IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 2, February 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202039 |

Breast Cancer Prediction Using Histopathological Image Classification

J.Jayashree, Mr.S.Karthick, Mr.G.Raghul, Mr.M.Maniraj

Assistant Professor, Master of Computer Application, Muthayammal Engineering College, Rasipuram, India

Student, Department of CSE (Cyber Security), Muthayammal Engineering College, Rasipuram, India

Student, Department of CSE (Cyber Security), Muthayammal Engineering College, Rasipuram, India

Student, Department of CSE (Cyber Security), Muthayammal Engineering College, Rasipuram, India

ABSTRACT - Breast cancer (BC) is closely linked with the maximum mortality rate for cancer detection across the globe and has become a predominant public health issue. Earlier detection might increase the possibility of survival and successful treatment. However, it is a time-consuming and very challenging task that depends on the diagnostician's experience. For patients and their prognosis, it is essential that BC cancer can be automatically detected by the analysis of histopathological images. Conventional feature extraction method extracts some lower-level features of images, and preceding knowledge is essential for selecting suitable features that could be heavily impacted by human beings. The deep learning (DL) technique extracts higher-level abstract features from an image automatically. Therefore, this study develops a new Ensemble Learning with Symbiotic Organism Search Optimization Algorithm for Breast Cancer Classification (ELSOSA-BCC) technique on Histopathological Images. In the ELSOSA-BCC technique, the noise is removed using Gabor filtering (GF). In addition, the ELSOSA-BCC technique employs the EfficientNet-B0 model for feature extraction and optimal hyperparameter tuning using the SOS algorithm. Finally, the ensemble learning-based classification process is performed by three classifiers namely deep stacked autoencoder (DSAE), kernel extreme learning machine (KELM), and bidirectional long short-term memory (BiLSTM). In this study, ELSOSA-BCC simulation values are tested on a medical dataset. ELSOSA-BCC has been shown to perform better than other models in the experimental results.

KEYWORDS: Breast Cnacer, Histopathological Image Classification, Feature Extraction, Deep Learning, Ensemble Learning with Symbiotic Organism Search Optimization Algorithm for Breast Cancer Classification(ELSOSA-BCC), Gaborr Filterring(GF), EfficientNet-B0 model, SOS Algorithm, Deep Stacked Auto-Encoder{DSAE).

I. INTRODUCTION

Breast cancer (BC) is the major factor in higher mortality rates in women all around the world. The heterogeneous nature of BC makes its initial representation a crucial step in treatment planning and decision-making. The routine clinical analysis of BC can be performed by using several radiology images, involving Magnetic Resonance Imaging (MRI), ultrasound, and mammography. Nonetheless, this non-invasive methodology might not effectively represent the heterogeneous behavior of BC. Hence, the pathological study is followed as a benchmark to comprehend the pathophysiology of BC. In the presented technique, the tissue sample is collected and mounted on glass slides and then stained this slide for the best description of immunophenotypic and tumoral morphological features. In addition, the lack of labeled and extensive datasets has made another serious problem for the abovementioned challenges. In the context of DL, features can be extracted and retrieved data automatically, and abstract representations can be learned automatically. In the fields of biomedical science, computer vision, etc., they could solve the problem of conventional feature extraction. Parallel to this, the unprecedented advances in machine learning enable diagnosis based on image analysis, previously only possible in certain specialties, through the synergy of artificial intelligence and digital pathology. In comparison with human pathologists evaluating options, computer-aided image analysis allows for a more thorough identification, extraction, and quantification of features. An artificial intelligencebased diagnosis of skin

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202039 |

cancer is compared with datasets widely used and prevalent reviews. Using deep learning and machine learning techniques, the research will provide a deeper understanding of skin cancer diagnosis.

II. LITERATURE SURVEY

Several studies have been reported on breast cancer prediction. The studies from different approaches are applied to the given problem statement and showed high classification accuracies. Details of some of the previous research works are given in the following Table 1.

Table 1. Overview of papers dealing with Various methods and findings

Title	Author	Concept	Limitation
Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques (2020)	Md. Milon, Islam	Presented a comparative study of five machine learning techniques for the prediction of breast cancer.	Outliers may be occurred
Prediction of Breast Cancer using Machine Learning Approaches (2022)	Reza Rabiei	Models were initially trained with demographic and laboratory features (20 features). The models were then trained with all demographic, laboratory, and mammographic features (24 features) to measure the effectiveness of mammography features in predicting breast cancer	False positive rate is high
Breast cancer prediction using machine learning (2020)	Rawal, Ramik	The proposed work can be used to predict the outcome of different technique and suitable technique can be used depending upon requirement.SVM has the highest value of correctly classified instances.	Only support labeled datasets
Breast cancer prediction using varying parameters of machine learning models	Gupta, Puja, and Shruti Garg	Plan to apply and other machine learning algorithms using new parameters on larger data sets. It presents six supervised machine learning algorithms such as k-Nearest Neighborhood, Logistic Regression, Decision Tree, Random forest, Support Vector Machine with radial basis fuctional kernel.	Computational complexity is high
Breast Cancer Prediction using varying Parameters of Machine Learning Models (2020)	Mohamed Amine Naji, Sana El Filali, Kawtar Aarika, EL Habib Benlahmar, Rachida Ait Abdelouhahid, Olivier Debauche	The main objective of this research paper is to predict and diagnosis breast cancer, using machine-learning algorithms and find out the most effective with respect to confusion matrix, accuracy and precision.	Computational complexity is high

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202039 |

III. MACHINE LEARNING METHODS

A. Feature Extraction using CNN:

In contrast to hand-crafted features, deep learning attempts to learn features directly from input data. Achieving many levels of feature representation from an old one for more abstract semantics of the data is the primary goal of deep learning. Convolutional neural networks are a specific type of deep learning approach that has gained interest in the fields of image categorization and feature extraction. A few mathematical functions are combined to create a CNN model. Convolution is a CNN function that generates an output by taking inputs, like an image, and applying a filter or kernel. Trainable layers are placed on top of each other to create CNN. A supervised classifier and a few sets of arrays known as features connect these layers.

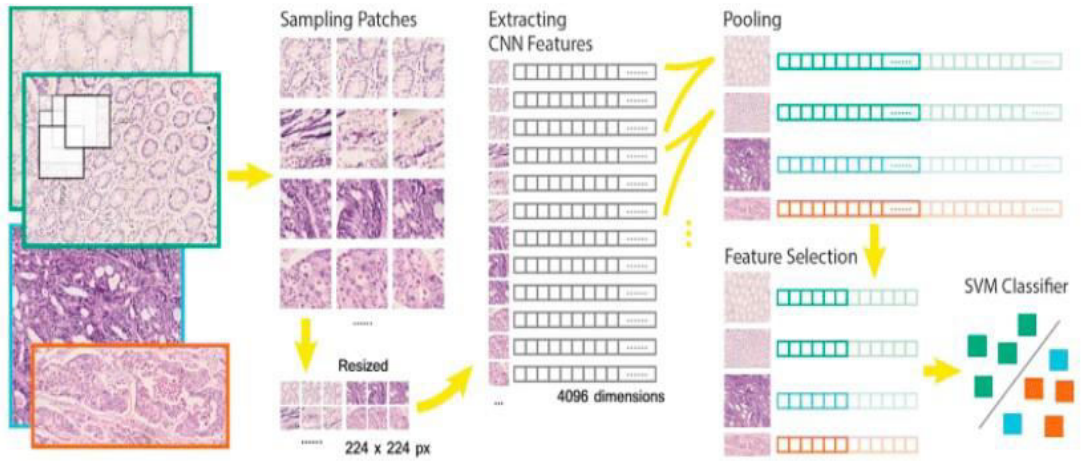


Figure 1. Feature extraction using CNN

CNN is made up of three primary components: the FC layer, the pooling layer, and the convolutional layer. The Convolutional Layer, which is the primary building element of CNN, uses a feature detector with an input signal, as illustrated in Fig. 1.This layer computes the output of every neuron connected in local regions. It then computes the dot product of the weights for the next input. The term "filter" or "kern" refers to the area where neurons are attached to the input. Rectified Linear Units, or ReLUs, are used to create non-linearity in the provided image. A pool layer, which is frequently used in the trade, is positioned in between two successive convolutional layers. The pool layer is typically employed to produce a lower progressive spatial size of representation by resizing the input size for the subsequent convolution. A fully connected (FC) layer is one that is connected to every activation function of the layer that came before it. Because it is supplied in the FC Layer, a classifier is typically utilised in the end. Finding the appropriate class for an image based on attributes that have been recognised is the primary goal of employing classifiers.

B. Machine Learning Classifier

Classification is the process of determining which specific class a given data piece belongs to. Targets, labels, and categories are used to refer classes in classification tasks. Predictive modelling uses a mapping function to classify different data points by mapping input variables (X) to output variables (Y).

C. Support Vector Machine

Using a separating hyperplane, this supervised learning system can create precise classifications. Other methods involve the use of trained labelled data (supervised learning), from which the algorithm generates an ideal separation hyperplane for the purpose of classifying fresh samples. The hyperplane in 2D space splits a 2-D plane into segments, each of which, based on the labelled data, defines a class.

$$1/2 \|w\|2 + C (\sum\delta(i) i)$$

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202039 |

In the above equation, w represents the weights of class, C is the regularization parameter for the SVM and δ is the slack variable.

D. Logistic Regression

It is a Classification model which is very simple to implement and performs efficiently on linear separable classes. It is a Binary classification algorithm but can be implemented for multi-class classification by using OvR Method. Logistic Regression uses Sigmoid function with parameter z as net input: -

$$\emptyset(z) = 1/1+e^{-z}$$

IV.DATASET DESCRIPTION

In this proposed approach, we have used BreakHis Dataset which is also known as Breast Cancer Histopathological Images[19]. The BreakHis dataset contains overall 7909 images of histopathological Samples stained with Hematoxylin and Eosin. The Dataset has been divided into two categories i.e Benign and Malignant with respective to sample 2,480 and 5429. For implementation. our dataset is partitioned into the following manner: 80% for the training phase and 20% for testing phase as shown in Table 2.

Table 2. Training and Testing Samples used for the experiment

Cancer Class	Training Samples(80%)		Testing Samples(20%)	
	Benign	Malignant	Benign	Malignant
Number of Samples	1683	4637	792	788

These image samples are compiled from 82 patients tissue samples by using different Magnification factors (40x,100x,200x and 400x) as shown in Table 3.

Table 3. Number of samples in Benign and Malignant

Magnification	40X	100X	200X	400X
Number of total samples	1596	1687	1617	1420
Benign	424	446	424	389
Malignant	1172	1241	1193	1031

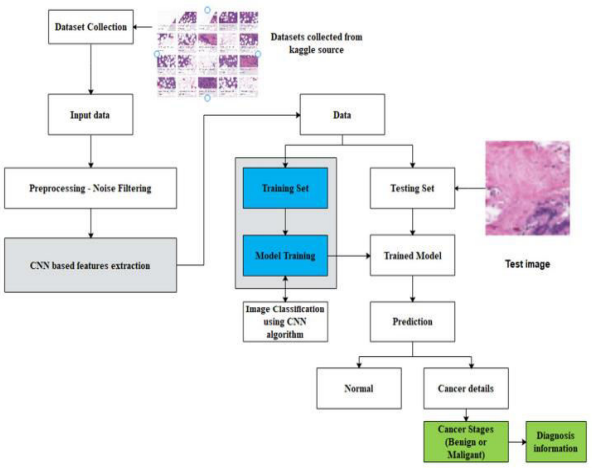
V. PROPOSED METHODOLOGY

Cancer is a potentially fatal disease caused mainly by environmental factors that mutate genes encoding critical cell regulatory proteins. Early detection and prevention of cancer plays a very important role in reducing deaths caused by cancer. Identification of genetic and environmental factors is very important in developing novel methods to detect and prevent cancer. Implement the framework to classify the cancer from histopathological images and implement deep learning based neural network algorithm can be used to predict the cancer diseases with improved accuracy. Neural Networks has emerged as an important method of classification. Convolutional neural network algorithm has been employed as the training algorithm in this work. This project proposes a diagnostic system for predicting cancer disease with improved accuracy. And also specify the level cancer whether it is benign or Malignant.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

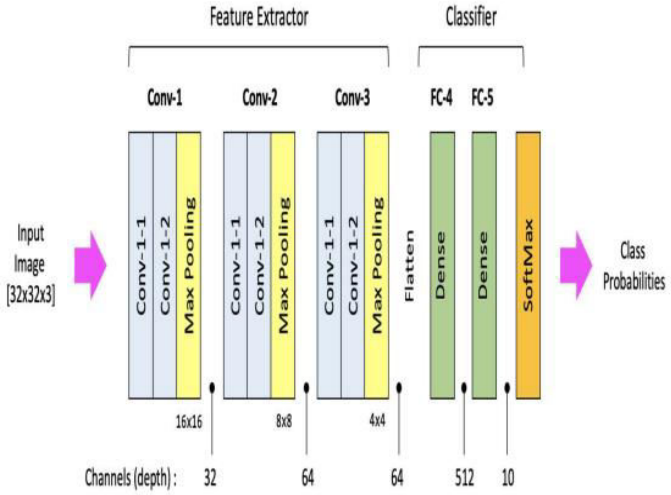
|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202039 |



A. Model Definiton

Convolutional Neural Network (CNN) is a type of artificial neural network designed for processing structured grid data, such as images. CNNs have proven to be highly effective in tasks like image classification, object detection, and image generation. The key idea behind CNNs is the use of convolutional layers to automatically and adaptively learn spatial hierarchies of features from input data. Define the architecture of the CNN, including the number of layers, the types of layers, and the hyperparameters. This may involve choosing the appropriate combination of convolutional, pooling, and fully connected layers.



B. Model Training:

Train the CNN on the prepared data using an appropriate optimization algorithm, such as stochastic gradient descent or Adam. This involves adjusting the weights and biases of the network to minimize the loss function and improve its performance on the training data.

C. Model Evaluation:

Evaluate the performance of the trained model on a separate validation set to prevent overfitting and to tune the hyperparameters.

D. Model Inference:

Use the trained model to make predictions on new, unseen data.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1202039 |

VI.OUTPUT

The final output is the predicted class labels or regression values for each sample in the input data. This not only facilitates accurate cancer identification but also minimizes the risk of misdiagnosis, thereby improving patient outcomes. In conclusion, the application of Convolutional Neural Networks (CNNs) in histopathological image analysis for breast cancer prediction represents a significant stride towards enhancing diagnostic accuracy and efficiency. The CNN model's ability to automatically learn hierarchical representations from raw pixel data enables it to discern intricate structures and abnormalities within histopathological images. This not only facilitates accurate cancer identification but also minimizes the risk of misdiagnosis, thereby improving patient outcomes.

REFERENCES

1. Islam, Md Milon, et al. "Breast cancer prediction: a comparative study using machine learning techniques." SN Computer Science 1 (2020): 1-14.
2. Rawal, Ramik. "Breast cancer prediction using machine learning." Journal of Emerging Technologies and Innovative Research (JETIR) 13.24 (2020): 7.
3. Rabiei, Reza, et al. "Prediction of breast cancer using machine learning approaches." Journal of Biomedical Physics & Engineering 12.3 (2022): 297.
4. Naji, Mohammed Amine, et al. "Machine learning algorithms for breast cancer prediction and diagnosis." Procedia Computer Science 191 (2021): 487-492.
5. Gupta, Puja, and Shruti Garg. "Breast cancer prediction using varying parameters of machine learning models." Procedia Computer Science 171 (2020): 593-601.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details