Research article

A hybrid deep learning model for breast cancer diagnosis based on transfer learning and pulse-coupled neural networks

Meteb M. Altaf*

National Center for Robotics Technology and Internet of Things, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia

* Correspondence: Email: maltaf@kacst.edu.sa; Tel: +966114883555.

Abstract: Radiology experts often face difficulties in mammography mass lesion labeling, which may lead to conclusive yet unnecessary and expensive breast biopsies. This paper focuses on building an automated diagnosis tool that supports radiologists in identifying and classifying mammography mass lesions. The paper's main contribution is to design a hybrid model based on Pulse-Coupled Neural Networks (PCNN) and Deep Convolutional Neural Networks (CNN). Due to the need for large datasets to train and tune CNNs, which are not available for medical images, Transfer Learning (TL) was exploited in this research. TL can be an effective approach when working with small-sized datasets. The paper's implementation was tested on three public benchmark datasets: DDMS, INbreast, and BCDR datasets for training and testing and MIAS for testing only. The results indicated the enhancement that PCNN provides when combined with CNN compared to other methods for the same public datasets. The hybrid model achieved 98.72% accuracy for DDMS, 97.5% for INbreast, and 96.94% for BCDR. To avoid overfitting, the proposed hybrid model was tested on an unseen MIAS dataset, achieving 98.77% accuracy. Other evaluation metrics are reported in the results section.

Keywords: breast cancer diagnosis; PCNN; CNN; deep learning model; transfer learning

1. Introduction

In 2018, approximately 18 million cancer cases entered the existing population living with cancer worldwide, and over half of them resulted in death [1]. According to statistical evidence, breast cancer

is one of the most common types of cancer among women. Around 1 in 8 women in the USA, approximately 12%, will belong to the breast cancer population over the course of their lifetime [2]. Studies and statistics indicate that if breast cancer patients can be diagnosed at an early stage, the rate of five-year survival may reach 90%, while it would not exceed 20% in the terminal stage [1]. Generally, "mammography" provides high-resolution, clear, and accurate imaging for the purpose of breast cancer examination. Accordingly, an intelligent computer-aided diagnosis system assists radiologists in benefiting from mammography to make a precise diagnosis. Previous research trials indicate that computer-aided diagnosis systems can enhance the diagnostic success rate and eliminate the probability of misdiagnosis [2–5]. The main goal for most computer-aided diagnosis platforms is to perform labeling and to differentiate between malignant and benign lesions.

Recently, Deep Learning (DL) systems have outperformed other Machine Learning (ML) systems in computer vision problems [6–9]. This is especially the case for Convolutional Neural Networks (CNN) [10,11]. DL techniques have been exploited in diverse medical domains and applications, including pulmonary peri-fissural classification [12] and interstitial lung disease lymph node identification [13]. Most of the implemented systems that have used CNN have applied a "vanilla" approach. In particular, the extracted CNN features are used solely or in combination with other hand-crafted descriptors to perform the classification [14,15]. Meanwhile, the most effective characteristic of employing CNNs is to neutralize feature engineering in the classification process and use raw images. CNN architecture is specially designed to benefit from the 2D structure of the input image. However, more importantly, it can generalize to other recognition problems [14,15]. To benefit from the characteristics of CNNs, large annotated datasets should be used in training, which are not available in the medical field, especially in breast cancer. Furthermore, when training CNNs from scratch, the training process consumes intensive computational power and memory. To overcome this issue, Transfer Learning (TL) may be used [16], where the idea is to leverage a pre-trained model from available images and exploit a fine-tuning for in-hand classes [17]. TL is widely applied in DL applications and it has shown effectiveness in the medical field, particularly in domains where data are normally limited. Another neural network that can be used as a transformation, which expresses image contents without segmentation, is known as the Pulse-Coupled Neural Network (PCNN) [18]. The PCNN can convert a 2D image into a 1D signal that can be perceived as an "Image Signature" [19]. The PCNN is a neural network model that was first proposed by Eckhorn [19] at the beginning of that century. The core idea of PCNN comes from the study of synchronous pulse bursts in the cat's visual cortex.

1.1. Why PCNN?

- Models involving only CNNs may be sensitive to image acquisition setting, scanning type, and pre-processing steps. PCNNs can neutralize these effects through the iterative pulses.
- PCNNs offer a method for prioritizing the features and can act as a feature selector. The main contributions of our work are as follows:
- To extract mammogram features more precisely, capturing both local and global features, and use different filter scales.

- Combining a PCNN model with a CNN pre-trained model to extract and select features before classification in the dense layer.
- The proposed model neutralizes the problem of overfitting due to relatively small size datasets.
- The proposed model guarantees the generalization by being exposed to multiple datasets.

The rest of this paper is organized as follows: Section 2 discusses related work published in the scientific community; Section 3 presents the scientific background and describes the datasets; Section 4 discusses the proposed system; and Section 5 presents the experimental results and concludes the findings.

2.   Related work

In the medical community, examinations and screening tests for detecting and diagnosing breast cancer by domain experts are parts of an extremely sensitive, time-consuming, and costly process. The efficiency of the diagnosis procedure can be increased by exploiting technology and intelligent software components. Hence, using these tools causes direct cost reduction, along with eliminating diagnosing effort. Accordingly, a substantial number of studies have been undertaken based on AI and ML approaches. One of these approaches, presented in Akay [20], used Support Vector Machines (SVM) based on a feature selection algorithm to detect breast cancer. The implemented system achieved 99.51% classification accuracy. Karabatak [21–25] proposed a hybrid system using an expert system, association rules, and feed forward neural network to detect breast cancer. The role of the association riles was to eliminate the neural network input features, while the neural network was trained for the classification task. The system's overall classification accuracy was greater than 95%. A hybrid approach was proposed in [26] combining both evolutionary techniques and fuzzy systems. One advantage of this system was that it provided explainable rules for the domain experts. Recently, Deep Neural Networks (DNNs) have been introduced in the biomedical field to detect cancers, including breast cancer. Table 1 illustrates previous trials that have been undertaken using deep learning.

In past few years, exploiting ML and DL in breast cancer is world widely shining, many researchers have used different techniques and achieved a competitive results [29–34].

Transfer Learning (TL) is needed in the medical field because data are rare and costly, and few datasets are publicly available. Furthermore, it is a very time-consuming task for radiology experts to collect and annotate datasets [35]. At the same time, the task of training a CNN from scratch is computationally extensive and costly in terms of memory resources. Extracting features from the first layers of a pre-trained CNN should embrace generic features such as edge detectors. Those are useful for many requirements, but in the later layers, these features are merged in such a way as to become more focused on the details of the label classes in the training set.

Until a solution is devised to the problem of limited availability of large-scale medical datasets for mammography, the methodology of merging Transfer Learning and data augmentation seems to be a promising approach for CNN training.

The extracted features can show whether a model has successfully learned or not, allowing a decision to be made about whether to halt the training process early. This is referred to as early stopping.

Table 1. Sample of deep neural network research in cancer diagnosis.

Authors	Description	Results
Dontchos et al. [27]	It uses CNN vanilla architecture to identify the existence of breast cancer. The model outputs are verified by neurology experts from both Academia and Community.	94.9% (academic acceptance) 90.7(community acceptance) ($p< 0.001$)
Spanhol et al [28]	Two tissue types: Epithelial and stromal have been segmented in histological images. The segmentation was performed using CNN. The classification also was performed using pre-trained CNN model named as AlexNet. Multiple experiments were executed on multiple datasets.	90% accuracy
F. F. Ting et al [10]	A modified training algorithm for convolutional neural network It detects breast cancer.	90.50% accuracy
Lee. et al [11]	They proposed incremental boosting CNN for providing an accurate fast diagnosis for breast cancer from histopathological. images. The model.	Accuracy of 96.4 and 99.5% for 4 and 2 classes classification jobs

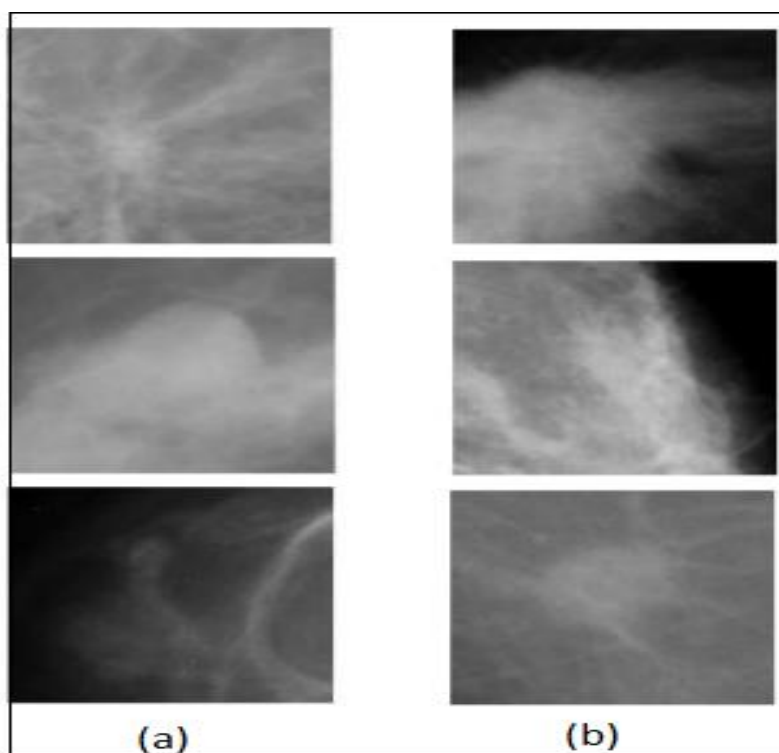## 3. Methods and datasets

3.1. Datasets

Figure 1. Image samples from DDMS dataset (a) Benign masses; and (b) Malignant masses.

In this work, four publicly available datasets were used: 1 for testing only and 3 for training and testing. The main reason why one uses a dataset for testing only is to ensure that the proposed model does not overfit. One way to boost the learning task is to merge all 3 datasets into a single one, DDMS [36]. The DDMS dataset includes more than 2650 instances ordered by the severity degree of the case findings. Each

case instance has 4 view mammograms for the same person with the associated ground truth (Figure 1), along with other related information on the following:

- Patient's age.
- Exact exam timestamp.
- Breast density (determined by the specialist both for benign and malignant cases).

3.1.1. INbreast

INbreast contains a total of 115 cases [37]. Both craniocaudal and mediolateral oblique views are included (Figure 2). Overall, it contains a total of 410 images. INbreast contains the following information:

- Acquisition date.
- Patient's age.
- Existence of previous cases in family.
- Breast density.

INbreast contains a wide diversity of cases, including different types of lesions. In this work, we focus on benign and malignant cases.
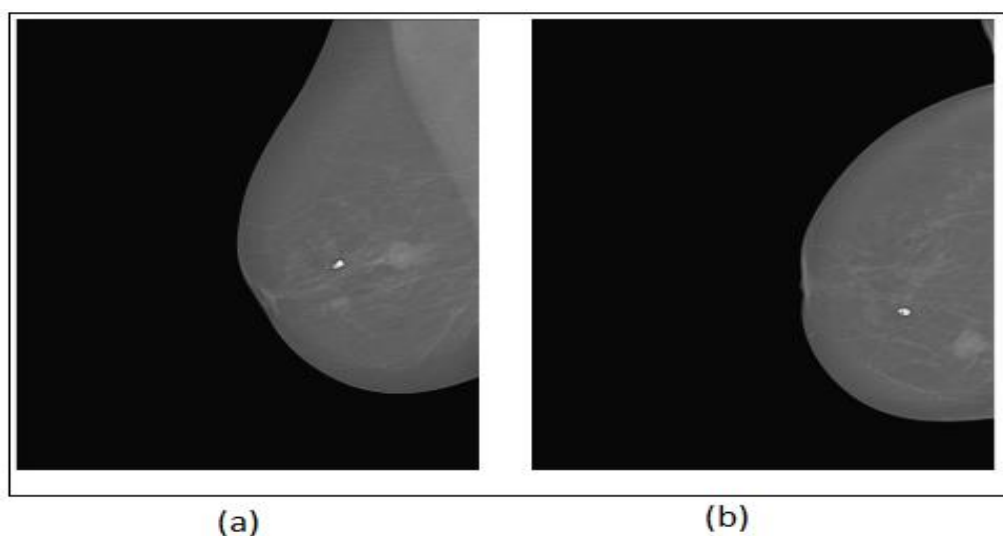

Figure 2. Image samples from INbreast dataset (right breast views) (a) Mediolateral oblique view; and (b) Craniocaudal view.

BCDR is a new database of film mammography. It contains 736 lesions in 344 patients. The lesions were verified by biopsy (Figure 3). The available pieces of information for each case are the following:

- Clinical data.
- Craniocaudal.
- Mediolateral oblique.
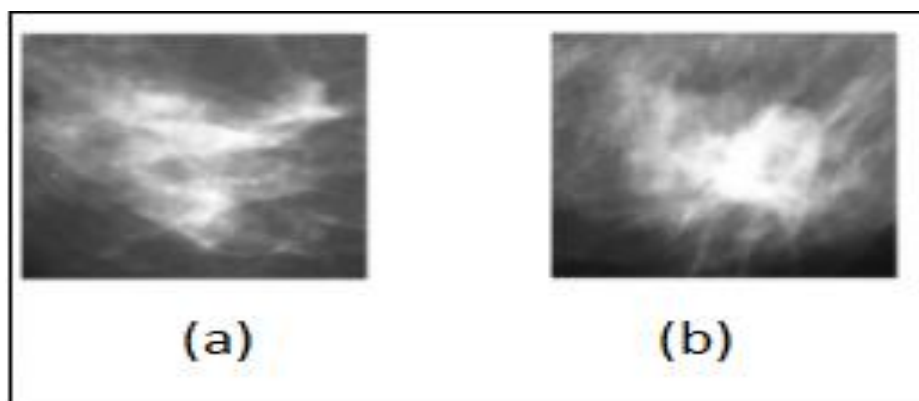- Lesion contour coordinates.

Figure 3. Image samples from BCDR dataset (a) Benign masses; and (b) Malignant masses.

3.1.2. MIAS

Mammographic Image Analysis Society (MIAS) is a commonly known dataset that includes 322 film mammograms [38]. It also contains expert validation ground truth (Figure 4).
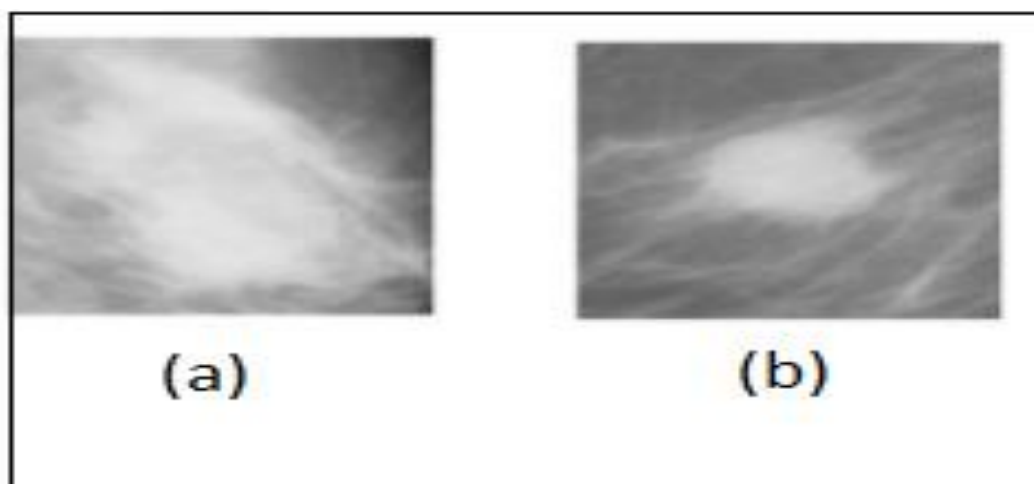

Figure 4. Image samples from MIAS dataset (a) Benign masses; and (b) Malignant masses.

3.2. Convolutional neural network (CNN)

Convolutional Neural Networks (CNN) were proposed by Yann LeCun et al. in 1998 [39]. A CNN is a multistage combination of convolutional layers and fully-connected layers. Generally, a convolution process is used to minimize the memory required, and it is computed on limited regions to enhance performance.

There are four main operations in a CNN [15]:
- Convolution.
- Non-linearity.

- Pooling (subsampling).
- Classification (fully-connected layer).

Starting from a 2D input matrix such as an image, a CNN is composed of layers. Each layer contains various filters (kernels) (Figure 5). Input through the forward pass, it is convolved with each filter and dot products between the entries of the filters and the input are computed.
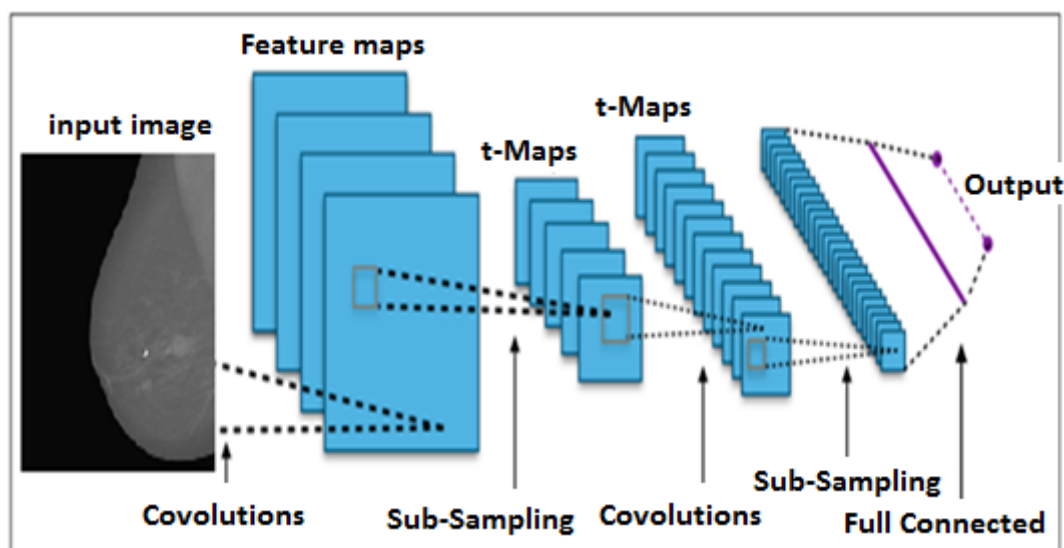

Figure 5. Example of CNN architecture.

A feature map is obtained by repeatedly applying a function across sub-regions of the entire image (convolution) with a linear filter, adding a bias term, and then applying a non-linear function. If we denote the K_{th} feature map at a given layer as h^k, whose filters are determined by the weights W^k and bias b_k, then the feature map h^k is obtained as follows [40]:
CNN Architecture Example

$$h_{ij}^k = tanh(W^k * x)_{ij} + b_k \tag{1}$$

To capture a richer representation of the input, each hidden layer is composed of many feature maps, $\{h^{(k)}, k = 0..K\}$

3.3. Transfer learning

Most computer vision studies that have used TL have customized the ImageNet dataset for pre-training tasks [41]. There are several commonly known pre-trained CNN models, including Alex-Net [42], VGG16 [43], ResNet50 [44], and Goog LeNet [45]. Each model was pre-trained using ImageNet, and each one is tended for a 1K-class classification job. Until high volume datasets for mammography medical images are available, the concept of combining TL and data augmentation is an extremely promising CNN training methodology.
AlexNet:

- Proposed in 2012
- 8 layers (5 convolution-pool layers and 3 fully-connected)
- Filter sizes are 3, 5, and 11
- 96-384 filters
- 4096-4096-1000 fully-connected layer sizes respectively
- Input is 227 x 227

VGG-Net

- Proposed in 2014
- 19 layers (16 convolution-pool layers and 3 fully-connected)
- Filter size is 3
- 64-512 filters
- 4096-4096-1000 fully-connected layers sizes respectively
- Input is 224 x 224

Res-Net

- Proposed in 2015.
- 152 layers (151 convolution-pool layers and 1 fully connected)
- Filter sizes are 1, 3, and 7
- 64-2048 filters
- 1000 fully-connected layer size is 1000
- Input is 224 x 224

Google-Net

- Proposed in 2014.
- 22 layers (21 convolution-pool layers and 1 fully connected)
- Filter sizes are 1, 3, 5, and 7
- 64-384 filters
- 1000 fully-connected layer size is 1000
- Input is 224 x 224

As mentioned before, these TL architectures were trained on the 1 K object type identification on the ImageNet database. ImageNet includes 1.2 million images as a training set, 50,000 images for validation purposes, and 100,000 images as a test set. All of these models use a data augmentation methodology besides dropout layers to minimize the probability of overfitting.

3.4. Pulse-coupled neural networks

PCNNs constitute a biologically-inspired model based on a cat's sight model. The model consists of a single layer of neurons, each of which is connected directly to one input image pixel. The neuron accepts 2 different inputs: the linking and the feeding inputs. The feeding input embraces an external signal besides an internal one meanwhile the linking only accepts local signal [18]. The internal stimulus is perceived from the feeding cycle radius around the neurons. Meanwhile, the external stimulus comes from the intensity of the associated pixel in the input image. Afterward, the linking and feeding inputs are combined to construct the internal potential that determines the output firing of the neuron (Figure 6) [18].
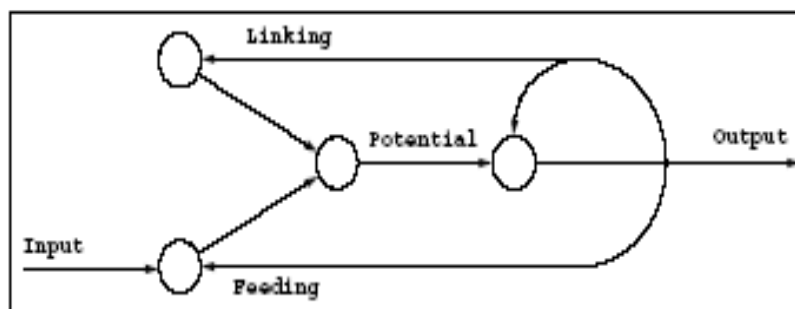
Figure 6. Original PCNN neuron [18].

3.4.1. Modified PCNN

There are many variations for PCNN training algorithms. These variations seek to mimic the physiological pulse-coupled neuron. They express different simplifications made to the computations and still retain the main features of the theory. One of these variations is "Modified PCNN". In this model, each neuron is governed by the following sequence of equations [18,19]:

$$L(i) = L(i-1)e^{-\propto L} + Vl.(W * Y_{sur}(i-1)) \tag{2}$$

$$F(i) = P + F(i-1) + e^{-\propto F} + VF.(W * Y_{sur}(i-1)) \tag{3}$$

$$Y(i) = F(i).[1 + \beta.L(i) \tag{4}$$

$$\theta(i) = \theta(i-1)e^{-\alpha q} + V\theta.YO(i-1) \tag{5}$$

$$Y(i) > \theta\ (i)\ \ then\ Yo(i) = 1$$

$$\text{otherwise}: Yo(i) = 0 \tag{6}$$

where L(i) represents input linking, F(i) represents feeding, P is the image's pixel intensity, U(i) is the neuron activation potential, and θ(i) is the neuron threshold. There are multiple hyperparameters, including (αL, αF and αq, as well as the decay coefficients besides (β) which represent the linking coefficient. Finally, VL and VF are the linking and threshold potential hyperparameters. Regarding the neuron output, Y presents the firing information regarding the surrounding neurons, while Yo conveys the neuron firing information. Additionally, W is the weight matrix. The "Modified PCNN" neuron architecture is illustrated in Figure 7 [18].
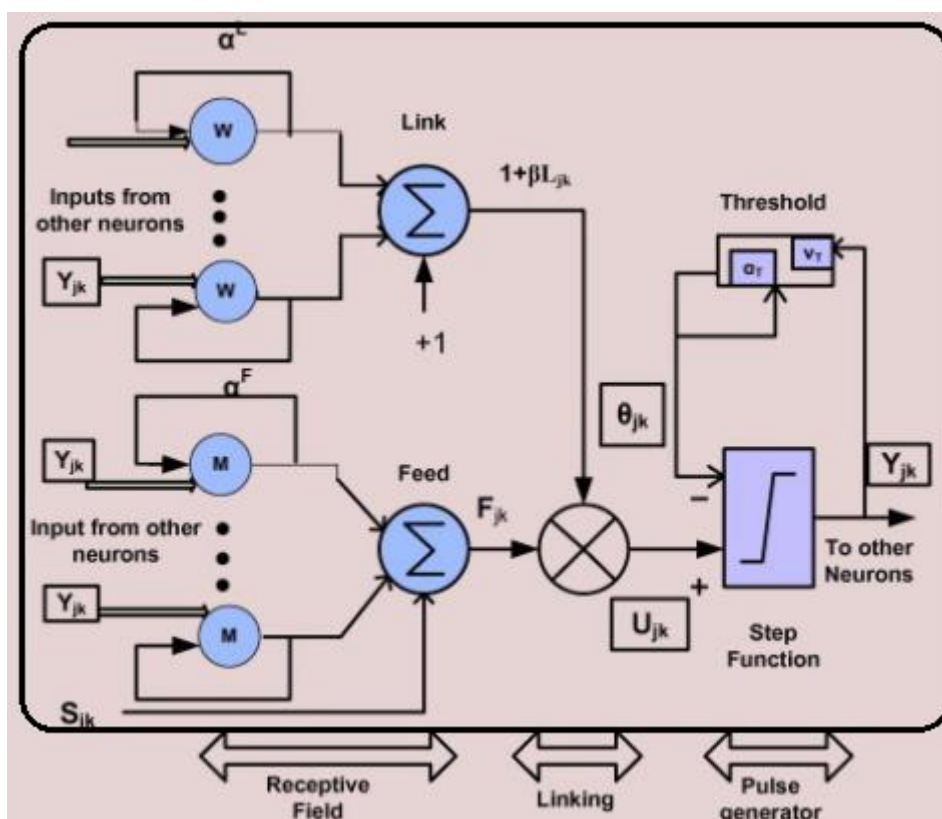
Figure 7. Modified model for PCNN neuron [18].

The output of each iteration (the firing status of each neuron) is used to form a time-series-like signature. The standard feature generation formula G(n) for iteration (i) is computed as [46–48]:

$$G(n) = \sum_{i=0}^{n} Y(i) \qquad (7)$$

In this work, we use the enhanced formula proposed by Elons [18]. This improves signature quality by adding a weighting factor known as a "Continuity Factor".

4. Proposed system

The aim of this system is to build a classification model to classify the input images based on one of the following possible labels: Normal, Abnormal Benign, and Abnormal Malignancy. Different Transfer Learning (TL) models will be exploited to provide enhanced features for dense layers to classify the input image. The main challenge is to neutralize the effect of acquisition quality on the model, which is the main motivation for using PCNNs. PCNNs generate image signatures without the need for segmentation or pre-processing steps. This signature is fed into feedforward neural layers besides CNN features (Figure 8). As mentioned above, different TL models were tested on different datasets. Both models represent a feature extractor models, both features vectors are fused to represent a unified feature vector.
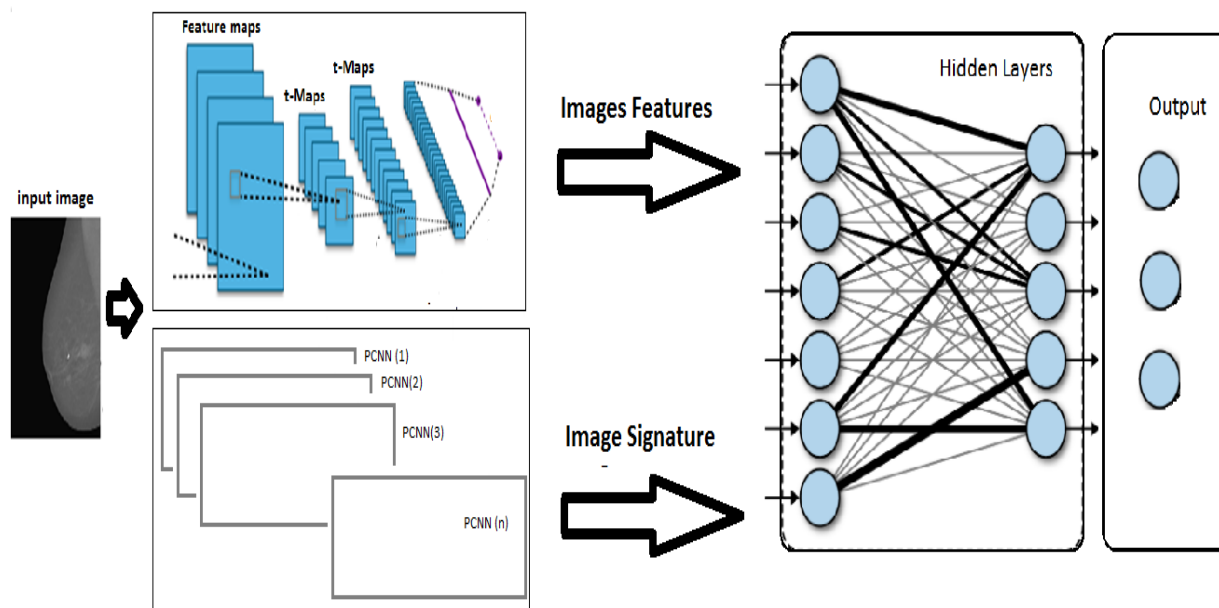
Figure 8. Proposed system modules.

The fused feature vector is embraced into a deep feedforward network for the purpose of classification.

Most previous researchers have used a segmentation step to identify the Region of Interest (ROI). The main motivation for this is to reduce the computational effort of the CNN and eliminate the impact of limited training data size. Meanwhile, the proposed system accepts the complete image without segmentation or any pre-processing steps. This idea mainly depends on the PCNN's ability to generate image signatures that are invariant to acquisition quality, scaling, rotation, and translation.

The initial layers of a CNN seek to capture generic features and typically conduct tasks such as edge detectors. At the same time, the successive layers pay attention to more specific details in the classes.

Due to differences between the ImageNet dataset and the hand-crafted datasets, model fine-tuning is proposed to adjust the features and to consider the designated datasets.

The superiority of DL and PCNN compared to the other ML methods stems from the capability of multi-level feature representation in any hidden layer. Additionally, DL and PCNN can significantly neutralize the effects of low-quality imaging resolution and expected noise.

5. Results and configurations

This research was conducted using the PyTorch platform and the Python environment. For the DDMS dataset, this was a subset of 900 images from 450 patients, including 150 malignant cases, 150 benign cases, and 150 are normal cases. Meanwhile, we selected 75 cases from INbreast containing a total of 300 images, 100 images from 25 normal cases, 25 benign cases, and 100 from malignant. From BDCR, a subset of 450 patients was selected with the same distribution. The final dataset was a combination of all the subsets selected. The input was "augmented" using a sequence of random transformations, ensuring that the designated models would never receive the same image as an input multiple times. Shifting was

conducted with a fraction of 0.3 from the image's overall height and width. Also, a rotation was randomly applied in a range between 0 degrees and 45 degrees. Shear and zoom were applied in the ranges of 0.5 (0.5 1.5) respectively. In the preparation phase, the training procedure considered balanced class preserving, and the dataset was separated into a training set (80%) and a test set (20%).

Since the data has no impactful imbalance, accuracy can be represent a key performance indicator. The simulation environment can be represented as a grid search process to select the optimal set of hyper-parameters. The model is evaluated for each dataset separately and on the merged dataset.

Since the model final output is a probability distribution and the data do not suffer from severe imbalance so we count on the default classification certain threshold. Meanwhile, we should consider the model uncertainty for biased or imbalance classes situation.

Selecting the correct hyperparameters when conducting fine-tuning is a sensitive process. The optimization step was performed using Stochastic Gradient Descent (SGD) rather than any other optimizer. The main motivation for this was to ensure that the learning rate magnitude was limited. The feedforward layers exploited adaptive ADAM. Optimizer as a training algorithm.

Moreover, to enhance the experimental results and to neutralize the probability of overfitting, data augmentation was exploited with L2 regularization, and dropout was added. At the outset, the dataset (training and testing) was got encoded on the label using one-hot encoding. This meant that the last layer in the feedforward layers consisted of 3 neurons. A softmax activation function was used in that layer, while ReLU was used in all hidden layers. Selecting hyperparameters for a PCNN is a complex process; the experiments simulations were conducted on the setting multiple values for hyper-parameters as following:

-β = (0.05, 0.1, 0.5, 0.9)
-αL = (0.1, 0.2, 0.5, 0.9, 1)
-αq = (0.1, 0.4, 0.8, 0.85, 0.86, 0.9)
-VL = (0.1, 0.3, 0.5, 0.9)
-VF = (0.1, 0.2, 0.5, 0.7, 1)
-Vθ = (1, 3, 5, 7, 10, 12)

After running grid search and evaluating the results over validation and testing sets, the optimal combinations of hyper-parameters are:

-β = 0.1
-αL = 0.9
-αq = 0.86
-VL = 0.3
-VF = 0.7
-Vθ = 10.

$$W = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

To perform the model performance evaluation, it was tested both on the test split from the merged dataset and the completely unseen dataset, the MIAS dataset.

To perform the model performance evaluation, it was tested both on the test split from the merged dataset and the completely unseen dataset, the MIAS dataset. Models were trained and tested using

stratified 10-fold cross-validation. Different measurements were collected, including mean accuracy, standard deviation, and training time for each TL model. Results on a merged dataset testing set are illustrated in Table 2. To verify the main idea of this research, the experiments were conducted twice: one does not consider segmentation before the model (the main idea) and the second conducts segmentation before the model.

Table 2. Merged dataset testing results.

Without Segmentation		
Model	Accuracy %	Standard Deviation
AlexNet	97.3	0.3
VGG-Net	97.7	0.32
Res-Net	97.92	0.52
Google-Net	98.9	0.76
With Segmentation		
AlexNet	97.24	0.38
VGG-Net	97.81	0.3
Res-Net	97.73	0.53
Google-Net	98.81	0.81

Table 2 illustrates that the model can operate without any pre-processing steps better and the standard deviation optimal value lays in VGG-Net and AlexNet.

When the proposed model was trained for each dataset separately, it achieved 98.72% accuracy for DDMS, 97.5% for INbreast, and 96.94% for BCDR.

These results were obtained when 3 hidden layers were used in the feedforward network and the sizes were 1000×1000×200. After the first experiment was conducted, the TL models were scored and tested on an unseen dataset, MIAS. Table 3 provides an overview of the results.

The parameters of a PCNN are difficult to optimize and the grid search operation is a very intensive computation process. Therefore, the values were set based on the best practices established in other studies.

Table 3. MIAS dataset results.

Without Segmentation		
Model	Accuracy %	Standard Deviation
AlexNet	97.1	0.26
VGG-Net	97	0.4
Res-Net	98.1	0.45
Google-Net	98.77	0.7
With Segmentation		
AlexNet	97.11	0.24
VGG-Net	97.06	0.38
Res-Net	97.92	0.55
Google-Net	98.76	0.74

One method for assessing the proposed breast cancer diagnosis performance involves calculating both the AUC and accuracy for the MIAS unseen dataset. In this test, the results were 0.99 and 98.77% respectively.

Meanwhile, an evaluation of the same dataset by Peng et al. [9] yielded 96% accuracy and 0.92 AUC. Also, Alxander et al [49] worked on Molecular predictions reach balanced accuracies up to 78%, whereas accuracies of over 95% can be achieved for subgroups of patients. Yu-Dong Zhang et al [50] ran BDR-CNN-GCN algorithm 10 times on the breast mini-MIAS dataset and achieved an accuracy of 96.1%.

The experimental results lead to two important insights:

- Using PCNN can neutralize the effect of image differences and can eliminate any preprocessing steps.
- GoogleNet outperforms other TL models as a pre-trained model due to its ability to generalize.

Based on the hyperparameter values, the overall system training time was measured between 20 and 30 hours, considering the grid search process to optimize model hyper-parameters.

Finally, the advantages and disadvantages can be summarized as:

Advantages:

- Model does not require any preprocessing or segmentation steps.
- Proved the generalization capability by testing on multiple datasets.
- The model proved success for not falling into overfitting problem.
- Superiority in results over previous research trials.

Disadvantages:

- Complex model contains multiple models.
- Hard to maintain in production environment.

The concept of "Uncertainty quantification" is a key factor in uncertainties reduction whether in optimization or decision making processes. UQ can be exploited to resolve a diverse set of practical applications hence, in will be addressed in focus in future while exploring well-established techniques [51,52]

6. Conclusions

In this research, different deep learning pre-trained models were exploited in combination with PCNNs to classify breast cancer candidate images. The model was trained and tested on different datasets, achieving 0.99 AUC and 98.77% accuracy. The implementation of the proposed model proves that using PCNNs can eliminate the segmentation preprocessing step, and the PCNN image signature presents an important discriminant feature to classify. The model was tested on a completely unseen dataset to ensure it does not fall to the overfitting problem due to limited dataset size. One of the major beneficial aspects of the proposed model is that it can be generalized to any computer vision application. PCNN and CNN merging can be used as complementary feature extractors for images and videos. In future work, new ideas to enhance the performance can be introduced. One of them is exploring other datasets and expand the merged dataset. Also, another idea is introduce other complex and deeper models. The last thing the authors will keep eyes on, is to ensure that the model avoids to fall in overfitting.

Acknowledgements

Conflict of interest

The authors declare that they have no conflicts of interest to disclose.

References

1. S. Negoita, E. Feuer, A. Mariotto, K. A Cronin, V. I Petkov, S. K Hussey, et al., Annual report to the nation on the status of cancer, part II: Recent changes in prostate cancer trends and disease characteristics, *Cancer*, **124** (2018), 2801–2814.

2. R. L. Siegel, K. D. Miller, A. Jemal, Cancer statistics, 2016, *CA. Cancer J. Clin.*, **66** (2016), 7–30.

3. R. L. Siegel, K. D. Miller, A. Jemal, Cancer statistics, 2017, *CA. Cancer J. Clin.*, **67** (2017), 7–30.

4. L. Berlin, Radiologic errors, past, present and future, *Diagnosis*, **1** (2014), 79–84.

5. Y. Guo, X. Shang, Z. Li, Identification of cancer subtypes by integrating multiple types of transcriptomics data with deep learning in breast cancer, *Neurocomputing*, **324**, (2019), 20–30.

6. B. Sahiner, A. Pezeshk, L. M. Hadjiiski, X. Wang, K. Drukker, K. H. Cha, et al., Deep learning in medical imaging and radiation therapy, *Med. Phys.*, **46** (2019), e1–e36.

7. H. M. Ahmad, M. J. Khan, A. Yousaf, S. Ghuffar, K. Khurshid, Deep learning: a breakthrough in medical imaging, *Curr. Med. Imaging*, **16**, (2020), 946–956.

8. G. Litjens, T. Kooi, B. Bejnordi, A. Setio, F. Ciompi, M. Ghafoorian, et al., A survey on deep learning in medical image analysis, *Med. Image Anal.*, **42** (2017), 60–88.

9. A. A. Mohamed, W. A. Berg, H. Peng, Y. Luo, R. C. Jankowitz, S. Wu, A deep learning method for classifying mammographic breast density categories, *Med. Phys.*, **45** (2018), 314–321.

10. F. F. Ting, Y. J. Tan, K. S. Sim, Convolutional neural network improvement for breast cancer classification, *Expert Syst. Appl.*, **120** (2019), 103–115.

11. D. M. Vo, N. -Q. Nguyen, S.-W. Lee, Classification of breast cancer histology images using incremental boosting convolution networks, *Inf. Sci. (Ny).*, **482** (2019), 123–138.

12. F. Ciompi, B. de Hoop, V. R Sarah, K. Chung, E. Scholten, M. Oudkerk, et al., Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box, *Med. Image Anal.*, **26** (2015), 195–202.

13. H. Shin, H. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, et al., Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, *IEEE Trans. Med. Imaging*, **35** (2016), 1285–1298.

14. A. Krizhevsky, I. Sutskever, G. E. Hinton, Machine learning and computer vision group deep learning with tensorflow, 2012.

15. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in *Computer Vision and Pattern Recognition*, 2014.

16. Z. Cömert, A. F. Kocamaz, Fetal hypoxia detection based on deep convolutional neural network with transfer learning approach, *Comput. Sci. On-line Conf.*, (2018), 239–248.

17. M. Oquab, L. Bottou, I. Laptev, J. Sivic, Learning and transferring mid-level image representations using convolutional neural networks, in *Conference on Computer Vision and Pattern Recognition*, (2014), 1717–1724.

18. M. S. Abdel-Wahab, M. Aboul-Ela, A. Samir, Arabic sign language recognition using neural network and graph matching techniques, *Proceed. Int. Conf. App. Inf. Commun.*, (2006), 163–168.

19. A. SamirElons, M. Abull-ela, M. F. Tolba, Pulse-coupled neural network feature generation model for Arabic sign language recognition, *IET Image Process.*, **7** (2013), 829–836.

20. M. F. Akay, Support vector machines combined with feature selection for breast cancer diagnosis, *Expert Syst. Appl.*, **36** (2009), 3240–3247.

21. M. Karabatakm, M. C. Ince, An expert system for detection of breast cancer based on association rules and neural network, *Expert Syst. Appl.*, **36** (2009), 3465–3469.

22. M. B. Rodrigues, R. D. NóBrega, S. Alves, P. Fillho, J. Durate, A. Sangaiah, et al., Health of things algorithms for malignancy level classification of lung nodules, *IEEE Access*, **6** (2012), 18592–18601.

23. N. Arunkumar, M. A. Mohammed, S. A. Mostafa, D. A. Ibrahim, J. Rodrigues, V. H. C. de Albuquerque, Fully automatic model-based segmentation and classification approach for MRI brain tumor using artificial neural networks, *Concurr. Comput. Pract. Expert*, **32** (2020), e4962.

24. C. M. J. M. Dourado, S. P. P. D. Silva, R. V. M. D. Nóbrega, P. P. R. Filho, K. Muhammad, V. H. C. D. Albuquerque, An open IoHT-based deep learning framework for online medical image recognition, *IEEE J. Sel. Areas Commun.*, **39** (2020), 541–548.

25. K. Muhammad, S. Khan, J. D. Ser, V. H. C. de Albuquerque, Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey, *IEEE Trans. Neural Networks Learn. Syst.*, (2020), 1–16.

26. C. A. Pena-Reyes, M. Sipper, A fuzzy-genetic approach to breast cancer diagnosis, *Artif. Intell. Med.*, **17** (1999), 131–155.

27. B. N. Dontchos, A. Yala, R. Barzilay, J. Xiang, C. D. Lehman, External validation of a deep learning model for predicting mammographic breast density in routine clinical practice, *Acad. Radiol.*, **28** (2020), 475–480.

28. T. D. Truong, H. T. -T. Pham, Breast cancer histopathological image classification utilizing convolutional neural network," in *International Joint Conference on Neural Networks (IJCNN)*, **69** (2019), 531.

29. M. Abdar, M. Zomorodi-Moghadam, X. Zhou, R. Gururajan, X. Tao, P. D. Barua, et al., A new nested ensemble technique for automated diagnosis of breast cancer, *Pattern Recognit. Lett.*, **132** (2020), 123–131.

30. P. Ferroni, F. M. Zanzotto, S. Riondino, N. Scarpato, F. Guadagni, M. Roselli, Breast cancer prognosis using a machine learning approach, *Cancers (Basel)*, **11** (2019), 328.

31. D. A. Omondiagbe, S. Veeramani, A. S. Sidhu, Machine learning classification techniques for breast cancer diagnosis, in *IOP Conference Series: Materials Science and Engineering.*, **495** (2019), 12033.

32. Y Tseng, C. Huang, C. Wen, P. Lai, M. Wu, Y. Sun, et al., Predicting breast cancer metastasis by using serum biomarkers and clinicopathological data with machine learning technologies, *Int. J. Med. Inform.*, **128** (2019), 79–86.

33. M. Abdar, V. Makarenkov, CWV-BANN-SVM ensemble learning classifier for an accurate diagnosis of breast cancer, *Measurement*, **146** (2019), 557–570.

34. E. H. Houssein, M. M. Emam, A. A. Ali, P. N. Suganthan, Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review, *Expert Syst. Appl.*, (2020), 114161.

35. J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, *Adv. Neural Inf. Process. Syst.*, **4** (2014), 3320–3328.

36. K. Bowyer, D. Kopans, W. Kegelmeyer, R. Moore, M. Sallam, K. Chang, et al., The digital database for screening mammography, in *Proceedings of the Fifth International Workshop on Digital Mammography*, **58** (1996), 27.

37. I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, J. S. Cardoso, INbreast: toward a full-field digital mammographic database, *Acad. Radiol.*, **19** (2012), 236–248.

38. J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, et al., Mammographic Image Analysis Society (MIAS) Database v1. 21[Dataset], 2015.

39. Y. LeCun, The MNIST Database of Handwritten Digits, 1998.

40. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick , et al., Caffe: Convolutional architecture for fast feature embedding, in *Proceedings of the 22nd ACM international conference on Multimedia*, (2014), 675–678.

41. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., ImageNet large scale visual recognition challenge, *Int. J. Comput. Vis.*, **115** (2015), 211–252.

42. B. Q. Huynh, H. Li, M. L. Giger, Digital mammographic tumor classification using transfer learning from deep convolutional neural networks, *J. Med. Imaging*, **3** (2016), 034501.

43. J. D. Gallego-Posada, D. A. Montoya-Zapata, O. L. Quintero-Montoya, D. A. Montoya-Zapa, Detection and diagnosis of breast tumors using deep convolutional neural networks, in *Conference Proceedings of XVII Latin American Conference in Automatic Control*, (2016), 17.

44. P. Xi, C. Shu, R. Goubran, Abnormality detection in mammography using deep convolutional neural networks, *Comput. Sci.*, 2018.

45. D. Lévy, A. Jain, Breast mass classification from mammograms using deep convolutional neural networks, *Comput. Sci.*, 2016.

46. H. S. Ranganath, G. Kuntimad, J. L. Johnson, Pulse coupled neural networks for image processing, in *Proceedings IEEE Southeastcon 95 Visualize the Future*, (1995), 37–43.

47. H. Jia, X. Peng, L. Kang, Y. Li, Z. Jiang, K. Sun, Pulse coupled neural network based on Harris hawks optimization algorithm for image segmentation, *Multimed. Tools Appl.*, **79** (2020), 28369–28392.

48. T. Kalaiselvi, K. Rahimunnisa, A. S. Begum, Department of electronics and instrumentation engineering, easwari, *Pulse*, **29** (2020), 4411–4415.

49. A. Binder, M. Bockmayr, M. Hägele, S. Wienert, D. Heim, K. Hellweg, et al., Morphological and molecular breast cancer profiling through explainable machine learning, *Nat. Mach. Intell.*, **3** (2021), 355–366.

50. Y. -D. Zhang, S. C. Satapathy, D. S. Guttery, J. M. Górriz, S. -H. Wang, Improved breast cancer classification through combining graph convolutional network and convolutional neural network, *Inf. Process. Manag.*, **58** (2021), 102439.

51. M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, et al., A review of uncertainty quantification in deep learning: techniques, applications and challenges, *Inf. Fus.*, 2021.

52. R. Alizadehsani, M. Roshanzamir, S. Hussain, A. Khosravi, A. Koohestani, M. H Zangooei, et al., Handling of uncertainty in medical data using machine learning and probability theory techniques: a review of 30 years (1991–2020), *Ann. Oper. Res.*, (2021), 1–42.