Breast Cancer Detection Using Nanoparticle Sensor with Machine Learning Algorithms

J.N.V.R. Swarup Kumar
Dept. of CSE, GST
GITAM Deemed to be University
Visakhapatnam, India
sjavvadi2@gitam.edu

Charishma Karri
Dept. of CSE, GST
GITAM Deemed to be University
Visakhapatnam, India
charishma497@gmail.com

Naga Srihitha Vatsavayi
Dept. of CSE, GST
GITAM Deemed to be University
Visakhapatnam, India
nagasrihithavatsavayi@gmail.com

Harshitha Lekkala
Dept. of CSE, GST
GITAM Deemed to be University
Visakhapatnam, India
harshithalekkala27@gmail.com

Saaketh Choudarapu
Dept. of CSE, GST
GITAM Deemed to be University
Visakhapatnam, India
saakethsays@gmail.com

***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 using urine sample test data collected from the patients, utilizing advanced technology to access DNA signatures associated with breast cancer biomarkers. Our approach involves the barcode of nanoparticle sensors, which matches 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.**

***Index Terms*—breast cancer, nanoparticle sensor, machine learning, SQL, urine test**

I. 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 [16]. There are four stages, each defining how much the breast is affected. Besides these types, there are subtypes 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 [4]. 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 [10]. 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 I: 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.

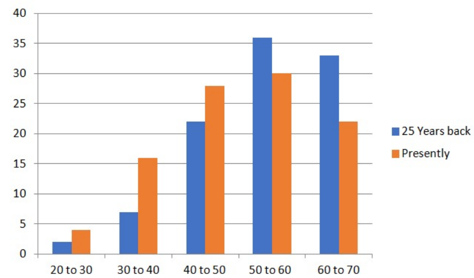


Fig. 1: Breast Cancer Analysis in India

The above 1 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.

II. Literature Review

Our area of research mainly focuses on early detection, which should be simple, like a COVID-19 rapid test. There are many possible ways to detect breast cancer, such as urine samples [12], 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 [1]. 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 [5]. In machine learning, techniques are also used, such as decision tree (DT), K-nearest neighbor (KNN), and Naive Bayes (NB) [15], 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 nanoparticles undergoing (MIT) [3]. 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 II: Existing systems

Researcher(s)	Method and Outcomes
Shoko Kure et al	Training for sniffer dogs using VOCs [2]. Dogs can detect cancer by sniffing. **Limitations:** Time-consuming, not machine learning-based, limited to detecting without the stage. **Outcomes:** Proposed method is quicker, stages cancer, and uses machine learning for analysis.
Jee Yee Kim	Blood tests to identify Ribonucleic Acid (RNA) [13]. Predicted values similar to mammograms. **Limitations:** Does not specify cancer stage due to poor outcomes. **Outcomes:** Proposed method can specify cancer stage and has high accuracy.
Julia Beretov et al	Urine sample analysis for proteins [9]. Monitors non-invasive breast cancer possibility. **Limitations:** Requires surveillance and prevention for early diagnosis, no rapid test. **Outcomes:** Similar method enhanced with nanoparticle biosensors for fast detection.
MIT students	Sensors to detect cancer enzymes [2]. Shows potential for tumor detection. **Limitations:** Real-time sensor creation challenging. **Outcomes:** Uses prepared a database for fast prediction without time-consuming sensor creation.

In the following research, we proposed the following Methodology [3]: 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 [14] 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[6], 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 the detection of breast cancer in its early stages.

A. *Nano-particle-integration*

Nanotechnology holds significant promise for the future of cancer detection, treatment, and diagnosis. Nanosensors have the potential to revolutionize cancer detection by serving as highly sensitive tools for early diagnosis. Our research indicates that these nanosensors can function as versatile sensing tools, offering a wide range of applications in the field of oncology.

Our findings suggest that developing a rapid diagnostic test using nanoparticle sensors is feasible. These sensors are coated onto a tissue-like paper, onto which a sequence of DNA barcodes and biomarkers specific to certain cancers are embedded. When urine comes into contact with this nanoparticle-coated paper, a reaction occurs that is observable under sunlight or white light. This reaction causes a color change in the paper, typically from white to light pink or violet. However, if the paper turns black, it indicates the presence of unhealthy habits or potentially cancerous conditions.

The process of utilizing these nanoparticle sensors in cancer detection involves:

Coating nanosensors onto tissue-like paper: Nanosensors are applied onto a paper-like substrate, creating a platform for interaction with biological samples.

Embedding DNA barcodes and biomarkers: Specific DNA barcodes and biomarkers indicative of various cancers are incorporated into the nanosensor coating, enabling targeted detection.

Interaction with urine samples: Urine samples from individuals are introduced onto the nanoparticle-coated paper, allowing for the detection of specific biomarkers present in the samples.

Observation of color change: Under sunlight or white light, the nanoparticle-coated paper undergoes a color change reaction upon interaction with biomarkers in the urine samples.

Interpretation of results: The observed color change in the paper serves as an indicator of potential cancerous conditions or unhealthy habits, enabling rapid and non-invasive diagnosis.

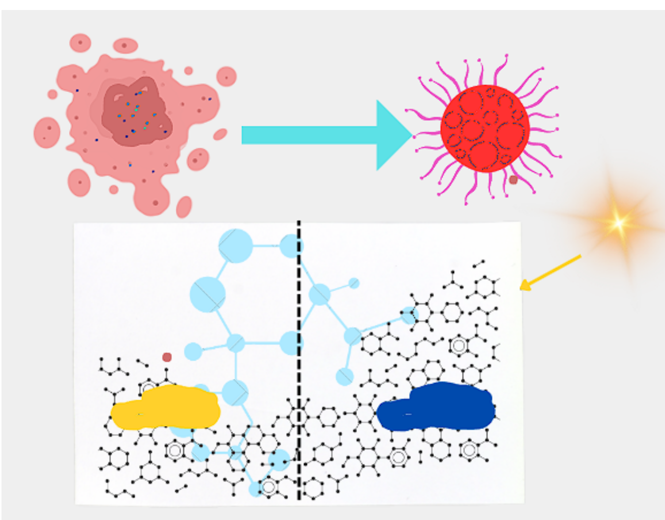


Fig. 2: rapid test results

The development of such a rapid diagnostic kit holds immense potential for improving cancer detection and diagnosis, offering a convenient and accessible method for early intervention and treatment. Further research and refinement of this technology are necessary to optimize its sensitivity, specificity, and clinical utility.

III. Methodology

Initialize Tables: Begin by creating four tables: Patient, DNA_Barcode, Nano_Sensor, and Urine_Sample, each containing relevant attributes.

Combine Tables: Merge the four tables based on the common attribute Patient_ID to create a comprehensive dataset D.

Consolidate Additional Attributes: Integrate D with a local database containing additional attributes such as symptoms, tumor structure, and classification, which are relevant for cancer detection.

Final Database Set: After removing unwanted attributes and null values in the database, our final data is ready to predict breast cancer.

Generate Plots: Use D' to generate plots and correlation heatmaps to identify the relations among variables.

Split Database: Split the database into X and Y based on defined outcomes for sensor data analysis.

Interface of Website: Create a website using (HTML, CSS, JavaScript) interface that allows users to input urine_value, biomarker_value, and sensor_value. Design thresholds for these input values to predict the presence of cancer based on the trained model.

Deployment: Deploy the web interface linked to the trained machine learning model, enabling users to receive real-time predictions based on their input.

End Algorithm: The result is a web-based prediction system where users can input relevant data, and the system provides predictions for cancer detection based on the trained model.

IV. Proposed Model

The proposed model is represented in a systematic diagram, as shown in Figure 3.

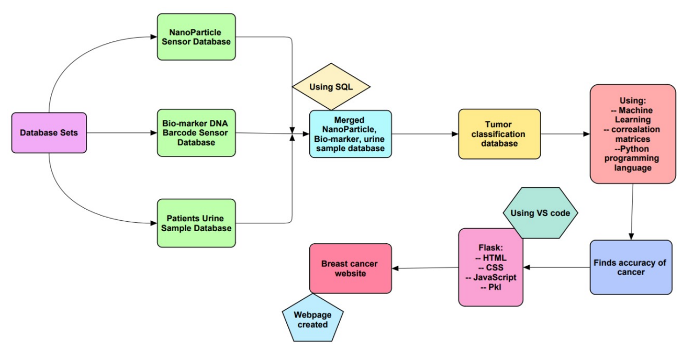
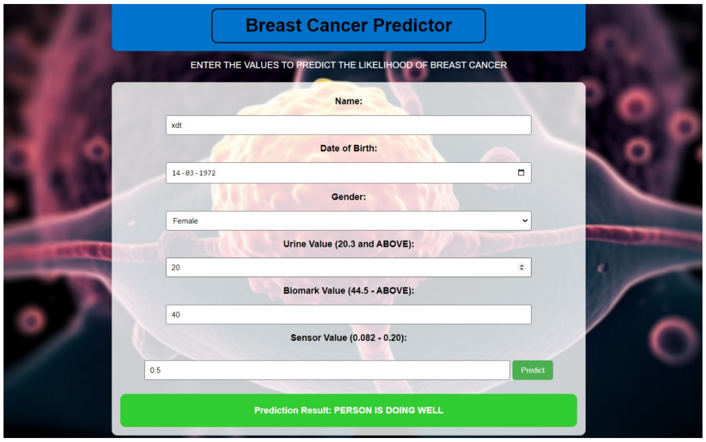
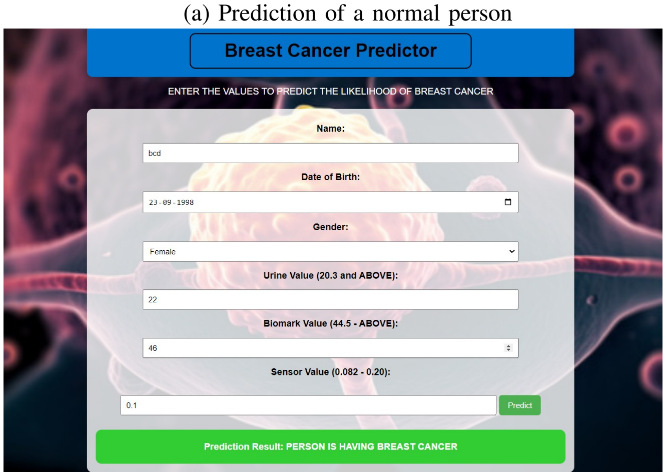


Fig. 3: Process flow diagram for proposed system

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 [11]. 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. This website is about early breast cancer prediction



(a) Prediction of a normal person



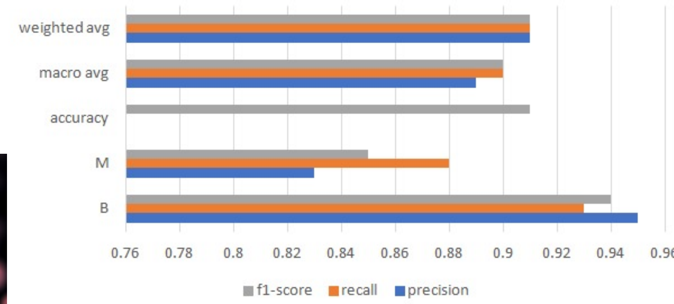
(b) Prediction of a Breast cancer person

Fig. 4: Comparison of predictions

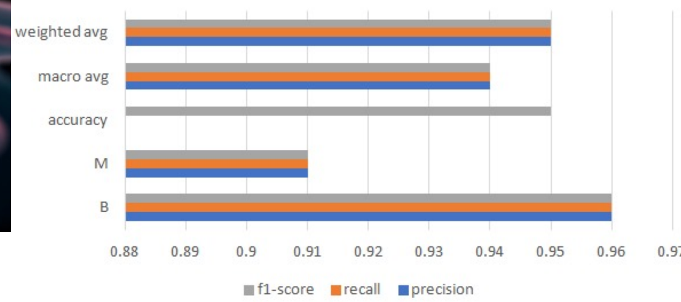
form, and this tool is built to predict cancer at its early stages by taking inputs in three categories: Urine value, Biomark value, and sensor value, which includes basic details like (Name, Date of Birth, and Gender). This is a user-friendly platform that can be accessible on the internet, and this webpage is built with a machine learning model [8].

V. Results & Discussion

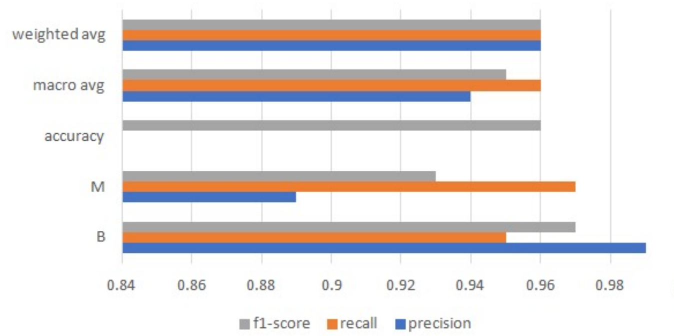
Accuracy reports classifiers which we got in predictions of breast cancer here are the analysis of the results in bar graph form of variant classifiers In the Bagging classifier(a); the scores are precisely achieved with 93% for class Benign and 88% for class Malignant. This precise score results in F1-scores with 94% for class Benign and 85% for class Malignant. This is the second most KNN classifier(b) reveals both the recall scores for class B, with 96% and 91%, and for class M, with 91% and the Random classifier(c) reveals both the recall scores for class B, with 96% and 91%, and for class M,



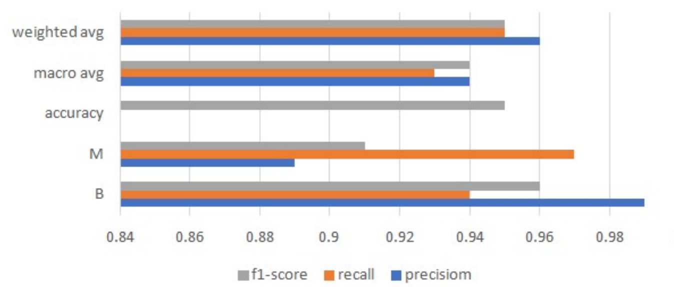
(a) Bagging Classifier Report



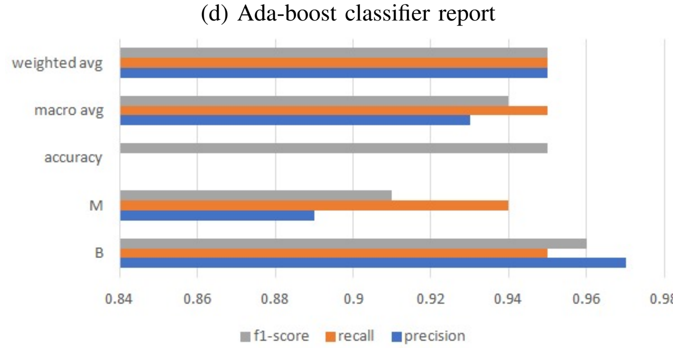
(b) KNN Classifier Report



(c) Random Forest Classifier Report



(d) Ada-boost classifier report



(e) Gradient-boot Classifier Report

Fig. 5: Comparison of Classifiers

with 91% The Ada-boost classifier(d) 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%. The Gradient Boosting(e) 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 III: 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 achieved an overall accuracy of 95% in Breast cancer detection with precision scores of 99%. The model demonstrates strong capability in identifying both cases, so for the prediction of Breast cancer, we are choosing (MLP) [7].

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.

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

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

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

[4] Cleveland Clinic. Nipple discharge. https://my.clevelandclinic.org/health/symptoms/21014-nipple-discharge, n.d. Accessed: August 6, 2024.

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

[6] A. Elouerghi, L. Bellarbi, A. Afyf, and T. F. Talbi. 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)*, pages 1–4, 2020.

[7] GitHub User: Charishma116. Early detection of breast cancer repository. https://github.com/Charishma116/early_decetion_of_breastcancer, n.d. Accessed: August 6, 2024.

[8] GitHub User: srihitha2004. Breast cancer repository. https://github.com/Srihitha2004/Breast-cancer-, n.d. Accessed: August 6, 2024.

[9] S. Kure, S. Iida, M. Yamada, H. Takei, N. Yamashita, Y. Sato, and M. Miyashita. 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.

[10] 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., n.d. Accessed: August 6, 2024.

[11] Nancy Alaswad. Breast cancer dataset. https://www.kaggle.com/datasets/nancyalaswad90/breast-cancer-dataset/discussion/334181, n.d. Accessed: August 6, 2024.

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

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

[14] A. Rovshenov and S. Peker. 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)*, pages 1–6, 2022.

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

[16] World Health Organization. Breast cancer. https://www.who.int/news-room/fact-sheets/detail/breast-cancer, n.d. Accessed: August 6, 2024.