Assessment 1: Research Proposal

Intelligent Identification of Breast Cancer Empowered with Machine Learning

|  |  |
| --- | --- |
| Name: Satnam Singh Brar | Supervisor: |
| University ID: 20116668 |  |
| Word Count: 3037 |  |
| Date: 04 August 2021 |  |

Table of Contents

[Abstract 3](#_Toc69506452)

[Introduction 3](#_Toc69506453)

[Aim/Objective 4](#_Toc69506455)

[Research Questions 5](#_Toc69506456)

[Ethical Considerations 5](#_Toc69506457)

[Literature Review 5](#_Toc69506458)

[Project Timeline… 7](#_Toc69506462)

[Methodology 7](#_Toc69506463)

[Project Evaluation 8](#_Toc69506464)

[Conclusion 8](#_Toc69506465)

[References 8](#_Toc69506466)

# Table of Figures

[Figure 1: Project Timeline 7](#_Toc69506491)

[Figure 2: Proposed Model for Breast Cancer Prediction System 8](#_Toc69506492)

# Abstract

Every year many people die due to the late diagnosis of cancer. Early detection of this deadly disease can reduce mortality rates. Breast cancer is one of the significant types of cancer and regularly recognized malignant growth among women and necessary explanation behind expanding death rate among women. The only way to decrease it is to diagnose it at early stages. Diagnostic errors by medical practitioners are considered one of the leading causes of death. The advent of technology has made it easier for humans to automate patient prognosis and reduce disease from symptoms and report analysis. Although modern era technology makes it easy to detect cancers early, a lot of work is still needed in this field. In this research, fusion-based, different machine learning algorithms will apply to diagnose breast cancer early. For this purpose, two datasets (images and Features based) will collect from Kaggle and IEEE-dataport. The proposed model will be trained on an image-based dataset using a convolutional neural network, and the features dataset will train using deep extreme machine learning. In the proposed fusion-based model, accuracy will be better as compared to the state-of-the-art previous published approaches.

**Keywords:** Artificial Intelligence, Machine Learning, Convolutional Neural Network, Deep Extreme Machine Learning.

# Introduction

Artificial intelligence (AI) is a subfield of computer science. It emphasizes the creation of intelligent machines that work, behave, and react like humans. It has become an essential part of the technology industry nowadays. In recent years, there has been a magnified focus on using artificial intelligence in several domains to resolve complex issues, such as urban infrastructure, resident living environment, transport management, medical treatment, shopping, security assurance, and so forth. Further, the endorsement of artificial intelligence in health care is growing whereas evolving the face of health care delivery.

Early detection of cancer is one of the main approaches to prevent death. Breast cancer is the second-biggest cause of mortality between forty to fifty-five years (Zaheer et al. 2019). Globally, the underlying cause and diagnosis at the initial stage of breast cancer are considered significant problems that produce a drastic effect on mental and social dimensions in human life. The medical practitioners recommend several tests to perform. The underlying symptom of breast cancer typically involves forming a lump in the breast or under the armpit; modification, irritation, and dimpling in skin size and nipple become pulled in and reddish in coloration (Momenimovahed and Salehiniya, 2019).

The abnormal development of human cells in an organ is termed as tumors that may be cancerous. There are two types of tumors; benign and malignant. Benign tumors are localized, surgically treated, and do not life-threatening, such as fibroids and lipomas. They can only be dangerous when developing in the brain and interacting with typical structures or block terminals in any organ. Besides, various benign tumors like internal organ polyps are precancerous and are surgically eliminated to prevent their transformation into malignant. These tumors generally don’t reappear once eliminated, but if they do, they will generally be in the same position (Wan et al. 2016).

Malignant tumors are cancerous and may lead to death if untreated. The cells of the tumors can invade surrounding tissues or enter into the bloodstream or lymph nodes and influence physiological regulations. They can progress in any part of the body, such as the breast, intestines, lungs, reproductive organs, blood, and skin (Schmid et al. 2018).

The mortality rate of breast cancer is remarkably high. As WHO (world health organization) demonstrated, breast cancer is the most frequent cancer among women, and almost 2.1 million women are affected each year. In 2018, around 627,000 women were no more due to breast cancer development (Jafari et al. 2018).

Various investigational techniques are used to decrease the pathological state. Information technology such as data mining and different artificial intelligence methods is supported to diagnose diseases, like machine learning, back-propagation neural network, and classification and support vector machine, and so forth (Khan et al. 2020).

Machine Learning techniques are recently used in various emerging fields of science. Like, medical domain, Smart city, Health, etc. (Kalantari et al. 2018). Machine learning is a subfield of artificial intelligence (AI) where software programs can classify and predict results accurately without programming them explicitly. The learning process of ML software modules involves providing some data for those models, allowing those models to look for patterns into data and make better decisions in the future based on the data provided. The main aim of ML is to enable software programs to be learned directly from provided data and adjust their results according to this data without the aid or interference of humans.

Deep extreme machine learning is a subfield of machine learning. Deep extreme machine learning generally uses sequences of several layers to accomplish the feature extraction and classification tasks. Layers used in deep extreme machine learning are connected in a cascade manner so that the output of each layer is connected to the input of the following layer. With deep extreme machine learning, software modules can be learned and trained to accomplish classification and prediction tasks from images, sounds, videos, or text data. The performance and accuracy of deep extreme machine learning models can be very excellent and exceed human beings' performance. Deep extreme machine learning models are trained to accomplish classification or regression tasks using many datasets (data with labels) and robust neural network structures (Nguyen et al. 2019).

The artificial neural network model is based on various layers that are connected like neurons. An artificial neural network is another subfield of machine learning that is briefly stimulated by the human neural network, and it employs different neurons to perform amassed tasks. Moreover, Deep Neural Network frequently constructs an effective model for diagnostics breast cancer disease that provides better accuracy in data (Shen et al. 2019).

Late diagnose of breast cancer is one of the leading causes of death. This research proposed an intelligent system that will predict breast cancer empowered with fusion-based machine learning algorithms. The two different types of datasets (images and Features based) will collect from Kaggle and IEEE-dataport for this research. The proposed model will be trained on an image-based dataset using a convolutional neural network and a features dataset trained using deep extreme machine learning.

## Aim(s)

To design an effective and efficient health care model to predict breast cancer disease based on a fused Machine Learning approach.

## Objectives

This research aims to overcome the banking loan defaulter and predict it in the early stages.

* Literature Review work on state of the art and popular approaches already used for breast cancer prediction.
* Analyzing the datasets for better accuracy in minimum time to predict breast cancer in the health care sector.
* Devising a new probabilistic approach fused based model for breast cancer prediction using machine learning.

## Research Questions

The question that motivates the project’s progress forward is as follows:

* What is the significance of predicting breast cancer stages by using a fused Machine Learning model?
* How does the Machine Learning model can help to extend the accuracy of the proposed systems?
* What are the parameters that involve in the evaluation of the breast cancer prediction system?

## Ethical Considerations

Ethics is a complicated subject that has only become more prominent during the advent of Big Data. The UK Data Service department also provides guidelines for ethical research with specific relation to Big Data. These guidelines will form the basis for this report’s ethical approach.

## Literature Review

A strong Literature Review provides good guidance from the experiences of fellow researchers working in the same domain. It also includes the quality and validation of the research being done. The following papers have been finalized for Literature Review in this Research.

These days hazardous development is an enormous trendy scientific difficulty all completed. As indicated through the International Agency for Research on Cancer (IARC) is the part of the World Health Organization (WHO), there were 8.2 million passing's performed by compromising development in 2012, and 27 million of the latest activities of this disorder are relied on to take place until 2030.

(Berbar, 2018) were focused on methods of feature extraction to diagnose the malignant masses in mammograms and classify them. They worked to enhance the performance of the GLCM method by introducing seven texture features and applied them on sub-images. By merging two types of different components, they proposed three hybrid methods; Wavelet-CT1,-CT2, and ST-GLCM. They suggested that all these methods outperform compared previous feature extraction methods according to the AUC measure. They concluded that the GLCM or ST-GLCM extracted a small number of features relative to multi-resolution parts.

(Aliev et al. 2017) explored the use of neural techniques to determine heart activity and ensure their outcomes using smart devices. They strongly recommended the use of the neural approach in the field of medical diagnosis.

(Khan et al. 2020) proposed a cloud-based intelligent BCP-T1F and BCP-SVM expert system that specifically diagnoses the breast cancer stage and the type in an infected person. The expert system will elaborate on the grievous stages of cancer, to which extent a patient has suffered or not. This BCP-SVM system gives the higher precision 97.06 % of the breast cancer detection model, and the BCP-T1F expert system provides 96.56% accuracy of breast cancer at an initial stage.

Another study introduced a technique for the classification of mammograms which was consisted of 4 stages. The preprocessing phase was applied in the median filter to upgrade the picture's nature and exclude clamor in the image. They analyzed the difference in the average of the mammograms and explored the artificial neural network classifier to group the image into a fitting class. The outcome of sensitiveness, particularity, and exactness activity in their model were 72.72%, 93.6%, and 88.66%, respectively (Hamad et al. 2018).

(Chiao et al. 2019) proposed a framework using a convolutional neural network to observe breast malignancy. The various systems have been introduced to recognize breast malignancy in which a convolutional neural network is used to make a quick explanation and diagnose breast malignancy disease.

Identifying breast cancer at early stages is an effective way to decrease the associated mortality rate. Specialized radiologists often misconduct the examination or misinterpreted the vital information related to the diagnosis (Muramatsu et al. 2016).

## Project Timeline

This process of assigning deadlines to the modules of the project is called creating a project timeline. A Gantt chart is the tool preferred worldwide to visualize the project timeline. The project timeline for this project is shown in Figure 1.

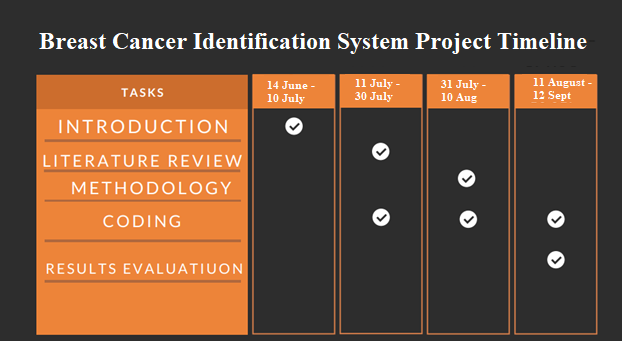


Figure 1: Project Timeline

# Methodology

The proposed model of predicting breast cancer using fused based machine learning model will explain in this section. Breast most cancers are the most painful sickness for women in the entire globe. Mostly we used mammography to come across the presence of breast cancer. But we've got different methods to discover this disorder with the aid of only artificial intelligence methods.

Artificial intelligence plays a vital role in every field of life, and computer intelligence's rapid growth has transformed into the digital world. In the Intelligent Health care system, early diagnosis of diseases can decrease the death rate worldwide. The proposed model consisted of fusion-based machine learning techniques.

The proposed model will collect data from two sources: medical examination and electronic medical record, shown in Figure 2. The data fusion approach will apply to the medical examination dataset and medical examination data stored in a big database using different sensors. On the other hand, EMR data is also stored in a big database. In the next step, the data fusion approach will apply to different data types like structured and unstructured.

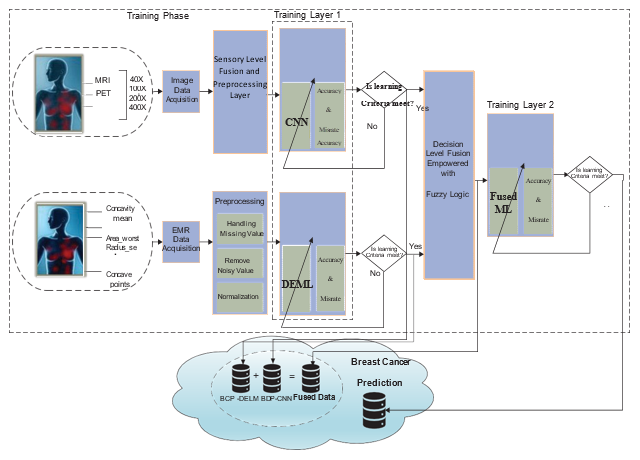


Figure 2: Proposed Model for Breast Cancer Prediction System

The preprocessing layer activates and mitigates the noise as well as missing values using the moving average method. ML approaches will apply for the prediction of disease. If the results show positive, then refer to Hospital for further treatment, and if the results show negative, then no disease is detected. In both cases, the data is stored in big data.

The proposed model consisted of two phases training and validation. In the training phase, medical examination and electronic medical records will use to collect data to detect diseases. Various smart city devices like IoMT will collect data in the form of images using different sensors. In IoT, electronic medical records will obtain in the form of text. Data obtained from IoMT is called unstructured, and data obtained from IoT is called structured data. Pre-processing handles the noise removal and resizing images, and the Data fusion approach technique will apply to structured and unstructured data in structured data. There are further two layers named training layer one and training layer two. In training layer 1, there are two sub-layers, namely the application layer and the performance layer. Machine learning techniques will apply to predicting disease and performance layers to evaluate the proposed model's performance. If the learning criteria don’t perform well, the model needs to retrain. Otherwise, the following training layer 2 activates for predicting disease stages and data store in the cloud to predict diseases. In training layer 2, machine learning techniques will apply to data for predicting disease stages, and a performance layer used to evaluate the performance of the proposed model using some statistical parameters like accuracy and miss rate. If the learning criteria don’t perform well, the model needs to retrain. Otherwise, the data will store in the cloud for the prediction of disease stages.

In the validation phase, IoMT and IoT use data to predict diseases and their stages. The data fusion approach is applied to structured and unstructured data and becomes structured data. After data fusion, the Preprocessing layer is used to mitigate noise, missing values, and resizing images. The proposed model imports data from the cloud for the prediction of diseases. If diseases do not detect, then it’s called negative. Otherwise, it is called positive. After predicting conditions, the proposed model imports data from the cloud to predict disease stages, and the proposed model will recommend a doctor or Hospital.

# Project Evaluation

In this portion, all experimental procedures, tests, and results will be discussed. It will start with describing the image-based and features-based dataset used for training and testing the proposed model. Next, the achieved accuracy and performance of the segmentation CNN and DEML will be introduced to the breast cancer prediction system. After that, the accuracy and implementation of the whole prediction system will be presented. This fusion-based model based on machine learning techniques will provide the much better accuracy than the traditional method that has already been applied.

# Conclusion

Breast cancer is one of the significant types of cancer and a substantial cause of death rate among women. The only way to decrease it is to diagnose it at early stages. The preliminary determination of this research to design the automated system assists medical practitioners in diagnosing breast cancer disease by using machine learning techniques. The proposed method will use to analyze the stage of breast cancer quickly. The methodology in the present study will show the highest accuracy compared with previously reported studies on breast cancer diagnosis. This system will extensively change the conventional method, making the system more efficient, convenient, and personalized.

# References

* Aliev, T.A., Rzayeva, N.E. and Sattarova, U.E., 2017. Robust correlation technology for online monitoring of changes in the state of the heart by means of laptops and smartphones. *Biomedical signal processing and control*, *31*, pp.44-51.
* Berbar, M.A., 2018. Hybrid methods for feature extraction for breast masses classification. *Egyptian informatics journal*, *19*(1), pp.63-73.
* Chiao, J.Y., Chen, K.Y., Liao, K.Y.K., Hsieh, P.H., Zhang, G. and Huang, T.C., 2019. Detection and classification the breast tumors using mask R-CNN on sonograms. *Medicine*, *98*(19).
* Hamad, Y.A., Simonov, K. and Naeem, M.B., 2018, November. Breast cancer detection and classification using artificial neural networks. In *2018 1st Annual International Conference on Information and Sciences (AiCIS)* (pp. 51-57). IEEE.
* Jafari, S.H., Saadatpour, Z., Salmaninejad, A., Momeni, F., Mokhtari, M., Nahand, J.S., Rahmati, M., Mirzaei, H. and Kianmehr, M., 2018. Breast cancer diagnosis: Imaging techniques and biochemical markers. *Journal of cellular physiology*, *233*(7), pp.5200-5213.
* Kalantari, A., Kamsin, A., Shamshirband, S., Gani, A., Alinejad-Rokny, H. and Chronopoulos, A.T., 2018. Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions. *Neurocomputing*, *276*, pp.2-22.
* Khan, F., Khan, M.A., Abbas, S., Athar, A., Siddiqui, S.Y., Khan, A.H., Saeed, M.A. and Hussain, M., 2020. Cloud-based breast cancer prediction empowered with soft computing approaches. *Journal of Healthcare Engineering*, *2020*.
* Momenimovahed, Z. and Salehiniya, H., 2019. Epidemiological characteristics of and risk factors for breast cancer in the world. *Breast Cancer: Targets and Therapy*, *11*, p.151.
* Muramatsu, C., Hara, T., Endo, T. and Fujita, H., 2016. Breast mass classification on mammograms using radial local ternary patterns. *Computers in biology and medicine*, *72*, pp.43-53.
* Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., García, Á.L., Heredia, I., Malík, P. and Hluchý, L., 2019. Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, *52*(1), pp.77-124.
* Schmid, I., Klenk, A.K., Sparber-Sauer, M., Koscielniak, E., Maxwell, R. and Häberle, B., 2018. Kaposiform hemangioendothelioma in children: a benign vascular tumor with multiple treatment options. *World Journal of Pediatrics*, *14*(4), pp.322-329.
* Shen, L., Margolies, L.R., Rothstein, J.H., Fluder, E., McBride, R. and Sieh, W., 2019. Deep learning to improve breast cancer detection on screening mammography. *Scientific reports*, *9*(1), pp.1-12.
* Wan, Z., Yin, T., Chen, H. and Li, D., 2016. Surgical treatment of a retroperitoneal benign tumor surrounding important blood vessels by fractionated resection: A case report and review of the literature. *Oncology letters*, *11*(5), pp.3259-3264.
* Zaheer, S., Shah, N., Maqbool, S.A. and Soomro, N.M., 2019. Estimates of past and future time trends in age-specific breast cancer incidence among women in Karachi, Pakistan: 2004–2025. *BMC public health*, *19*(1), pp.1-9.
* <https://ieee-dataport.org/documents/mammograms-breast-cancer-images#files>
* <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>