Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 3459 - 3466
Received 25 April 2021; Accepted 08 May 2021.

Classification of Thermographic Images for Breast Cancer Detection Based on Deep Learning

Shilpa R. Lahane[1], Prof. Nekita Chavan[2], Prof. Mangala Madankar[3]
[1]Research Scholar, Computer Science and Engineering Department, G. H. Raisoni College of Engineering, Nagpur, Maharashtra, India.
[2,3]Assistant Professor, Computer Science and Engineering Department, G. H. Raisoni College of Engineering, Nagpur, Maharashtra, India.
Shilpa.lahane.mtechcse@ghrce.raisoni.net[1] , Nekita.chavan@raisoni.net[2], Mangalaa.madankar@raisoni.net[3]

ABSTRACT

As per statistics, around the world, breast cancer is the most common cancer, today it kills around 2.5 million people annually. Early diagnosis of this condition, along with accurate classifications promote more favorable outcomes for the people who already have it and save premature victims from developing it. Difficulties face cancer researchers with the task of differentiating between benign and malignant tumors, as well as making conclusions on mild and advanced breast cancer. In machine learning, using algorithms that are able to find and identify patterns is used for the detection of all cancers. However, as previously discussed, all of them center on "binary grouping" (cancer and no-cancer; benign and malignant). In this research, we propose a Computer-aided Diagnosis (CAD) method for diagnosis and classification of patients (beneath the computer-based DMR) into three distinct groups (cancer, no cancer, and non-cancerous) under database management. Three exceptional classifiers include Convolution Neural Network (CNN), Support Vector Machine (SVM), and Random Forest (RF).For the classification process, we investigate and study three powerful classifiers: Convolution Networks (SVM), and Random Forest (RF). We also review the effects of prior to the pre-processing of the mammogram images, which enables higher successful classification.

Keywords
Breast Cancer; Infrared Thermography; Machine Learning, Classification, Segmentation

Introduction

Cancer is generally considered to be one of the world's most complex and deadly illnesses. It causes a significant amount of fatalities per year. Cancer claimed the lives of 8.8 million people worldwide in 2015, based on the current World Health Organization (WHO) [1]. Harmful neoplasms or cancer are considered an inherited issue because of specific cell modifications induced by gene expression or developmental changes [2].

This genetic change is usually triggered by a cell division defect during the optimization procedure, and it may have been acquired or caused by DNA twisting as a result of the following exposures. DNA repair qualities, tumor suppressor characteristics, and proto-oncogene qualities are the three main categories of cancer drivers that researchers usually report [3]. These features are responsible for proto-oncogenes forcing cells to expand and isolate when they shouldn't (proto-oncogenes), tumour silencers leading cells to go rogue in their division method (tumour silencer), and damaged DNA not being repaired properly (DNA fix). Cancer may spread through one tissues and organs to someone else in the body; in these circumstances, the cancer is referred to as metastatic breast cancer, and it has the same characteristics no matter where everything spreads [4]. The forms of cancer differ based on when they occur and how severe they are. About 100 different forms of cancer exist, each with its own name depending on the organ or tissue in which it manifests [3]. In essence, understanding the trajectory of cancer, the organ by which it is created, and the mechanisms of cancer progression helps in treatment disclosure.

Despite current research and pharmaceutical market advances, the ideal cancer drug has yet to be developed and tested. Guidance of different cell genetic characteristics, specifically those necessary for their structure [5], is perhaps one of the most significant disclosures unique to cancerous forms. In general, this approach is still a long way from being a successful treatment that is accessible to individuals from all socioeconomic backgrounds [6]. Cancer affects men and women of all ages, both young and old, although it is more frequent in senior citizens than in their childhood spouses [7]. There are many reasons for the recurrent forms of cancers, with geography and sexual identity also playing a significant part in their manifestation. Female breast cancer is amongst the most widespread cancer types [8]. As a result, the exploration of both discovery and recovery mechanisms for it has been sparked. For even the most part, breast cancer causes aren't well understood. Until now, researchers and specialists are still unable to provide definitive reasons on whether some women are more likely to develop breast cancer than others [9]. There are clear realities and symptoms that suggest the presence of breast cancer in

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 3459 - 3466
Received 25 April 2021; Accepted 08 May 2021.

any case. Armpits, a difference in the colour of the breast skin (redness or orange shading), an abnormal reorganize and assorted variety of the areola, which is followed by releases that can contain blood, as well as a substantial improvement in the size or presentation of the breast are a few of the most common side effects of pancreatic cancer [10].

Any single symptom may or may not suggest the existence of breast cancer in general. In this case, the discovery of either of the side effects should greatly persuade the woman to continue with the conventional breast cancer screening method. In the early stages of improvement, much, if not any, of these metrics do not exist. As a result, it would be difficult to diagnose cancer in the early stages.

One of the most often used methods in medicine for treating cancer is screening the patient for cancer before the onset of serious symptoms. The screening method is built on a recognized technology for diagnosing people and early cancer detection. Early detection of a tumor, as in other types of cancer, aids healing and decreases the amount of weight that patients must bear. In standard breast cancer screening, computerized tomography (CT), magnetic resonance imaging (MRI), mammography, thermography, and ultrasound are typically used. These techniques often have their own set of tools and methodologies; the standard results of these procedures vary depending on the materials used, and it is advised that the results be accepted with more than one approach [11]. Despite the fact that many doctors and clinicians consider mammography to be the safest tool for breast cancer screening, there is an increasing need for a more reliable treatment.

Thermography has become increasingly common in recent years, especially for cervical cancer screening [12]. This is due to the alluring realities of its own generally safe invention, as well as the possibility of future upgrades with cutting-edge mechanical advancement. Present research in this field is aiming to come up with a more definitive and widely agreed tumor result that can be used as a guideline for breast cancer screening. Likewise, overcoming the recently established deterrents of the time-consuming screening process, particularly when picture preparation is required.

The science of thermographic imaging and its uses has been revived thanks to continuous ground-breaking steps forward. Thermography is used to screen for breast cancer, and are among the most well-known applications. Regardless, thermography is yet to be recognized as the optimal method for this application. Furthermore, given the fact that thermography is far from a risk-free procedure, physicians would prefer mammography results to thermography outcomes. As a result, if thermal imaging breast cancer screening is improved to a sufficient degree, it will become a viable replacement option. The fundamental issue to be discussed here is picture preparing the job for doing so. This research recommends using a Convolutional Neural Network (CNN) for thermographic testing and treatment to address the disadvantages mentioned above.

Proposed System

We present an accurate and efficient approach for segmenting thermographic breast images and identifying breast cancer for classification as normal or pathological, i.e. without cancer or with cancer, in this research. We suggested using a convolutional neural network to analyze and classify the segmented thermographic images (CNN).

This approach divides the picture inputs into image pre-expanding RGB and Gray channels, then combines the mechanism of feature extraction using the classification model on the n with an automatic image denoising and classification; the end result is feeding both processes the breast image processing and feature learning network a two-extraction to one nest, which will determine whether the image is benign or malignant.

Following are the steps involves in execution of our proposed system.

1. Input image dataset to the system.
2. Pre-processing is performed to enhance image quality.
3. Several features are extracted from input image dataset from which training file is generated.
4. Generated training file dataset and new test input images are pass to improved CNN classification algorithm.
5. We also implement SVM and Random Forest for analytical comparison.
6. The output of CNN algorithm is cancer detection i.e the input test shows normal, benign or malignant.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 3459 - 3466
Received 25 April 2021; Accepted 08 May 2021.

7. At the end graphical evaluation is perform to check the performance of proposed system.

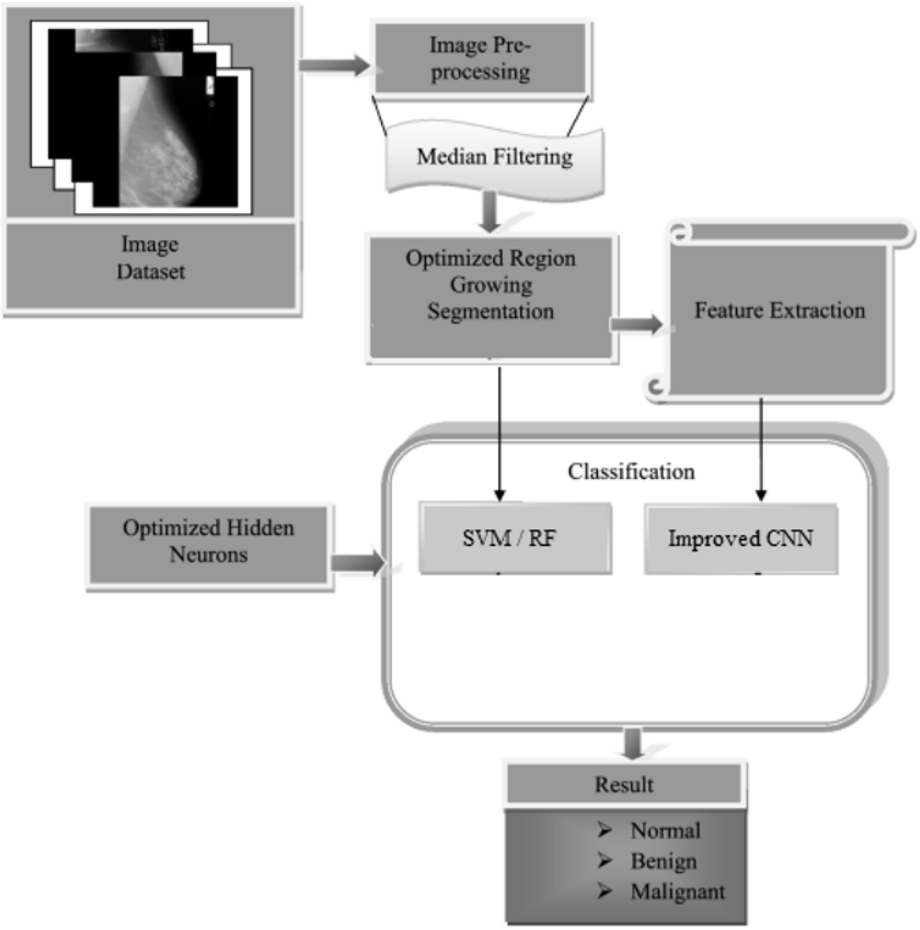

Figure 1 System Architecture

Implementation Details

The methods presented in this paper are meant to analyze infrared photographs of healthy individuals and those with cancer, after which there are various features to extract various classes of the patients. The rest of the photos are placed in a special classification folder, where a small portion of them are used to train classifiers and the other folders'/files are examined to test classifiers. It can now be used to identify any new picture, regardless of the content classifications it was trained with.

A. Dataset

The dataset used includes images of about over 150 patients (around 1000 images) with and without breast cancer. These photos were found on Kaggle. Since the images in the other positions were not uniform, only the frontal images with raised arms were included in this work. The area of interest (ROI), i.e., the region containing only the patient's breast, was then extracted from each image in the database. The files were also resized to 128X128 pixel dimensions.

B. Image Pre-Processing

The image pre-processing stage is critical for producing direct and distinct images. The classifying step is allowed by the image pre-processing step. First, the data augmentation procedure was used.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 3459 - 3466
Received 25 April 2021; Accepted 08 May 2021.

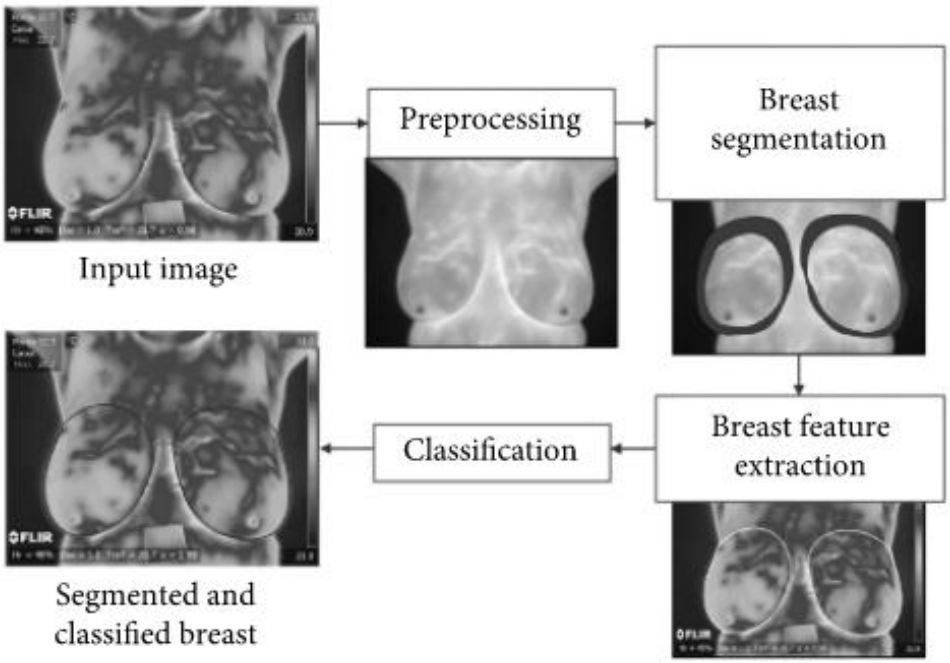

Figure 2 Block Diagram for Breast Cancer Identification

By applying multiple conversions to the original input, this step contributes to increasing the volume of the dataset. The input was repeated using various conversions such as translations, symmetries, and rotations. The steps taken for pre-processing augmentation are described below;
1) Translation – the image can be converted to a certain number of pixels in a specific direction.
2) Centring – the rows and columns on the sides of each image were cut. As a result, photographs of various sizes may be obtained. Later, the whole number of rows and columns is cropped, and the number of pictures is counted. The files are then leveled into a single size before being resized. Following the pre-processing, the randomized images for normal and ill patients are chosen based on transparency.

C. Segmentation

To enable the best results, the objective and segmentation processes had to be completed prior to extraction and classification of high-specific high-resolution files. With spectral issues being more common, delineation taking a hit, resolution was diminished and the image detail in the segmented, higher resolution files. To build on this, the object-oriented image segmentation technique was used, which eliminated the salt and noise while still increasing the precision of the image through the use of object shape and spectral signature.

D. Feature Extraction

The attribute extraction aids in the scaling of the method and the development of more sensible datasets of larger and superior images. There are other attribute extraction approaches that need to process the images first, but the CNN can immediately retrieve the characteristics of the input data. Using convolution, this method of extraction of features allows for the segmentation from different parts of an image. This is so because natural images are usually motionless, implying that the quantification and characteristics of one side of the image are identical to those of the other side.

E. Classification

During the classification process, the instability matrix was used to provide a general explanation of the how the classifier completed the classification operation. Typically, the overview concentrates on the performance of particular individuals. A test series of pre-determined tags that have been considered to be correct were loaded into the matrix. The information was fed into a CNN classifier, which produced predictions.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 3459 - 3466
Received 25 April 2021; Accepted 08 May 2021.

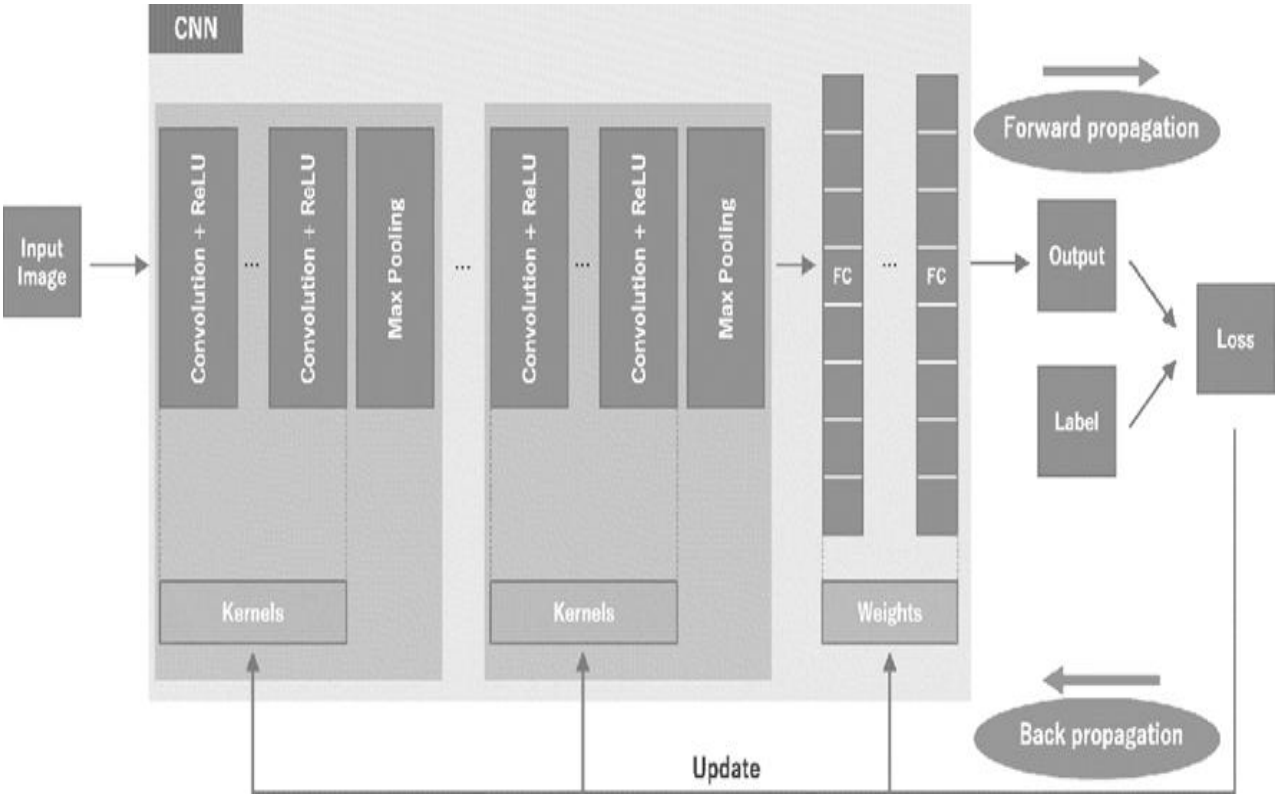

Figure 3 Convolutional Neural Networks Architecture

Mathematical Formulation

System S is represented as
S = {ID, P, F, T, CNN, M}

1. Input Dataset

ID = $\{i_1, i_2, i_3 \ldots i_n\}$
Where ID is the input image dataset and i_1, $i_2 \ldots$in are the number of images.

2. Preprocessing

PR = $\{pr_1, pr_2, pr_3\}$
PR is preprocessing and pr1, pr2 and pr3 are the steps to be carried out during preprocessing.

- pr1 be the reading of input dataset
- pr2 be the enhancement of image input and
- pr3 be the removal of hair from image.

3. Feature Extraction

F= $\{f_1, f_2, f_3 \ldots f_n\}$
Where F is the set of features extracted from the image and f_1, f_2, $f_3 \ldots$ f_n are the extracted features such as border, thickness, color, etc.

4. Training and Testing file generation

T = $\{T_1, T_2\}$

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 3459 - 3466
Received 25 April 2021; Accepted 08 May 2021.

Where T is the set of Training and Testing file and T1 is Training file and T2 is Testing file both the files contains various extracted features values while training file contains class of each image as 0 or 1.

5. Convolutional Neural Network (CNN).

CNN = {C, RL, PO, FC, LS}
Where CNN is algorithm consisting of various stages as
C is convolutional operation
RL be the ReLU activation layer
PO be the Pooling layer
FC be the Full Connection layer and
LS be the Loss function.

6. Cancer Detection

M= {0, 1}
M is the set of Class having value 0 or 1
0 be the absent of Cancer and
1 be the present of Cancer

Result Analysis

Some measurements are collected here during the pre-training of the network in order to determine the training success. Training precision, training error, validation accuracy, and elapse time are the measurements that are measured. The training accuracy shows the classification accuracy for each individual mini-batch, while the validation accuracy shows the overall classification accuracy for the dataset. The loss of each mini-batch as a result of cross-entropy loss is known as training loss. Figures 4 and 5 depict the training outcomes for CNN classification over seven epochs.

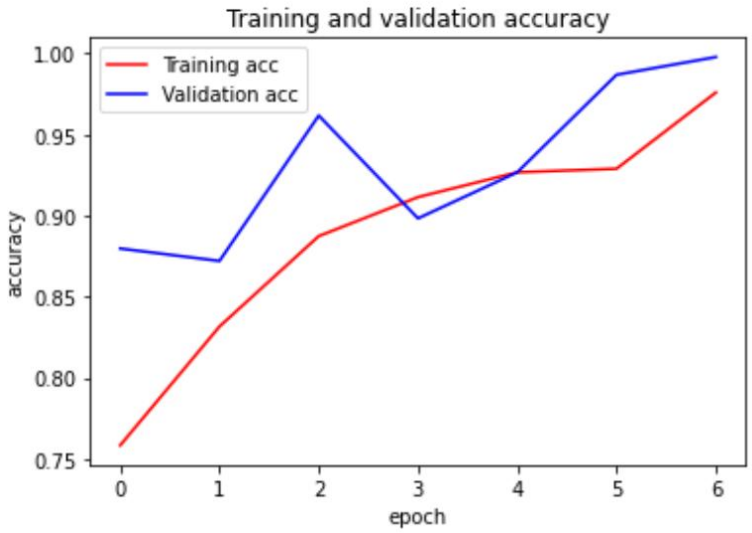

Figure 4 Training and Validation Accuracy

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 3459 - 3466
Received 25 April 2021; Accepted 08 May 2021.

Figure 5 Training and Validation Loss

The overall accuracy of the CNN algorithm for the experiment comes out to be 99.65% with the loss of 0.0067%. The confusion Matrix for the CNN algorithm is shown in figure 6.

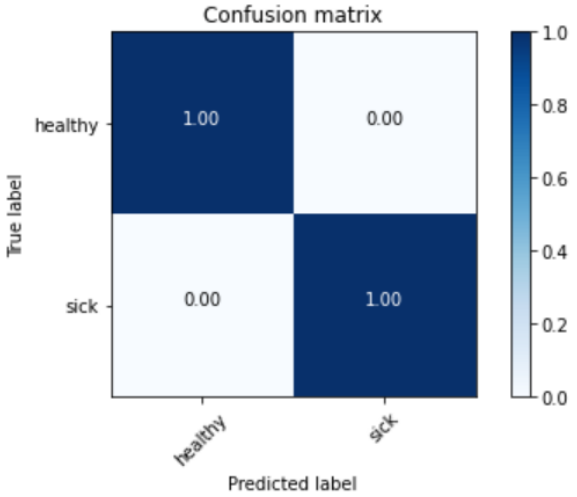

Figure 6 Confusion Matrix for CNN Algorithm

For evaluation purpose, we calculated the accuracy of SVM and Random Forest algorithm on the same dataset which come out to be 89.84% and 90.55% respectively. The comparison of Accuracy for CNN with SVM and RF are presented in figure 7.

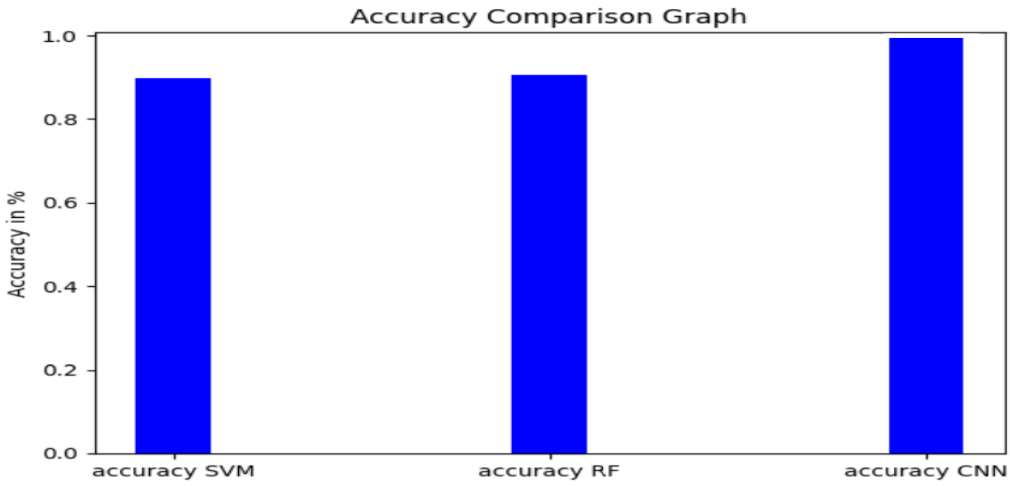

Figure 7 Accuracy Comparison of CNN with SVM and RF

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 3459 - 3466
Received 25 April 2021; Accepted 08 May 2021.

Conclusion

Although work within this project will help the advancement of thermographic imaging of the breast for research purposes, its key goal is to increase the capability of breast imaging through image processing. With the help of Segmentation, this indicates an increased risk of breast cancer before a medical screening does. The subject of this paper posits that there is an improved way of determining areas of initial significance in the patient's chests: comparing cvt k in the right and left. Using the Segment technique, the initial concentrate will start at the breasts and then progress outward, targeting different areas as the breasts mature to deliver specific characteristics of specific segmentation information. As the cases that never produce positive or negative urine and/frequent positive urine under study are categorized as pathological (with cancer).

This work demonstrates that a stable and effective classification technique can be implemented using the use of CNN is provided with the addition of CNN segmentation. For the most part, our main contributions are the novel region of interest (ROI) for thermographic sinus images, where we segmented them; using ground-truth experts' images to generate representative data sets to train CNN on; and providing input data to CNN. To expand on that, future studies will provide a method to scientifically evaluate our findings.

References

[1] Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, Franceschi S, et al. a synthetic analysis. Lancet Glob Heal 2012; 4(2016):e609–16. https://doi.org/10.1016/S2214-109X (16)30143-7.

[2] Chakraborty S, Rahman T. The difficulties in cancer treatment, E cancer medical science. 6 (2012) ed16. doi:10.3332/ecancer.2012.ed16.

[3] National Cancer Institute, What Is Cancer? (2019). https://www.cancer.gov/aboutcancer /understanding/what-is-cancer (accessed September 26, 2019).

[4] Memorial Sloan Kettering Cancer Center, What Is Metastatic Cancer? Answers to Six Common Questions, (2019). https://www.mskcc.org/blog/what-metastatic-canceranswers-six-common-questions (accessed September 26, 2019).

[5] Roake CM, Artandi SE. Control of cellular aging, tissue function, and cancer by p53 downstream of telomeres. Cold Spring Harb Perspect Med 2017; 7: a026088 https://doi.org/10.1101/ cshperspect.a026088.

[6] Nixon NA, Khan OF, Imam H, Tang PA, Monzon J, Li H, et al. Drug development for breast, colorectal, and non-small cell lung cancers from 1979 to 2014. Cancer 1979; 123(2017):4672–9. https://doi.org/10.1002/cncr.30919.

[7] Cancer Research UK, Age and Cancer, (2019). https://www.cancerresearchuk.org/about-cancer/causes-of-cancer/age-and-cancer (accessed September 26, 2019).

[8] American Cancer Society, How Common Is Breast Cancer?, (2019). https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html (accessed September 26, 2019).

[9] National Breast Cancer Foundation, Causes of Breast Cancer - National Breast Cancer Foundation, (2019). https://www.nationalbreastcancer.org/causes-of breast-cancer (accessed September 26, 2019).

[10] Travis RC, Balkwill A, Fensom GK, Appleby PN, Reeves GK, Wang X-S, Roddam AW, Gathani T, Peto R, Green J, Key TJ, Beral V. Night shift work and breast cancer incidence: three prospective studies and meta-analysis of published studies. J Natl Cancer Inst 2016; 108:djw169. https://doi.org/10.1093/jnci/djw169.

[11] National Breast Cancer Foundation, What Types Of Cancer Are Diagnosed As Stage 0 And 1 Breast Cancer?, (2019). https://www.nationalbreastcancer.org/breastcancer-stage-0-and-stage-1 (accessed September 26, 2019).

[12] Kandlikar SG, Perez-Raya I, Raghupathi PA, Gonzalez-Hernandez J-L, Dabydeen D, Medeiros L, et al. Infrared imaging technology for breast cancer detection – current status, protocols and new directions. Int J Heat Mass Transf 2017; 108:2303–20. https://doi.org/10.1016/ J.IJHEATMASSTRANSFER.2017.01.086.