IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

Intelligent-Breast Abnormality Detection (I-BAD) framework and Risk Classification using Machine Learning Techniques

Dr M Kavitha [1], M Kalyani [2]

[1] Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation (KLEF), Vaddeswaram, Green fields, Guntur, Andhra Pradesh, India -522302
[2] Department of Mathematics, PACE Institute of Technology and Sciences, Ongole, AP, India
Email: mkavita@kluniversity.in

DOI : 10.48047/IJFANS/11/Sp.Iss5/061

AbstractBreast Cancer (BC) is the second leading cause of death among women throughout the world. Early-stage identification of breast abnormality helps the people to attend better treatment at a premature stage of tumour. Breast abnormality detection and risk rate prediction will support the people to increase the survival rate of a patient. Machine learning (ML) techniques have a long track record in the healthcare domain and especially in disease risk classification. In this article, an innovative Internet of Things (IoT) based intelligent- breast abnormality detection (I-BAD) framework to monitor and collect various breast health vital parameters is proposed and evaluated the efficiency of different machine learning techniques in breast abnormality classification.

Keywords: Breast Cancer, Healthcare, Internet of Things, Machine Learning.

1. Introduction

Internet of Things (IoT) provides an environment that relates various physical objects like sensors and actuators to the Internet and allows these objects to gather and transfer the data over a network. It offers services in different fields like healthcare, agriculture, transport, industry control, water management, environment safety, security, object tracking [1-2]. Patient health care becomes easy and smart due to IoT emerging in the medical field. It enables more individualized care where users are self-handling and self-examining their health, and doctors can improve the experience of care [3]. Smart health care systems allow remote communications among patients and doctors to provide necessary services like health monitor, health conditions tracking, and recording of vital medical information. The patient's health conditions can track using sensors, and the reports are generated to the corresponding authorized clinician or care team automatically. It supports the clinician and care team to take the best preventive decisions to words patient therapy and individual care at the right time [4].
Breast cancer is the major cause of death among women for the last 65 years. Usually, prevention is better than cure. But for the case of breast cancer disease prevention is not applicable to all categories of people. Because this disease will be contaminated due to family history means a genetic factor is one of the major risk factors [27]. Premature identification of breast tumour and attending the best treatment at right time is the only way to reduce the breast cancer death rate in the world [28-29]. So many researchers came with different solutions for this problem but still, it is a major research area.
Technological advances like IoT intervention in breast health assessment will provide a better solution for this issue. In this connection, as part of this article, we are designing an IoT-based

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **© 2012 IJFANS. All Rights Reserved**, **Volume 11, Sp.Iss 5 , 2022**

intelligent - breast abnormality detection framework (I-BAD) for continuous breast health monitor and proposing various machine learning approaches to classify the breast abnormality risk rate. In continuation, the performance of all the identified machine learning approaches is evaluated on the breast cancer data set.

Coming to the organization of this article, in the section - I, we discussed the introduction of IoT and the impact of breast cancer on women. In section II, we are going to discuss breast cancer and various researchers' contributions to early-stage detection of breast tumour. In section III, the design of the proposed framework is discussed. In section IV, machine learning approaches supported to classify the data are discussed. In section V, the performance of all the identified approaches is discussed. Finally, the article concludes with a clear outcome explanation in the conclusion heading.

2. Related Work

In this section, various researchers' contributions on designing technological tools for Patient health monitor and ML support in breast cancer prediction.

Megalingam, R. K et.al. [5], proposed a sensor-enabled electronic device to monitor the bed assisted patient health. It comprises sensors to screen the pulse of patients at home. The sensors enclosed to the body can sense the events and send signals to smartphones through Bluetooth. If any emergency event is noticed by the system, it will automatically generate a warning notification to the caretaker [26].

Vuppalapati et al., [6], proposed a sensor coordination structure with Electronic Health Records (EHR) for patient health care. He discussed how wearable gadgets record various essential health parameters data and coordinates these parameters data to EHR to find accurate health condition of the patient and offers a better medical service by interfacing with the doctor.

Kulkarni, C et al., [7], proposed a smart self-administrative health framework to limit doctor efforts and give a consistent understanding to the patient. He showed an automatic transmission between smart gadgets and an android portable application using Bluetooth. This framework confirms by an automatic control over an oxygen cylinder utilizing the mobile application.

Karthicraja, V. M et al., [8], proposed a wearable smart framework using IoT, which emulates the idea of personal care focusing on health and security administrations. Smart wearable security frameworks can screen people day in and day out in their place out-of-doors, as per prudent security strategies. This framework involves an across the board class of wearing or implanted devices for information fit, pre and post handling, and conclusion creation frameworks. For human interface, it has multimedia gadgets with GUI as isolated elements.

Kumar, S. P et al.,[9], examined how patients manage abnormal health using IoT. He talked about how IoT helps to manage health through wearable devices. The wearable device smartwatch is utilized here to monitor various parameters. If the patient is having an irregular pulse or any sort of coronary illness trigger had been sent to the caretaker or the local doctor as a precaution.

The health care sector has been benefitted more through the ML prediction frameworks. From the past few decades, ML techniques are widely used in developing effective predictive models to support dynamic decision making.

In Milon Islam et al.,[10], the authors focused on the prediction of breast cancer using ML techniques. The algorithm's results are compared using different performance measures such as

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **© 2012 IJFANS. All Rights Reserved**, **Volume 11, Sp.Iss 5 , 2022**

sensitivity, specificity, accuracy, correlation coefficient, false discovery, and omission rates. These measures are verified for both the training and testing phases of the experiment.

In Douangnoulack et al., [11], the authors focused on building the best performance classifier with a minimum number of classification rules for breast cancer diagnosis. The author uses principle component analysis (PCA) and he also applied the J48 decision tree, Reduced error pruning a tree, and random tree classifiers and he noticed that the J48 decision tree gives best compare others in this experiment.

In Turgut et al., [12], the authors work out on breast cancer classification using eight machine learning algorithms named DT, RF, LR, SVM, KNN, MLP, Adaboost, and gradient boosting. Finally, their experiment on the Wisconsin breast cancer dataset results as SVM gives the best result compare to others.

In Gupta et al., [13], the authors compare linear regression (LR), random forest (RF), multi-layer perceptron (MLP), and decision trees (DT) machine learning techniques on breast cancer diagnosis using Wisconsin breast cancer dataset. A comparison of these algorithms is done in terms of precision, recall, accuracy, and R2. Their results evidence that MLP is given the best result on the identified data set.

In Wadkar et al., [14], the authors discussed breast cancer detection in the ANN network, and the performance of the system is analyzed using the SVM technique. The accuracy of the system is compared with CNN, KNN, SVM by constructing a confusion matrix. The experiment results in ANN gives better accuracy.

3. System Design

The main objective of this proposed Internet of Things (IoT) based Intelligent Breast Abnormality Detection (I-BAD) framework is to detect breast abnormalities at an early stage. It motivates the patient to take better statements at the premature stage of the tumour. In some cases, it leads to a permanent cure also. Nowadays, IoT based health applications [15] provides cost-effective and high-quality remote health care services. Continuous observing of patient health using IoT devices causes an enormous amount of data generation. The processing of this data will support to predict the health risks at the earliest. Figure 1 displays an illustration of the three-tier architecture of the proposed I-BAD framework. This includes a mobile app, a wearable technical tool, cloud, processing techniques, and communication techniques. Tier 1 refers to the patient environment, where patient breast health parameters are monitored using different tools. Tier-2 refers to storage and data processing mechanisms. Tier -3 refers to the user environment. Communication mechanisms are supported to relate all the components in the proposed framework.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

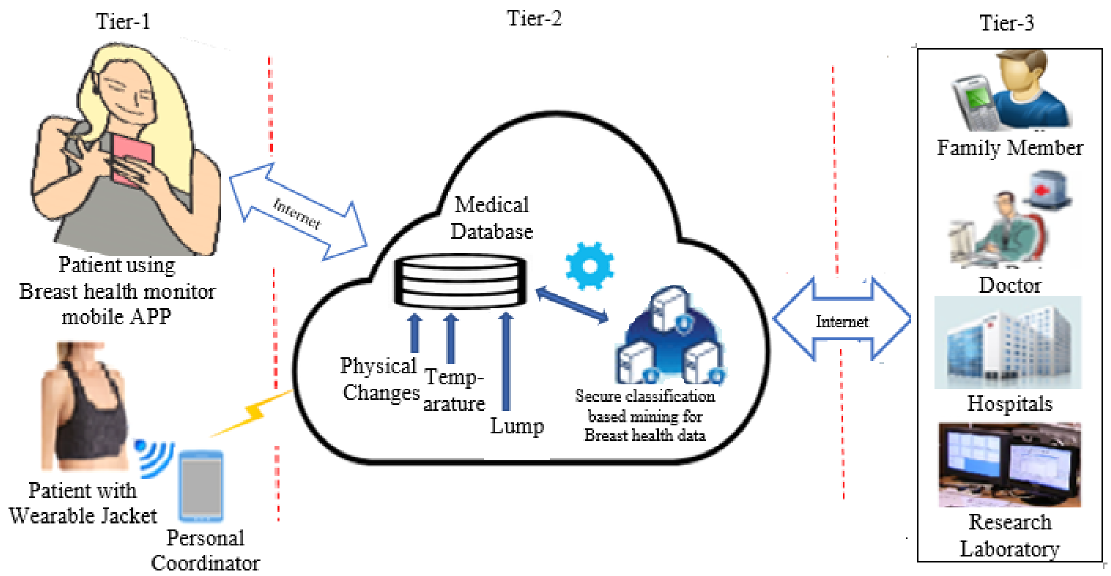

Figure 1. Intelligent-Breast Abnormality Detection (I-BAD) System Framework

3.1. Breast Health Parameters

The tumour in the breast affects to change the internal and physical structure of the breast. Eight parameters are considered here to evaluate breast health. Table 1 shows the parameter description which supports the evaluation of breast health in the proposed framework. Family history is one of the parameters we considered because the breast cancer occurrence rate is high for the breast cancer family history women. Five parameters are related to appearance, feel, and size changes of the breast. Two more parameters we considered are breast temperature and lump. These parameters are monitored using different tools to fix breast abnormalities at an early stage. P1 to P6 parameters are monitored using android based recommender system. It is a simple mobile App with a user-friendly questioning system. The design and evaluation of this recommender system are discussed in chapter-4 with neat sketches. The last two parameters P7 and P8 are monitored using a sensor-enabled wearable jacket [25]. In this, the sensors closely monitor the specified parameters continuously. This continuous data is processed using statistical measures and localization techniques.

Table 1. Breast health measurable parameters

Notation	Description
P1	Breast cancer family history
P2	Hard spots on breast surface
P3	Swelling of all or one part of breast

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Sp.Iss 5 , 2022

P4	Skin irritation or dimpling
P5	Change in color and appearance of areola
P6	Unusual discharge from nipple other than milk
P7	Temperature
P8	Lump

3.2. Storage and Processing

Cloud is an alternative mechanism for expensive data warehouses at hospitals. It supports the storing and process of patient records. Cloud service allows the user to access data based on the requirement at any time [16-17]. In the I-BAD framework, breast health parameters of a patient are monitored through various proposed technological techniques, and the sensed data can store in the cloud for permanent record [24]. On top of stored breast health parameters, machine learning (ML) techniques are applied to detect the breast health risk rate at an early stage. ML methods support the process of healthcare-related datasets and enable the prediction to empower accurate decision making[30]. From the last few decades, ML plays a vital role in classifying tumors [18-19]. A hybrid model designed using three ML techniques (Decision Tree (DT), Naive Bayes (NB), Random Forest (RF)) is proposed in I-BAD for breast health parameters processing. These ML classifiers support to analyse the breast cancer risk rate. To enhance the functioning of the system deep learning model convolution neural network (CNN) is also applied, and its performance is verified in terms of accuracy, precision, recall, F1- Score.

4. Implementation

The related work reveals that ML classification techniques have a history in breast abnormality prediction. The functioning of the system is measured using a synthetic dataset. The data set consists of 10 columns (attributes). The first attribute refers to patient id, the next eight columns as input, and the last column refers to the output label. The data set consists of 4 categories of patients. That means normal patients with number 0, low risk with number 1, a medium risk with number 2, and high risk with number 3. That means the ML technique can classify the patient into any one of these four categories by processing input attribute values. The dataset pre-processed to remove outliers. That means unrelated data or missing values are handled in the pre-processing phase. The data set is divided into the training part and the testing part. In the training phase, the system can learn the hidden patterns by processing the training part of the data set. In the testing phase, the system evaluation is done by using the testing part of the data set. The general workflow of the ML classification technique is given in figure 2. A hybrid model with three ML classification techniques Naive Bayes (NB), Decision Trees (DT), Random Forest (RF) is applied on the data set. The performance of these algorithms is compared with one and another. Although ML techniques are commonly used in breast cancer detection, its accuracy

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, Volume 11, Sp.Iss 5 , 2022

still needs to be improved. In this connection, a deep learning method - convolution neural network (CNN) is applied to the data set. The breast health parameters are evaluated using CNN and this result is compared with ML techniques.

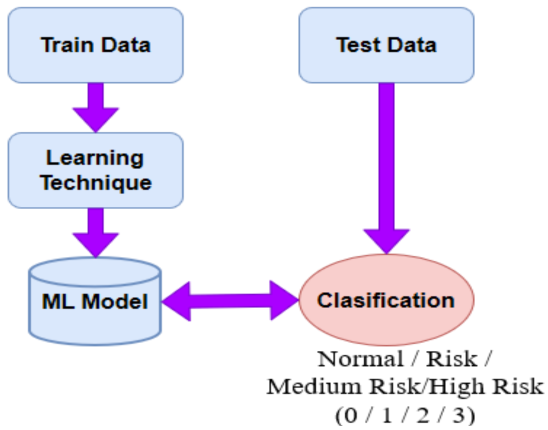

Figure 2. Workflow of ML technique [13]

4.1. Naive Bayes (NB)

Naïve Bayes (NB) is quick and performs well even with small datasets. It is a probabilistic classifier and applies Bayes statement with strong independent hypotheses [20].

Algorithm: Naïve Bayes

Input: Dataset
Output: Predicted result
Procedure:
BEGIN
Divide dataset into two parts as training, testing sets and four classes with sets of features.
Mean and standard deviation calculation on each class features
Calculate summary for each feature of a class
Calculate the probabilities of features using normal distribution
Calculate probability of each class by multiplying probabilities of its features
In test dataset, the instance class is predicted by calculating probability.
END

4.2. Decision Tree (DT)

Decision Tree (DT) algorithm works on the principle of divide and conquers methodology and it looks in the form of a tree. Generally, C 4.5 decision tree algorithm is used in research works. The attributes are represented as nodes in the tree.[21].

Algorithm: Decision Tree

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

Input: Dataset
Output: Predicted result
Procedure:
BEGIN
Divide dataset into two parts as training, testing sets
Select the train dataset to do the process of learning
Prepare a chart of each attribute to corresponding classes
Catch all possible values for each individual attribute that correlate with feasible classes
Calculate values of every attributes which belongs to different classes
The attribute which is having minimum number of values which reside in the unique class is taken as root node
Based on least number of values which has different classes the next level attributes in the tree are selected
END

4.3. Random Forest (RF)

Random Forest (RF) builds a various number of decision trees by considering random samples. Each tree classifies its examples and maximum votes decision is chosen. It works well even for small data sets also. If any new example enters it does not affect entire structure, it affects only one part of subtree.

Algorithm: Random Forest

Input: Dataset
Output: Predicted result
Procedure:
BEGIN
Select N random observations from dataset
Construct a decision tree with these N records
Choose number of trees needed in the experiment and repeat steps 2 and 3
Each tree in the forest predicts the category to which the new record belongs
Depends on maximum vote the new record is assigned to that class like decision tree algorithm procedure.
END

4.4. Convolution Neural Network (CNN)

In CNN, during network training, the data set was divided into two groups, 70% training data, and 30% testing data. The randomly selected training data was fed into the network to train the model, and the testing data was applied to figure out the accuracy of the pre-trained network. Usually, CNN increases the accuracy of prediction because the CNN framework processes the input in two phases one is feature learning and the other is a classification [22].

4.5 Performance Measures of proposed techniques

ISSN PRINT 2319 1775 Online 2320 7876
 Volume 11, Sp.Iss 5 , 2022

There are various metrics to measure the performance of ML techniques. Five measures accuracy, precision, recall, F1 measure, and support are considered here. In [23], the authors defined these measures as follows.

- Accuracy: The accuracy of the technique is calculated as the number of correct predictions divided by the total number of examples in the data set.
- Precision: It is also called confidence and it is the rate of both true negatives and true positives that have been identified as true positives.
- Recall: Recall is also called sensitivity and it is the number of positive observations that are correctly predicted.
- F1 Score: It is also called F- measure and it is the harmonic mean precision and recall.
- Support: It is defined as the number of samples of the true response that lies in each class of target values.

5. Experimental Results

GUI experienced Anaconda navigator and python notebook IDE is used here to execute the ML and CNN techniques on the data set. Anaconda is an open-source tool to run machine learning, deep learning, and data science projects in python and R programming languages with a vast set of supportable packages. It provides a user-friendly environment and easy installation of needed packages like pandas, mat plot lib, NumPy, sci-kit-learn, SciPy, tensor flow, and so on.There are ten columns in the data set. The data set is randomly divided (70:30) into two parts for training and testing. The data set having eight input attributes to classify the breast abnormality risk rate. Figure3 refers to the classification of patients. 0 means normal, 1 means risk, 2 means medium risk, and 3 means high risk. Figure 4 shows how the attributes in the data set are positively correlated.

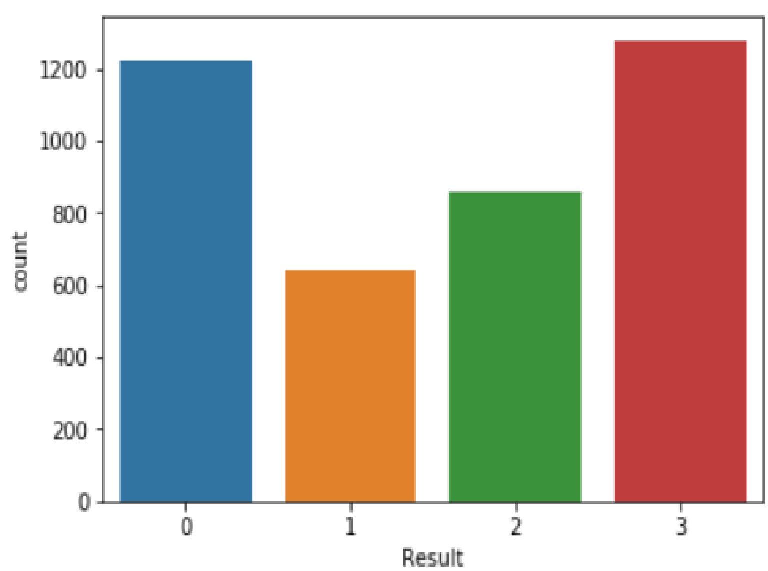

Figure. 3. Classification of patient breast health abnormality rate

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, Volume 11, Sp.Iss 5 , 2022**

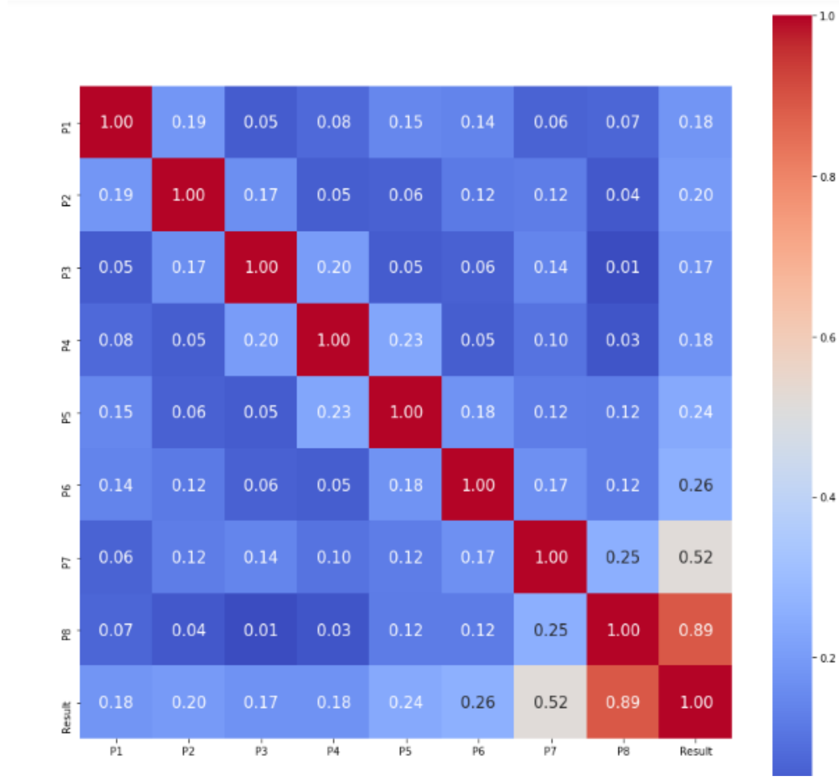

Figure 4. Correlation matrix of attributes of dataset

Gaussian naive bayes, random forest, decision tree and convolutional neural network techniques are applied on the dataset and table 2 shows the accuracy of each model. Figure 5 shows accuracy comparison of all the discussed learning models. This evidence that deep learning CNN model will give more accurate compare to ML techniques.

Table 2. Breast health measurable parameters

Algorithm	Accuracy
Convolutional neural network	**99.83**
Decision Tree	**99.25%**
Random Forest	**93.16%**
Naïve Bayes	**85.25%**

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **Volume 11, Sp.Iss 5 , 2022**

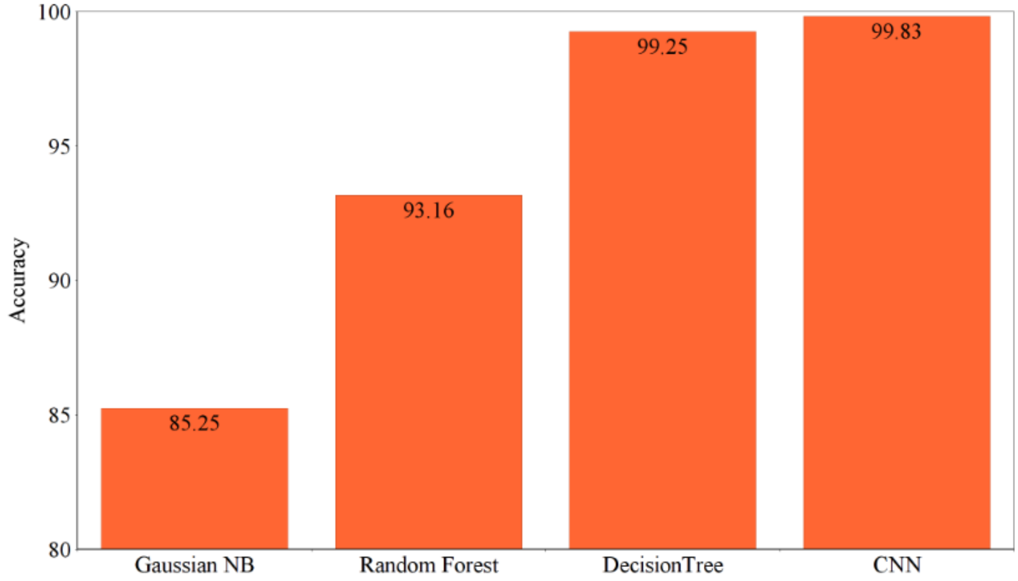

Figure 5. Classifiers accuracy comparision

6. Conclusion

Internet of Things (IoT) based intelligent- breast abnormality detection (I-BAD) framework is proposed in this article to monitor various breast health vital parameters, and breast health risk rate is classified using ML and DL techniques. All ML supervised classification algorithms are applied to the dataset and among these three algorithms named decision tree, random forest, naive bayes are given good results. The accuracy of ML techniques in breast health risk classification is good even though there is a need for improvement in this accuracy. So deep learning model CNN is also applied, it evidences that CNN gives better accuracy compared to ML techniques.

7. Future Work

In health care, machine observation is more accurate than manual observation, and this is proven in many cases. Designing a sensor to measure all parameters of breast health is a crucial task. If such a sensor is designed, it might increase the performance of the existing intelligence system.

References

[1] Dorsemaine, B., Gaulier, J. P., Wary, J. P., Kheir, N., &Urien, P. (2015, September). Internet of things: a definition & taxonomy. In 2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies (pp. 72-77). IEEE.

[2] Kramp, Thorsten, Rob Van Kranenburg, and Sebastian Lange. "Introduction to the Internet of Things." Enabling Things to Talk. Springer, Berlin, Heidelberg, 2013. 1-10.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **© 2012 IJFANS. All Rights Reserved**, **Volume 11, Sp.Iss 5 , 2022**

[3] Islam, SM Riazul, Daehan Kwak, MD Humaun Kabir, Mahmud Hossain, and Kyung-Sup Kwak. "The internet of things for health care: a comprehensive survey." IEEE Access 3 (2015): 678-708.
[4] Metcalf, David, et al. "Wearables and the internet of things for health: Wearable, interconnected devices promise more efficient and comprehensive health care." IEEE pulse 7.5 (2016): 35-39.
[5] Megalingam, R. K., Unnikrishnan, D. K. M., Radhakrishnan, V., & Jacob, D. C. (2012, December). HOPE: An electronic gadget for home-bound patients and elders. In 2012 Annual IEEE India Conference (INDICON) (pp. 1272-1277). IEEE.
[6] Vuppalapati, Chandrasekar, Anitha Ilapakurti, and Santosh Kedari. "The role of big data in creating sense ehr, an integrated approach to create next generation mobile sensor and wearable data driven electronic health record (ehr)." 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService). IEEE, 2016.
[7] Kulkarni, C., Kenjale, S., Patil, M., Kulkarni, C., &Hiray, S. R. (2017, May). Smart self-regulatory health system. In 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT) (pp. 796-800). IEEE.
[8] Karthicraja, V. M., Prasad, V. K., Kruthika, U., & Sivakumar, R. (2018, March). Smart Wearable Systems for Personal Care with Internet of Things. In 2018 Fourth International Conference on Biosignals, Images and Instrumentation (ICBSII) (pp. 206-212). IEEE.
[9] Kumar, S. P., Samson, V. R. R., Sai, U. B., Rao, P. M., & Eswar, K. K. (2017, February). Smart health monitoring system of patient through IoT. In 2017 international conference on I-SMAC (IoT in social, mobile, analytics and cloud) (I-SMAC) (pp. 551-556). IEEE.
[10]Md. Milon Islam, Hasib Iqbal, Md. Rezwanul Haque and Md. Kamrul hasan, " Prediction of Breast Cancer Using Support Vector Machine and K-Nearest Neighbors", Region 10 Humanitarian Technology Conference(R10-HTC), pp. 226-229, IEEE 2017.
[11]Douangnoulack, P., &Boonjing, V. (2018). Building Minimal Classification Rules for Breast Cancer Diagnosis. In 2018 10th International Conference on Knowledge and Smart Technology (KST) (pp. 278-281). IEEE.
[12]Turgut, S., Dağtekin, M., & Ensari, T. (2018, April). Microarray breast cancer data classification using machine learning methods. In 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT) (pp. 1-3). IEEE.