J, Livaudais-Toman J, Karliner L, Kerlikowske K, Tice JA, Quinn J, Ozanne E, Kaplan CP. Perceived versus objective breast cancer risk in diverse women. J Womens Health (Larchmt). 2014 May;23(5):420-7. doi: 10.1089/jwh.2013.4516. Epub 2013 Dec 28. PMID: 24372085; PMCID: PMC4011422.

[10] https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[11] http://www.desiree-project.eu/content/general-objetives

[12] https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[13] Thomas DB. Breast cancer in men. Epidemiol Rev. 1993;15(1) 220-231. doi:10.1093/oxfordjournals.epirev.a036108. PMID: 8405206.

[14] Yavuz, E., Eyupoglu, C. An effective approach for breast cancer diagnosis based on routine blood analysis features. Med Biol Eng Comput 58, 1583–1601 (2020). https://doi.org/10.1007/s11517-020-02187-9

[15] Institute of Medicine (US) and National Research Council (US) National Cancer Policy Board; Curry SJ, Byers T, Hewitt M, editors. Washington (DC): National Academies Press (US); 2003.

[16] Sunny, Jean & Rane, Nikita & Kanade, Rucha & Devi, Sulochana. (2020). Breast Cancer Classification and Prediction using Machine Learning. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS020280.

[17] Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x

[18] Sunny, Jean & Rane, Nikita & Kanade, Rucha & Devi, Sulochana. (2020). Breast Cancer Classification and Prediction using Machine Learning. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS020280.

[19] Murali, Swapana & Padmavathy, C.. (2015). A Critical Review on Breast Cancer Literature: Screening, Awareness and Preventive Measures. Mediterranean Journal of Social Sciences. 6. 10.5901/mjss.2015.v6n4s3p256.

[20] Carter, Jeanne & Park, Elyse & Moadel, Alyson & Cleary, Sean & Morgan, Carole. (2002). Cancer knowledge, attitudes, beliefs, and practices (KABP) of disadvantaged women in the South Bronx. Journal of cancer education : the official journal of the American Association for Cancer Education. 17. 142-9. 10.1080/08858190209528822.

[21] Ms. M. Swapana, Dr. C. Padmavathy. A Critical Review on Breast Cancer Literature:Screening, Awareness and Preventive Measures.Mediterranean Journal of Social Sciences MCSER Publishing, Rome-Italy.

[22] Özaras, Gözde & Durualp, Ender & Civelek, Fethiye & Uensal, Münevver. (2010). Analysis of Breast Self-Examination Training Efficiency in Women between 20-60 Years of Age in Turkt. Asian Pacific journal of cancer prevention : APJCP. 11. 799-802.

[23] Dandash, Khadiga & Al-Mohaimeed, Abdulrahman. (2007). Knowledge, Attitudes, and Practices Surrounding Breast Cancer and Screening in Female

Teachers of Buraidah, Saudi Arabia. International journal of health sciences. 1. 61-71.

[24] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," Procedia Comput. Sci., vol. 83, no. Fams, pp. 1064–1069, 2016.

[25] Hassan, S.A., Sayed, M.S., Abdalla, M.I. et al. Breast cancer masses classification using deep convolutional neural networks and transfer learning. Multimed Tools Appl 79, 30735–30768 (2020). https://doi.org/10.1007/s11042-020-09518-w

[26] Łukasiewicz S, Czeczelewski M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast Cancer-Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies-An Updated Review. Cancers (Basel). 2021 Aug 25;13(17):4287. doi: 10.3390/cancers13174287. PMID: 34503097; PMCID: PMC8428369.

[27] Arooj, Sahar Rahman, Atta Zubair, Muhammad Farhan Khan, Muhammad & Alissa, Khalid & Khan, Muhammad & Mosavi, Amir. (2022). Breast Cancer Detection and Classification Empowered With Transfer Learning. Frontiers in Public Health. 10.3389fpubh.2022.924432.

[28] Vyas, Sarthak & Chauhan, Abhinav & Rana, Deepak & Ansari, Mohd. (2022). Breast Cancer Detection Using Machine Learning Techniques. International Journal for Research in Applied Science and Engineering Technology. 10. 10.22214/ijraset.2022.43055.

[29] M. A. Mohammed, M. K. A. Ghani, R. I. Hamed, and D. A. Ibrahim, "Analysis of an electronic methods for nasopharyngeal carcinoma: prevalence, diagnosis, challenges and technologies," Journal of Computational Science, vol. 21, pp. 241–254, 2017.

[30] F. Moreno-Seco, L. Micó, and J. Oncina, "A modification of the LAESA algorithm for approximated k-NN classification," Pattern Recognit. Lett., vol. 24, pp. 47–53, 2003.

[31] M. A. Mohammed, M. K. A. Ghani, R. I. Hamed, and D. A. Ibrahim, "Review on Nasopharyngeal Carcinoma: concepts, methods of analysis, segmentation, classification, prediction and impact: a review of the research literature," Journal of Computational Science, vol. 21, pp. 283–298, 2017.

[32] V. Chaurasia and S. Pal, "Data mining techniques: to predict and resolve breast cancer survivability," International Journal of Computer Science and Mobile Computing, vol. 3, no. 1, pp. 10–22, 2014.

[33] Lewis, Roger. (2000). An Introduction to Classification and Regression Tree (CART) Analysis.

[34] Peterson, Leif. (2009). K-nearest neighbor. Scholarpedia. 4. 1883. 10.4249/scholarpedia.1883.

[35] Wibawa, Aji Kurniawan, Ahmad Murti, Della Adiperkasa, Risky Perdana Putra, Sandika Kurniawan, Sulton Nugraha, Youngga. (2019). Naïve Bayes Classifier for Journal Quartile Classification. International Journal of Recent Contributions from Engineering, Science IT (iJES). 7. 91. 10.3991/ijes.v7i2.10659.

[36] Evgeniou, Theodoros & Pontil, Massimiliano. (2001). Support Vector Machines: Theory and Applications. 2049. 249-257. 10.1007/3-540-44673-7$_1$2.