Multimedia Tools and Applications (2022) 81:13139–13157
https://doi.org/10.1007/s11042-020-09600-3

Thermal infrared imaging based breast cancer diagnosis using machine learning techniques

Samir S. Yadav[1] · **Shivajirao M. Jadhav**[2]

Received: 29 January 2020 / Revised: 1 June 2020 / Accepted: 12 August 2020 /
Published online: 25 August 2020

Abstract

The human's temperature is little known and important to the diagnosis of diseases, according to most researchers and health workers.In ancient medicine, doctors may treat patients with wet mud or slurry clay. The part that would dry up first was considered the diseased part when either of these spread over the body. This can be done today with thermal cameras generating pictures with electromagnetic frequencies. Inflammation and blockage areas that predict cancer without radiation or touch may be detected by thermography. It can be used before any visible symptoms occur as a great advantage in medical testing. Machine learning (ML) is used in this paper as statistical techniques to give software programs the capacity to learn from information without being directly coded. ML can help to do so by learning these thermal scans and identifying suspected areas where a doctor needs to research more. Thermal photography is a comparatively better alternative to other methods that need sophisticated equipment, enabling machines to provide an easier and more effective approach to clinics and hospitals.

Keywords Breast cancer · Machine learning · Thermal infrared image · CNN

1 Introduction

Breast cancer is the most commonly diagnosed cancer in woman and its death rates are among the highest of any cancer [35]. The irregular and accelerated development of breast cells (tumor) is responsible for breast cancer. These tumors ultimately develop and form a malignant, lumpy tumor. The World Health Organization (WHO) reported a minimum of 10 lakh breast cancer patients each year, in India. Five lakh patients died of all 10 lakh cases

✉ Samir S. Yadav
ssyadav@dbatu.ac.in

Shivajirao M. Jadhav
smjadhav@dbatu.ac.in

[1] Dr. Babasaheb Ambedkar Technological University, Raigad, Lonere, India

[2] Department of Information Technology, Dr. Babasaheb Ambedkar Technological University, Raigad, Lonere, India

reported [60] and by the year 2025, this figure could be raised to five times with an increase in women of 23 percent and men of 19 percent [24]. It Also detects that breast cancer at the late stage for a significant proportion of women, reducing the rate of survival according to World Health Organisation 2019 [6]. In both advanced and underdeveloped countries, the occurrence of breast cancer is widely studied. Breast cancer survival rates vary widely from 80 percent to an approximate 60 percent in an average income countries and less than 40 percent in the developing and low-income countries worldwide. Within these low income nations, the absence of pre-diagnostic services and appropriate diagnostics and treatment services are attributed to the lower rate of survival. [32]. There is an estimate that more than 508,000 women all over the world died of breast cancer alone in 2011. Although breast cancer is thought to have occurred only in the developed world, it is alarming that around 50% of cases in breast cancer and 58% in deaths come from low-income and developed countries [62]. Early diagnosis of the symptoms reduces treatment costs substantially. However, most of the public will not exercise caution until more severe symptoms begin to show. Individuals refrain from exams for the cost and complexity of their examinations, the fear of suffering and irritation and the long-term nature of most examinations. In contrast, relative to developed cities, rural areas remain mostly oblivious [13]. In an article regarding Indian villagers, the Medibulletin reports that “Villagers do not have access to information, but say that there are 200-250 people with cancer in the village among the 1,700 families” [49]. With such large numbers, it is essential to save countless lives by a faster and less expensive test. The key approaches for early breast cancer screening are generally imaging methods such as mammography, magnetic resonance imaging(MRI) and ultrasound. Mammography is a screening process for the use of low-dose x-rays to classify malignant tumors [22]. It uses x-rays to detect high density regions in the breasts. Women aged 45 years should be screened every one to two years for prevention of breast cancer. Sadly, every ten years in mammograms, the number of misplaced positives is 49.1% [58]. The cost of treatment can be significantly reduced if the tumor is discovered in the early stages [3]. Regardless of the appropriate test incidence of breast cancer, the expense of such tests in poorer countries prevents women from doing regular tests. In comparison, the invasive nature of these examinations often leads to the belief that mammograms are rejected until a doctor prescribes them. Although mammography is a common approach to early breast cancer diagnosis, its main disadvantage is that it may yield a significant number of false-positive results. [65]. Also, Mammography is not recommended for young patients who have dense breasts, as there is difficulty in detecting nodules in this type of breast, which are hidden in the breast tissue. Radiation also is a factor that makes mammography an exam that must be applied with caution. In effect, the MRI technology for women with genetic abnormalities is prescribed in addition to mammography [5]. But the major drawback of MRI is that its spatial resolution is limited, resulting in low sensitivity to the loss of the subcentimeter. Thus it is necessary for the researches and organisations to study alternative methods like thermography for diagnosis. Thermography can solve the weaknesses of these imaging techniques of diagnosis because breast thermography is small in cost, has minimal harmful consequences, free of radiations and is capable of identifying tumors even before they become palpable [30]. With the development of computer technology and statistical machine learning algorithms, computer-aided diagnosis (CAD) has achieved rapid development in the field of medical imaging. Clinical findings and CAD systems are often used for the different diseases diagnosis [19]. Breast cancer screening based on thermal imaging is currently the most widely used field of CAD. The research focus is also mainly on improving the detection accuracy of tumors and classification.

Doctors used slurry clay or muddy mud as a diagnostic tool during early days of medicine [39]. A part that dried up the fastest was considered as the diseased part once the clay was applied. Temperature as a proxy of identifying potential diseases is a lesser-known but instrumental variable. Using current technology i.e., thermography that allows visualizing the rays of the infrared spectrum in order to map the temperature of an object. Infrared breast thermography is a type of examination that detects infrared radiation emitted by the breast surface, producing a temperature map known as a thermogram. These thermogram can be obtained with the aid of a thermal scanner [31]. Such scanners generate images of temperature pattern with a broad infrared range measured in the six electromagnetic spectrums. Certain cancer precursors can be identified much earlier than the standard indications without the need for any contact or radiation, including lymph swelling and inflammatory areas. The use of thermal components can be an excellent way in which possible causes can be identified before harmful symptoms occur. Since it only requires an infrared camera to capture thermal photographs, it costs significantly less than the necessary equipment for a mammogram [48]. Therefore, from last twenty years many researcher started working on thermography for breast cancer screening using machine learning algorithms. Some scientists concentrate their work on the position and size of vanishing tumors and simulation models , while others focus on features such as breast segments, menstruation, and protocols for acquisition.

In recent years, researcher moved to deep learning with regard to breast cancer diagnostic. Deep learning is a subfield of machine learning with algorithms focused on the structure of the human brain. Deep learning can abstract the high-dimensional expression of the image through convolution, down-sampling, etc., so as to effectively classify the images. In any case, the obvious features of the picture that can be distinguished by the naked eye are the key to designing the model. Recently, several deep learning models have been developed and implemented in medical image applications [7, 29, 54, 64]. Convolutional Neural Network(CNN) is a popular deep learning model in image classification due to it's quick and easy implementation, it has capacity to extract small bit of information from large amount of data, learned automatically significant features hierarchies from raw data given directly to it, it can achieve better accuracy with large datasets and many more [26, 34, 63]. To detect most important or noticeable part of the images i.e., their salient features detection, CNNs are very useful. Recently deep learning researchers have used CNN based salient feature detection methods in images to achieve better performance than conventional techniques that use manual crafting of salient features [11, 15, 16, 18]. In last decades CNNs are successfully used by many researcher for breast cancer diagnosis based on different imaging techniques [20, 27, 33, 54], unfortunately very few of them have used CNN for breast cancer screening using thermographic images in the pasts. The main reason for this might be the CNN's two major drawbacks in radiology i.e., unavailability of large dataset for thermography and overfitting problem. Nevertheless, in the last two years or so the CNNs are most used models in small datasets and image pattern recognition because they can greatly improve the recognition rate, and at the same time reduce the requirements for the quality of the original pictures, and at the same time reduce the requirements for the number of training samples. Therefore, in this paper, we are developing the CNN based CAD model for breast cancer diagnosis using thermal infrared images database [8]. We compare state-of-the-art pre-trained CNN transfer learning models of VGG16 and InceptionV3 models for performance measurement. Also, to improve the performance of our CNN model, an analysis of different factors like network complexity, data augmentation, fine-tuning of model parameter in convolutional layers are performed for screening of breast cancer dataset [8].

The remainder of this paper is structured as: The literature review in this direction is given in Section 2. Section 3 explains about data and methods used in this paper. In Sections 4 and 5 the detailed results are discussed and the finally, Section 6 summarizes the conclusion and suggests further improvements in this direction.

2 Literature review

Several early detection methods for breast cancer screening, such as mammography and MRI, have been used over the last decades [4, 20, 50, 54, 61]. However, these methods are having many drawbacks as discussed in Section 1. Hence, in this section our focus is only on those researches that have used thermography technique for breast cancer diagnosis.

Digital infrared thermal imaging (thermography) was used for breast cancer diagnosis in 1960 and for pain management in 1980 [14]. The authors of [41] demonstrate a method based on techniques of extraction and image segmentation to identify and to diagnose irregular patterns in breast thermograms. The authors of [25] proposed an automated region of interest (ROI) for breast thermograms by considering both the lateral view and the front view of the breasts. The ROIs obtained support medical professionals to discriminate between normal and abnormal biomarkers. The authors of [1] performed texture feature extraction for every infrared image and then used SVM algorithm to differentiate those images between normal and malignant cases. They achieved 85.71% mean sensitivity, 90.48% specificity, and 88.10% accuracy. In [53] statistical features from thermal images are extracted for breast cancer detection using fuzzy classifier. The accuracy obtained in this research was 80%. The authors of [23] used artificial neural network for diagnosis of breast cancer using thermal images. In addition, The authors of [10] have shown that it is important to extract the hottest/coldest regions from thermographic images, which they use to do so quickly and easily with a detailed adjustment using the Lazy snapping process algorithm. The authors of [47] have used computer system that uses processing of images and ML algorithms to automatically analyze thermographic images generated by Br-aster devices. In [17] dynamic thermographies are used to generate time series, where multifractal analysis is performed to check for differences between tissue behavior to healthy and a malignant tumor.

Numerous traditional and modern machine learning techniques have been used by researchers to diagnose breast cancer based on thermal images. Majority of them have used images from online project database PROENG http://visual.ic.uff.br/ [8]. These techniques include support vector machine(SVM)algorithm [1, 28, 38], K-nearest neighbors(KNN)algorithm [28, 28, 56], hidden Markov models(HMM) [37], K-means and clustering [55], BayesNet [56], Genetic Algorithm (GA) [56], Artificial Neural Networks (ANN) [28, 38], Multilayer Perception [52], Deep and Recurrent ANN [38], Random forest(RF)model [52, 56], Decision Tree(DT) [52]. There are many drawbacks in these traditional methods used in literature, such as lack of standard preprocessing techniques, image features are chosen beforehand and designed with help of experts and extracted using complex problem-dependent techniques, low ceiling etc.

On the other hand, only few researchers have used modern machine learning CNN [12, 38] using thermography to detect breast cancer. The authors of [38]used statistical analysis of thermal images for classification of breast cancer based on deep learning techniques and achieved accuracy result more than 90%. While, authors of [12] analysed Resnet and VGG architectures to measure predictive accuracy of their CNN model using thermography and achieved more than 95% accuracy. The major drawback of these research is they consider variable number of patients in their work from which we can conclude that the database they used is not consistent. Also, they have not discussed enough regarding separating the database, i.e. how they split the patients data during training and testing of model. This can affect the model complexity. It is therefore necessary to consider both the performance and the complexity of the model throughout model development. Our work in this study is different from these literature studies using the same thermal image database [8] that they have used in their work because, our main aim is to develop simple convolutional neural network which can overcome all drawbacks of previously develop methodology as well as less complicated, more comfortable to train and can be generalized for arrival of new patients. Another goal of this model is to resolve repetitive and costly clinical test approaches in economically disadvantaged nations and the rest of the world simply as we as quickly and autonomously. Our major work in this paper is: first we demonstrate a novel CNN model to address some of these issues from literature studies and have a higher performance. second we evaluate performance comparison of two widely used pre-trained CNN model of VGG16 and InceptionV3 and third perform an analysis of different factors like network complexity, data augmentation, fine-tuning of model parameter in convolutional layers for optimization of model.

3 Materials and methods

This section is splits in to three subsections, consisting of introduction to software and hardware required for the model development, introduction to Convolutional Neural Network(CNN)and the model design.

3.1 Software and hardware requirement

Jupyter Notebooks: The experiment is a synthesis of a widely practiced medical diagnostic procedure and machine learning system that helps the researcher to create a diagnostic model that can distinguish among normal and ill people. An environment appropriate for the specific nature of this assignment is essential for the execution of this project. For machine learning projects, the most popular language is python. The programming language Python and the experimental notebook named the Jupyter Notebook will be used for this project [40]. Jupyter Notebooks may be deployed individually or together with a full distribution specially designed for the science computing of open-source software known as the distribution of Anaconda [36].

Python Libraries: For machine learning developments Python is the most popular language. It can be difficult to work with the image dataset, but can also be implemented in python with one single line of code. For this project, the main libraries are numpy, pandas, sklearn, itertools, matplotlib, OpenCV, and Keras. [46]. Numpy is one of the most popular packages used for computational programming [43]. Pandas is an open-source framework that is used for data structures and interpretation [45]. Matplotlib is a python module that helps the user to create 2D diagrams and graphs like plots, histograms, power curves, confusion matrix, bar charts and dispersion charts, etc. [21]. OpenCV means Open Source Computer Vision is a library designed to get a popular technology in the field of computer vision support. It includes over 2500 algorithms use both classical and modern computer vision techniques. These algorithms can identify and are very efficient in a wide range of objects [44]. Keras is tensorFlow's(backend process) high-level network API for building and training CNN models which is used in this project.

Hardware: In this project,for training CNN we have used virtual machine in Google cloud GPU with 16 GB of RAM and NVIDIA Tesla K80.

3.2 Convolutional neural network

Artificial Neural Networks (ANN) is a subset of machine learning that has grown and improved over the years. Scientists found that these neural networks can predict outcomes and recognize patterns with excellent accuracy, specificity, and high rate by exploiting the abundance of big data. ANN of neurons, as shown in Fig. 1, is connected to each other. Signals from one neuron to the other can be transferred and guided to these neurons. The functionality of ANN is influenced by the architecture and functionality of the human brain which enables it to predict accurately, just as it does for human beings [9]. In contrast, Convolutional Neural Networks (CNN) (Fig. 2) are used for image classification and identification. D. H. Hubel and T. N. Wiesel proposed a method that separates animals from many of their artifacts. They discussed using a cortical membrane in the cortex. This theory motivated researchers to construct algorithms to mimic neuron contact for object detection. It was believed that simple cells" remained within this simulated cortex. Such cells produced simple reactions which in effect became complex functional reactions formed by "complex cells". It is reconstructed in the context of an algorithm of deep learning. This needs an input picture and learning weights to distinguish between pictures It passes through a six-important process which produces an outcome based on its training [2] A convolution layer is the first layer in the image that is penetrated. A convolution is an organized process of mathematics, in which two sources of information are related. The input value is combined

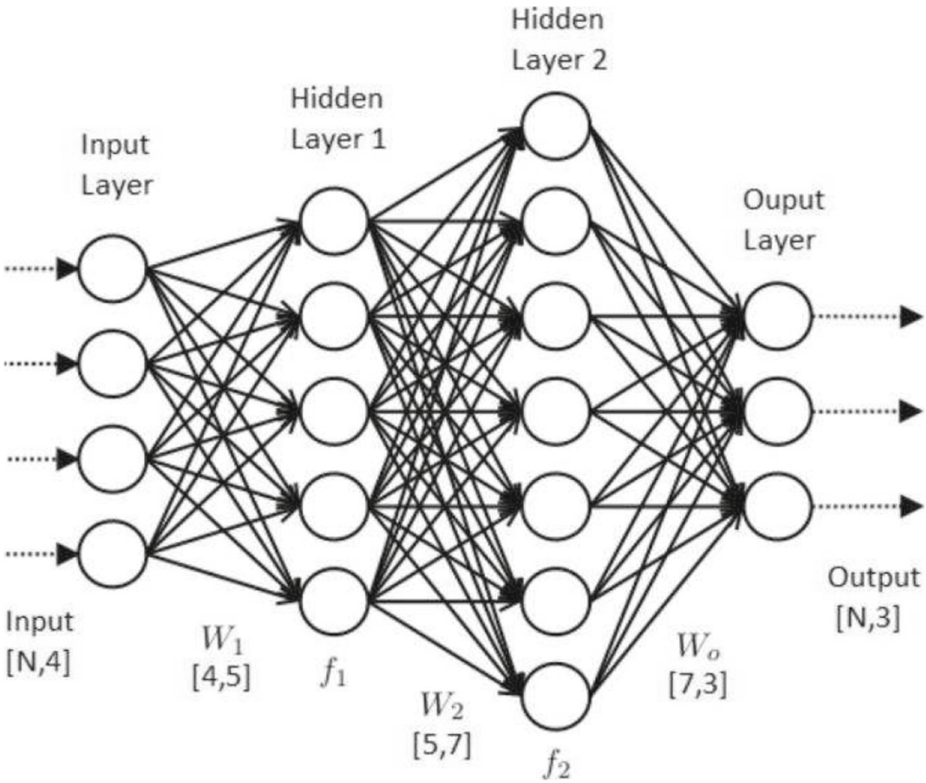

Fig. 1 Artificial neural network [9]

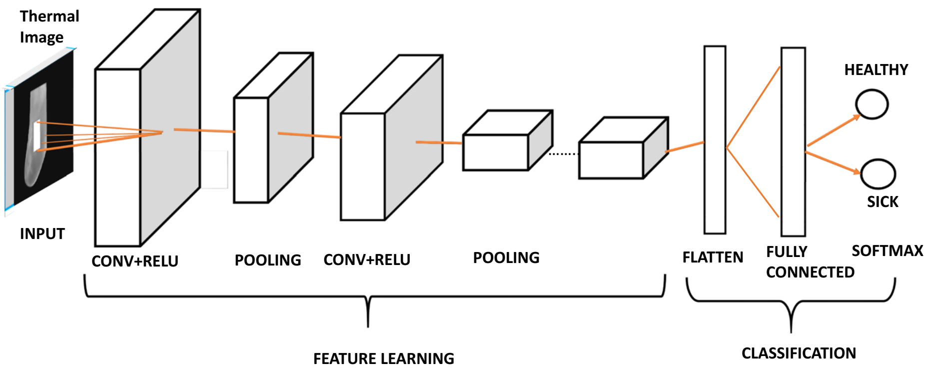

Fig. 2 Convolutional Neural Network(CNN)

to create activation maps with a smaller matrix defined as a kernel. The purpose of these maps is to define unique image features sthat are kernel-relevant. The dot product between the input matrix and the kernel is determined to attain a combined value that defines an independent input in the active matrix. The chosen pad will then be moved to the next patch and the process will continue until the matrix is completely triggered. [2] (Fig. 3). Next is poling layer. Each layer has the purpose of decreasing the spatial dimensions of the input which is handled in the next network layer. It decreases the spatial component, but does not change the volume's depth dimension. The optimum value is achieved by considering the values found. This technique is referred to as max pooling.

Negative figures are not suitable for this procedure and may break down the mathematics. The next layer named the normalization layer transforms all negative numbers in the equation into zeroes to prevent this from happening. Beyond that the convolutional layer and pooling layer are repeated several times until each iteration produces a smaller picture with a different activation chart.

After the neural network learns the features since incremental deep learning loops, the model may be too strong to match the training data and thus has no good prediction. This is avoided by dropouts. Dropout produces situations where the ANN is pressured by

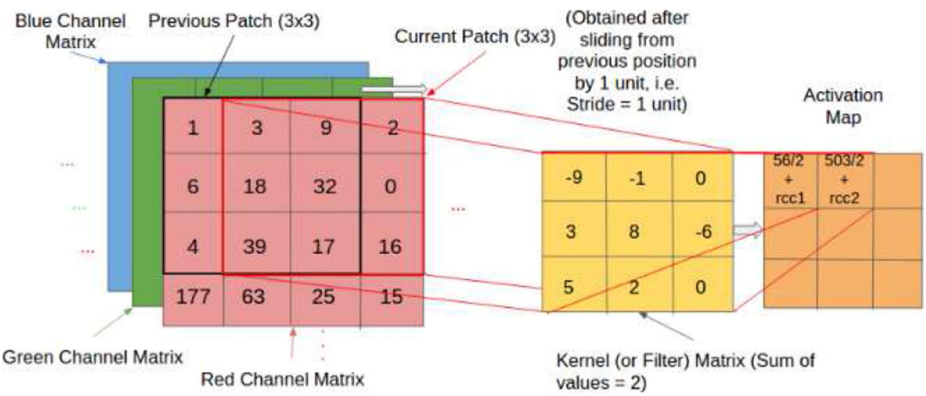

Fig. 3 The convolution layer of CNN

arbitrarily switching off neurons during the learner's stage to obtain many different representations of the same results. Many layer values are assigned randomly to zero during forward propagation.

Lastly, a SoftMax function is used when the classification phase ends, to transform the outputs into binary values or likelihood values. A SoftMax function is a mathematical function that incorporates real numbers into a matrix of probabilities.

3.2.1 Transfer learning and pre-trained CNN models

Until now, traditional machine learning and deep CNN models have been designed to work in isolation. These models are trained to solve specific tasks. When the feature-space distribution shifts, the models must be rebuilt from scratch. Transfer learning is the idea to overcome the method of isolated learning and to use the information obtained for one problem to solve similar ones. In many applications, previously learned model is used as a point of departure for another model to fine-tuned the model and to improve the performance. Several pre-trained models, trained by experts in the area, are available. Such previously learned models can be imported and fine-tuned in order to produce the appropriate classification. Two pre-trained models to study which one yields the best results will be tested in this project. These models can work efficiently with small as well as large database.

The VGG16 model The VGG16 is a pre-trained CNN model used to identify of large images. Tests show 92.7% precision on ImageNet, which represents a dataset of more than 14 million images. The suggestion was made by A. K. and Zisserman ,University of Oxford Simonyan [59]. This is one of the most famous convolutional neural network models that was sent to the ILSVRC-2014(ImageNet Large Scale Visual Recognition Challenge) competition. It is an improved version of the neural network AlexNet, which was the first convolutional neural network to win ILSVRC [57]. In VGG16 (Fig. 4) compared to the AlexNet network, large filters (sizes 11 and 5 in the first and second convolutional layers, respectively) were replaced by several 3x3 filters, one after the other. The VGG16 network for the competition was trained for several weeks using NVIDIA TITAN BLACK video cards.

InceptionV3 model InceptionV3 is further development idea of effective CNN from Google. This neural network achieves an accuracy of 92.8%, top 5 at ILSVRC-2015 [51]. The structure of the neural network InceptionV3 is shown in the Fig. 5. When designing InceptionV3, the basic principles of building the architecture were formulated:

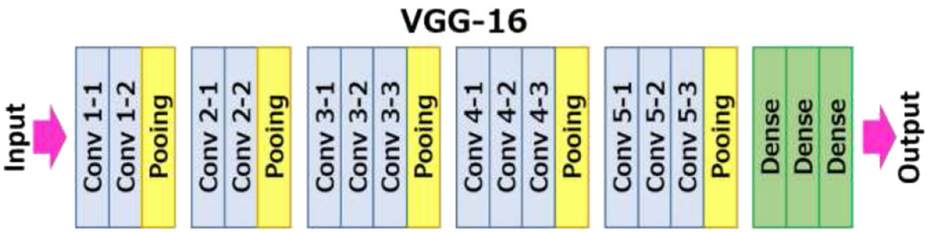

Fig. 4 VGG16 Model [59]

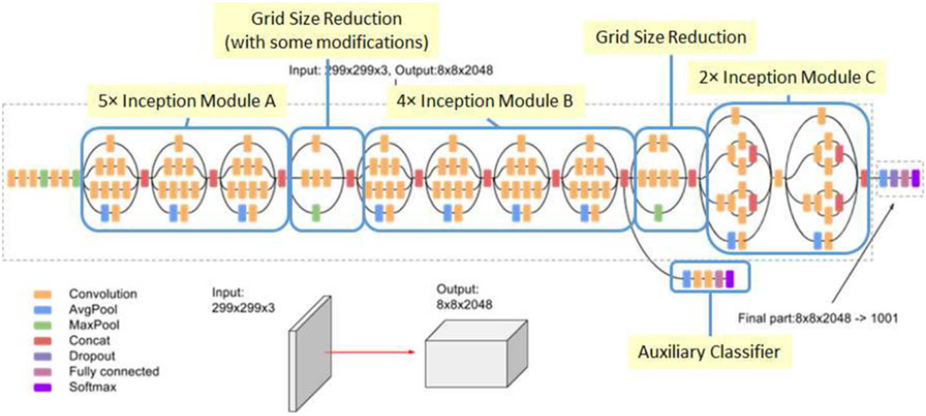

Fig. 5 Inception V3 Model

- A large number of signals are located in close proximity to each other. This can be used to make smaller convolutions. Neighboring signals are often correlated, which means that it is possible to reduce the dimension before convolution without loss of information;
- When increasing the free amount of resources for their effective use, it is necessary to increase the depth and width of the network at the same time;
- It is inefficient to use layers that sharply reduce the number of parameters, especially at the beginning of a neural network;
- "Wide" layers learn faster, which is especially important at high levels (but locally, that is, we can reduce the dimension after them).

3.3 Model design

3.3.1 Dataset

Work utilizing thermal imaging to identify breast cancer as a screening method is not new. In comparison to its mainstream replacement, mammography, several researchers in literature have used thermal images to show it's effectiveness. By contrast with a sick breast, differences between a healthy breast are very small but can be identified by a professional model. Thermal images used for this research are collected from online project database PROENG. A professional machine vision organization and data analysis have produced such images with a Support Vector Machine (SVM)algorithm for a computer-aided diagnostic device. CNN is much more advanced compared to SVM. SVM is based on a linear classification, while a CNN is a non-linear classification.

Thermal images that are stored in this database have been collected at 15 seconds intervals, and a total of 20 images have been recorded for each patient. The dimension of every infrared image has 640x480 pixels, i.e.307200 temperature points are given by one image. The detection method is greatly enhanced due to the number of layers introduced by the integration of the CNN. This database consists of total 67 patients aged between 29 to 85 years with a combination of 43 healthy and 24 breast cancer diagnosed patients [8].

Table 1 Classification of data

	Training	Testing	Validation
Healthy	479	119	20
Sick	371	89	20

3.3.2 Dataset structure

The data file was structured for each patient as healthy and ill patients with directories. Twenty thermal images were contained in the patient files. Most of these images appeared the same and might cause a problem for the neural network later.

The image data has to be structured before preparing the classification model. There are total 1140 images present in the model, However, due to the resemblance in thermal images few of them must be reduced to avoid model training issues. The images were eventually categorized within three folders with , training, testing and validation. Table 1 shows the how data is organize for the classification.

3.3.3 Preprocessing and data augmentation of thermal images

We preprocess thermal images by using contrast-enhancement, resizing, normalization and cropping in the same way that has been used in the literature study [52]. After preprocessing each thermal image get reduced to 224×224 dimesion from 640×480 dimension which help us to deals with least computations(reduction in computational cost by 75%).

For augmenting and preprocessing of images function of Keras *ImageDataGenerator()* is used. In the augmentation , we rotate, shear, shift zoom our image to increase our training data. This transformations are common when training CNN with small data sets. In principle, the test set should not be augmented as it is a proxy for real data. We evaluated three augmentation algorithms as shown in following Table 2. The augment0 model means there is no augmentation is applied i.e. original images are used for model. augment1 means thermal image will transform geometrically i.e., it will be flipped horizontally, rotated randomly within degrees of 0.05, shear horizontally by 0.03 times of original image and zoom 0.03 times than ordinal image. On the other hand the augment2 is quite complex transformation model. To avoid data exploration due to combination of different augmentation techniques and algorithms, in this work we evaluate the effects of these techniques only on InceptionV3 model. InceptionV3 is best model in this work.

Table 2 Augmentation design

Model	Parameters of Augmentation
augment0	No augmentation
augment1	rotation_range =5, shear_range=0.03, zoom_range=0.03, horizontal_flip=True
augment2	rotation_range=40, width_shift_range=0.2, rotation_range=40, width_shift_range=0.2,height_shift_range=0.2, rescale=1./255, shear_range=0.2, zoom_range=0.2, horizontal_flip=True, vericle_flip= true, fill_mode='nearest

3.3.4 Baseline CNN model design

Initially we build the baseline sequential CNN model by using Jupiter notebook in keras. This model consist of a convolutional layer(with 32 neurons and 3×3 kernel size). Next, the flatten layer which serves as a connection between the convolution and dense layers. The dense layer acts as output layer which will produce 2 outputs (healthy or sick patients). At the end of the dense layer, we use SoftMax activation function for converting the output in to binary values(0 or 1).

Next, we compile our model using three parameters: optimizer, loss and metrics. We use ‘adam’ optimizer for controlling learning rate. We use 10^{-4} learning rate for this model. To measure the loss of model we use ‘categorical_crossentropy’, A lower score of categorical_crossentropy indicates that the model is performing better. And, we use ‘accuracy’ matric to see the accuracy score on the validation set when we train the model. Finally we train our model using ’*fit()*’ function. As expected the accuracy produced by this model to be very low. Since, this model is not an ideal CNN model, the best CNN model has large numbers of previously pre-trained layers in it. Hence, to solve this problem we use transfer learning. As we discussed earlier transfer learning is a related trend where we train a CNN on data from a specific domain and later reuse this pre-trained model to extract image features in a different domain or as an initial network to fine-tune with new data. Several pre-trained models, trained by experts in the area, are available. Such pre-trained models can be imported and fine-tuned in order to produce the appropriate classification results. Therefor in this work, two pre-trained models, VGG16 and InceptionV3 are tested to find which one yields the best results.

3.3.5 Pretrained model design

We are using small number of thermal images for training our CNN model. To use this small dataset on pre-trained network might lead to problem of overfiitting of model, since these models are trained by using very large dataset from ImageNet. Hence to avoid this problem, we are going to use fine-tuning techniques on pre-trained models. We use three ways of fine-tuning on pre-trained models to improve accuracy :

i. Evaluate effect of size of classification layer on final classification accuracy: To experiment this, we use 5 different models on two pre-trained models, VGG16 and InceptionV3. The configuration of these models is shown in Table 3. The table consists

Table 3 Configuration of classification model

	Configuration	# Parameter VGG16	# Parameter InceptionV3
Model1	GAP → Softmax	8,194	1,152
Model2	GAP→FC(2048)→ FC(2048)→ Softmax	102,764,544	22,031,138
Model3	GAP → FC(512)→Dropout(0.5)→ FC(256) → Dropout(0.5)→ FC(128) → Dropout(0.5)→ Softmax	2,359,808	655,360
Model4	GAP → FC(512) → Dropout(0.5) → Softmax	1,678,131	473,440
Model5	GAP → FC(512)→ Dropout(0.5)→ FC(512) → Dropout(0.5)→ FC(256) → Dropout(0.5) → Softmax	1,180,160	262,272

of four columns representing the model type, configuration of classification layer, required number of parameter for VGG model and for InceptionV3 model respectively. For example the model4 in the 4th row has the four layers after convolutional layer which are, GAP means global average pooling layer, FC means fully connected layer with 512 neurons, dropout layer with 0.5 drop rate and finally softmax activation function.

ii. Freeze the weight of the first few layers in pre-trained Network: In the pre-trained network, the initial few layers take universal features like curves and edges, which are also relevant to our new problem. We want to keep these weights intact. Instead, we will get the network to focus on learning dataset-specific features in the subsequent layers. Therefore, to fine-tune our model in this way, we evaluate the number of convolution layer to be unfroze and untrained. We perform three different experiments shown in Table 4. In the first experiment, we unfreeze the last convolution layer(convolayer) of best model resulted in classification model experiments. In the next experiment to avoid overfitting due to the large amount of parameters of convolayer, we select the smaller model from the classification model experiments to unfreeze the last convolayer. And in the last experiment, the best model is chosen from the previous two experiments to unfreeze one more convolayer to measure the number of convolayers to be kept unfrozen.

iii. Fine-tunes hyperparameters: To make learning more stable and quicker, we select the best model from the previous two techniques of fine-tuning(i & ii), and then fine-tunes its other hyperameters like the learning rate, drop rate of dropout layer, the extension of the batch normalization layer, etc.

4 Result

A confusion matrix is used to check how well the model works for new data. It is a table that is used to define a classification model's efficiency and accuracy. A table is drawn from true positive(TP) means sick patients predicted to be sick, true negative(TN) means healthy patients predicted to be healthy, the false positive(FP) means sick patients predicted to be healthy and the false negative(FN) means healthy patients predicted to be sick. In addition to the confusion matrix, some metrics can be calculated to indicate how well a model classifies the information supplied. Accuracy, precision, recall and f1 score are the measures to be used. The f1 score might be the better metric to measure the success of the model's prediction. It is considered as the balance between precision and recall. Precision

Table 4 Configurations of CNN model after fine-tuning convolayer

Model	Configuration
convolayer Model1	Best CNN model with last convoLayer unfrozen
convolayer Model2	Smaller CNN model with last convoLayer unfrozen
convolayer Model3	Best CNN model in previous two models with one more unfrozen convolayer

Table 5 Results of Data Augmentation on InceptionV3 Model

Augmentation algorithms	Training Images	Accuracy InceptionV3
augment0	630	92.2
augment1	630	93.8
augment2	630	91.7
augment1	12,000	96.3
augment2	12,000	93.9

and recall are very simple and yet effective ways to measure prediction efficiency given as follows.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \cdot 100 \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$f1score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

4.1 Results of data augmentation algorithm

Results of data augmentation algorithms are given in Table 5. We can see that the augment1 algorithm gives better accuracy results for the InceptionV3 model as compare to the augment2; hence we use augment1 as the default model in the next experiments.

4.2 Results of classification model accuracy

The results of classification model accuracy of VGG16 and InceptionV3 are shown in Table 6, we can see that InceptionV3 with model3 is showing better accuracy results than VGG16. Therefore we fine-tuned model4 and InceptionV3 in the next experiments.

4.3 Results of fine-tuned model accuracy

From Table 7, it is found that the model1 with last unfrozen layer gives the best accuracy result.

Table 6 Classification Model Accuracy

	VGG16	Inception V3
Model1	69.7	72.9
Model2	73.1	77.8
Model3	84.8	91.4
Model4	89.7	95.9
Model5	85.5	92.1

Table 7 Results of Fine-tuned Model Accuracy with Unfrozen convolayer

Model	InceptionV3
Model4 by keeping last convolayer unfrozen	89.3
Model1 by keeping last convolayer unfrozen	95.4
Model1 by keeping last two convolayer unfrozen	91.1

4.4 Fine-tuned hyperparameters experimental results

Table 8 shows the results of the data exploration experiments. The best models resulted in last experiments (in Table 7) model1 and model4 with their last convolayer unfrozen are used for fine-tuning its hyperparameters. It can be seen that effect of adjustments in learning rate, dropout layers and many training images improve the accuracy of model1 and model4 with last unfrozen layer.

5 Discussion

5.1 Performance comparison

We have used different models in this paper for breast cancer screening using thermography. The performance of comparison of these models is summarize in Table 9 on the same dataset. The result obtained for InceptionV3 model are very high i.e 98.5% accuracy, 100% precision , 97.5% recall and 98.7% F1-score (Which is shown in the bold font in Table 9). The precision score of our model suggests that it can predict every breast cancer patient 100% correctly. VGG16 model is also performing well compare to our baseline CNN model developed initially. Earlier we have discussed that traditional machine learning algorithm

Table 8 Result of Fine-tuning hyperparameter on two best models: model1 and model4 by keeping their last convolayer unfrozen

Ex. No.	Model	Learning rate	Decay	# augment Training Images	Drop rate 1	Drop rate 1	Batch Normalization	InceptionV3 Accuracy
1	model1	0.001	0.8	630	0.7	NA	NO	89.3
2	by keeping	0.001	0.5	15000	0.6	NA	NO	90.5
3	last convolayer	0.0009	0.7	9000	0.6	NA	NO	91.3
4	unfrozen	0.0005	0.5	630	0.8	0.6	YES	93.4
5		0.0005	0.8	20000	0.6	NA	YES	94.3
6		0.005	0.8	8000	0.5	0.6	NO	97.4
1	model4	0.001	0.7	630	0.7	0.5	NO	87.9
2	by keeping	0.001	0.5	630	0.6	0.7	NO	89.3
3	last convlayer	0.0009	0.7	9000	0.6	0.5	YES	92.4
4	unfrozen	0.0005	0.7	20000	0.7	NA	YES	94.3
5		0.0005	0.8	15000	0.8	0.6	NO	95.4
6		0.005	0.8	8000	0.6	0.6	YES	95.7

Table 9 Impact of augmentation experiments on model

Model	Performance without augmentation				Performance with augmentation			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Baseline Model	59.4	0.590	0.667	0.626	62.9	0.661	0.677	0.669
VGG16	87.3	0.834	0.935	0.882	93.4	0.925	0.957	0.941
InceptionV3	93.1	0.933	0.949	0.941	**98.5**	**1**	**0.975**	**0.987**

have many disadvantages over modern machine learning techniques such as convolutional neural network because it is best feature learner, simple structure, low complexity.

5.2 Impact of data augmentation

The summary of all experiment testing results are shown in Table 9 with and without augmentation. We can see that the performance of the model improves with augmentation. That's because augmentation converts the image geometrically, which makes it easy for the CNN model to understand the underground feature without the influence of rotation and scale. Nonetheless, it can be shown from Table 5 that complex transformations aren't necessarily easier than basic ones. Large and complex transformations add a certain amount of noise to the feature that influences the learning process.

5.3 Analysis of model complexity

We analyse model complexity of our fine-tuned models using Tables 3 and 5. In left part of Fig. 6, first two columns of table represents the models from Table 3 which arranged in ascending order of their model parameters and their respective accuracy results are given in last two columns using Table 5. while from the right part of Fig. 6 we can see that very large number of parameters or very small number of parameters yields poor accuracies, where the best result of VGG16 and Inception V3 are of model4 that has the right number of parameters to train our database.

5.4 Analysis of techniques preventing overfitting

Table 8 shows the data exploration result of best model1 and model4 with keeping their last convolayer unfrozen using hyperparameters.The overall training tends to overfitting,

Model	VGG16 parameters	InceptionV3 parameters	VGG16 Acc	InceptionV3 Acc
Model1	8,194	1,152	69.7	72.9
Model5	1,180,160	262,272	85.5	92.1
Model4	1,678,131	473,440	89.7	95.9
Model3	2,359,808	655,360	84.8	91.4
Model2	102,764,544	22,031,138	73.1	77.8

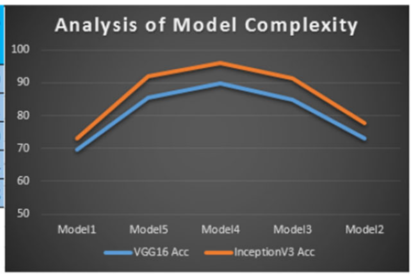

Fig. 6 Analysis of model Complexities

because there is no specific aspect that has a consistent and substantial effect on the overall accuracy. In contrast, model4 outcomes with particular conditions shown in Table 8, the drop rate and the number of changes in each learning iterations, consistently improve the accuracy. This is expected since there are so many parameters in the last convolayer.Hence training of model tends to overfitting.

6 Conclusion and future scope

In this article, the usefulness of thermal infrared imaging for the diagnosis of breast cancer is discussed over the existing mammograms source substitute. Invasive and costly testing has been reported for mammograms. Every 2 years, women in low-income countries tend to avoid breast cancer tests, and this lowers the rate of survival due to lack of awareness and early inspections. Mammograms were also still correct. A large number of false negatives and false positives regarding mammograms have been reported in the American cancer society. The alternative to mammograms is a non-invasive approach that can spot tumor growth well before mammograms are available. Numerous researchers are actively exploring the issue of using thermography as a diagnostic tool. In addition to research, in the United States, the method is even used as an alternative to mammograms in several health centers, including the O2 Wellness Centre. The thermography may be used for cardiovascular disease detection, cardiac diseases at the periphery, monitoring of human symptoms, early breast cancer markers, and many more, according to this article and institutes. This is considerable evidence that thermography is an important and underestimated method.

With the development of computer technology and statistical machine learning algorithms, CAD has achieved rapid development in the field of medical imaging. CAD technology can theoretically be applied to a variety of imaging techniques. In this work we have used modern machine learning approach CNN over traditional machine learning techniques to classify breast cancer using thermal images. We have used two pre-trained transfer learning model; VGG16 and InceptionV3 to improve performance of our baseline CNN model. The best result comes from the InceptionV3 model using data augmentation and fine-tuning of different several parameter. A collection of only 650 images is considered to be a very small data set for a convolutional neural network created through millions of picture training and testing. We discarded many thermal images of database due to resemblance which can cause an issue to CNN while training. Hence, in future it will be helpful to apply feature aggregation technique like multi-context ultra aggregation (MCUA) [42] to overcome the problem of resemblance and use more images of database for learning CNN model.

Funding Information No funding was received for this study.

Compliance with Ethical Standards

Conflict of interests The authors declare no conflict of interest, financial or otherwise.

Research involving human participants and/or animals This research paper does not contain any studies with human participants or animals performed by any of the authors.

References

1. Acharya UR, Ng EY-K, Tan J-H, Sree SV (2012) Thermography based breast cancer detection using texture features and support vector machine. J Med Syst 36(3):1503–1510
2. Agravat RR, Raval MS (2019) Prediction of overall survival of brain tumor patients. In: TENCON 2019-2019 IEEE Region 10 Conference (TENCON). IEEE, pp 31–35
3. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Luber B, Alani RM et al (2014) Detection of circulating tumor dna in early-and late-stage human malignancies. Sci Transl Med 6(224):224ra24–224ra24
4. Cano F, Madabhushi A, Cruz-Roa A (2018) A comparative analysis of sensitivity of convolutional neural networks for histopathology image classification in breast cancer. In: Romero E, Lepore N, Brieva J (eds) 14th international symposium on medical information processing and analysis. International society for optics and photonics, vol 10975. SPIE, pp 277–284. [Online]. Available: https://doi.org/10.1117/12.2511647
5. Cho N, Han W, Han B-K, Bae MS, Ko ES, Nam SJ, Chae EY, Lee JW, Kim SH, Kang BJ et al (2017) Breast cancer screening with mammography plus ultrasonography or magnetic resonance imaging in women 50 years or younger at diagnosis and treated with breast conservation therapy. JAMA Oncol 3(11):1495–1502
6. de Oliveira NPD, dos Santos Siqueira CA, de Lima KYN, de Camargo Cancela M, de Souza DLB (2020) Association of cervical and breast cancer mortality with socioeconomic indicators and availability of health services, vol 64
7. Dong Y, Jiang Z, Shen H, Pan WD, Williams LA, Reddy VV, Benjamin WH, Bryan AW (2017) Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In: 2017 IEEE EMBS international conference on Biomedical and Health Informatics (BHI). IEEE, pp 101–104
8. Elias T (2019) Visual lab - a methodology for breast disease computer-aided diagnosis. [online] visual.ic.uff.br. http://visual.ic.uff.br/en/proeng/thiagoelias/. accessed: 2020-01-20
9. Erdoğan H (2019) Trajectory Analysis and control using recurrent neural networks
10. Etehadtavakol M, Emrani Z, Ng E (2019) Rapid extraction of the hottest or coldest regions of medical thermographic images. Med Biol Eng Comput 57(2):379–388
11. Fan D-P, Cheng M-M, Liu J-J, Gao S-H, Hou Q, Borji A (2018) Salient objects in clutter: bringing salient object detection to the foreground. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 186–202
12. Fernández-Ovies FJ, Alférez-Baquero ES, de Andrés-Galiana EJ, Cernea A, Fernández-Muñiz Z, Fernández-martínez JL (2019) Detection of breast cancer using infrared thermography and deep neural networks. In: International work-conference on bioinformatics and biomedical engineering. Springer, pp 514–523
13. Foa EB, Chrestman KR, Gilboa-Schechtman E (2008) Prolonged exposure therapy for adolescents with PTSD emotional processing of traumatic experiences, therapist guide. Oxford University Press, Oxford
14. Frize M, Herry C, Scales N (2003) Processing thermal images to detect breast cancer and assess pain, in 4th International. In: IEEE EMBS special topic conference on information technology applications in biomedicine, 2003. IEEE, pp 234–237
15. Fu K, Zhao Q, Gu IY-H, Yang J (2019) Deepside: a general deep framework for salient object detection. Neurocomputing 356:69–82
16. Fu K, Fan D-P, Ji G-P, Zhao Q (2020). Jl-dcf:, Joint learning and densely-cooperative fusion framework for rgb-d salient object detection, arXiv:2004.08515
17. Gerasimova E, Audit B, Roux SG, Khalil A, Gileva O, Argoul F, Naimark O, Arneodo A (2014) Wavelet-based multifractal analysis of dynamic infrared thermograms to assist in early breast cancer diagnosis. Front Physiol 5:176
18. Gong C, Tao D, Liu W, Maybank SJ, Fang M, Fu K, Yang J (2015) Saliency propagation from simple to difficult. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2531–2539
19. Hafiz AM, Bhat GM (2020) A survey of deep learning techniques for medical diagnosis. In: Information and communication technology for sustainable development. Springer, pp 161–170
20. Hamidinekoo A, Denton E, Rampun A, Honnor K, Zwiggelaar R (2018) Deep learning in mammography and breast histology, an overview and future trends. Med Image Anal 47:45–67
21. Hunter JD (2007) Matplotlib: a 2d graphics environment. Comput Sci Eng 9(3):90–95
22. Jaeger BM, Hong AS, Letter H, Odell MC (2016) Advancements in imaging technology for detection and diagnosis of palpable breast masses. Clin Obstet Gynecol 59(2):336–350

23. Janghel R, Shukla A, Tiwari R, Kala R (2010) Breast cancer diagnosis using artificial neural network models. In: The 3rd international conference on information sciences and interaction sciences. IEEE, pp 89–94
24. Jones GW (2007) Delayed marriage and very low fertility in pacific asia. Popul Dev Rev 33(3):453–478
25. Josephine Selle J, Shenbagavalli A, Sriraam N, Venkatraman B, Jayashree M, Menaka M (2018) Automated recognition of rois for breast thermograms of lateral view-a pilot study. Quant Infrared Thermogr J 15(2):194–213
26. Joshi AV (2020) Deep learning, in machine learning and artificial intelligence. Springer, Berlin, pp 117–126
27. Kadam VJ, Jadhav SM, Vijayakumar K (2019) Breast cancer diagnosis using feature ensemble learning based on stacked sparse autoencoders and softmax regression. J Med Syst 43(8):263
28. Karim CN, Mohamed O, Ryad T (2018) A new approach for breast abnormality detection based on thermography. Med Technol J 2(3):245–254
29. Khan S, Yong S-P (2017) A deep learning architecture for classifying medical images of anatomy object. In: 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, pp 1661–1668
30. Köşüş N, Köşüş A, Duran M, Simavlı S, Turhan N (2010) Comparison of standard mammography with digital mammography and digital infrared thermal imaging for breast cancer screening. J Turk Ger Gynecol Assoc 11(3):152
31. Kremer JM (2017) Characterization of Axenic Immune Deficiency in Arabidopsis thaliana. Michigan State University
32. Leaf C (2004) Why we're losing the war on cancer (and how to win it). Fortune-European Edition-149(5):42–55
33. Liu K, Kang G, Zhang N, Hou B (2018) Breast cancer classification based on fully-connected layer first convolutional neural networks. IEEE Access 6:23722–23732
34. Lu Y, Zhang L, Wang B, Yang J (2014) Feature ensemble learning based on sparse autoencoders for image classification. In: 2014 International Joint Conference on Neural Networks (IJCNN), pp 1739–1745
35. Lunenfeld B, Stratton P (2013) The clinical consequences of an ageing world and preventive strategies. Best Pract Res Clin Obstet Gynaecol 27(5):643–659
36. Maestre CR, Gregori FA, López MP, Aldeguer RR Jupyter notebook: theory and practice of mathematical models in engineering and architecture
37. Mahmoudzadeh E, Montazeri M, Zekri M, Sadri S (2015) Extended hidden markov model for optimized segmentation of breast thermography images. Infrared Phys Technol 72:19–28
38. Mambou SJ, Maresova P, Krejcar O, Selamat A, Kuca K (2018) Breast cancer detection using infrared thermal imaging and a deep learning model. Sensors 18(9):2799
39. Maxim LD, Niebo R, Utell MJ (2014) Screening tests: a review with examples. Inhal Toxicol 26(13):811–828
40. McKinney W (2012) Python for data analysis: data wrangling with pandas. NumPy, and IPython. O'Reilly Media Inc.
41. Milosevic M, Jankovic D, Peulic A (2014) Thermography based breast cancer detection using texture features and minimum variance quantization. EXCLI J 13:1204
42. Nie G-Y, Cheng M-M, Liu Y, Liang Z, Fan D-P, Liu Y, Wang Y (2019) Multi-level context ultra-aggregation for stereo matching. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3283–3291
43. Oliphant TE et al (2019) Numpy: N-dimensional array package for python
44. Opencv documentation (2019) https://opencv.org/about/, accessed: 2020-01-20
45. Python data analysis library. [Online]. Available: https://pandas.pydata.org/
46. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2011) Scikit-learn: Machine learning in python. J Mach Learn Res 12:2825–2830
47. Qi H, Head JF (2001) Asymmetry analysis using automatic segmentation and classification for breast cancer detection in thermograms. In: 2001 conference proceedings of the 23rd annual international conference of the IEEE engineering in medicine and biology society, vol 3. IEEE, pp 2866–2869
48. Qi H, Kuruganti PT, Snyder WE (2012) Detecting breast cancer from thermal infrared images by asymmetry analysis. Med Med Res 38
49. Rahman MM, Saha KC, Mukherjee SC, Pati S, Dutta RN, Roy S, Quamruzzaman Q, Rahman M, Chakraborti D (2015) Groundwater arsenic contamination in bengal delta and its health effects. In: Safe and sustainable use of arsenic-contaminated aquifers in the gangetic plain. Springer, pp 215–253
50. Rampun A, Scotney BW, Morrow PJ, Wang H, Winder J (2018) Breast density classification using local quinary patterns with various neighbourhood topologies. J Imaging 4(1):14

51. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M et al (2015) Imagenet large scale visual recognition challenge. Int J Comput Vis 115(3):211–252
52. Santana MAd, Pereira JMS, Silva FLd, Lima NMd, Sousa FNd, Arruda GMSd, Lima RdCFd, Silva WWAd, Santos WPd (2018) Breast cancer diagnosis based on mammary thermography and extreme learning machines. Res Biomed Eng 34(1):45–53. https://doi.org/10.1590/2446-4740.05217
53. Schaefer G, Závišek M, Nakashima T (2009) Thermography based breast cancer analysis using statistical features and fuzzy classification. Pattern Recogn 42(6):1133–1137
54. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W (2019) Deep learning to improve breast cancer detection on screening mammography. Sci Rep 9(1):1–12
55. Silva LF, Sequeiros GO, Santos MLO, Fontes CA, Muchaluat-Saade DC, Conci A (2015) Thermal signal analysis for breast cancer risk verification. In: MedInfo, pp 746–750
56. Silva LF, Santos AAS, Bravo RS, Silva AC, Muchaluat-Saade DC, Conci A (2016) Hybrid analysis for indicating patients with breast cancer using temperature time series. Comput Methods Programs Biomed 130:142–153
57. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556
58. U of Ottawa Evidence-Based Practice Center, Moher D, Schachter HM (2004) Measuring the quality of breast cancer care in women. Agency for Healthcare Research and Quality
59. ul Hassan M (2018) Vgg16–convolutional network for classification and detection, Neurohive. Dostopno na: https://neurohive.io/en/popular-networks/vgg16/[10. 4. 2019]
60. Upadhyay RP (2012) An overview of the burden of non-communicable diseases in india. Iran J Public Health 41(3):1
61. Vikhe P, Thool V (2018) Morphological operation and scaled réyni entropy based approach for masses detection in mammograms. Multimed Tools Appl 77(18):23777–23802
62. W.H. Organization (2019) Global action plan on physical activity 2018-2030: more active people for a healthier world. World Health Organization, 2019
63. Wu G, Lin Z, Han J, Liu L, Ding G, Zhang B, Shen J (2018) Unsupervised deep hashing via binary latent factor models for large-scale cross-modal retrieval. In: Proceedings of the twenty-seventh international joint conference on artificial intelligence, IJCAI-18. International joint conferences on artificial intelligence organization, pp 2854–2860. [Online]. Available: https://doi.org/10.24963/ijcai.2018/396
64. Yadav SS, Jadhav SM (2019) Deep convolutional neural network based medical image classification for disease diagnosis. J Big Data 6(1):113
65. Yao X, Wei W, Li J, Wang L, Xu Z, Wan Y, Li K, Sun S (2014) A comparison of mammography, ultrasonography, and far-infrared thermography with pathological results in screening and early diagnosis of breast cancer. Asian Biomed 8(1):11–19