DOI : https://doi.org/10.56452/7-2-546

ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

Histopathological image classification for breast cancer using deep transfer learning techniques

Kamred Udham Singh
Asst. Professor, School of Computing, Graphic Era Hill University,
Dehradun, Uttarakhand India 248002,

Abstract:

Among female cancers, breast cancer has the highest incidence rate. The survival rate may be improved by early diagnosis of breast cancer. In order to reduce the burden on pathologists and enhance the quality of interpretation, automatic image analysis approaches are required. The first step in cell identification is finding their nuclei. Histopathological imaging may also reveal single-cell division and the metastasis of cancer from one organ to another. Detecting breast cancer from histopathology pictures is reviewed here..

Keywords: Histopathology images, Cell nuclei, Metastasis, Mitosis, Convolutional Neural Network (CNN)

Introduction

Malignant tumours, which may cause cancer, are the result of aberrant cell proliferation that has the ability to infiltrate or spread to other regions of the body. Malignant tumours may spread to other parts of the body, whereas benign ones don't. Breast cancer begins in the breast cells themselves. While both men and women are at risk for developing breast cancer, women are more typically diagnosed. After skin cancer, breast cancer is the most often diagnosed malignancy among women. It's possible for cancer to spread from almost any source. Metastatic tumours are those that have spread beyond the site where the initial tumour originated. By undergoing mitosis, a single cell may produce two daughter cells. Mitosis is the process by which a malignant cell duplicates itself and gives rise to two identical, cancerous daughter cells. Cancer metastasizes when it spreads to other organs or parts of the body. Metastasis is the leading cause of mortality from cancer. Long-term survival rates may be increased with early cancer identification. The use of medical imaging is crucial for the early diagnosis of cancer. Manually interpreting a large volume of medical photographs may be a time-consuming and error-prone process. Computer-aided diagnostic (CAD) systems were therefore developed to aid medical professionals in their interpretation of medical pictures. During a biopsy, a small bit of tissue is removed for further analysis under a microscope. The tissue sample is subsequently sent to the laboratory for microscopic examination and analysis by a pathologist. Histopathology refers to the microscopic analysis and research of living organisms and their tissues. Histopathology derives from the Greek words for "tissues" (histo), "disease" (patho), and "study" (logo). Histopathology, then, refers to the study of tissues with the aim of diagnosing illness. In histopathology, hematoxylin and eosin are the most often utilised stains. When hematoxylin is in contact with DNA, the nuclei take on a purple or blue hue. Eosin binds to proteins, giving those

structures a pink hue. The nuclei of the cells may be located with the use of these various dyes. Cancer is an umbrella term for a group of illnesses characterised by the presence of aberrant cells with the ability to proliferate uncontrolled and destroy normal human tissue.Cancer may metastasize to distant organs and tissues [1]. Cancer comes in many forms, the most common of which being skin cancer, breast cancer, and brain tumours. For instance, magnetic resonance imaging (MRI) is used to identify brain tumours, whereas dermatologist-level categorization using Deep Learning is utilised to diagnose skin cancer. As the second largest cause of cancer mortality in women, breast cancer is the most frequent invasive cancer in females.Breast cancer begins in the lining of the milk ducts and develops into a lump or tumour over time [2]. Clinicians have long believed that it takes a very long time for a tumour to develop from a single cell to a size of 1 centimetre. A malignant tumour in the breast has the potential to metastasize, or spread to other regions of the body. Both aggressive and noninvasive forms of breast cancer have been identified. Cancer that has metastasized from the milk duct or lobule to other breast tissues is considered invasive. Cancers that aren't invasive can't spread to other parts of the breast. Cancer that originates in a milk duct and has spread to the surrounding fatty and fibrous tissue is known as invasive ductal carcinoma (IDC) or infiltrating ductal carcinoma. Most cases of breast cancer are invasive ductal carcinoma. It accounts for 80% of all cases of breast cancer.[2] Detecting invasive breast cancer may be difficult and time consuming. The pathologist looks at the tissue under a microscope. In order to locate cancerous tissue, the pathologist must visually examine extensive areas that are cancer-free. In order to save the most lives possible, it is crucial to create a system that aids in the early identification of cancer. CEO and Founder of Enlitic Jeremy Howard said in a TEDx 2014 address that early cancer detection increases a patient's chance of survival by a factor of 10. In order to classify the elements of an input picture, a Convolutional Neural Network (ConvNet/CNN) uses learnable weights and biases to give different parts of the image more or lesser priority.It may be used to determine [4] what category a picture most closely fits within. A CNN's structure is like the interconnections between neurons in the human brain. The convolutional neural network (CNN) is responsible for simplifying the processing of pictures without sacrificing information crucial to making an accurate forecast. Cancer in humans is a complex disease with several causes and symptoms, including genetic instability and the accumulation of many different types of molecules. The diagnostic classifications currently in use are inadequate for predicting treatment success and patient outcomes due to their inability to capture the whole clinical heterogeneity of malignancies. Most standard anti-cancer treatments do not differentiate between cancerous and healthy tissue. To add insult to injury, cancer is often diagnosed and treated too late. The cancer cells have metastasized and are now present in all organs. At the time of clinical manifestation, the vast majority of people with breast, lung, colon, prostate, and ovarian cancers already have over and concealed metastatic colonies. There is a present limitation on the efficacy of therapeutic methods. In our proposed effort, we will use deep learning algorithms to identify breast cancer in mammograms.

Histopathology photos showing staining for benign and malignant tumours are shown in Fig. 1. While a benign tumour may not pose an immediate danger to life, it might cause problems for important organs or tissues. Most benign tumours are composed of clusters of cells that are very similar to the tissue's regular cells. When compared to the healthy cells around it, the cells in a malignant tumour seem and act normally. Malignant tumours are cancerous growths that behave deliberately, invade neighbouring tissues, and ultimately cause death if left untreated.

DOI : https://doi.org/10.56452/7-2-546

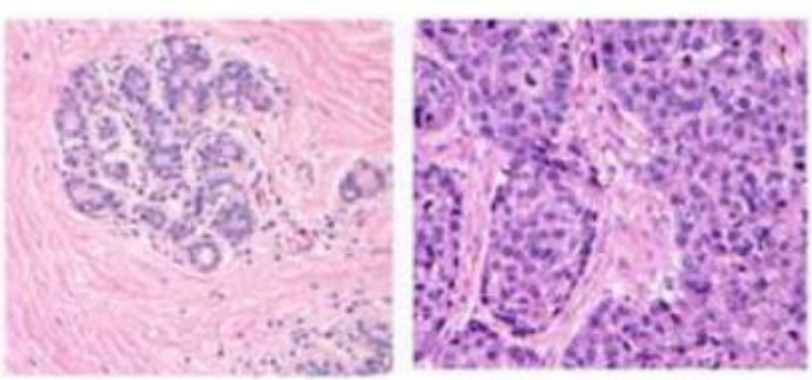


Fig-1: Hematoxylin and Eosin Images of breast cancer (a): Benign (b): Malignant

Literature Review

Alice M.L. Santilli et.al., (2021) Removing tumours from the breast while leaving good tissue intact is the goal of breast conserving surgery, a common kind of cancer therapy. Reoperation rates might reach 35% owing to the challenges of detecting malignancy in the surgical margins. By measuring the tissue's molecular signature in real time, REIMS is a mass spectrometry technique that may help with this problem. However, training a cancer detection algorithm from scratch requires a huge sample size, which is impractical given the time and effort required to collect breast spectra. To address the challenge of improving cancer diagnosis near surgical margins with a small sample of labelled data, we suggest using self-supervised learning. To accomplish the intermediate goal—capturing latent properties of REIMS data without using cancer labels—a deep model is trained. The model makes up for the limited amount of information by randomly rearranging the spectra into novel configurations. The model learns the order of the data's patches via inquiry, therefore capturing the data's features. After the model has been trained, the weights are passed to a second network, which is then optimised for cancer detection. Using information from 144 cancer and normal REIMS samples, the suggested technique obtained 97% accuracy, 91% sensitivity, and 100% specificity.

Uswatun Khasanaet.al.,(2020) The development of malignant cells in the breast tissue is known as breast cancer. Breast cancer may also develop in adipose or connective tissue. Is a lethal kind of cancer that has historically been difficult to identify in its early stages, but because to advances in medical technology, there are now several tools at doctors' disposal for doing so, including USG

(Ultrasonography). Ultrasound is a method of producing pictures of internal body structures including organs and soft tissues utilising high-frequency sound wave technology. Due to the poor quality of the ultrasound test findings, doctors often disagree over the diagnosis. The concept for developing automated breast cancer detection sprang from these issues. The segmentation procedure uses the watershed transform method, which is capable of distinguishing objects depending on their backgrounds and thereby producing the location of the malignancy. When employing watershed for segmentation, the next step is to use thresholding binaries to isolate the cancer in the picture. Area of cancer calculation is the last stage. This research compares hospital data and trial findings for calculating cancer area, finding that the approach is accurate to within 11.35 percentage points across 88.65 percent of all data evaluated. These results suggest that the watershed transform technique is suitable for segmenting the breast ultrasound picture and so should be employed..

Jalaluddin Khan et.al.,(2020) Worldwide, women between the ages of 40 and 65 have breast cancer diagnosed more often than any other age group. It has remained difficult for doctors and radiologists to determine whether a breast cancer tumour is benign or malignant in its early stages. When a tumour is quickly diagnosed and its nature is established, doctors may begin therapy in the proper way and with the right drugs. Therefore, precisely diagnosing the kind of tumour is crucial for saving the patient's life. Using fine needle aspirate (FNA) of a breast mass and optimised fuzzy sets, this research provides an intelligent diagnostic technique for breast cancer tumour type identification and classification. The created technology analysed a digitised picture of a FNA using image processing and feature selection methods. Nine calculated characteristics from a digital picture of a fine-needle aspiration of a breast mass are used as the first input of a neuro-fuzzy system in the created approach. The Association Rules (AR) were used in the suggested technique to determine which of the retrieved characteristics were the most useful and informative. In addition, the Chaotic Bat Optimisation method (CBOA) is implemented as a learning method to improve neuro-fuzzy system performance. Breast Cancer Wisconsin (Diagnostic) data is used to evaluate the effectiveness of the suggested technique. Based on the results of the obtained numerical analysis, the created diagnostic system has high precision and great performance..

Types of Breast Tumors Breast cancer may be classified according to its stage of development and the kind of breast tissue from which it originated (gland, duct, fat, or connective tissue). Early-stage carcinoma (invasive malignant tumour formed by muted epithelial cells) that has not yet invaded surrounding tissue is called a carcinoma insitu tumour. Cancer is diagnosed when malignant cells from glands or ducts invade and destroy neighbouring healthy tissue. The symptoms of this kind of cancer vary from patient to patient (Eastman & Crosin, 2006). Depending on the tumor's location, in situ and infiltrating breast cancers may have ductal and lobular features. Ductal Insitu Carcinoma (DCIS) is a non-invasive form of breast cancer that develops from abnormal cells in the duct lining. The abnormal cells inside the tubes did not spread to the rest of the breast. However, ductal in carcinoma in situ may spread to nearby breast tissue if not treated or diagnosed in time. Figure 1: The medical term "carcinoma in situ," which signifies cancer that has not spread from its original site (National basis for breast cancer)

DOI : https://doi.org/10.56452/7-2-546

Figure 1: Invasive Ductal Carcinoma Showing Microlobulated Borders and Micro Calcifications

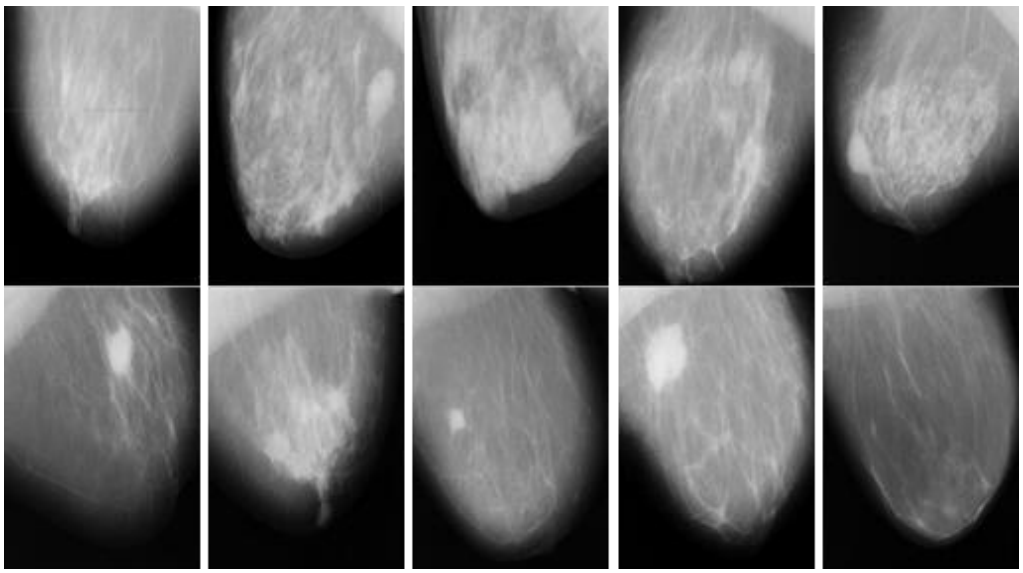
Feature Selection And Extraction

Even though there are many potential models for the feature reduction and selection process, only a subset of them will be practical for an actual implementation framework. Therefore, it is important to investigate the potential of the mammography dataset for attribute selection techniques. For each given collection of characteristics in the data, the goal of feature selection approaches is to identify the subset of features that best facilitates categorization. The filter is made by keeping just the most crucial filtering features and discarding the rest. This reduces the amount of feature vector features.

This study introduces a performance-based approach to feature extraction and selection for a variety of classification algorithms. The Mini-MIAS database is mined for mammograms for the trial..

SAMPLE IMAGE

The mini-MIAS database, the DDSM database, and the city's hospitals all contributed to the sample pictures. In this case, data is made available via The Cancer Image Archive (TCIA) at the National Cancer Institute (NCI). The DICOM formatted pictures in this collection cover the spectrum from healthy to cancerous. Both the training set and the test set each consisted of 86 different picture pairings. All of the CT images in the database have a resolution of 512 pixels on a side and a bit depth of 12 bits.



DOI : https://doi.org/10.56452/7-2-546

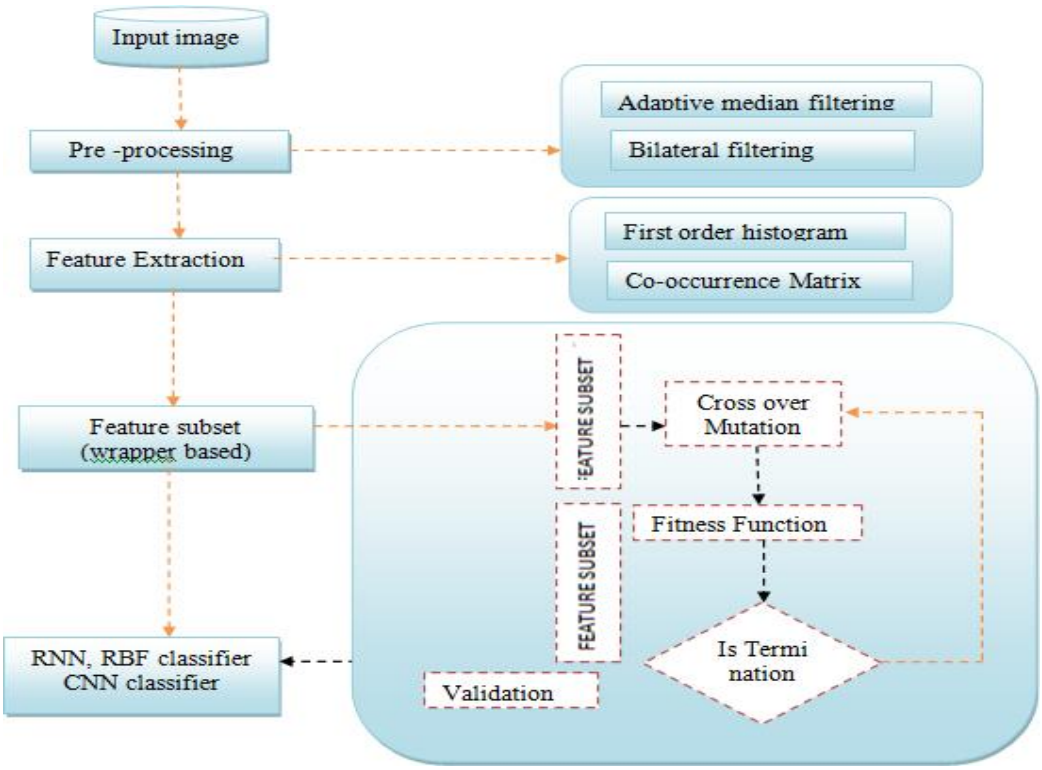


Figure 2: Detailed Block Diagram for the Proposed Method

To save time and memory processing massive amounts of data, feature extraction techniques are often utilised. An image's relevant details are those that have been eliminated. These characteristics, when extracted, are considered to have the image-defining attributes. It's also a key piece of information for addressing the analytical difficulty of a few select applications. The feature is being extracted so that additional features may be calculated using a smaller dataset. Both of these extracted characteristics are provided as input to an image classifier. Intensity-based extraction, shape-based extraction, and texture-based functionality are all types of feature extraction. The research strategy involves the extraction of a texture-dependent element using the GLCM algorithm. The feature extraction model's Block Diagram is shown in detail in Figure 2.

Texture Based Feature Extraction

A lot of work goes into dimensionality reduction from fields including image processing, pattern recognition, and feature extraction. Input data is reduced to a more manageable collection of feature descriptions if its size makes it likely to be repetitious. Extraction of features is the process of transforming raw data into a collection of useful functions. Colour, shape, texture, and even the existence of a backdrop may all be considered features. Transform features, spatial features, colour features, edge and boundary features, shape features, texture features, etc. are some of the most often used techniques for feature extraction.

DOI : https://doi.org/10.56452/7-2-546

The surface qualities and appearance of an artefact are described by its texture, which is determined by the size, arrangement, density, and ratio of its constituent pieces. To collect these characteristics using texture analysis, texture feature extraction is a crucial first step. In order to extract texture characteristics, the suggested approach uses a combination of the first order histogram and the co-occurrence matrix based technique.

Feature Selection

However, not all characteristics contribute economically to the categorization process. The purpose of feature selection is to locate the optimal collection of features. The genetic algorithm is used to choose features for the proposed system. Genetic algorithms are a kind of empirical search model that utilises the natural selection process in humans. A GA uses repeated procedures to generate a new population from an existing population of chromosomes (solution candidates). This is achieved via the use of mutation and crossover in genetics. It's quite similar to how Charles Darwin explained how the best survive via a process of genetic recombination, reproductive evolution, and natural selection. A function often termed fitness function is used to estimate the fitness of a solution candidate. A chromosome's position in the population may be calculated using a numerical value derived from the fitness function. The fitness function is constructed in accordance with the nature of the issue.

GA Based Feature Selection

To improve the classifier's efficiency, a Genetic algorithm is used to shrink the feature space down to a more manageable size. Fitness assessment, chromosomal encoding, selection method, genetic operators, and conditional stops the repetition are all crucial steps in a Genetic algorithm. Genetic algorithms do their searches in a binary space, treating chromosomes as a string of bits. At the outset, a primary population is generated at random and evaluated using the fitness function. When testing, the feature indexed at location '1' in the bit string is compared to the chromosome with the value '1'. The ranking is based on the reliability of tried and true categorization data. According to the ranking, the chromosomes are chosen that have the most advantageous fitness attribute. A new, more likely chromosome can only be created by mutation and crossing on the existing chromosome. Figure 3 depicts this procedure being repeated until a fitness function is attained.

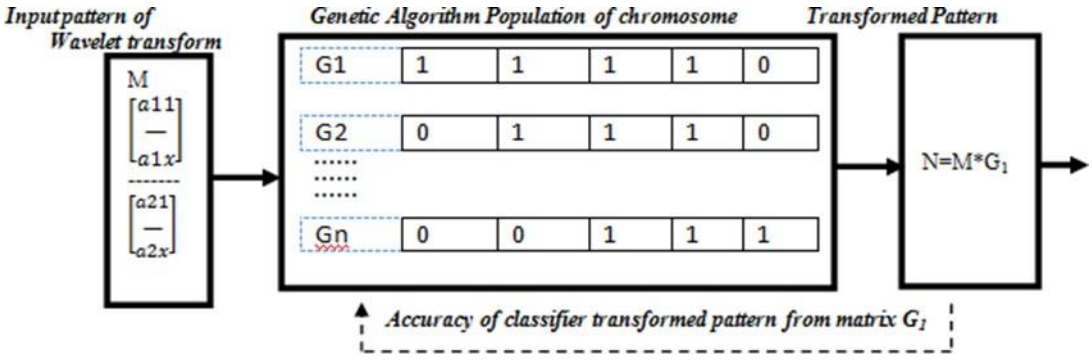


Figure 3: Matrix Representation Indicating Chromosome Bit Value in Genetic Algorithm

Gathering Data for Neural Networks

When employing neural networks to solve a problem, it is crucial to gather training data. Training data, consisting of examples with values for various input and output variables, must be compiled in many cases. Prior choices must be made, such as which variables to use and how many (and what) examples to gather. Even neural networks can only deal with numbers in a limited range. If the data is

DOI : https://doi.org/10.56452/7-2-546

not numeric, falls outside of a typical range, or is incomplete, this process becomes problematic. Scaling numerical data or using the mean or other statistics to fill in missing numbers is one approach to this issue.

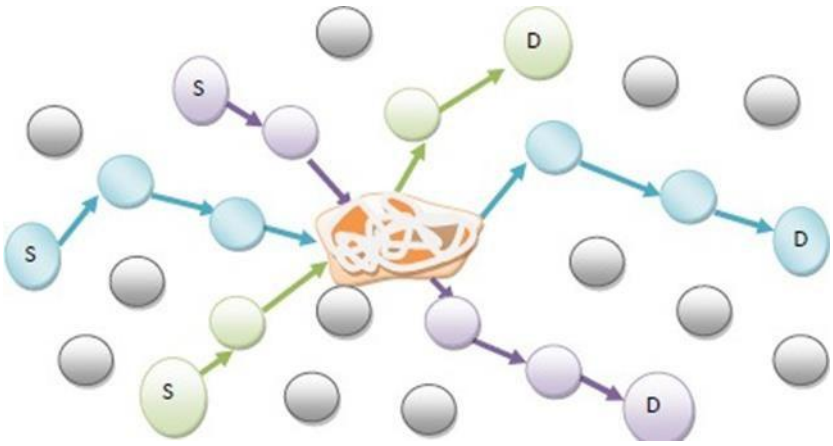


Figure 4: Data Gathering Neural Network

Pre-processing and Post-processing

The neural networks take in data in numerical form and output new data based on the data. The transfer function of a device is often set up to accept input from everywhere but to only generate output within a narrow band. The input range for the logistic transfer function is (-1, + 1), while the output is only between 0 and 1. Due of the limited variation in the statistical data, pre- and post-processing are used in real-world applications. While most individual neural networks specialise in either regression or classification, they may execute many tasks simultaneously. Consequently, the network would typically share an output variable, albeit in the event of other state categorization issues, this would only apply to certain output nodes. Converting variables from their output units into their input units is a job for the post-processing phase. A cross-talk issue might develop when more than one output variable is provided for a given network. Separate networks are trained for each performance and then combined into an ensemble to eliminate this interference.

The problem states the required input and output quantities. The total number of combined units may cause some misunderstanding. However, for the time being, it is believed that the right choice of input variables has been chosen. The amount of elements in a hidden layer is used to determine how many hidden layers to use.

The mini-MIAS database, the DDSM database, and the city's hospitals all contributed to the sample pictures. In this case, data is made available via The Cancer Image Archive (TCIA) at the National Cancer Institute (NCI). The DICOM formatted pictures in this collection cover the spectrum from healthy to cancerous. Both the training set and the test set each consisted of 86 different picture pairings. All of the CT images in the database have a resolution of 512 pixels on a side and a bit depth of 12 bits.

DOI : https://doi.org/10.56452/7-2-546

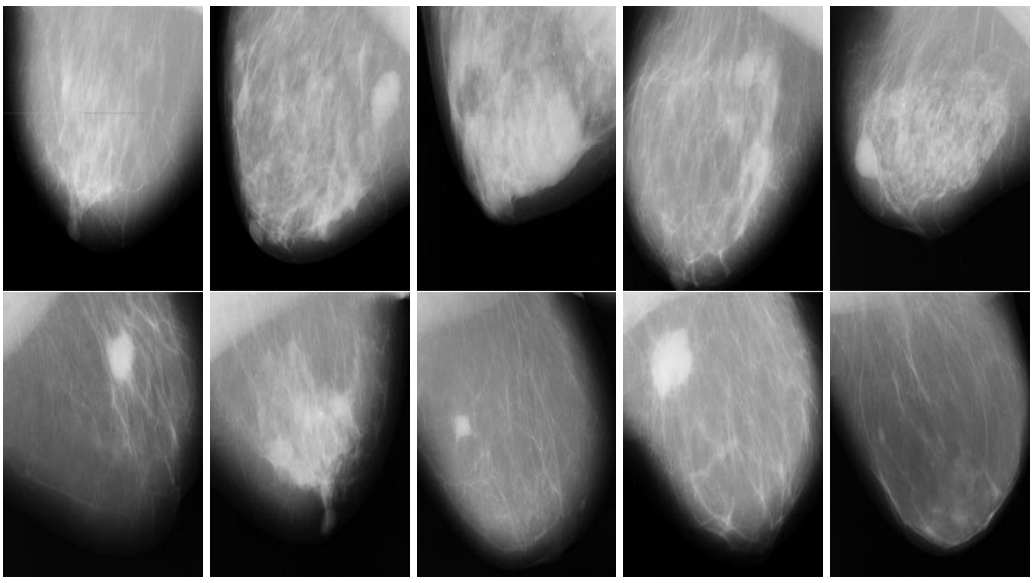


Figure 5. Benign Samples

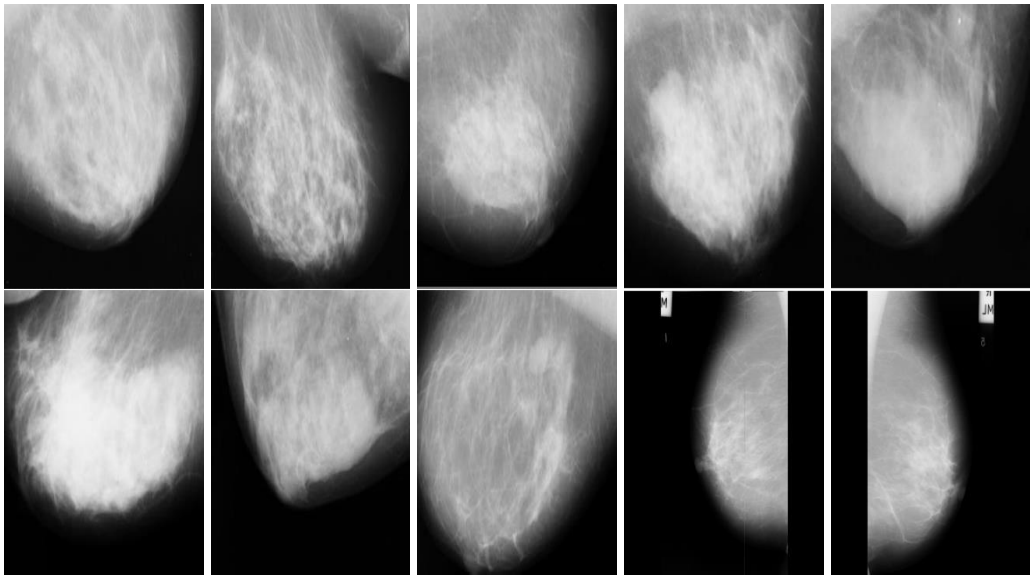


Figure 6. Malignant Samples

Images of benign and malignant breast cancer samples from the public mini-MIAS databases are shown in Figures 5 and 6, respectively. These CT scans of the lungs are used to determine the disease's progression by using a categorization system. In all, 86 samples from each category are collected, however only 5 output examples from each category are shown here. MATLAB, version 2016a, is used to analyse these pictures.In this article, we talk about the findings for the various types of breast tumours. Breast mammography images are analysed using a CNN classifier, which has a 95.8% success rate in detecting and classifying breast tumours based on size. This improves the accuracy of the CAD system's mammography image categorization method. Results acquired by providing the original breast mammography picture retrieved from the database are shown in the following stages. Noise in the sample photos is reduced and the overall image quality is improved

DOI : https://doi.org/10.56452/7-2-546

via the use of preprocessing methods. Next, the photos are processed to isolate the tumour regions. The resulting table lists the retrieved features from the example photos. The categorization method is then used to determine severity levels with improved accuracy.

CLASSIFICATION STAGE

Classification of lung pictures is performed using RNN, RBF, and CNN classifiers after feature extraction. Mammograms are analysed and categorised according to whether or not the tumour is cancerous. The results are graded according to their accuracy, sensitivity, and specificity. Finally, the results of many classifiers are compared, and it is shown that CNN provides the best results.

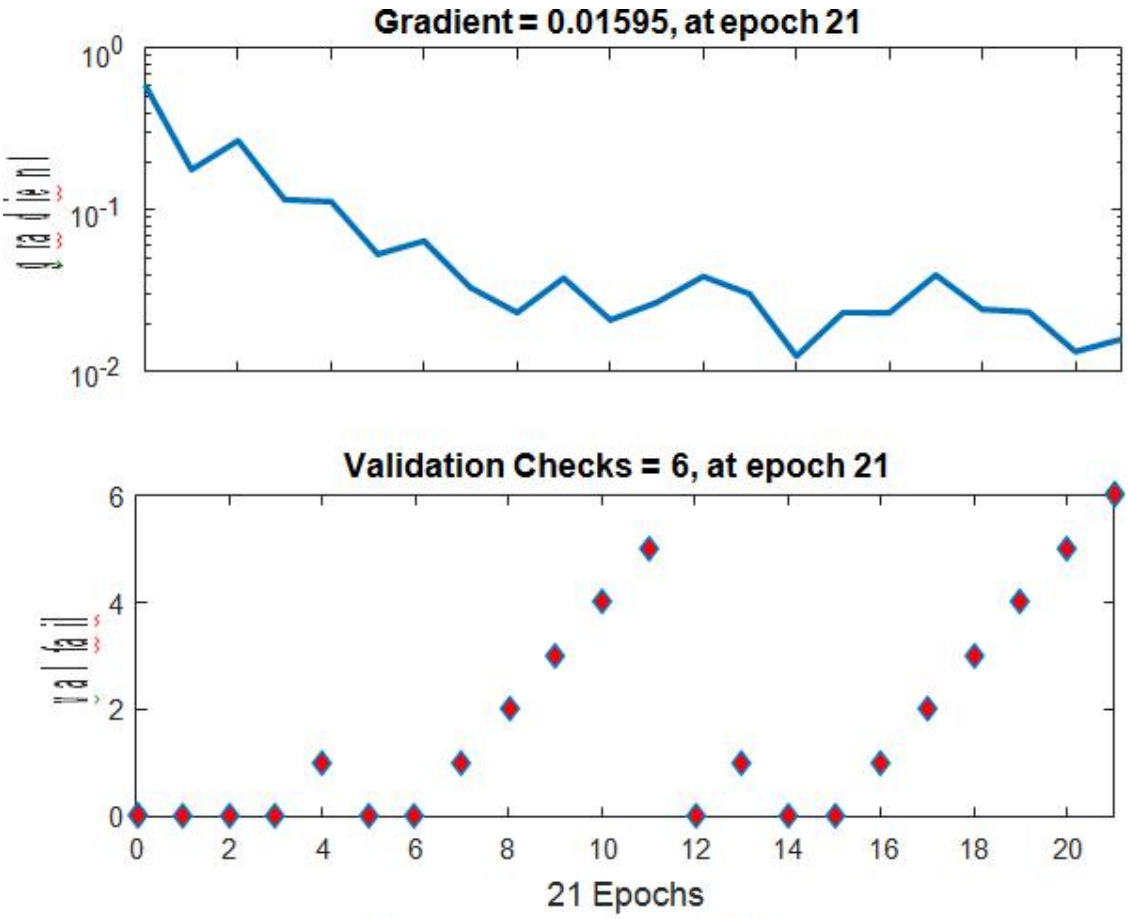


Figure 8. Gradient and Validation Plot

The Figure 8 shows the gradient and validation checks of RNN. In Figure 8 plot at epoch 21 the gradient is 0.01595, the validation check is 5

DOI : https://doi.org/10.56452/7-2-546

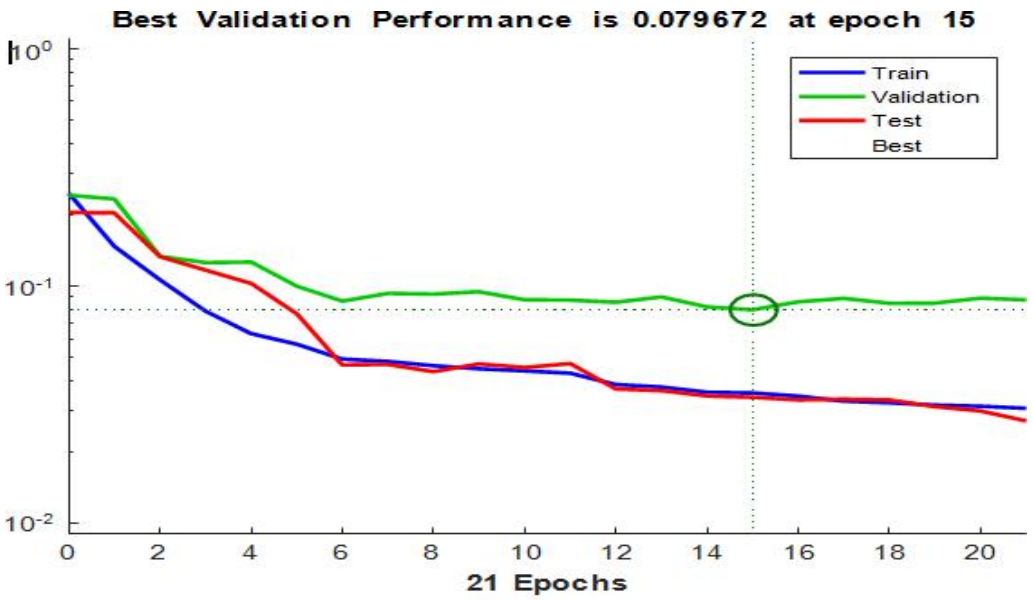


Figure 9: Performance Plot

The Figure 9 shows the performance plot of recurrent neural network. In the lowest validation error the best performance will occur when training stops or increase the validation error**.**

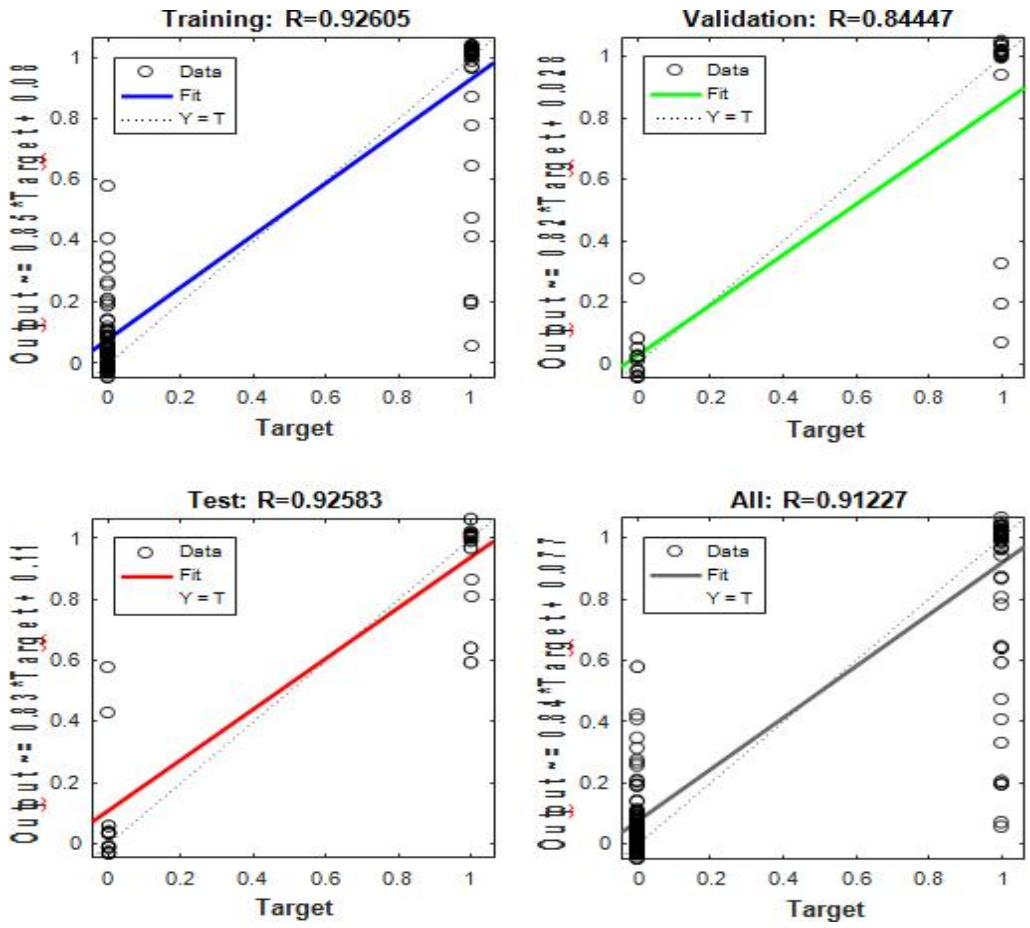


Figure 10. Overall Performance of Neural System

Figure 10 shows the artificial neural network's training plot of the error histogram. In the histogram, we see the occurrence of erroneous values for both the targets and the yield. The blue bar represents the training data, the green bar represents the validation data, and the red bar represents the testing data.

Conclusion

For each amplification factor, we used a variety of criteria for performance assessment, including accuracy, precision, sensitivity (recall), specificity, and F1-score. For both binary and multiclass classification, we also created confusion matrices. Due to the high degree of difficulty in classifying medical pictures, we used a transfer learning approach that involves retraining a network by swapping out its last three layers from three previously trained networks.

References

1. X. Jia, W. Meng, S. Li, Z. Tong and Y. Jia, "A rare case of intracystic Her-2 positive young breast cancer," *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Houston, TX, USA, 2021, pp. 2598-2602, doi: 10.1109/BIBM52615.2021.9669897.
2. M. L. Santilli *et al*., "Self-Supervised Learning For Detection Of Breast Cancer In Surgical Margins With Limited Data," *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, Nice, France, 2021, pp. 980-984, doi: 10.1109/ISBI48211.2021.9433829.
3. B. Bılgıç, "Comparison of Breast Cancer and Skin Cancer Diagnoses Using Deep Learning Method," *2021 29th Signal Processing and Communications Applications Conference (SIU)*, Istanbul, Turkey, 2021, pp. 1-4, doi: 10.1109/SIU53274.2021.9477992.
4. M. Li, "Research on the Detection Method of Breast Cancer Deep Convolutional Neural Network Based on Computer Aid," *2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, Dalian, China, 2021, pp. 536-540, doi: 10.1109/IPEC51340.2021.9421338.
5. V. E. Orel *et al*., "Computer-assisted Inductive Moderate Hyperthermia Planning For Breast Cancer Patients," *2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO)*, Kyiv, Ukraine, 2020, pp. 474-477, doi: 10.1109/ELNANO50318.2020.9088908.
6. J. Khan, N. A. Golilarz, J. P. Li, P. Kuzeli, A. Addeh and A. U. Haq, "Breast Cancer Diagnosis using Digitized Images of FNA Breast Biopsy and Optimized Neurofuzzy System," *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, Chengdu, China, 2020, pp. 286-290, doi: 10.1109/ICCWAMTIP51612.2020.9317387.
7. Easson, A. Pandya, J. Pasternak, N. Mohammed and A. Douplik, "Improving the patient cancer experience: Multispectral (White Light/Autofluorescence/Raman) Needle Endoscopy for cancer diagnostics in breast and thyroid," *2020 Photonics North (PN)*, Niagara Falls, ON, Canada, 2020, pp. 1-2, doi: 10.1109/PN50013.2020.9166986
8. K. Karasawa *et al*., "A Phase I clinical trial of carbon ion radiotherapy for Stage I breast cancer: clinical and pathological evaluation," in *Journal of Radiation Research*, vol. 60, no. 3, pp. 342-347, May 2019, doi: 10.1093/jrr/rry113.

9. X. Feng *et al*., "Accurate Prediction of Neoadjuvant Chemotherapy Pathological Complete Remission (pCR) for the Four Sub-Types of Breast Cancer," in *IEEE Access*, vol. 7, pp. 134697-134706, 2019, doi: 10.1109/ACCESS.2019.2941543.
10. N. Aibe *et al*., "Results of a nationwide survey on Japanese clinical practice in breast-conserving radiotherapy for breast cancer," in *Journal of Radiation Research*, vol. 60, no. 1, pp. 142-149, Jan. 2019, doi: 10.1093/jrr/rry095.
11. H. Song, H. Watanabe, X. Xiao and T. Kikkawa, "Influence of Air-gaps between Antennas and Breast on Impulse-Radar-Based Breast Cancer Detection," *2019 13th European Conference on Antennas and Propagation (EuCAP)*, Krakow, Poland, 2019, pp. 1-2.
12. S. Ghoul, D. M. K. Nashawati, M. A. Hamaly, S. Mutlaq, A. Mansour and A. Nofal, "Streamlining The Interventional Breast Imaging Workflow by Lean Methodology Implementation," *2018 1st International Conference on Cancer Care Informatics (CCI)*, Amman, Jordan, 2018, pp. 41-46, doi: 10.1109/CANCERCARE.2018.8618178.
13. A. A. Hmaidan, E. Boutou, K. Jamal and A. Al-Omari, "Availability and Usability of the Hospital-based Cancer Registry Data for Measuring the Quality Outcome Indicators of Healthcare Provided to Breast and Colorectal Cancer Patients at King Hussein Cancer Center," *2018 1st International Conference on Cancer Care Informatics (CCI)*, Amman, Jordan, 2018, pp. 195-204, doi: 10.1109/CANCERCARE.2018.8618213.
14. K. Park, W. Chen, M. A. Chekmareva, D. J. Foran and J. P. Desai, "Electromechanical Coupling Factor of Breast Tissue as a Biomarker for Breast Cancer," in *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 1, pp. 96-103, Jan. 2018, doi: 10.1109/TBME.2017.2695103.
15. N. Khuriwal and N. Mishra, "Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm," *2018 IEEMA Engineer Infinite Conference (eTechNxT)*, New Delhi, India, 2018, pp. 1-5, doi: 10.1109/ETECHNXT.2018.8385355.