Project Report

On

Study of Breast cancer detection using different CNN models

Guided By

Dr. Ram Sarkar

Professor of Computer Science & Engineering, Jadavpur University

Group Members

Name	Roll Number
Pulak Kumar Ghosh	001910501042
Aryapriyo Mandal	001910501033

Acknowledgement

We are indebted to my project guide, Dr. Ram Sarkar, Professor, Dept. of Computer Science and Engineering, Jadavpur University for his constant support, prompt help and unwavering encouragement which helped me a lot in completing this project. Under his guidance We have been able to do fruitful research and learn new concepts at hand. We are also thankful to Prof. Nandini Mukherjee, Head of the Dept. of Computer Science and Engineering, Jadavpur University for allowing us to carry out our research in the department.

Pulak Kumar Ghosh
001910501042

Aryapriyo Mandal
001910501033

4th year B.E. Student

Department of Computer Science and Engineering

Jadavpur University

Certificate

This is to certify that the project entitled "Breast cancer detection using artificial intelligence techniques" has been carried out by Pulak Kumar Ghosh and Aryapriyo Mandal under my guidance and supervision, and be accepted in partial fulfilment of the requirement for the degree of Bachelor of Computer Science and Engineering.

Dr. Ram Sarkar

Professor

Jadavpur University

Bachelor of Computer Science & Engineering

Abstract: Cancer is one of the most dangerous diseases to humans, and yet no permanent cure has been developed for it. Breast cancer is one of the most common cancer types. According to the National Breast Cancer Foundation, in 2020 alone, more than 276,000 new cases of invasive breast cancer and more than 48,000 non-invasive cases
were diagnosed in the US. To put these figures in perspective, 64% of these cases are diagnosed early in the disease's cycle, giving patients a 99% chance of survival. Artificial intelligence and machine learning have been used effectively in detection and treatment of several dangerous diseases, helping in early diagnosis and treatment, and thus increasing the patient's chance of survival. Deep learning has been designed to analyse the most
important features affecting detection and treatment of serious diseases. For example, breast cancer can be detected using genes or histopathological imaging. Analysis at the genetic level is very expensive, so histopathological imaging is the most common approach used to detect breast cancer. In this research work, we systematically reviewed previous work done on detection and treatment of breast cancer using genetic sequencing or histopathological imaging with the help of deep learning and machine learning. We also provide recommendations to researchers who will work in this field.

Keywords: BREAST CANCER, MACHINE LEARNING, DEEP LEARNING

RELATED WORKS :

Many studies have been conducted about breast cancer detection through imaging or through genomics. However, to the best of our knowledge, no research has been conducted including both techniques.

Some authors summarised the various techniques used to classify breast cancer using histopathological image analysis (HIA) based on different architectures of artificial neural networks (ANN). The authors grouped their work according to the applied dataset. They arranged it in ascending chronological order. This work found that ANNs were first used in the field of HIA around 2012. ANNs and PNNs were the most frequently applied algorithms. However, in feature extraction, most of the work used textural and morphological features. It was clear that deep CNNs were quite effective for early detection and diagnosis of breast cancer, leading to more successful treatment. Prediction of non-communicable diseases (NCDs) was conducted using many algorithms.

In some books, the authors compared the performance of various classification algorithms. The classification algorithm were performed on eight NCD datasets using eight classification algorithms and 10-fold cross-validation method. These were evaluated using AUC as an indicator of accuracy. The authors stated that the NCD datasets have noisy data and irrelevant attributes. KNN, SVM and NN proved to be robust to this noise. In addition, they stated that the irrelevant attribute problem can be handled with some pre-processing techniques to improve the accuracy rate.

Natural inspired computing (NIC) algorithms have been designed and applied to diagnose various human disorders. Some authors introduced five insect-based NIC algorithms used for diagnosis diabetes and cancer. The authors

found that it achieved a high level of performance in detecting different types of cancer (breast, lung, prostate and ovarian). To be more specific, breast cancer was detected using a hybridisation of the guided ABC and neural networks.

The authors also developed a highly effective methodology of detecting diabetes and leukaemia. They concluded that the hybridisation of NICs with other classification algorithms produces more precise and promising results. They mentioned that more work is required to detect different stages of diabetes and cancer.

Some authors demonstrated the effectiveness of NNs, in the classification of cancer diagnoses, especially in the initial stages. According to their study, the majority of NNs have shown promise in detecting tumour cells. However, the imaging approach required high computational capacity to pre-process the images.

Some authors reviewed different machine learning, deep learning and data mining algorithms related to breast cancer prediction. Several research papers on breast cancer were reviewed, with a total of 27 papers in machine learning, 4 papers in ensemble techniques and 8 papers in deep learning techniques. The authors mentioned that most of the papers used imaging, while only a few papers and genetics. The main algorithms used to predict breast cancer using genetics over SVM, decision tree and random forest. However, imaging techniques used several algorithms such as CNN and Naïve Bayes.

On the other hand, some authors focused on gene mutation for detecting breast cancer. They mentioned that the gene prediction classification phase aims to carry out gene annotation, gene finding and gene mutation detecting

to ascertain the presence or absence of a cancer. They concluded that several methods can be used including regression, probability models, SVMs, NNs and deep learning. They also mentioned the many opportunities available to capture the relationship between nucleotide and feature extraction, since DNA sequencing involves a large amount of data in the form of a string sequence.

Some authors examined recent studies applying deep learning to breast cancer with different imaging modalities. They organised these studies using the aspects of dataset, architecture, application and evaluation. They focused on deep learning frameworks developed in three breast cancer imaging utilizing DLR-based CAD systems. Their study included private datasets and classification using CNNs.

After studying these surveys, our contribution will involve studying imaging to predict breast cancer and to get more information that can help early diagnosis and treatment of breast cancer. We will also provide recommendations to researchers who wish to conduct research in this area.

Introduction

Breast Cancer is one of the major causes of death in women around the world. According to the American cancer society, 41760 women and more than 500 men died from breast cancer recently. Breast cancer occurs in four main types: normal, benign, in-situ carcinoma and invasive carcinoma. A benign tumour involves a minor change in the breast structure. It is not harmful and does not classify as a harmful cancer. In cases of in-situ carcinoma, the cancer is only in the mammary duct lobule system and does not affect other organs. This type is not dangerous and can be treated if diagnosed early. Invasive carcinoma is considered to be the most dangerous type of breast cancer, as it can spread to all other organs. According to the scientists, breast cancer can be detected using several methods including X-ray mammography, ultrasound (US), computed tomography (CT), portion emission tomography (PET), magnetic resonance imaging (MRI) and breast temperature measurement. Usually, the golden standard is a pathological diagnosis for detecting breast cancer. This involves an image analysis of the removed tissue, which is stained in the lab to increase visibility. Haematoxylin and eosin (H&E) are commonly used for staining process. Breast cancer can be diagnosed using one of two approaches: histopathological image analysis or genomics. Histopathological images are microscopic images of breast tissue that are extremely useful in early treatment of the cancer. As for genomics, the scientists stated that radio-genomics is an emerging research field focusing on multi-scale associations between medical imaging and gene expression data.

Radio-genomics provide both radiological and genetic features that can enhance diagnosis. It can analyse tissues at the molecular level, helping with prediction and early detection of cancer. The main difference between

imaging information and radio-genomics is the critical knowledge gap between imaging at the tissue level and analysing the underlying molecular and genetic disease biomarkers. As imaging is less precise, it may lead to over- or under-treatment. While radio-genomics is much more effectively than histopathological imaging, it is rarely used because the process involves datasets that are expensive and require high computational power. As a result, a limited number of labs conduct radio-genomics experiments. The research paper addresses the following research questions and highlights the deep learning models looking at their performance, the datasets used and possibilities for breast cancer classification and detection.

1. **Which deep learning models perform most effectively?**

   We will compare deep learning models with classical machine learning models to compare their performance. We will also list the performance metrics used.

2. **What are the most used features for breast cancer classification? How are these features selected and extracted?**

   We will observe the most important features that contribute to breast cancer classification, and the methods used to extract these features.

3. **What datasets are available for both gene sequencing and MRI? What feature selection and extraction methods are used?**

   We will list and discuss all public and private datasets for gene sequencing and MRI imaging data. We will also list some of the methods used to select and extract the features.

4. **Comparing gene sequence data with image data, for breast cancer detection problem, what are the drawbacks, challenges, and advantages?**

We will compare imaging and gene sequencing as they relate to breast cancer detection, using a tabular presentation to highlight the main differences between the two approaches.

Dataset

This action lists the most common datasets containing mammograms used for the breast cancer detection and classification. There are five datasets presented which are: DDSM, CBIS-DDSM, MIAS, Inbreast and BCDR datasets.
We have using CBIS-DDSM datasets for our work.

CBIS-DDSM Dataset: Curated Breast Imaging Subset of DDSM i.e. CBIS-DDSM is an updated and standardised version of the Digital Database for Screening Mammography (DDSM). The CBIS-DDSM collection is available as well which includes a subset of the DDSM data selected and curated by a trained mammographer. The CBIS-DDSM contains 6775 studies. The images have been decompressed and converted to DICOM format. Updates ROI segmentation and bounding boxes, and pathologic diagnosis for training data are also included.

Dataset Link: https://www.kaggle.com/code/khizarkhan/breast-cancer-detection-using-resnet50/data (accessed on 1st November, 2022)

Dataset was divided into 3 columns: Normal, Benign, Malignant

Methodology

Our target topic is breast cancer detection using deep learning. We ended up using around 80 of the most recent papers related to breast cancer treatment and diagnosis. Some of the papers examined only deep learning, while others used a combination of machine learning and deep learning.

In our search process, we mainly used the Scopus database to obtain the articles. This is to exclude non-refereed publications. However, in Fig. 1, we state the distribution of selected papers among the existing databases. The top five databases are PubMed, ScienceDirect (Elsevier), IEEE, Springer and Nature.

We used the following search statement: ("breast cancer") AND ALL ("deep learning" OR "deep neural network") AND ALL ("gene" OR "genome" OR "microarray" OR "DNA" OR "X-ray" OR "mammography" OR "MRI" OR "ultrasound"). More than 1000 papers were found that were published between January 2010 and May 2020. The following figure shows the distribution of paper publication during this period (Fig. 2).

We can see that breast cancer study publications peaked around 2019 and 2020. We reduced the number of papers to 80 by including only papers using genetic expression and imaging and by focusing on journal and conference articles only. The imaging modalities we considered were ultrasound, radiography, mammography and magnetic resonance imaging (MRI), as well as various types of gene expression and gene sequencing. In this research, we focused on papers that implement the breast cancer detection using the techniques of AI, as well as papers that predict breast cancer using both gene data and image data. We applied the following eligibility criteria on each paper: (1) The language is English; (2) the topic is related to breast cancer

detection and treatment; (3) the paper discussed hybrid models of machine learning; (4) the paper discussed genetic expression data; (5) the paper purely discusses deep learning; (6) the paper discusses imaging data; (7) only journal and conference publications are retained; (8) only medical or biomedical engineering publications are kept which are related to the topic. Please note that non-refereed publications were excluded from the study.

Firstly, we recorded the main information such as the name of the paper, year of the publication, the list of authors and the publisher. Then, we included some information to conduct the systematic review, such as the algorithm used and whether the paper discusses only deep learning or a hybrid between DL and ML, the recorded accuracy and other performance evaluation parameters, the dataset, the features, and many other columns. We answered our research questions using these criteria.

We have used 4 different CNN models to check their accuracy and time-consume. They are: ResNet, VggNet, Inception, MobileNet

ResNet(Residual Networks): In order to solve the problem of the vanishing/exploding gradient, this architecture introduced the concept called Residual Blocks. In this network, we use a technique called skip connections. The skip connection connects activations of a layer to further layers by skipping some layers in between. This forms a residual block. Resnets are made by stacking these residual blocks together.

The approach behind this network is instead of layers learning the underlying mapping, we allow the network to fit the residual mapping. So, instead of say H(x), initial mapping, let the network fit,

$F(x) := H(x) - x$ which gives $H(x) := F(x) + x$.

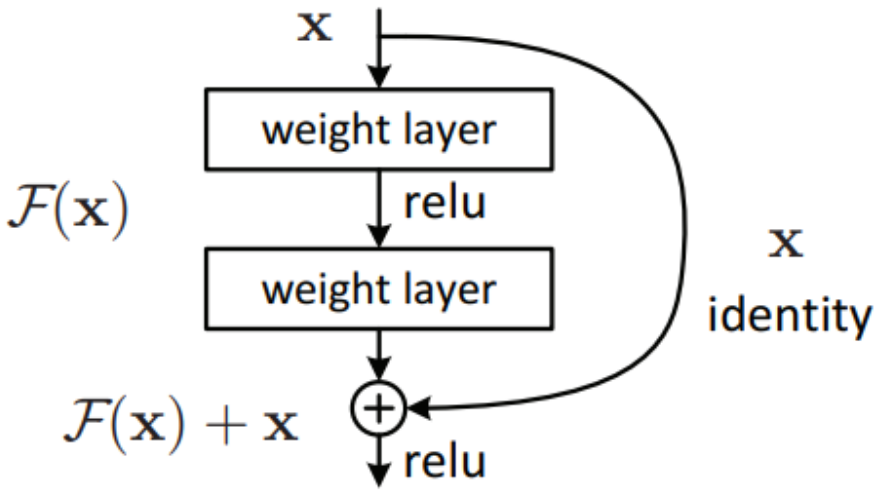

Fig 1 : Skip (Shortcut) connection

The advantage of adding this type of skip connection is that if any layer hurt the performance of architecture then it will be skipped by regularization. So, this results in training a very deep neural network without the problems caused by vanishing/exploding gradient.

VGG Net : The input to this network is an image of dimensions (224, 224, 3). The first two layers have 64 channels of a 3*3 filter size and the same padding. Then after a max pool layer of stride (2, 2), two layers have convolution layers of 128 filter size and filter size (3, 3). This is followed by a max-pooling layer of stride (2, 2) which is the same as the previous layer. Then there are 2 convolution layers and a max pool layer. Each has 512 filters of (3, 3) size with the same padding. This image is then passed to the stack of two convolution layers. In these convolution and max-pooling layers, the filters we use are of the size 3*3 instead of 11*11 in AlexNet and 7*7 in ZF-Net. In some of the layers, it also uses 1*1 pixel which is used to manipulate the number of input channels. There is a padding of 1-pixel (same padding) done after each convolution layer to prevent the spatial feature of the image.

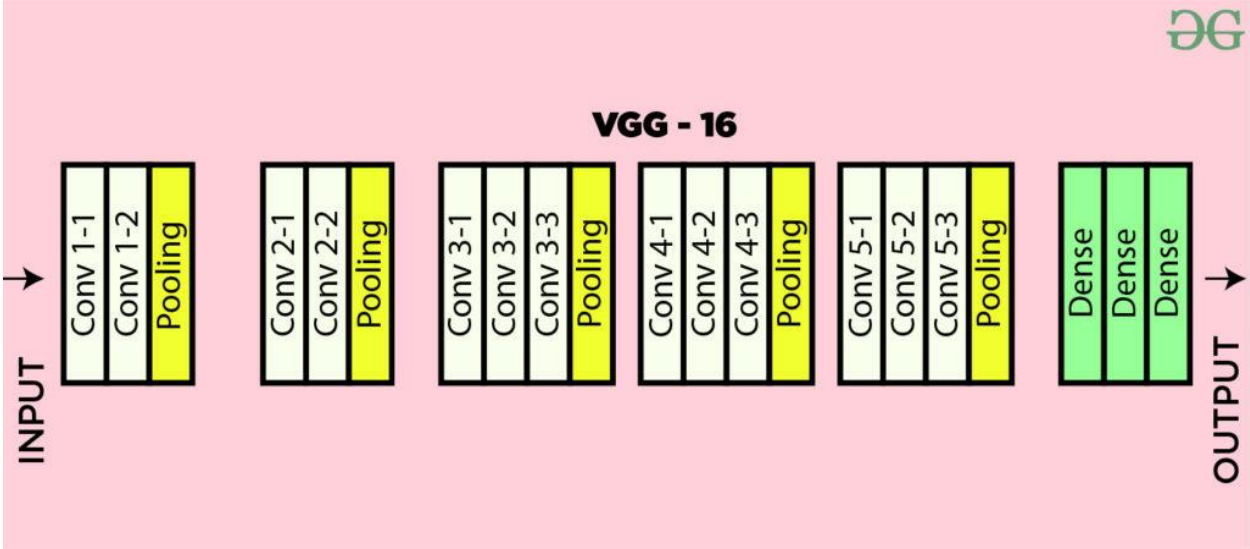

Fig 2 : VGG-16 Architecture Map

After the stack of convolution and max-pooling layer, we got a (7, 7, 512) feature map. We flatten this output to make it a (1, 25088) feature vector. After this there is 3 fully connected layer, the first layer takes input from the last feature vector and outputs a (1, 4096) vector, the second layer also outputs a vector of size (1, 4096) but the third layer output a 1000 channels for 1000 classes of ILSVRC challenge i.e. 3rd fully connected layer is used to implement softmax function to classify 1000 classes. All the hidden layers use ReLU as its activation function. ReLU is more computationally efficient because it results in faster learning and it also decreases the likelihood of vanishing gradient problems.

MobileNet : MobileNet is a model which does the same convolution as done by CNN to filter images but in a different way than those done by the previous CNN. It uses the idea of Depth convolution and point convolution which is different from the normal convolution as done by normal CNNs. This increases the efficiency of CNN to predict images and hence they can be able to compete in the mobile systems as well. Since these ways of convolution reduce the

comparison and recognition time a lot, so it provides a better response in a very short time and hence we are using them as our image recognition model.

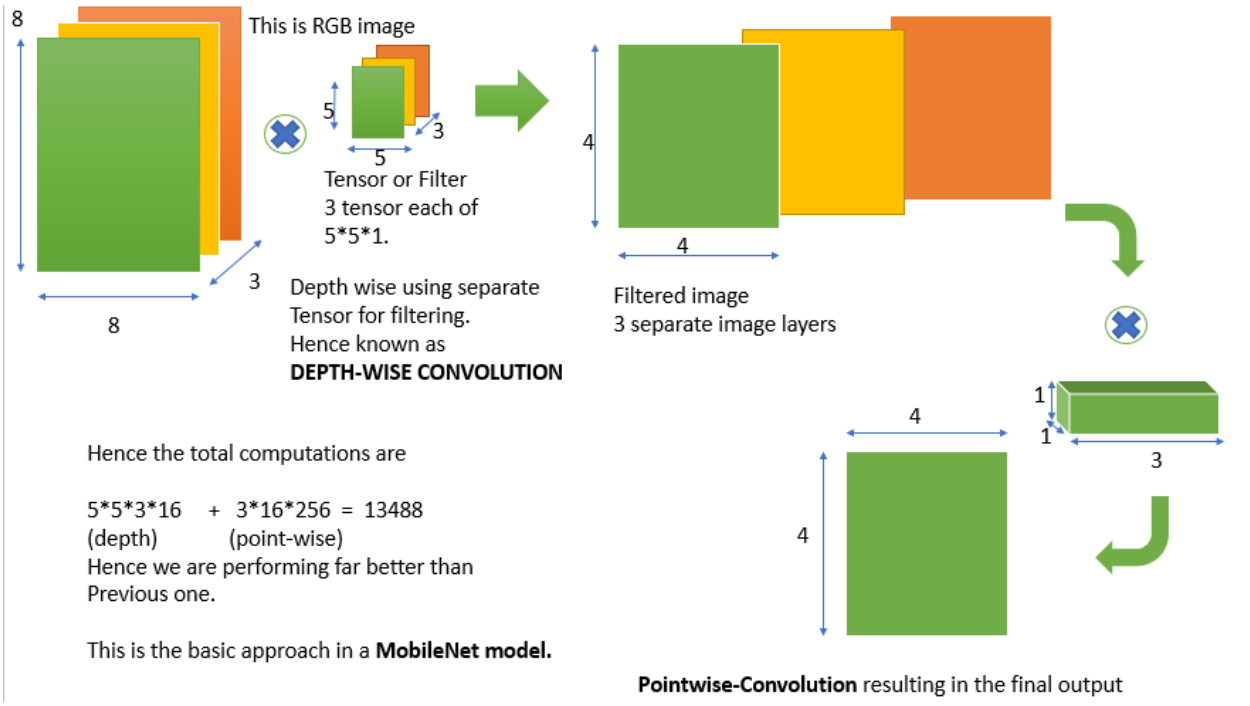

Fig 3 : MobileNet Architecture

Inception: Inception V1 have sometimes use convolutions such as 5*5 that causes the input dimensions to decrease by a large margin. This causes the neural network to uses some accuracy decrease. The reason behind that the neural network is susceptible to information loss if the input dimension decreases too drastically. Furthermore, there is also a complexity decrease when we use bigger convolutions like 5x5 as compared to 3x3. We can go further in terms of factorization i.e. that we can divide a 3x3 convolution into an asymmetric convolution of 1x3 then followed by a 3x1 convolution. This is equivalent to sliding a two-layer network with the same receptive field as in a 3x3 convolution but 33% cheaper than 3x3. This factorization does not work well for early layers when input dimensions are big but only when the input size mxm (m is between 12 and 20). According to the Inception V1

architecture, the auxiliary classifier improves the convergence of the network. They argue that it can help reduce the effect of the vanishing gradient problem in the deep networks by pushing the useful gradient to earlier layers (to reduce the loss). But, the authors of this paper found that this classifier didn't improve the convergence very much early in the training.

Inception V3 is similar to and contains all the features of Inception V2 with following changes/additions:

- Use of RMSprop optimizer.
- Batch Normalization in the fully connected layer of Auxiliary classifier.
- Use of 7x7 factorized Convolution
- **Label Smoothing Regularization:** It is a method to regularize the classifier by estimating the effect of label-dropout during training. It prevents the classifier to predict a class too confidently. The addition of label smoothing gives 0.2% improvement from the error rate.

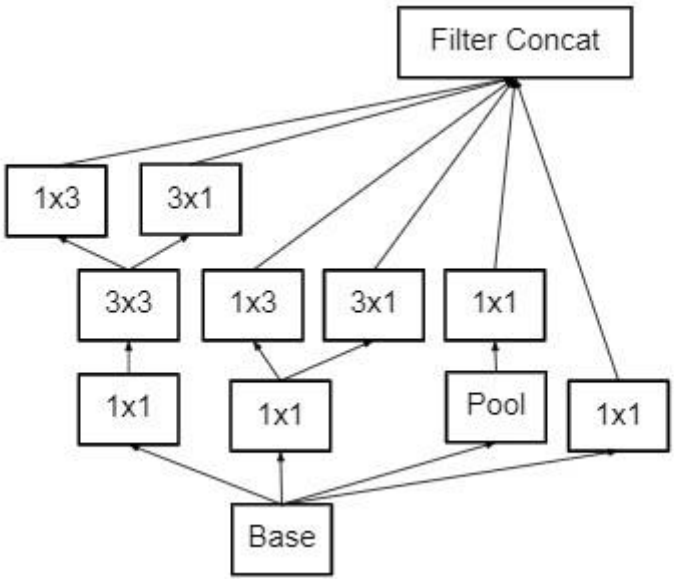

Fig 4 : Inception V3 Architecture

We have run all four types of models, and their results are shown in the result section.

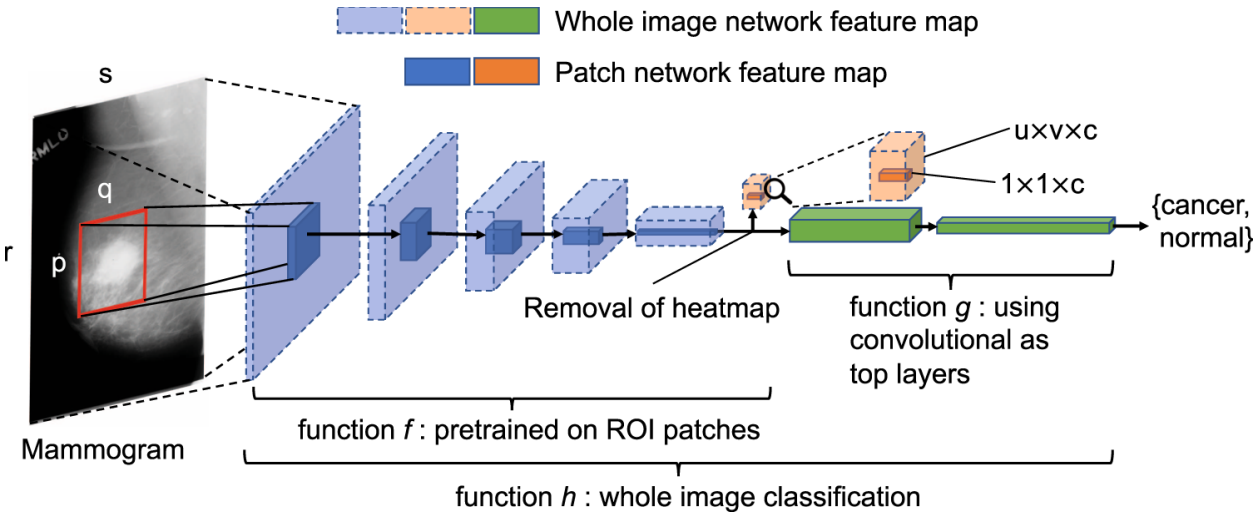

Fig 5 : CNN Diagram

Resnet50 was used instead of convolutions that are shown

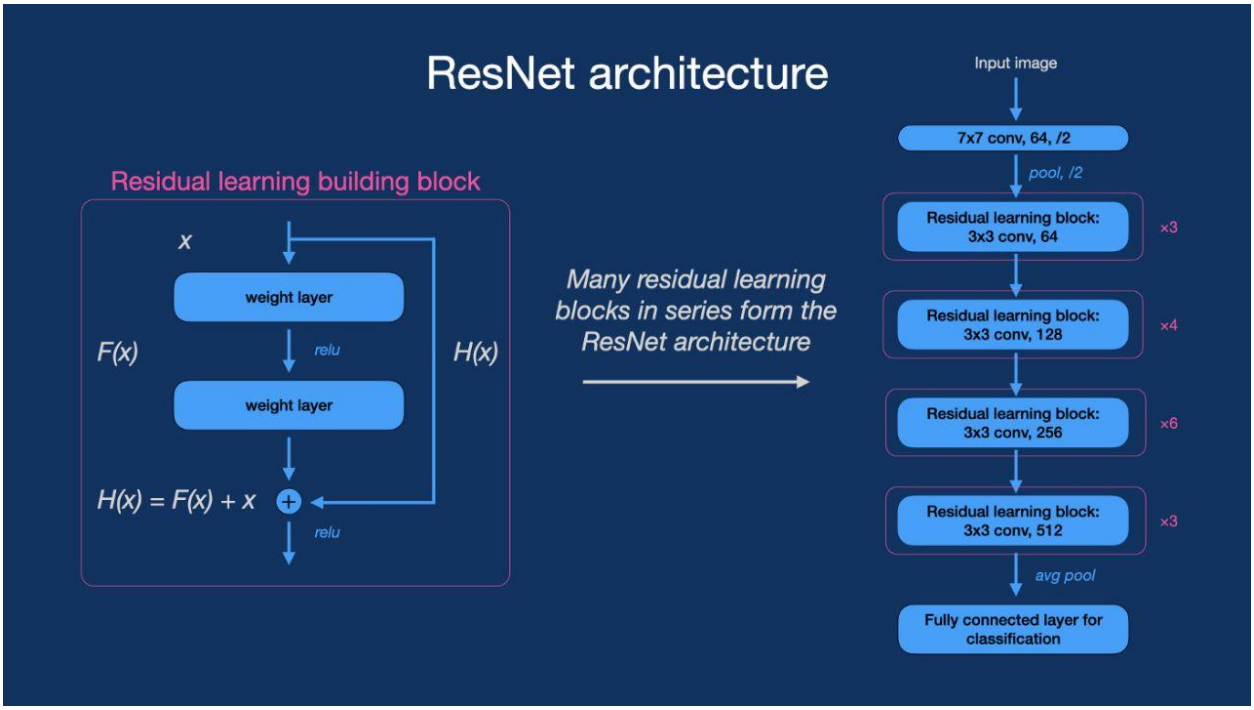

Fig 6 : ResNet Architecture Diagram

Results and Discussion:

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 resnet50 (Functional)       (None, 4, 4, 2048)        23587712

 flatten (Flatten)           (None, 32768)             0

 batch_normalization (BatchN  (None, 32768)            131072
 ormalization)

 dense (Dense)               (None, 256)               8388864

 dropout (Dropout)           (None, 256)               0

 dense_1 (Dense)             (None, 128)               32896

 dropout_1 (Dropout)         (None, 128)               0

 dense_2 (Dense)             (None, 64)                8256

 dropout_2 (Dropout)         (None, 64)                0

 dense_3 (Dense)             (None, 3)                 195

=================================================================
Total params: 32,148,995
Trainable params: 8,495,747
Non-trainable params: 23,653,248
```

Fig 7 : Resnet50 & 3 Dense Layers

Fig 8 : ResNet Model Metrics Visualization

Accuracy after 50 epochs : 0.75 & Loss after 10 epochs : 1.00.

Training accuracy : 0.97 & Training Loss: 0.1

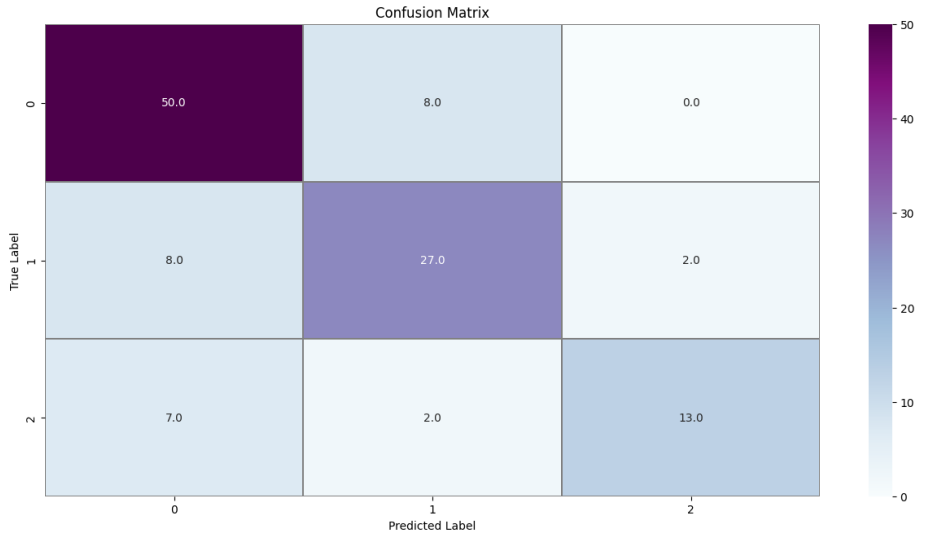

Fig 9 : ResNet Confusion Matrix

For MobileNet Model :

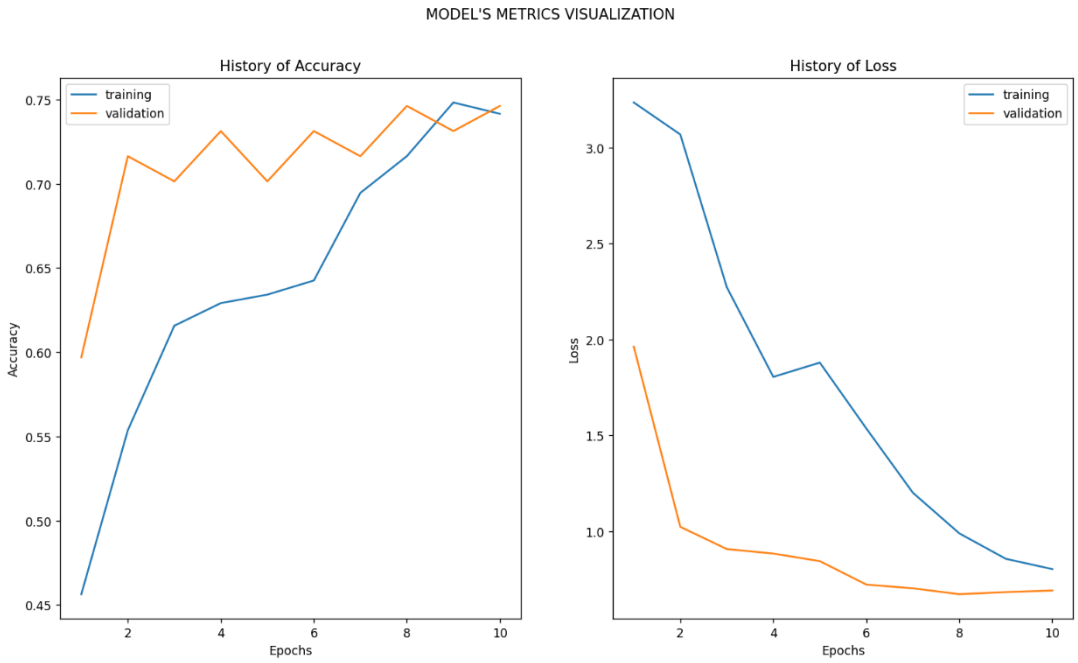

Fig 10 : MobileNet Model Metrics Visualization

After 50 epochs Accuracy is 0.74 and Loss is 0.4

Training Accuracy is 0.75 and Training loss is 0.6

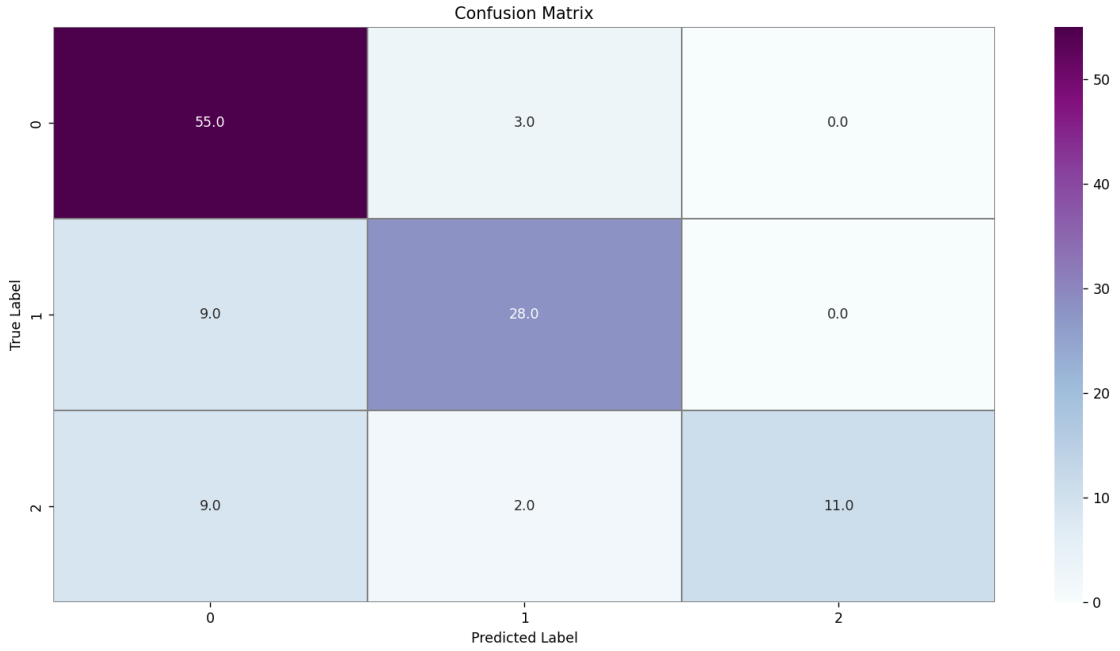

Fig 11 : MobileNet Confusion Matrix

For VGG Model:

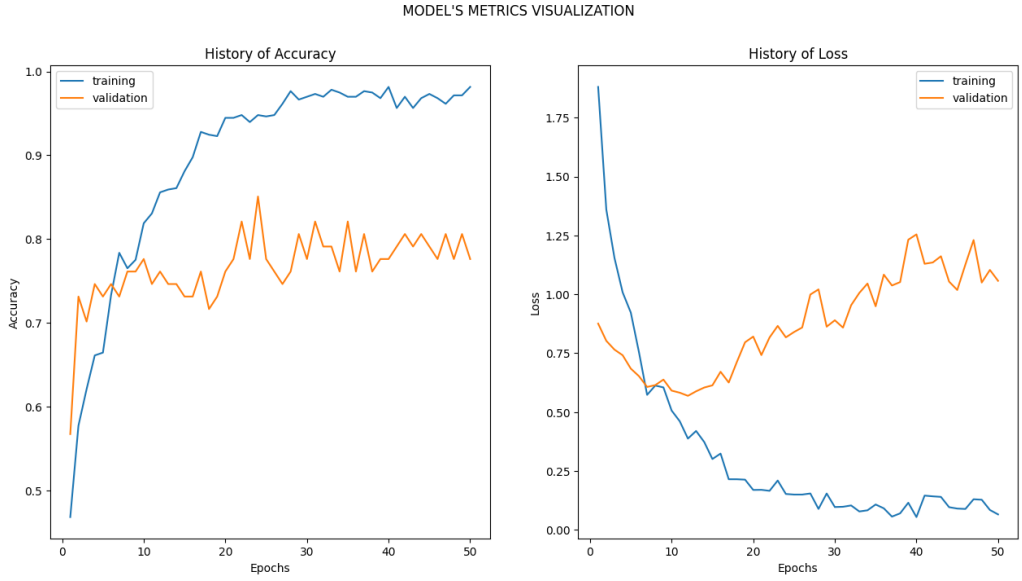

Fig 12 : VGG Net Model Metrics Visualization

After 50 epochs Accuracy is 0.75 and Loss is 1.00

Training Accuracy is 0.98 and Loss is 0.01

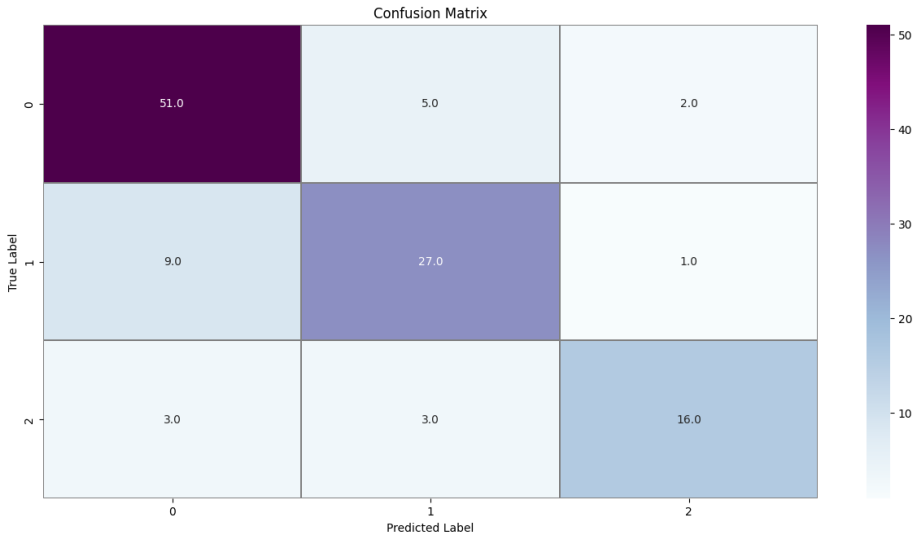

Fig 13 : VGG Net Confusion Matrix

For Inception Model:

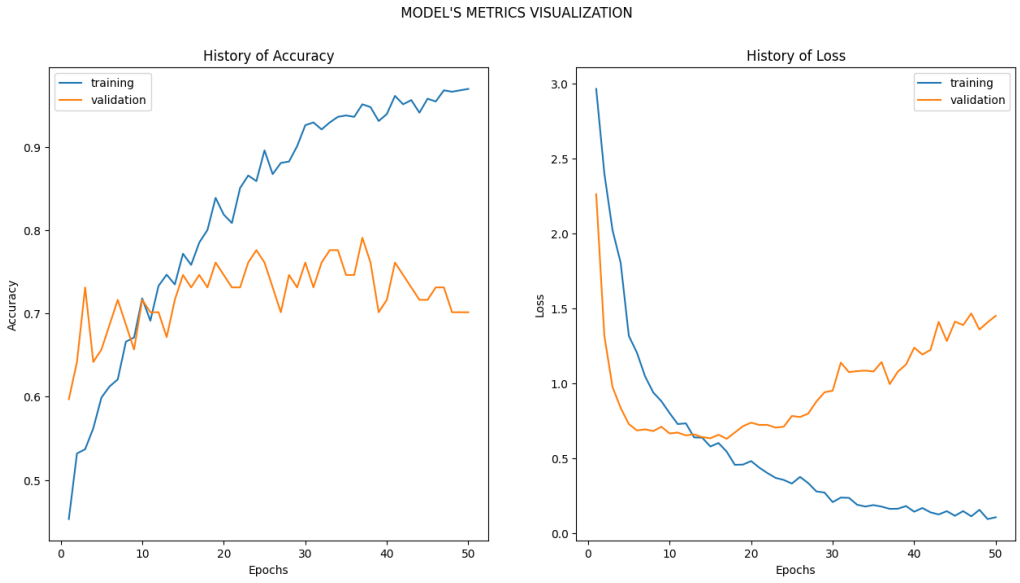

Fig 14 : Inception Model Metrics Visualization

After 50 epochs Accuracy is 0.78 and Loss is 1.4

Training Accuracy is 0.98 and Loss is 0.02

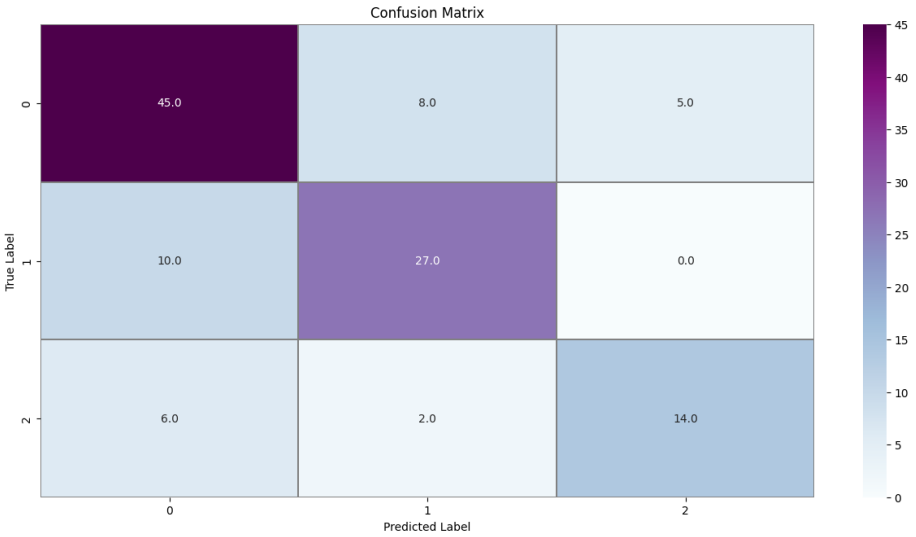

Fig 15 : Inception Confusion Matrix

#	Accuracy	Loss	Training Accuracy	Training Loss
ResNet	0.75	1.00	0.97	0.1
VGG Net	0.75	1.00	0.98	0.01
MobileNet	0.74	0.4	0.75	0.06
Inception	0.78	1.4	0.98	0.02

Competitive Analysis of Accuracy & Loss of different CNN Models

In this systematic study, our initial search turned up 1000 conference and journal papers. After eliminating duplicated papers and unrelated studies that were “purely medical or about cancer in general”, we ended up with 80 papers related to both ML and DL.

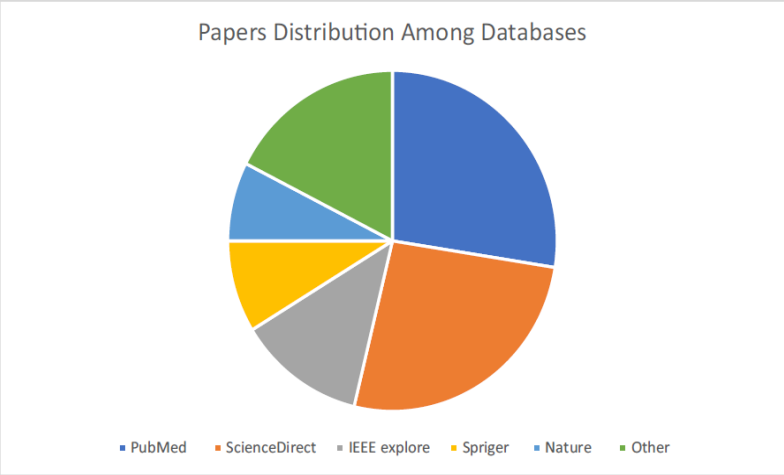

Fig 16 : Papers Distribution Among Databases

We wanted to focus on DL approaches or DL-ML hybrid models, so only papers related to DL were selected. Fig. 3 explains our search methodology. The following tables (Table 1 & 2) summarize the algorithms used and their performance, with some details related to both genetic and imaging data.

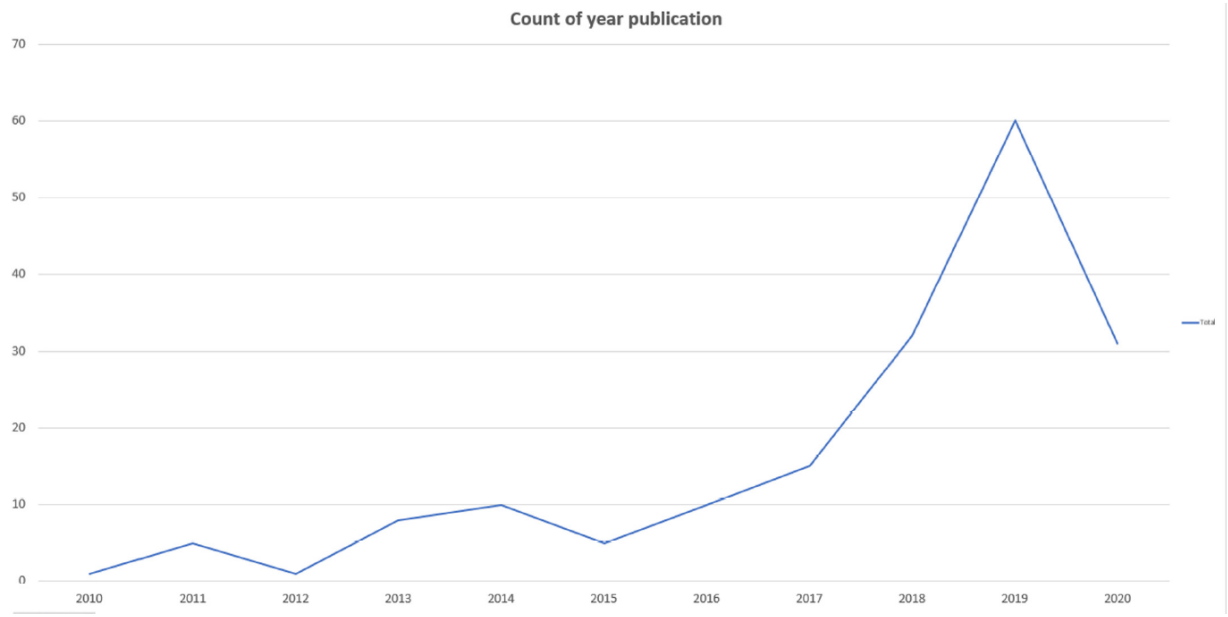

Fig 17 : Number of Publications between 2010 and 2020

In the previous table, generally, the work in this area is divided into two main groups, the first involving binary classification (whether or not breast cancer

is present) and the second classifying breast cancer types. We noticed that the use of binary classification results in the highest accuracy, and that it is generally more accurate than multiclass classification. In a paper, the author used the BPNN algorithm in a highly effective way. We also see that most selected papers with binary classifications performed well in terms of most evaluation parameters. However, the best performance for multiclass categorization or breast cancer subtype classification obtained 97% accuracy. The author of the paper basically compared machine learning and deep learning on the task of breast cancer subtype classification. Most of the papers were strictly focused on accuracy and did not mention other parameters such as precision, recall, AUC, etc. This is an important limitation when we are talking about medical projects. Many models have been used including CNN, DNN with attention mechanisms. Because not all papers took into consideration the confusion matrix parameters, we will include only accuracy in our graphs.

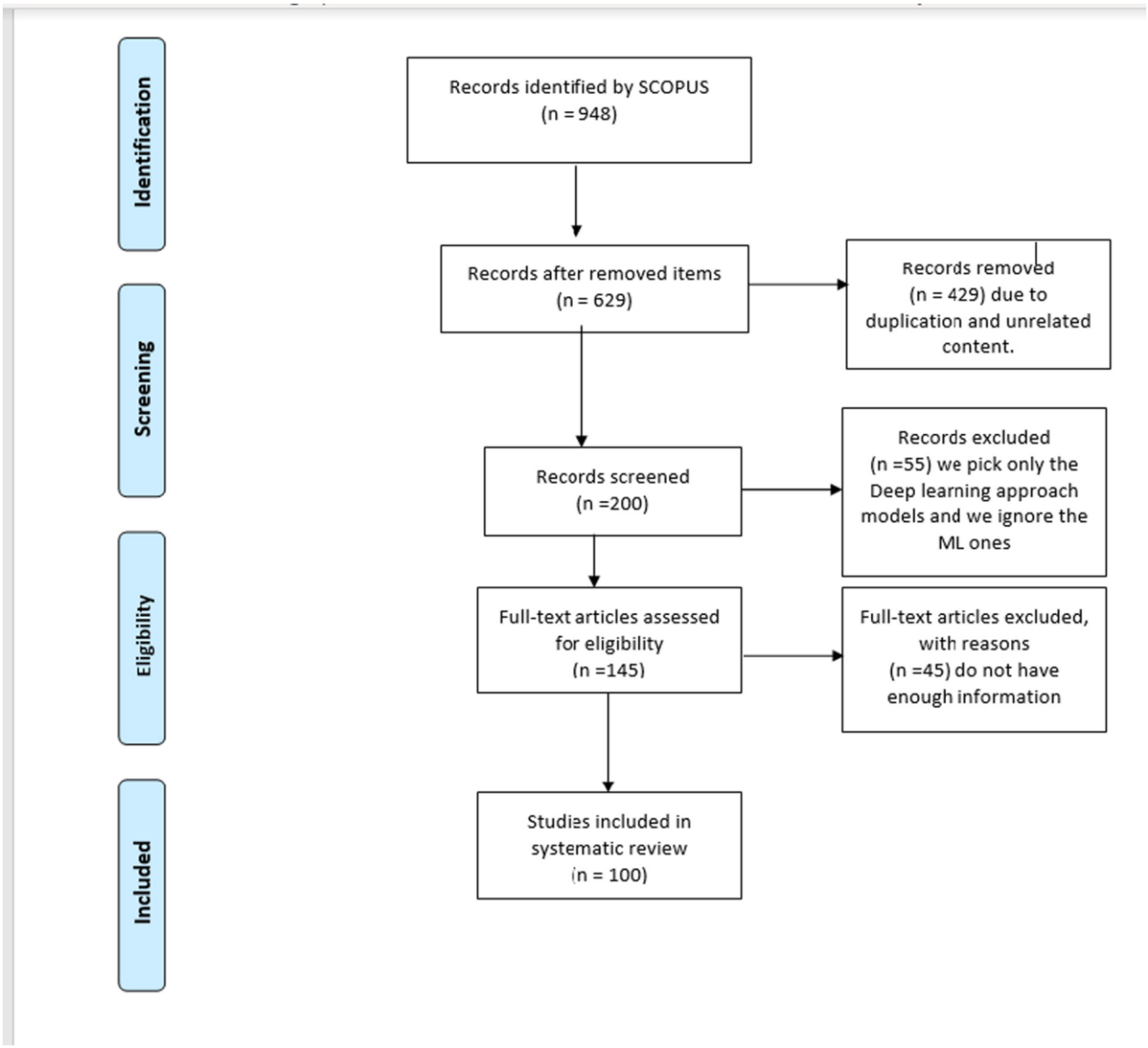

Fig 18 : Flow of information through the phases of a systematic review

Models, classes, and performance for gene sequencing data in selected papers.

Paper reference	Models/algorithm	Binary or multiclass	Classes	Accuracy	Other performance evaluation parameters	Anomaly application/task
[8]	IABC-EMBOT, IHM-FFNN, PSO-RM, ABCO-BCD and DNN-BCD	Binary	Negative, positive	0.975	–	BC detection
[9]	FNN, ANFIS, ANNFIS	Binary	Negative, positive	0.92	Precision: 0.944, recall: 0.944, F1:0.944	BC detection
[10]	Deep type, state-of-the-art	Multiclass	Normal, luminal A, luminal B, basal and HER2	–	–	Identifying cancer subtypes
[11]	BPNN	Binary	Mutant and non-mutant sequences	0.998	Sensitivity = 1 Specificity = 0.9985	Predication and classification of cancer
[12]	CNN	Multiclass	–	0.956	–	BC subtype classification
[13]	DA	Multiclass	–	0.95	–	BC detection
[14]	DNN + attention mechanism (hybrid)	Binary	–	0.87	–	BC detection
[15]	GCN	Multiclass	–	0.919	AUC = 0.84	Synergistic drug combinations
[16]	DNN + SVM (separately)	Multiclass	Binary, miotic/non	0.94	F-score: 0.556 Accuracy: 0.8319	Detection
[17]	DNN	Multiclass	4 classes - Basal-like, HER2-enriched, luminal A, and luminal B And binary (basal, non basal)	0.83	AUC: 0.82 Accuracy: 0.8682	Identify risk categories
[18]	DL + ML (separately) CNN, SVM, random forests, boosting	Multiclass	Axillary lymph node status, binary, cancer or not	0.97 by SVM	Accuracy: 0.98 AUC: 0.93	Cancer subtype classification
[19]	DL + ML (hybrid) Genetic algorithm (GA) based MLP, multilayer perceptron (MLP), logistic regression (LR)	Multiclass	Binary, cancer or not	0.84	AUC: 0.84	Prediction of axillary lymph node status in breast cancer
[20]	Feed forward neural network (FFNN)	Binary	Binary, cancer or not	0.983	Only accuracy	BC detection

Table 1

Model, classes, and best performance for MRI imaging data.

Paper reference	Models/algorithm Image	Binary or multiclass	Classes	Accuracy	Other performance evaluation parameters	Anomaly application
[73]	CNN	Multiclass	Luminal A Luminal B HER2	0.70	ROC = 0.85	BC subtype classification
[74]	CNN	Multiclass	Complete, partial, no response	0.88	Specificity of 0.951, sensitivity of 0.739	Predict breast tumor, Response to chemotherapy
[75]	CNN	Binary	Negative and positive	0.972	Sensitivity 0.983, and Specificity 0.965	BC detection
[76]	ANN, NB, K-NN, DT, RF	Multiclass	–	0.90	–	BC subtype classification
[77]	CNN	Binary	Negative and positive	0.873		
[78]	DL + ML (hybrid) Logistic regression, random forest and deep neural network	Binary	Negative and positive	0.98	–	BC detection
[79]	2CNN, 3CNN	Binary	Negative and positive	0.705	AUC = 0.763, sensitivity = 0.805, specificity = 0.618	BC detection
[80]	DL + ML (hybrid) Model 1: DBN-ELM- BP Model 2: DBN-BP-ELM Model 3: DBN + GA	Binary	Negative and positive	0.9975	–	BC detection

Table 2

To state the findings from Table 1, CNN model seems to be the number one used model among the papers for both binary and multiclass classification. A hybrid between machine learning and deep learning also seems to be affective

for such models. We can see a hybrid between MLP and LR. As well as SVM algorithm from ML has been used a lot in these papers.

Basically, when it comes to breast cancer detection using imaging data, we see high performance on binary classification. However, when it comes to classification of types, imaging did not produce results as accurately as gene expression data.

For multiclass differentiation or breast cancer subtype classification, the highest accuracy obtained using imaging data was 90% in a paper that uses ML. We can therefore see in Table 1 that in general, the use of CNN results in excellent performance for both gene expression and imaging. An example of this paper produced 97% accuracy for binary classification. The models used include both deep learning standalone models and hybrid models consisting of both machine learning and deep learning algorithms.

Comparing between hybrid and deep learning standalone models in gene expression, we can see that the standalone deep learning models obtain consistently higher accuracy. The hybrid technique in some paper obtained 87% accuracy, which was the lowest accuracy in all the models. The CNN in paper obtained an accuracy of 95%, while the algorithm BPNN achieved 99.8% accuracy. Many papers did not mention confusion matrix parameters. In fact, very few papers mentioned these parameters in which the authors mentioned sensitivity and specificity. Among the imaging papers, the highest accuracy was 99.7% for binary classification in a paper, which used a hybrid model with both DL and ML.

According to Table 2, the mostly used algorithm for breast cancer detecting and subtype classification is CNN, and that does make sense, since our data is MRI images and CNN considered the best for computer vision problems.

To conclude, many algorithms were used in the studies. Some papers used several models in series, while others used only one model. ANN and CNN were the most widely used algorithms in both gene sequence and images data. Many other algorithms were used, such as DNN and SVM, but most of the papers used CNN and ANN with various parameters and properties.

References:

[1] Zhou X, Li C, Rahaman MM, Yao Y, Ai S, Sun C, et al. A comprehensive review for
breast histopathology image analysis using classical and deep neural networks.
IEEE Access 2020;8:90931–56.
https://doi.org/10.1109/ACCESS.2020.2993788.
[2] Sutanto DH, Ghani MKA. A benchmark of classification framework for noncommunicable
disease prediction: a review. 2015.
[3] Gautam R, Kaur P, Sharma M. A comprehensive review on nature inspired computing algorithms for the diagnosis of chronic disorders in human beings. Prog
Artif Intell 2019;8:401–24. https://doi.org/10.1007/s13748-019-00191-1.
A.B. Nassif et al.
Artificial Intelligence In Medicine 127 (2022) 102276
12
[4] Mahmood M, Al-Khateeb B, Alwash WM. A review on neural networks approach on
classifying cancers. IAES Int J Artif Intell 2020;9:317–26. https://doi.org/10.11591/ijai.v9.i2.pp317-326.
[5] Fatima N, Liu L, Hong S, Ahmed H. Prediction of breast cancer, comparative review
of machine learning techniques, and their analysis. IEEE Access 2020;8: 150360–76. https://doi.org/10.1109/ACCESS.2020.3016715.
[6] Wisesty UN, Mengko T, Purwarianti A. Gene mutation detection for breast cancer
disease: a review. IOP Conf Ser Mater Sci Eng 2020;830:32051. https://doi.org/

10.1088/1757-899X/830/3/032051.
[7] Pang T, Wong JHD, Ng WL, Chan CS. Deep learning radiomics in breast cancer with
different modalities: overview and future. Expert Syst Appl 2020;158:113501. https://doi.org/10.1016/j.eswa.2020.113501.
[8] Punitha S, Amuthan A, Joseph KS. Enhanced monarchy butterfly optimization
technique for effective breast cancer diagnosis. J Med Syst 2019;43:206. https://
doi.org/10.1007/s10916-019-1348-8.
[9] Umamaheswari TS, Sumathi P. Enhanced firefly algorithm (EFA) based gene
selection and adaptive neuro neutrosophic inference system (ANNIS) prediction
model for detection of circulating tumor cells (CTCs) in breast cancer analysis. Cluster Comput 2019;22:14035–47. https://doi.org/10.1007/s10586-018-2183-2.
[10] Chen R, Yang L, Goodison S, Sun Y. Deep-learning approach to identifying cancer
subtypes using high-dimensional genomic data. Bioinformatics 2019;36:1476–83.
https://doi.org/10.1093/bioinformatics/btz769.
[11] Dhannoon BN. Predication and classification of cancer using sequence alignment
and back propagation algorithms in Brca1 and Brca2 genes. Int J Pharm Res 2019;
11. https://doi.org/10.31838/ijpr/2019.11.01.062.
[12] Mostavi M, Chiu Y-C, Huang Y, Chen Y. Convolutional neural network models for
cancer type prediction based on gene expression. BMC Med Genomics 2020;13.
https://doi.org/10.1186/s12920-020-0677-2.
[13] Liu Q, Hu P. Association analysis of deep genomic features extracted by denoising
autoencoders in breast cancer. Cancers (Basel) 2019;11. https://doi.org/10.3390/
cancers11040494.
[14] Chen H, Gao M, Zhang Y, Liang W, Zou X. Attention-based multi-NMF deep neural

network with multimodality data for breast cancer prognosis model. Biomed Res
Int 2019;2019:9523719. https://doi.org/10.1155/2019/9523719.
[15] Jiang P, Huang S, Fu Z, Sun Z, Lakowski TM, Hu P. Deep graph embedding for
prioritizing synergistic anticancer drug combinations. Comput Struct Biotechnol J
2020;18:427–38. https://doi.org/10.1016/j.csbj.2020.02.006.
[16] Danaee P, Ghaeini R, Hendrix DA. A deep learning approach for cancer detection
and relevant gene identification. Biocomput 2017:219–29. https://doi.org/
10.1142/9789813207813_0022.
[17] Romo-Bucheli D, Janowczyk A, Gilmore H, Romero E, Madabhushi A. A deep
learning based strategy for identifying and associating mitotic activity with gene
expression derived risk categories in estrogen receptor positive breast cancers.
Cytometry A 2017;91:566–73. https://doi.org/10.1002/cyto.a.23065.
[18] El-Nabawy A, El-Bendary N, Belal NA. A feature-fusion framework of clinical,
genomics, and histopathological data for METABRIC breast cancer subtype
classification. Appl Soft Comput 2020;91:106238. https://doi.org/10.1016/j.
asoc.2020.106238.
[19] Karakıs¸ R, Tez M, Kılıç YA, Kuru Y, Güler ˙I. A genetic algorithm model based on
artificial neural network for prediction of the axillary lymph node status in
breastcancer. Eng Appl Artif Intel 2013;26:945–50. https://doi.org/10.1016/j.
engappai.2012.10.013.
[20] Zhang F, Chen J, Wang M, Drabier R. A neural network approach to multibiomarker
panel discovery by high-throughput plasma proteomics profiling of
breast cancer. BMC Proc 2013;7:S10. https://doi.org/10.1186/1753-6561-7-S7-
S10.
[21] Alsaleem MA, Ball G, Toss MS, Raafat S, Aleskandarany M, Joseph C, et al. A novel
prognostic two-gene signature for triple negative breast cancer. Mod Pathol 2020;

33:2208–20. https://doi.org/10.1038/s41379-020-0563-7.
[22] Lancashire LJ, Powe DG, Reis-Filho JS, Rakha E, Lemetre C, Weigelt B, et al.
A validated gene expression profile for detecting clinical outcome in breast cancer
using artificial neural networks. Breast Cancer Res Treat 2010;120:83–93. https://
doi.org/10.1007/s10549-009-0378-1.
[23] Rawat RR, Ruderman D, Macklin P, Rimm DL, Agus DB. Correlating nuclear
morphometric patterns with estrogen receptor status in breast cancer pathologic
specimens. Npj Breast Cancer 2018;4:32. https://doi.org/10.1038/s41523-018-
0084-4.
[24] Zhang D, Zou L, Zhou X, He F. Integrating feature selection and feature extraction
methods with deep learning to predict clinical outcome of breast cancer. IEEE
Access 2018;6:28936–44. https://doi.org/10.1109/ACCESS.2018.2837654.
[25] Beykikhoshk A, Quinn TP, Lee SC, Tran T, Venkatesh S. DeepTRIAGE: interpretable
and individualised biomarker scores using attention mechanism for the
classification of breast cancer sub-types. BMC Med Genomics 2020;13:20. https://
doi.org/10.1186/s12920-020-0658-5.
[26] Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih N, et al.
High-throughput adaptive sampling for whole-slide histopathology image analysis
(HASHI) via convolutional neural networks: application to invasive breast cancer
detection. PLoS One 2018;13. https://doi.org/10.1371/journal.pone.0196828.
[27] Paul A, Mukherjee DP. Mitosis detection for invasive breast cancer grading in
histopathological images. IEEE Trans Image Process 2015;24:4041–54. https://doi.
org/10.1109/TIP.2015.2460455.
[28] Ehteshami Bejnordi B, Mullooly M, Pfeiffer RM, Fan S, Vacek PM, Weaver DL, et al.

Using deep convolutional neural networks to identify and classify tumor-associated
stroma in diagnostic breast biopsies. Mod Pathol 2018;31:1502–12. https://doi.
org/10.1038/s41379-018-0073-z.
[29] Simidjievski N, Bodnar C, Tariq I, Scherer P, Andres Terre H, Shams Z, et al.
Variational autoencoders for cancer data integration: design principles and
computational practice. Front Genet 2019;10:1205. https://doi.org/10.3389/
fgene.2019.01205.
[30] Shimizu H, Nakayama KI. A 23 gene–based molecular prognostic score precisely
predicts overall survival of breast cancer patients. EBioMedicine 2019;46:150–9.
https://doi.org/10.1016/j.ebiom.2019.07.046.
[31] Xu J, Luo X, Wang G, Gilmore H, Madabhushi A. A deep convolutional neural
network for segmenting and classifying epithelial and stromal regions in
histopathological images. Neurocomputing 2016;191:214–23. https://doi.org/
10.1016/j.neucom.2016.01.034.
[32] Jaber MI, Song B, Taylor C, Vaske CJ, Benz SC, Rabizadeh S, et al. A deep learning
image-based intrinsic molecular subtype classifier of breast tumors reveals tumor
heterogeneity that may affect survival. Breast Cancer Res 2020;22:12. https://doi.
org/10.1186/s13058-020-1248-3.
[33] Simsek S, Kursuncu U, Kibis E, AnisAbdellatif M, Dag A. A hybrid data mining
approach for identifying the temporal effects of variables associated with breast
cancer survival. Expert Syst Appl 2020;139:112863. https://doi.org/10.1016/j.
eswa.2019.112863.
[34] Sun D, Wang M, Li A. A multimodal deep neural network for human breast cancer
prognosis prediction by integrating multi-dimensional data. IEEE/ACM Trans
Comput Biol Bioinforma 2019;16:841–50. https://doi.org/10.1109/

TCBB.2018.2806438.
[35] Inan O, Uzer MS, Yilmaz N. A new hybrid feature selection method based on
association rules and PCA for detection of breast cancer. 2012.
[36] Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NNC, et al.
Accurate and reproducible invasive breast cancer detection in whole-slide images:
a deep learning approach for quantifying tumor extent. Sci Rep 2017;7. https://
doi.org/10.1038/srep46450.
[37] Mert A, Kılıç N, Akan A. An improved hybrid feature reduction for increased breast
cancer diagnostic performance. Biomed Eng Lett 2014;4:285–91. https://doi.org/
10.1007/s13534-014-0148-9.
[38] Jafari-Marandi R, Davarzani S, Soltanpour Gharibdousti M, Smith BK. An optimum
ANN-based breast cancer diagnosis: bridging gaps between ANN learning and
decision-making goals. Appl Soft Comput J 2018;72:108–20. https://doi.org/
10.1016/j.asoc.2018.07.060.
[39] Turkki R, Linder N, Kovanen PE, Pellinen T, Lundin J. Antibody-supervised deep
learning for quantification of tumor-infiltrating immune cells in hematoxylin and
eosin stained breast cancer samples. J Pathol Inform 2016;7. https://doi.org/
10.4103/2153-3539.189703.
[40] To˘gaçar M, Ergen B, C¨omert Z. Application of breast cancer diagnosis based on a
combination of convolutional neural networks, ridge regression and linear
discriminant analysis using invasive breast cancer images processed with
autoencoders. Med Hypotheses 2020;135:109503. https://doi.org/10.1016/j.
mehy.2019.109503.
[41] Tripathy RK, Mahanta S, Paul S. Artificial intelligence-based classification of breast
cancer using cellular images. RSC Adv 2014;4:9349–55. https://doi.org/10.1039/
c3ra47489e.

[42] Motalleb G. Artificial neural network analysis in preclinical breast cancer. Cell J
2014;15:324–31.
[43] Khameneh FD, Razavi S, Kamasak M. Automated segmentation of cell membranes
to evaluate HER2 status in whole slide images using a modified deep learning network. Comput Biol Med 2019;110:164–74. https://doi.org/10.1016/j.compbiomed.2019.05.020.
[44] Giordano A, Giuliano M, De Laurentiis M, Eleuteri A, Iorio F, Tagliaferri R, et al.
Artificial neural network analysis of circulating tumor cells in metastatic breast
cancer patients. Breast Cancer Res Treat 2011;129:451–8. https://doi.org/10.1007/s10549-011-1645-5.
[45] Romo-Bucheli D, Janowczyk A, Gilmore H, Romero E, Madabhushi A. Automated
tubule nuclei quantification and correlation with oncotype DX risk categories in ER
+ breast cancer whole slide images. Sci Rep 2016;6. https://doi.org/10.1038/srep32706.
[46] Mert A, Kiliç N, Bilgili E, Akan A. Breast cancer detection with reduced feature set.
Comput Math Methods Med 2015;2015. https://doi.org/10.1155/2015/265138.
[47] Shallu Mehra R. Breast cancer histology images classification: training from scratch
or transfer learning? ICT Express 2018;4:247–54. https://doi.org/10.1016/j.icte.2018.10.007.
[48] Cui X, Li Z, Zhao Y, Song A, Shi Y, Hai X, et al. Breast cancer identification via
modeling of peripherally circulating miRNAs. PeerJ 2018;2018:e4551. https://doi.org/10.7717/peerj.4551.
[49] Murtaza G, Shuib L, Mujtaba G, Raza G. Breast cancer multi-classification through
deep neural network and hierarchical classification approach. Multimed Tools Appl
2020;79:15481–511. https://doi.org/10.1007/s11042-019-7525-4.

[50] Turkki R, Byckhov D, Lundin M, Isola J, Nordling S, Kovanen PE, et al. Breast
cancer outcome prediction with tumour tissue images and machine learning. Breast
Cancer Res Treat 2019;177:41–52. https://doi.org/10.1007/s10549-019-05281-1.
[51] Bardou D, Zhang K, Ahmad SM. Classification of breast cancer based on histology
images using convolutional neural networks. IEEE Access 2018;6:24680–93.
https://doi.org/10.1109/ACCESS.2018.2831280.
[52] Saba T, Khan SU, Islam N, Abbas N, Rehman A, Javaid N, et al. Cloud-based
decision support system for the detection and classification of malignant cells in
breast cancer using breast cytology images. Microsc Res Tech 2019;82:775–85.
https://doi.org/10.1002/jemt.23222.
[53] Hajiabadi H, Babaiyan V, Zabihzadeh D, Hajiabadi M. Combination of loss
functions for robust breast cancer prediction. Comput Electr Eng 2020;84:106624.
https://doi.org/10.1016/j.compeleceng.2020.106624.

[54] Agrawal U, Soria D, Wagner C, Garibaldi J, Ellis IO, Bartlett JMS, et al. Combining
clustering and classification ensembles: a novel pipeline to identify breast cancer
profiles. Artif Intell Med 2019;97:27–37. https://doi.org/10.1016/j.artmed.2019.05.002.
[55] Smolander J, Dehmer M, Emmert-Streib F. Comparing deep belief networks with
support vector machines for classifying gene expression data from complex
disorders. FEBS Open Bio 2019;9:1232–48. https://doi.org/10.1002/2211-5463.12652.
[56] Zemouri R, Omri N, Morello B, Devalland C, Arnould L, Zerhouni N, et al.
Constructive deep neural network for breast cancer diagnosis. IFAC-PapersOnLine
2018;51:98–103. https://doi.org/10.1016/j.ifacol.2018.11.660.

[57] Denkiewicz M, Saha I, Rakshit S, Sarkar JP, Plewczynski D. Identification of breast
cancer subtype specific micrornas using survival analysis to find their role in transcriptomic regulation. Front Genet 2019;10. https://doi.org/10.3389/fgene.2019.01047.
[58] Guo Y, Shang X, Li Z. Identification of cancer subtypes by integrating multiple
types of transcriptomics data with deep learning in breast cancer. Neurocomputing
2019;324:20–30. https://doi.org/10.1016/j.neucom.2018.03.072.
[59] Tellez D, Balkenhol M, Otte-H¨oller I, Van De Loo R, Vogels R, Bult P, et al. Wholeslide
mitosis detection in h&e breast histology using phh3 as a reference to train
distilled stain-invariant convolutional networks. IEEE Trans Med Imaging 2018;37:
2126–36. https://doi.org/10.1109/TMI.2018.2820199.
[60] Umme Salma M, Doreswamy D. Hybrid BATGSA: a metaheuristic model for
classification of breast cancer data. Int J Adv Intell Paradig 2020;15:207–27. https://doi.org/10.1504/IJAIP.2020.105144.
[61] Luo J, Ning Z, Zhang S, Feng Q, Zhang Y. Bag of deep features for preoperative
prediction of sentinel lymph node metastasis in breast cancer. Phys Med Biol 2018;
63. https://doi.org/10.1088/1361-6560/aaf241.
[62] Rawat RR, Ortega I, Roy P, Sha F, Shibata D, Ruderman D, et al. Deep learned
tissue “fingerprints” classify breast cancers by ER/PR/Her2 status from H&E images. Sci Rep 2020;10:1–13. https://doi.org/10.1038/s41598-020-64156-4.
[63] Elbashir MK, Ezz M, Mohammed M, Saloum SS. Lightweight convolutional neural
network for breast cancer classification using RNA-Seq gene expression data. IEEE
Access 2019;7:185338–48. https://doi.org/10.1109/ACCESS.2019.2960722.
[64] Varma C, Sawant O. In: An Alternative Approach to Detect Breast Cancer Using
Digital Image Processing Techniques. Proc. 2018 IEEE Int. Conf. Commun. Signal

Process. ICCSP 2018, Institute of Electrical and Electronics Engineers Inc; 2018.
p. 134–7. https://doi.org/10.1109/ICCSP.2018.8524576.
[65] Lim J, Bang S, Kim J, Park C, Cho J, Kim S. Integrative deep learning for identifying
differentially expressed (DE) biomarkers. Comput Math Methods Med 2019;2019.
https://doi.org/10.1155/2019/8418760.
[66] Kumar A, Singh SK, Saxena S, Lakshmanan K, Sangaiah AK, Chauhan H, et al. Deep
feature learning for histopathological image classification of canine mammary
tumors and human breast cancer. Inf Sci (Ny) 2020;508:405–21. https://doi.org/
10.1016/j.ins.2019.08.072.
[67] Jin L, Huang Z, Chen L, Liu M, Li Y, Chou Y, et al. Modified single-output
Chebyshev-polynomial feedforward neural network aided with subset method for
classification of breast cancer. Neurocomputing 2019;350:128–35. https://doi.
org/10.1016/j.neucom.2019.03.046.
[68] Sur C. GSIAR: gene-subcategory interaction-based improved deep representation
learning for breast cancer subcategorical analysis using gene expression, applicable
for precision medicine. Med Biol Eng Comput 2019;57:2483–515. https://doi.org/
10.1007/s11517-019-02038-2.
[69] Saha M, Arun I, Ahmed R, Chatterjee S, Chakraborty C. HscoreNet: a deep network
for estrogen and progesterone scoring using breast IHC images. Pattern Recognit
2020;102:107200. https://doi.org/10.1016/j.patcog.2020.107200.
[70] Jain S, Kumar P. An efficient neural network based approach for the detection of
breast cancer. Int J Adv Sci Technol 2020;29:1605–13.
[71] Mostavi M, Chiu YC, Huang Y, Chen Y. Convolutional neural network models for
cancer type prediction based on gene expression. BMC Med Genomics 2020;13:44.

https://doi.org/10.1186/s12920-020-0677-2.
[72] Lu Z, Xu S, Shao W, Wu Y, Zhang J, Han Z, et al. Deep-learning–based characterization of tumor-infiltrating lymphocytes in breast cancers from histopathology images and multiomics data. JCO Clin Cancer Informatics 2020: 480–90. https://doi.org/10.1200/cci.19.00126.
[73] Ha R, Mutasa S, Karcich J, Gupta N, Pascual Van Sant E, Nemer J, et al. Predicting breast cancer molecular subtype with MRI dataset utilizing convolutional neural network algorithm. J Digit Imaging 2019;32:276–82. https://doi.org/10.1007/s10278-019-00179-2.
[74] Ha R, Chin C, Karcich J, Liu MZ, Chang P, Mutasa S, et al. Prior to initiation of chemotherapy, can we predict breast tumor response? Deep learning convolutional neural networks approach using a breast MRI tumor dataset. J Digit Imaging 2019; 32:693–701. https://doi.org/10.1007/s10278-018-0144-1.
[75] 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 2020;8:96946–54. https://doi.org/10.1109/ACCESS.2020.2993536.
[76] Tewary S, Arun I, Ahmed R, Chatterjee S, Chakraborty C. AutoIHC-scoring: a machine learning framework for automated Allred scoring of molecular expression in ER- and PR-stained breast cancer tissue. J Microsc 2017;268:172–85. https://doi.org/10.1111/jmi.12596.
[77] Chen Y-C, Zhang Z, Yoon E. Early prediction of single-cell derived sphere formation rate using convolutional neural network image analysis. Anal Chem 2020;92: 7717–24. https://doi.org/10.1021/acs.analchem.0c00710.
[78] Priya G, Radhika A. An effective machine learning approach for disease predictive modelling in medical application. Int J Sci Technol Res 2020;9:2819–24.
[79] Fang Y, Zhao J, Hu L, Ying X, Pan Y, Wang X. Image classification toward breast

cancer using deeply-learned quality features. J Vis Commun Image Represent 2019;64:102609. https://doi.org/10.1016/j.jvcir.2019.102609.
[80] Ronoud S, Asadi S. An evolutionary deep belief network extreme learning-based for
breast cancer diagnosis. Soft Comput 2019;23:13139–59. https://doi.org/10.1007/s00500-019-03856-0.
[81] Machap L, Abdullah A. Functional analysis of cancer gene subtype from coclustering
and classification. Indones J Electr Eng Comput Sci 2020;18:343. https://doi.org/10.11591/ijeecs.v18.i1.pp343-350.
[82] Xiao Y, Xu C, Ping Y, Guan J, Fan H, Li Y, et al. Differential expression patternbased
prioritization of candidate genes through integrating disease-specific expression data. Genomics 2011;98:64–71. https://doi.org/10.1016/j.ygeno.2011.04.001.
[83] Lin C-Y, Ruan P, Li R, Yang J-M, See S, Song J, et al. Deep learning with evolutionary and genomic profiles for identifying cancer subtypes. J Bioinform Comput Biol 2019;17:1940005. https://doi.org/10.1142/S0219720019400055.
[84] Luo P, Ding Y, Lei X, Wu F-X. deepDriver: predicting cancer driver genes based on
somatic mutations using deep convolutional neural networks. Front Genet 2019;
10:13. https://doi.org/10.3389/fgene.2019.00013.
[85] Chen R, Yang L, Goodison S, Sun Y. Deep-learning approach to identifying cancer
subtypes using high-dimensional genomic data. Bioinformatics 2020;36:1476–83. https://doi.org/10.1093/bioinformatics/btz769.
[86] Tomczak K, Czerwi´nska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an
immeasurable source of knowledge. Contemp Oncol (Poznan, Poland) 2015;19:
A68–77. https://doi.org/10.5114/wo.2014.47136.
[87] Rahman A, Muniyandi R. An enhancement in cancer classification accuracy using a
two-step feature selection method based on artificial neural networks with 15 neurons. Symmetry (Basel) 2020;12:271. https://doi.org/10.3390/sym12020271.

[88] Ha R, Chang P, Mutasa S, Karcich J, Goodman S, Blum E, et al. Convolutional
neural network using a breast MRI tumor dataset can predict oncotype dx
recurrence score. J Magn Reson Imaging 2019;49:518–24. https://doi.org/
10.1002/jmri.26244.
[89] Mullooly M, Ehteshami Bejnordi B, Pfeiffer RM, Fan S, Palakal M, Hada M, et al.
Application of convolutional neural networks to breast biopsies to delineate tissue
correlates of mammographic breast density. Npj Breast Cancer 2019;5:43. https://
doi.org/10.1038/s41523-019-0134-6.
[90] Chougrad H, Zouaki H, Alheyane O. Multi-label transfer learning for the early
diagnosis of breast cancer. Neurocomputing 2020;392:168–80. https://doi.org/
10.1016/j.neucom.2019.01.112.
[91] Majeed H, Nguyen TH, Kandel ME, Kajdacsy-Balla A, Popescu G. Label-free
quantitative evaluation of breast tissue using spatial light interference microscopy
(SLIM). Sci Rep 2018;8:6875. https://doi.org/10.1038/s41598-018-25261-7.
[92] Hamidinekoo A, Denton E, Rampun A, Honnor K, Zwiggelaar R. Deep learning in
mammography and breast histology, an overview and future trends. Med Image
Anal 2018;47:45–67. https://doi.org/10.1016/j.media.2018.03.006.
[93] Li H, Giger ML, Huynh BQ, Antropova NO. Deep learning in breast cancer risk
assessment: evaluation of convolutional neural networks on a clinical dataset of
full-field digital mammograms. J Med Imaging (Bellingham, Wash) 2017;4:41304.
https://doi.org/10.1117/1.JMI.4.4.041304.
[94] Rauschert S, Raubenheimer K, Melton PE, Huang RC. Machine learning and clinical
epigenetics: a review of challenges for diagnosis and classification. Clin Epigenetics
2020;12:51. https://doi.org/10.1186/s13148-020-00842-4.