BREAST CANCER DETECTION USING NANOPARTICLE SENSOR WITH MACHINE LEARNING

J.N.V.R. Swarup Kumar, Charishma Karri,
Naga Srihitha Vatsavayi, Harshitha Lekkala, G. Srinivas

Department, Of CSE, GST, GITAM Deemed to be University,
Visakhapatanam, 530045, Andhra Pradesh, INDIA.

Contributing authors: sjavvadi2@gitam.edu; charishma497@gmail.com;
nagasrihithavatsavayi@gmail.com; harshi.april27@gmail.com;
sgorla@gitam.edu;

Abstract

This research is mainly focused on the early detection of breast cancer in women by their urine samples with nanoparticle sensors. It detects certain enzymes and proteins that can be the main cause of cancer by machine learning algorithms, correlation analysis, and logical regression methods and created a web-based breast cancer prediction website using started nanoparticle urine analysis data, which contains the DNA barcode sequences that match with DNA of patient urine which is user friendly to use for every person This study mainly focuses on non-invasive cancer. We analyze non-invasive cancer by urine sample test data collected from the patients by utilizing advanced technology to access DNA signatures associated with breast cancer biomarkers. Our approach involves the barcode of nanoparticle sensors and matches with the urine samples. After the urine samples match, they are applied to the sensors, translated into digital data, and transmitted to a centralized system. Now, the centralized system collaborates with datasets derived from previous breast cancer cases. The algorithm now analyses the urine data and identifies the patterns. Then, it correlates with different stages of breast cancer.

Keywords: breast cancer, nanoparticle sensor, machine learning, SQL, urine test.

1 Introduction

Detection of breast cancer in the early stages is still an issue worldwide. Out of many cancers, breast cancer is the 2nd top all over the world. Breast cancer is caused by unhealthy cells that change the shape or color of the breast. This change is called a tumor, and if this tumor is increased and not treated. Then it could called it has breast cancer. There are two types of breast cancer: invasive and non-invasive breast cancer. Most of the patients are affected by non-invasive breast cancer [1]. There are four stages, each defining how much the breast is affected. Besides these types, there are sub-types of breast cancer. They are Hormone receptor-positive or Hormone receptor-negative; Ancient Egyptians first detected HER2-positive or HER2-negative triple-negative breast cancer.

In 460 B.C., A Breast Lump is caused when the tumor (unhealthy cells/tissues) spreads and forms like a clot. Even the Breast becomes hardened and thicker around the nipple. It is caused by Paget's disease of the breast, known for changing the breast shape and the color of the nipple. It can also be altered after or before swelling. After the change of its shape and color, the outward nipple turns into an inward nipple [2]. The skin around the breast changes color into red or orange. There's even a possibility of blood discharge from the outward/inward nipple. This is mainly caused by reproductive history, which means the person can get breast cancer more than once.

There's even a high possibility of women getting breast cancer because of Genetics or Family History. Women who consume alcohol or smoke can develop breast cancer. A breast cancer patient should quit smoking and drinking [3]. Some women who have dense breasts might end up getting breast cancer, too. To remove this rampant breast cancer, there's a therapy called breast removal, where the surgeon removes both breasts. There are multiple therapies and surgeries such as Lumpectomy, Mastectomy, Sentinel node biopsy, Axillary lymph node dissection, Radiation therapy, Chemotherapy Immunotherapy, and much more to remove breast cancer.

Table 1: Symptoms, Causes, Treatment, Durability

Symptoms	Causes	Treatment	Durability
A Breast Lump	Fibroadenomas	Excisional biopsy	2 Weeks
Uneven skin	Inflammatory breast cancer	Chemotherapy	3 Months
Breast Swelling	Mastitis	Antibiotics/drugs	5 Days to 3 Months
Flaking of pigmented skin	Eczema	Emollients	3 to 4 Weeks
Pain in/around the nipple	Hormonal fluctuations	Antithyroid drugs	Depends upon surgery

In the above table, you can witness some of the other symptoms and their treatments, and the durability of breast cancer symptoms leads to early detection of cancer, which can be identified easily.
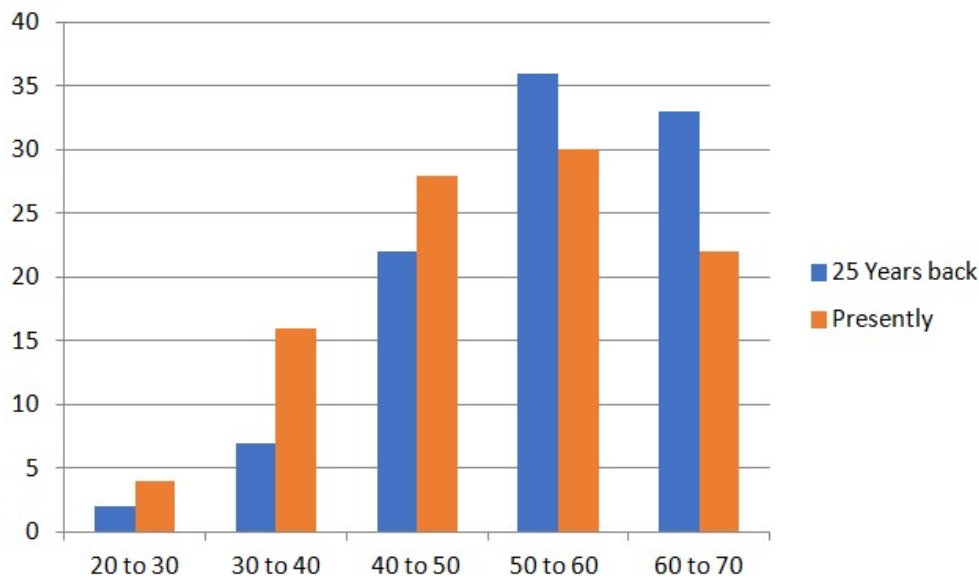
Fig. 1: Breast Cancer Analysis in India

The above bar graph represents the statistics of breast cancer patients from the past 25 years, which are in blue color and red in the present scenario where at the age of 50-60 years, we can witness more cases where more hormonal changes occur in the women due to menopause which comes with a lot of hormonal changes so early detection can be a huge advantage

2 Literature Review

Our area of research mainly focuses on early detection, which should be simple, like a COVID-19 rapid test, which we studied apart from the quick test] Shoko Kure 1, Shinya Iida 2,3,4 [4] researched by using volatile organic compounds (VOCs), we can train the sniffer dogs to detect breast cancer. The trained dogs acknowledge the potential of VOC analysis and conduct the research without using ML algorithms/devices. GCMS stands for gas chromatography-mass spectrometry, used when the dogs start to sniff, and VOC analyzes the potential. It is an entirely new approach to detecting breast cancer. Rapid tests can be done within the home in around 2 minutes.

In their researched way, it takes time and has potential only to certain limits. The conclusion is that dogs can only predict whether breast cancer is alive. Where as using machine learning model we can specify the stage. Jee Yee Kim [5] did research and found out that the detection of breast cancer can be done by using a blood test. The research aims to identify RNA results where they compare the blood of breast cancer patients and non-breast cancer patients by using developed machine learning techniques. The study predicts the values like mammograms. Then, the values are entered into the machine. The aim of the research is created by using prediction and

detection tests, but the results might turn into poor outcomes. In rapid tests, there needs to be more usage of developed tests and better outcomes. Only one outcome is breast cancer with the stage in the following research paper; Julia Beretov and Yong Li [6] used urine to detect breast cancer. The urine sample is monitored as potential for non-invasive breast cancer. The evaluation of urine protein specifies breast cancer tissues and tests the urine sample; first, it is kept under surveillance and prevention for early diagnosis. This research paper is similar to the rapid test. The only difference between those two is the nanoparticle sensor and biosensors. This nano biosensor value matches the value of urine protein.[4] did not use this nano biosensor concept; thus, it makes the detection of early breast cancer take time to detect cancer.

There are many possible ways to detect breast cancer, such as urine samples[7], blood samples, tears samples, and many more. Even dogs are used to sniff the person to identify breast cancer. Machine learning algorithms are used to identify the accuracy percentage of breast cancer with the help of support Vector Machines (SVM), Radical Basis Neural Networks, and Random Forests [8]. embedded sensors, which have ultrasensitive micro bio heat with IOT technology, are used with real-time data to achieve great accuracy in the detection of breast cancer. Digital images are used to detect and analyze the accuracy of breast cancer percentage[9]. In machine learning, techniques are also used, such as decision tree (DT), K-nearest neighbors (KNN), and Naive Bayes (NB) [10], which are tested to predict breast cancer with real medical images. However, this does not classify the stages of breast cancer. It even takes time to identify breast cancer. whereas Rapid tests can identify breast cancer at an early stage with the stages.

Compared to all the above, early detection can be possible by nanoparticles nearby, which are used mainly in rapid tests by comparing these research using nano particles undergoing (MIT)[11]. Some students from MIT have researched and shown that they could use the sensors to detect cancer tumors by five different enzymes. Still, creating a nanoparticle sensor in real-time takes so instead of that, we worked on a database that contains the biomarker values and DNA barcode values when the urine sample it predicts cancer as that database and symptoms and structure of tumor database is merged and provided to machine learning model using logical regression it shows accuracy and using flask we created a website where you enter the values of urine_sample, biomarker value, and sensor_value in specified range it predicts the person has cancer or not these make everyone's work simple as technology in our hand we provided a model made by machine learning to the website it gives 95 percent accuracy in every case which you can witness in our results.

Table 2: Existing systems

Researcher(s)	Method	Outcomes	Limitations	Comparison to NanoParticle sensor approach
Shoko Kure et al	Training for sniffer dogs using VOCs	Dogs can detect cancer by sniffing	Time consuming and not machine learning based, limited to detecting without the stage	Proposed method is quicker, stages cancer, and uses machine learning for analysis
Jee Yee Kim	Blood tests to identify Ribonucleic Acid (RNA)	The values that are predicted and are similar to mammograms	Does not specify the cancer stage because of poor outcomes	Proposed method can specify the cancer stage and has high accuracy
Julia Beretov et al	Urine sample analysis for proteins	Monitors the possibility of non-invasive Breast cancer	Requires surveillance and prevention for early diagnosis, no rapid test	Proposed method is similar but enhanced with nanoparticle biosensors for fast detection
MIT students	Sensors to detect cancer enzymes	Shows potential for the detection of tumor	Creation of sensors in real-time is challenging	Uses a prepared database for fast prediction without time-consuming sensor creation

In the following research, we proposed the following Methodology [11]: Early detection of breast cancer techniques Developed with the assistance of a Machine Learning model and the website which was created particularly for that purpose. The proposed model [12] is the main part of this research paper and is highlighted with a diagram that has a clear view of the prototype. In the finding section[13], the model is developed with results which are obtained from the Ada-boost classifier, Gradient Classifier, and MLP. These classifiers are known as Machine Learning algorithms. The results of urine samples between cancerous and non-cancerous people are included in the form of correlation matrices. The graph is even drawn to specify the difference between cancerous and non-cancerous people. Additionally, photocopies that distinguish the people present with or without Breast cancer by website function are added. This research paper concludes with aims to detect Breast cancer in the early stage.

3 Methodology

Table 3: Algorithm Breast Cancer Detection System

Steps	Description
Initialize four tables	Patient, DNA_Barcode, Nano_Sensor, and Urine_Sample with relevant attributes.[14]
Combined the tables	Merge the four tables on Patient_ID to form a comprehensive dataset D.
Consolidate Additional attributes	Integrate D with a local database containing additional attributes such as symptoms, tumor structure, and classification.
Reputable dataset	Clean D by removing unwanted attributes and null values to form a cleaned dataset D [15].
Generate pair plots	It Generates pair plots and a correlation heatmap from D' to identify relevant features.
Split the dataset	Split D into features X and target Y based on the defined outcome variable.
Split (X, Y)	Split (X, Y) into training and testing sets using a 75:25 ratio.
Normalize features	Normalize X to have a uniform scale across all features.
Train (X, Y)	Train a machine learning model using the training set (X_train, Y_train)
Evaluate Performance	Evaluate the model's performance on the testing set (X_test, Y_test) to obtain accuracy and other metrics.
Trained model	Serialize the trained model into a.Pkl file for deployment.
Website interface	Create an HTML, CSS, and JavaScript web interface that takes user input for urine_value, biomarker_value, and sensor_value [16].
Threshold values	Set thresholds for these values to predict the presence of cancer.
Website interface with model	Deploy the web interface linked to the machine learning model, allowing users to receive predictions based on their input.
End algorithm	Return Web-based Prediction System with the trained model

In the following table, we explain the algorithm that we implemented, specify the steps and methodologies of machine learning models, and initialize the database of urine analysis DNA barcodes and how machine learning splits the train database set, and this trained database is taken and builds a website using (HTML, CSS, JAVASCRIPT)

4 Proposed Model

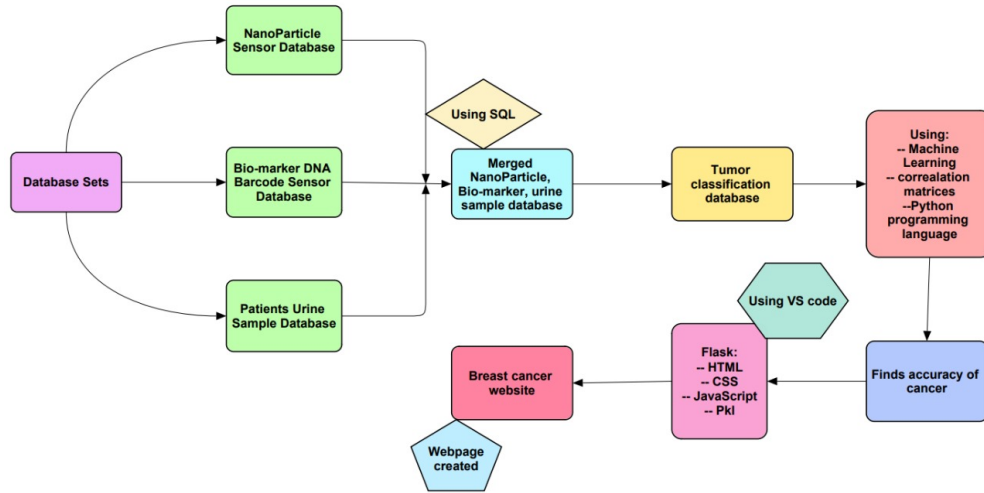The proposed model is represented in a systematic diagram, as shown in Figure **??**.

Fig. 2: Process flow diagram for proposed system

The flow chart defines a process from the start to the end. The process started with developing a database containing the proteins' biomarkers and DNA barcode values of a urine sample from Breast cancer patients. The urine samples are collected from clinical sources. The Biomarker values are collected from TCGA, EBI, and NCI. The collected datasets are merged and matched with the Kaggle symptom data using the machine learning model [14]. Later, the ML model is developed to Analyse and provide the matched data by the prediction. The website is developed to predict and generate the input data in real time. Here is the preview model of the website.

Fig. 3: Prediction of a normal person

Fig. 4: Prediction of a Breast cancer person

This website is about early breast cancer prediction form, and this tool is built to predict cancer at its early stages by taking inputs in three categories: Urine value, Biomark value, and sensor value, which includes basic details like (Name, Date of Birth, and Gender). This is a user-friendly platform that can be accessible on the internet, and this webpage is basically built with a machine learning model.

5 Results & Discussion

In the Bagging classifier, the scores are precisely achieved with 93% for class Benign and 88% for class Malignant. This precise score results in F1-scores with 94% for class Benign and 85% for class Malignant.
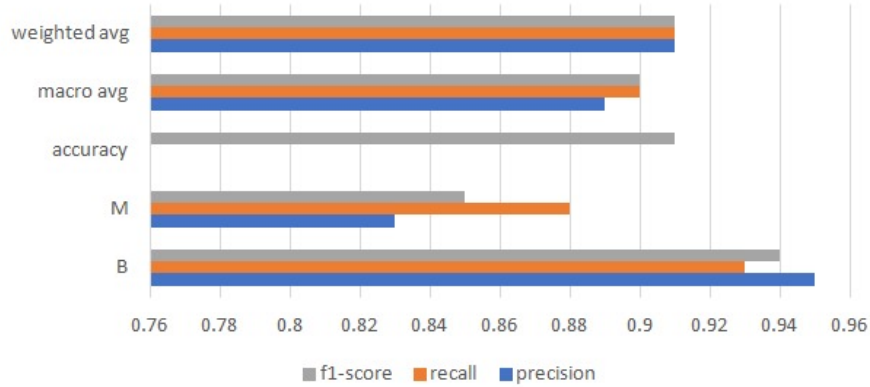

Fig. 5: Bagging Classifier Report

The above bar graph represents the precision, recall, and F1-score for class B and class M in breast cancer analysis.

The KNN classifier reveals both the recall scores for class B, with 96% and 91%, and for class M, with 91%.
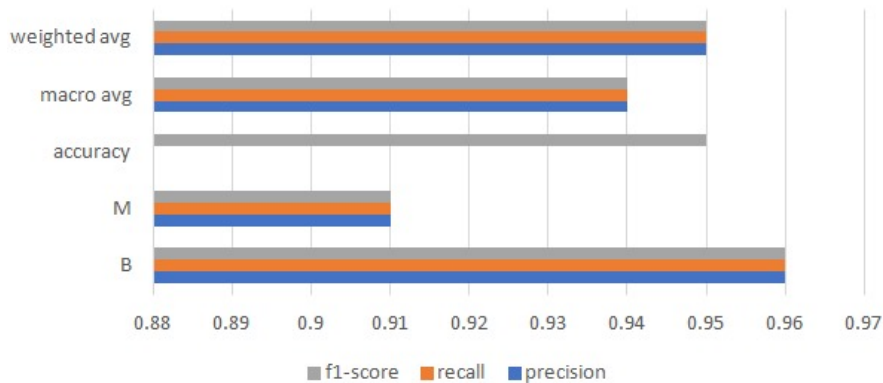

Fig. 6: KNN Classifier Report

The above bar graph represents the precision, recall, and F1-score for class B and class M in KNN Classifier Report.

The KNN classifier reveals both the recall scores for class B, with 96% and 91%, and for class M, with 91% and 91%
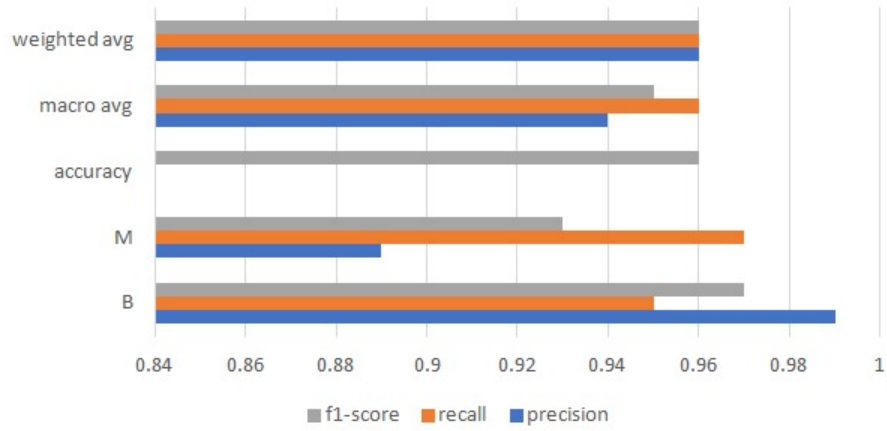

Fig. 7: Random Forest Classifier Report

The above bar graph represents the precision, recall, and F1-score for class B and class M in Random Forest Classifier Report.

The Ada-boost classifier ends the result with both precision and recall at 99% and 94% for class B, 86%, and 97% for class M. The accuracy of the results of both class B and class M is 95%
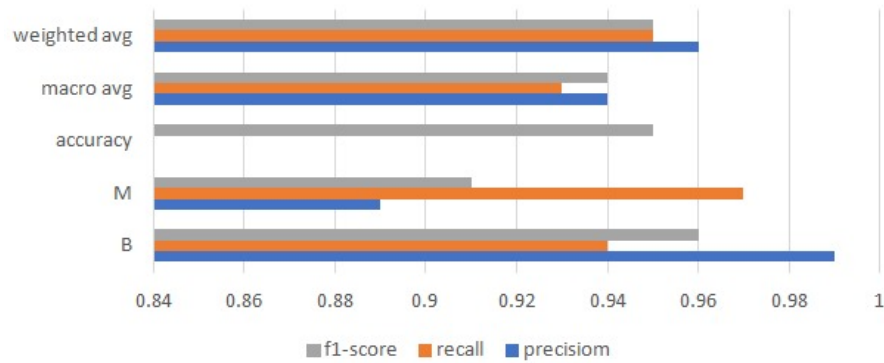

Fig. 8: Ada-boost classifier report

The above bar graph represents the Ada-boost classifier report's precision, recall, and F1-score for class B and class M.

Fig. 9: Gradient-boot Classifier Report

The above bar graph represents the Gradient-boot Classifier report's precision, recall, and F1-score for class B and class M.

The Gradient Boosting results showed 97% and 95% precision and recall for class B. For class M, the precision and recall scores are 89% and 94%. The F1-score resulted in 96% and 91% for classes B and M

Table 4: MLP final report

	Precision	Recall	F1-score	Support
Accuracy			0.95	144
B	0.99	0.95	0.96	81
M	0.86	0.94	0.91	33
Macro avg	0.93	0.95	0.94	114
Weighted avg	0.95	0.95	0.95	114

The multi-layer perceptron(MLP) classifier achevied and overall accuracy of 95% in Breast cancer detection with precision scores of 99%, the model demonstrates strong capability identifying in both the cases so, for prediction of Breast cancer we are choosing (MLP) [15]

There are a total five classifiers out of them, only the Random Forest Classifier and MLP Classifier have the highest accuracy of 96%, so considering some of the factors like database set and resource allocation comparing both the random and MLP and Random classifier, MLP would have highest accuracy percentage of 96% and high recall scores for both the cases the MLP report is as follows which we used for breast cancer prediction and using MLP report we imported into pkl doc to develop a website these are the values.

11

6 Conclusion

Based on our research findings and proposed model, which we implemented using a nanoparticle sensor database set and DNA barcode along with a machine learning model, we can further develop a comprehensive diagnosis tool using blood tests and VOC analysis because this could be easy to use. Results are predicted accurately. Nano biosensors not only help detect Breast cancer but can also work on many other cancers. It can also reduce detection time, which is very sensitive to use and development, and research should focus on real-time monitoring systems of the patient's biomarker values. This could be a great advantage in the treatment process, and these methods are very cost-effective and can be afforded by every single person.

References

[1] World Health Organization: Breast Cancer. https://www.who.int/news-room/fact-sheets/detail/breast-cancer. Accessed: February 21, 2024(n.d.)

[2] Cleveland Clinic: Nipple Discharge. https://my.clevelandclinic.org/health/symptoms/21014-nipple-discharge. Accessed: February 21, 2024(n.d.)

[3] Mayo Clinic: Mastectomy to Prevent Breast Cancer. https://www.mayoclinic.org/tests-procedures/mastectomy/about/pac-20394670#:~:text=Mastectomy%20to%20prevent%20breast%20cancer&text=A%20preventive%20mastectomy%20involves%20removing,mastectomy%20and%20risk%2Dreducing%20mastectomy. Accessed: February 21, 2024(n.d.)

[4] Benet, J.G., Seo, M.-K., Khine, M., Padró, J.G., Martnez, A.P., Kurdahi, F.J.: Breast cancer detection by analyzing the volatile organic compound (voc) signature in human urine. Scientific Reports **12** (2022)

[5] Park, S., Ahn, S., Kim, J.Y., Kim, J., Han, H., Hwang, D., Park, J., Park, H.S., Park, S., Kim, G.M., Sohn, J.H., Jeong, J., Song, Y.U., Lee, H., Kim, S.I.: Blood test for breast cancer screening through the detection of tumor-associated circulating transcripts. International Journal of Molecular Sciences **23** (2022)

[6] Kure, S., Iida, S., Yamada, M., Takei, H., Yamashita, N., Sato, Y., Miyashita, M.: Breast cancer detection from a urine sample by dog sniffing: A preliminary study for the development of a new screening device, and a literature review. Biology **10** (2021)

[7] Paci, E.: Mammography and beyond: developing technologies for the early detection of breast cancer. Breast Cancer Research **4**, 123–125 (2002)

[8] Amethiya, Y., Pipariya, P., Patel, S., Shah, M.: Comparative analysis of breast cancer detection using machine learning and biosensors. Intelligent Medicine (2021)

[9] D, H., S, A.D., Begum, A., Hemanth, P.: Early detection of breast cancer with iot: A promising solution. 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 1–6 (2023)

[10] Sandhu, J.K., Kaur, A., Kaushal, C.: A review of breast cancer detection using the internet of things and machine learning. 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 145–150 (2023)

[11] Beretov, J., Wasinger, V.C., Millar, E.K., Schwartz, P., Graham, P.H., Li, Y.: Proteomic analysis of urine to identify breast cancer biomarker candidates using a label-free lc-ms/ms approach. PLoS ONE **10** (2015)

[12] Rovshenov, A., Peker, S.: Performance comparison of different machine learning techniques for early prediction of breast cancer using wisconsin breast cancer dataset. 2022 3rd International Informatics and Software Engineering Conference (IISEC), 1–6 (2022)

[13] Elouerghi, A., Bellarbi, L., Afyf, A., Talbi, T.F.: A novel approach for early breast cancer detection based on embedded micro-bioheat ultrasensitive sensors: Iot technology. 2020 International Conference on Electrical and Information Technologies (ICEIT), 1–4 (2020)

[14] Nancy Alaswad: Breast Cancer Dataset. https://www.kaggle.com/datasets/nancyalaswad90/breast-cancer-dataset/discussion/334181. Accessed: February 21, 2024(n.d.)

[15] GitHub User: Charishma116: Early Detection of Breast Cancer Repository. https://github.com/Charishma116/early_decetion_of_breastcancer. Accessed: February 21, 2024(n.d.)

[16] GitHub User: srihitha2004: Breast Cancer Repository. https://github.com/Srihitha2004/Breast-cancer-. Accessed: February 21, 2024(n.d.)