Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

Effective IoT-Based Prediction Approach Using Machine Learning Algorithm for Breast Cancer

[1]Narsimha Chary Rajampetakasham,
ETL Developer., Information Technology (PhD in IT), University of the Cumberlands , Beaumont, Texas ,Rajampetakasham@gmail.com

[2]Mounika Nalluri,
Senior Software Engineer, Information Technology (Masters in Computer Information Systems), Murray State University, mounikanalluri89@gmail.com

[3]Chinna babu Mupparaju,
Software Engineer, Information Technology (Masters in Computer Science), University of Central Missouri, chinnababumupparaju@gmail.com

[4]Bhavishya Katta,
Senior System Analyst,
Information Technology(Masters in Computer Science, Masters in Information System Security , Masters in IT), University of the Cumberlands Austin, Texas , bhavishya0511@gmail.com

[5]Venkata Chaitanya Kumar Suram,
Senior Software Engineer., Information Technology(PhD in IT), University of the Cumberlands, Austin, Texas, Venkatsuram8@gmail.com

Abstract

Diagnosing breast cancer early and correctly is essential for survival and treatment in the modern healthcare system. The evolution of AI and ML has allowed researchers to analyze live and historical data from the Internet of Things (IoT). ML's progress has developed more sophisticated and self- sufficient CAD systems. We present a method for early breast cancer diagnosis using Internet of Things (IoT) devices and machine learning. In this study, we will investigate the feasibility of combining machine learning and IoT techniques for better breast cancer detection. This article describes a medical IoT diagnostic system that can determine the difference between cancer patients and healthy persons. In order to distinguish between malignant and benign tumours, an optimized artificial neural network (ANN) or convolutional neural network (CNN) is used as a benchmark classifier, joining forces with the Support Vector Machine (SVM) and Multilayer Perceptron (MLP). Hyperparameters play a crucial role in machine learning algorithms as they directly affect the performance of training algorithms and the final models. Particle swarm optimization (PSO) feature selection was utilized to improve MLP and SVM's classification accuracy. In order to find the best settings for the CNN and ANN models, a grid-based search was employed. The proposed technique was put through its paces using the WDBC dataset (Wisconsin et al.). The proposed model achieved 99.2% accuracy in classification using the ANN approach and 98.5% accuracy using the CNN method.

Keywords: Machine learning, Breast cancer, IoT, Healthcare.

1. INTRODUCTION

IoT (Internet of Things) and machine learning techniques have proven extremely useful in several areas of medical practice, including the detection and diagnosis of cancer. Because of the IoT, a variety of wearable devices and medical sensors can now gather and transmit data in real-time, which has had a revolutionary effect on healthcare [1]. The early detection of health issues, illness management, and remote patient monitoring are only some of the applications of IoT-based systems that have proven successful. Cancer
detection is considerably improved with the ability to continuously monitor patients' vital signs via the Internet of Things, such as temperature, pressure, heart rate, and electrocardiogram (ECG). This enables the medical personnel to recognize early warning indications of the disease.

Applying machine learning algorithms in the medical field, specifically in cancer diagnosis, has yielded positive results. These algorithms can discover complex patterns and relationships from massive datasets, making them superior to human analysis. To aid in cancer diagnosis, staging, and treatment planning, machine learning models may analyze complicated protein databases, genetic data, and medical imaging [2, 3].

K-nearest neighbours (KNN) is a widely used machine learning technique for cancer prediction. Patients are categorized into groups based on the degree to which the protein profiles of each patient are similar to those found in the training dataset. Another popular method, Naive Bayes, determines whether or not a person has cancer based on their protein profile [3]. In order to make inferences from feature values, Decision Tree algorithms build a tree-like model. The methods utilized by Support Vector Machines (SVMs) aim to locate the hyperplane that differentiates between malignant and benign conditions in the most accurate manner possible.

Machine learning, in conjunction with the Internet of Things, has the potential to improve cancer diagnosis significantly. First, the Internet of Things allows continuous patient health monitoring by collecting data in real-time. Machine learning algorithms can process and analyze this real-time data to find patterns indicative of malignancy, including vital signs and protein profiles. In addition, patients can benefit from less frequent hospital visits thanks to remote monitoring made possible by IoT devices.

Several terms are used interchangeably to describe IoT in the medical field. These include the Internet of Medical Things (IoMT), Health Things (IoHT), and Medical IoT. Connected digital health tools, such as apps, devices, and healthcare systems, are collectively called the "Internet of Medical Things" (IoMT). It tests the hardware of a network of sensors that take measurements all over a patient. Integrating AI allows for quick medical data analysis and diagnoses, while IoMT provides a way for wireless and far- flung devices to link securely via the Internet. When operating in the cloud, Internet of Things devices must adapt to various conditions, such as network topology, power transmission, and computational resources. Healthcare practitioners and patients alike have embraced telehealth services, allowing for remote patient monitoring, early disease diagnosis, and effective treatment.

2. LITERATURE REVIEW

Significant shifts are happening in the healthcare industry right now. The healthcare system has to be rethought in light of recent demographic, economic, social, and technological transformations that have far- reaching repercussions. This is true not just in America but everywhere. Everything from delivery techniques and strategies for allocating resources to financial models, the expansion of scientific knowledge, and the roles of physicians are changing as the healthcare system develops. The shift from paper to electronic patient medical records facilitates the expansion of healthcare information administration. Several potential sources for clinical data collection include patient charts, doctors' notes, electronic medical records, and imaging devices. Compared to other industries' databases, medical records could be more organized and cohesive [5].

Strategic extensive data use might save the healthcare business in the United States around USD 300 billion each year, according to some estimates. Reduced healthcare costs in the United States account for the bulk of the value here (66%). According to another research, the digital universe grows nearly forty per cent annually. This suggests that digital data about Medicare will expand quicker than other forms of data [7].

In order to prioritize solutions to the nation's most critical health challenges, new methods and administrative structures are required to replace the inefficient current system. Breast cancer has a disproportionately high incidence rate among females. Cancer is a disease that occurs in people when healthy cells grow out of control and metastasize to other organs. It was the leading cause of death worldwide in 2014, accounting for 8.2 million lives lost [8].

The researchers advocated for a medical diagnostic system based on the Internet of Things to identify cancerous tumours from benign ones. Researchers used an ANN and a CNN with tuned hyperparameters to distinguish between cancerous and benign tissue [9]. Our research used the support vector machine and the boosted decision tree model. Hyperparameters play a crucial part in creating machine learning algorithms since they immediately impact the efficiency of training algorithms and the quality of the resulting models. This is why studying and improving machine learning algorithms' hyperparameters is crucial.

The authors emphasized the importance of using machine learning-based models for early identification of breast cancer when the disease is most curable. The authors set out to devise a means by which a patient might calculate the probability that she would be diagnosed with breast cancer. The goal was to leave as few cancer cells as possible for subsequent therapies to destroy. The authors discovered that their CNN method achieved around 86% accuracy using the validation dataset.

Separately, the researchers developed a prototype Internet-of-Things device for screening for breast cancer. Thai resources were used for research and analysis. It was found that the prototype could track how long people spent staring at a screen and send out warnings if they went over the "golden" window [11]. The study's author agreed that the (DLA-EABA) was an excellent mathematical method for spotting breast cancer. By consciously blending standard computer vision techniques with the (DCNN), we may construct transfer learning for tumour classification. A fully connected layer and a softmax layer are necessary for implementing error estimates and classification. It was proposed to conduct a study in which machine learning methods were used to pick and extract the features and then to use segmentation and classification methods to evaluate the results and determine the best approach.

As one possible reason for spotting teeny cancer cells, [12] examines the use of deep learning. Histologic marking and diagnosis were both accomplished with the use of the BreCaHAD dataset. To prevent overfitting, they collect data using 20 distinct features. The proposed hybrid dilation deep learning method integrates supervised and unsupervised learning aspects.

3. SENSORS FOR BUILDING IOT SYSTEM FOR MONITORING PATIENTS

Figure 1 depicts the numerous sensors used in this study to keep tabs on patients' vitals and aid in cancer diagnosis. The first type of sensor is a temperature sensor, which determines a person's core temperature. This sensor, like the DS18B20, can measure temperatures from -55 degrees to +125 degrees with a high degree of precision of 0.1 degrees Celsius. It allows for the early diagnosis of infections and inflammations by continually monitoring temperature fluctuations.

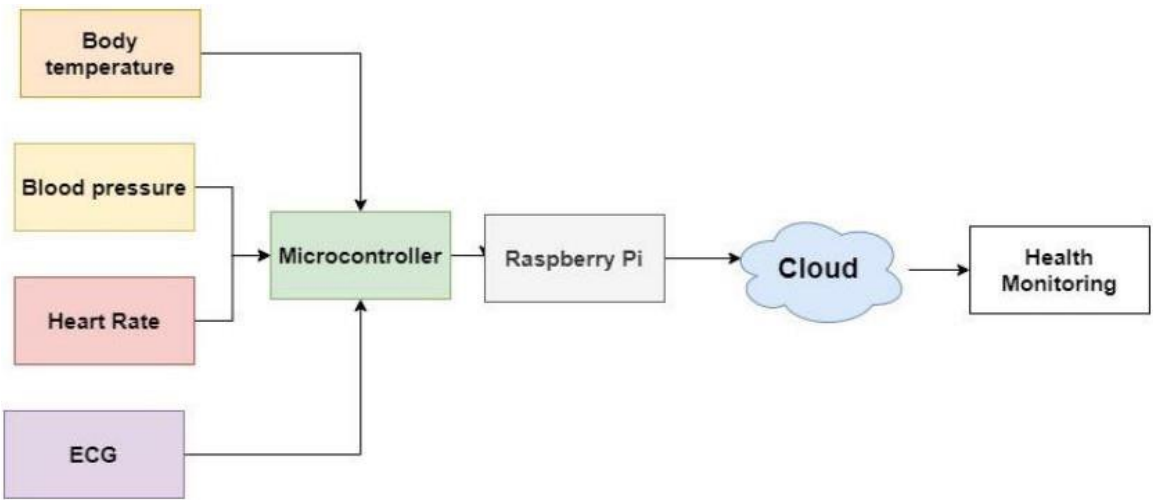


Figure. 1 Patient health monitoring

The second type of sensor is a pressure sensor, vital in keeping tabs on the patient's arterial pressure. The pressure range of this sensor (often 0 to 300 mmHg) must be measured with high accuracy (1 mmHg), and the sensor itself can be either piezoresistive or capacitive. Detecting aberrant oscillations and blood pressure patterns might help doctors diagnose serious diseases like cancer.

Finally, a pulse sensor is employed to keep tabs on the patient's blood pressure. With an accuracy of at least 1 BPM, this sensor, typically based on optical technologies like photoplethysmography (PPG), should be able to monitor heart rate between 40 to 200 BPM. The detection of arrhythmias through monitoring heart rate can shed light on one's overall cardiovascular health and may even provide clues as to the presence of cancer or other cardiac diseases.

3.1 Communication architecture

As seen in Figure 1, the research also uses cloud storage to facilitate the sensors' and healthcare providers' constant and uninterrupted flow of data. The sensors upload their data to the cloud, where it can be accessed and analyzed securely by authorized healthcare professionals in real-time. The cloud-based solution makes it easy for doctors to view the patient records assigned to them. Providers can track their patients' progress in real-time and use the resulting data to make educated decisions about their care. The ability to remotely access patient health data empowers physicians to more effectively track patterns, identify outliers, and intervene at the right time. Zigbee technology is used in this study to enable the wireless transfer of sensor data to keep tabs on the patient's well-being. Zigbee is an IoT-optimized wireless communication protocol with a low data rate and low power consumption.

4. PROPOSED METHODOLOGY

Invasive forms of breast cancer are relatively common in females. Predicting gene expression patterns for early clinical diagnosis is the primary focus of the current state-of-the- art investigation. Figure 2 depicts the proposed breast cancer diagnostic architecture.

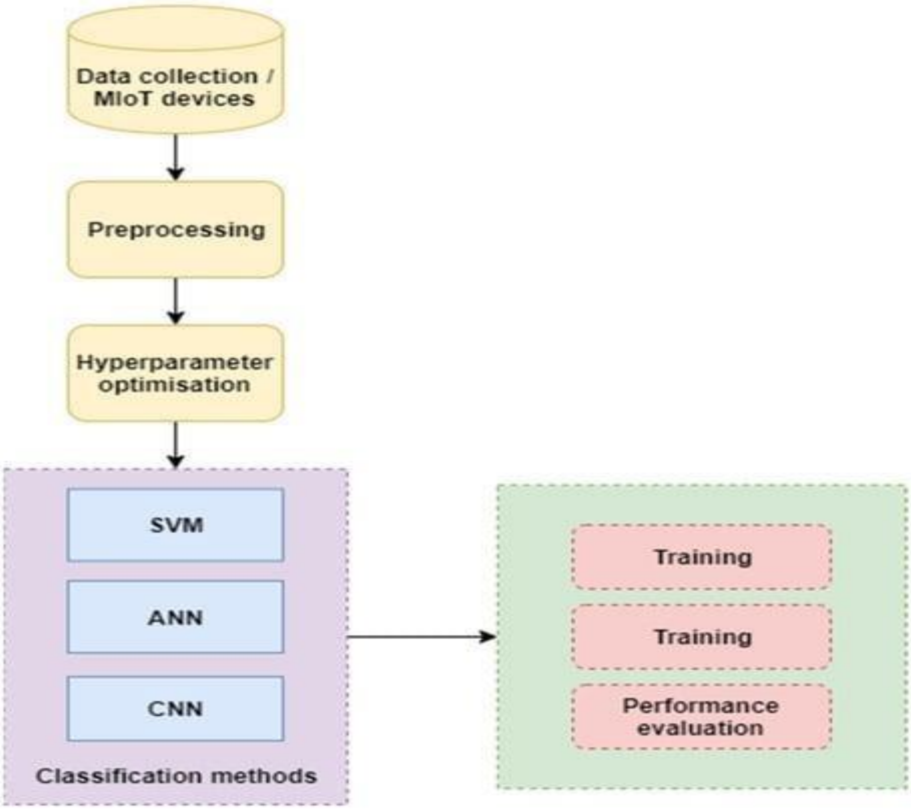


Figure 2. Proposed breast cancer classification architecture.

weight matrix. At each layer of the network,

4.1 Support vector machine

Supervised learning has led to the development of tools like support vector machines (SVM). Statistical learning theory lies at the heart of SVM. SVM is utilized for both single- and multi-class problems in classification. When applied to a multi- dimensional space, the SVM method generates enormous hyperplanes by optimizing the distance between data points and building a hyperplane using support vectors.

4.2 Artificial Neural Networks (ANNs)

To predict a class variable y based on some input data x, an ANN uses a series of ever more complex layers of fully connected neurons. The ANN can be considered a function that realizes the mapping for the given probability distribution p(y|x). We use an ANN with l hidden layers and an output layer to accomplish the function mapping. Weighted edges connect nodes in one layer to their counterparts in the next. This set of weights can be represented graphically as athere is also a bias vector, denoted by b. In order to perform any calculations at the highest levels, a non-linear function is required. Each hypothetical neuron is coupled to an activation function. This function converts the neuron's output into an integer between zero and one. The activation function was typically a softmax function in earlier papers.

$$\text{softmax}\,(x_i) = \frac{e^{(x_i)}}{\sum_j e^{(x_i)}}$$

4.3. Convolutional Neural Networks

Convolutional neural networks (a form of deep learning approach) use a multilayer, feed- forward neural network of perceptron units to learn from examples and analyze data under supervision. It is typically employed for visual data classification via pictures. CNN is widely employed in the medical imaging industry for data classification and prediction. CNN is not like other neural networks because of its unique architecture. In CNN, images are stored as higher-dimensional tensors or matrices. In order to create a tensor, arrays are nested inside other arrays

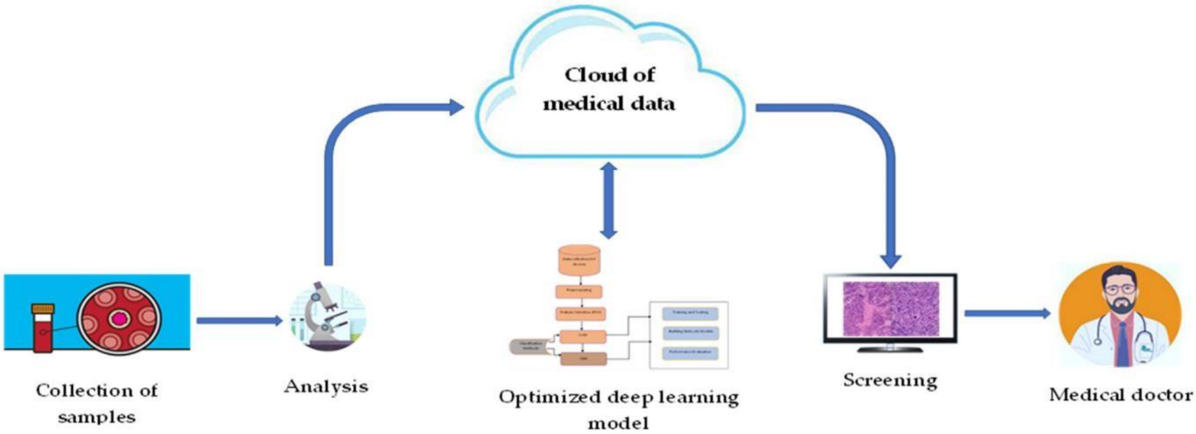


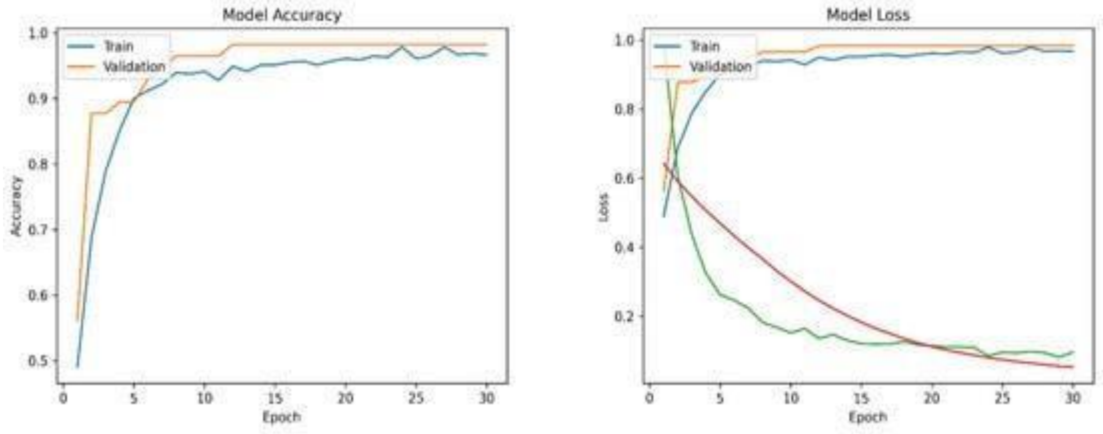
Figure 3. The proposed medical IoT-based system for breast cancer diagnosis uses an optimized convolutional neural network (CNN) classifier with hyperparameters, as depicted in a schematic architectural diagram.

Figure 3 is a schematic depiction of our recently proposed strategy for using IoMT in the diagnosis of breast cancer. The document is divided into three parts: Patient breast tissue samples are used as a starting point for histopathological analysis. The flexibility of the analysis is increased by storing microscopy images of cells on a cloud server. Histopathological imaging samples and related annotations are among the data types often stored in a patient's electronic health record (EHR). The sample data is then sent into our custom-built CNN classifier, which has fine- tuned hyperparameters. To save time and space, our suggested framework encourages using cloud computing services to classify submitted data. Finally, the therapist receives the breast cancer detection results on her computer or smartphone smartphone, where they are double-checked and finalized with medical supervision.

5. RESULTS AND DISCUSSION

5.1 Results of Top-Performing CNN Model

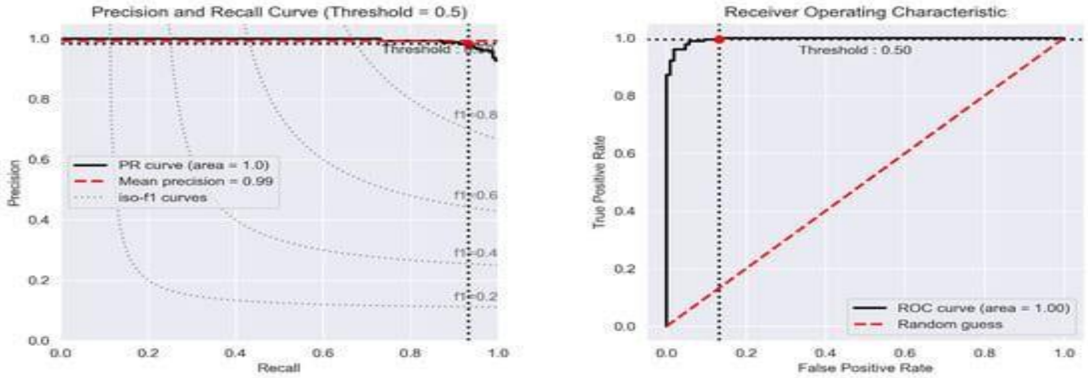
The independent test set determined the best-performing network model and its optimal set of hyperparameter values. These results showed that the proposed CNN model could reliably predict breast cancer. So, the proposed convolutional neural networks are a great alternative to the usual machine learning models that take much time. Figure 4 shows the model training results, which show the accuracy and loss.



(a) (b)

Figure 4. (a) Model accuracy after training and (b) model loss after training.

The ROC, Precision, and Recall curves can all be seen in Figure 5. People have shown that the best ANN model has an AUC of 1.00 and a mean accuracy of 99%.



(a) (b)

Figure 5. The best-performing model's (a) Precision-Recall Curve and (b) Receiver Operating Characteristic (ROC).

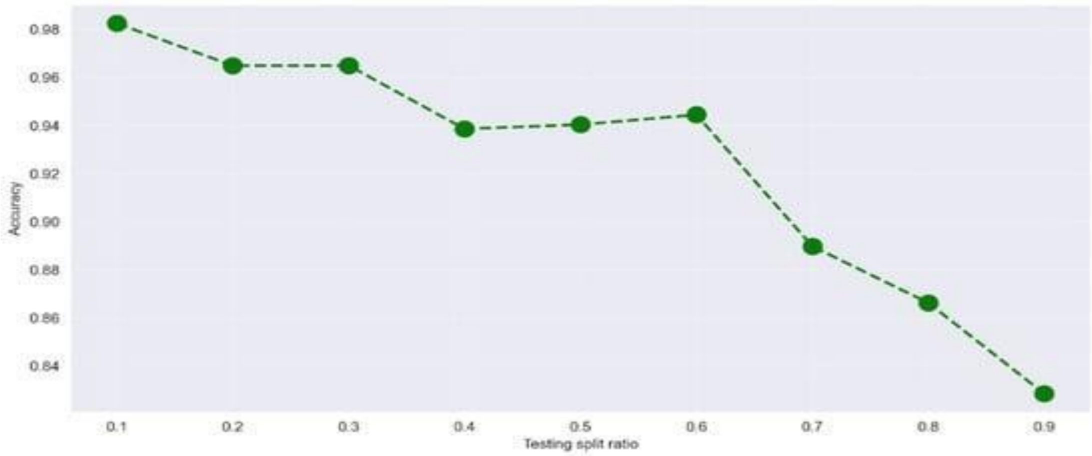


Figure 6. Study of the CNN model's testing/training split ratio.

Finally, we investigated how the training/testing split affects categorization precision. The findings are shown in Figure 6. The precision increases as the number of data samples used for training increases. 90% of the data was used for training and 10% for testing. This was the best way for the CNN model to work (0.98).

CONCLUSION

The most common type of cancer in the world is lung cancer, which affects more men than women. Models that use machine learning to look at, sort, and process breast cancer diagnostic factors. An essential part of non-invasively diagnosing, finding, and predicting breast cancer is imaging tools based on the Internet of Things. Detecting breast cancer at its earliest stages will be significantly easier if Internet of Things (IoT) technology is combined with screening approaches based on machine learning. This survey article discussed how the Internet of Things (IoT) and machine learning techniques can help with early breast cancer detection and prediction. The proposed model achieved a classification accuracy of 98.5% when using CNN and 99.2% when using ANN. This demonstrates that, on limited datasets, basic ANNs may still outperform CNNs. The receiver operating characteristics and precision-recall curves demonstrate that the optimal ANN model achieved an AUC of 1.00 and a mean accuracy of 0.99. The ANN model outperformed the CNN model with a 99.2 per cent success rate. The Mann-Whitney test demonstrated a statistically significant difference ($p = 0.018$) between the two groups. The feature selection process in healthcare research, especially when employing ML approaches and IoT, might yield varying outcomes depending on the dataset, the location of the sick persons, and their way of life. This study evaluated the diagnostic model for breast cancer for its performance in a clinical setting. The analysis of the models shows that the proposed model is good at finding the labels for benign and malignant cells.

REFERENCES

[1] R. Kaur et al., “Machine learning and price-based load scheduling for an optimal IoT control in the smart and frugal home,” Energy and AI, vol. 3, p. 100042, 2021, doi: 10.1016/j.egyai.2020.100042.

[2] C. Brewster, I. Roussaki, N. Kalatzis, K. Doolin, and K. Ellis, “IoT in Agriculture: Designing a Europe-Wide Large-Scale Pilot,” IEEE Communications Magazine, vol. 55, no. 9, pp. 26–33, 2017, doi:
10.1109/MCOM.2017.1600528.

[3] R. Kumar, A. Chug, A. P. Singh, and D. Singh, “A Systematic Analysis of Machine Learning and Deep Learning Based Approaches for Plant Leaf Disease Classification: A Review,” Journal of Sensors, vol. 2022, 2022, doi: 10.1155/2022/3287561.

[4] Dourado, C.M.J.M.; Da Silva, S.P.P.; Da Nobrega, R.V.M.; Filho, P.P.R.; Muhammad, K.; De Albuquerque, V.H.C. An Open IoHT- Based Deep Learning Framework for Online Medical Image Recognition. *IEEE J. Sel. Areas Commun.* **2020**, *39*, 541–548.

[5] Manyika J., Chui M., Brown B. Big Data: The Next Frontier for Innovation, Competition, and Productivity McKinsey Global Institute, San Francisco, CA, USA (2019) Google Scholar

[6] EMC with research & analysis by IDC The Digital Universe Driving Data Growth in Healthcare (2023) Google Scholar

[7] BridgeHead Software 2011 international healthcare data management survey (2019)

[8] LeCun Y., Bengio Y., Hinton G. Deep learning Nature, 521 (2018), pp. 436-444

[9] Shravya C., Pravalika K., Subhani S. Prediction of breast cancer using supervised machine learning techniques Int. J. Innov. Technol. Explore. Eng., 8 (2019), pp. 1106-1110

[10] 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, 8 (2020), pp. 96946-96954

[11] Khan M.N., Rahman H.U., Almaiah M.A., Khan M.Z., Khan A., Raza M., Al- Zahrani M., Almomani O., Khan R. Improving energy efficiency with content-based adaptive and dynamic scheduling in wireless sensor networks IEEE Access, 8 (2023), pp. 176496- 176521

[12] Bubukayr M.A., Almaiah M.A.Cybersecurity concerns in smartphonessmartphones and applications: A surveyProceedings of the 2021 International Conference on Information Technology (ICIT),Amman, Jordan, IEEE, Washington, DC, USA (2023), pp. 726-732

[13] Barracliffe, L.; Arandjelovic, O.; Humphris, G. A pilot study of breast cancer patients: Can machine learning predict healthcare professionals’ responses to patient emotions. In Proceedings of the International Conference on Bioinformatics and Computational Biology, Honolulu, HI, USA, 20–22 March 2017; pp. 20–22. [**Google Scholar**]

[14] Hassan, M.A.; Malik, A.S.; Fofi, D.; Karasfi, B.; Meriaudeau, F. Towards health monitoring using remote heart rate measurement using digital camera: A feasibility study. *Measurement* **2020**, *149*, 106804. [**Google Scholar**] [**CrossRef**]

[15] Al-Turjman, F.; Alturjman, S. Context- sensitive access in the industrial Internet of things (IIoT) healthcare applications. *IEEE Trans. Ind. Inform.* **2018**, *14*, 2736–2744.