The Study of Noise Effect on CNN-Based Deep Learning from Medical Images

Kittipat Sriwong, Kittisak Kerdprasop, and Nittaya Kerdprasop

Abstract—**Currently, computational modeling methods based on machine learning techniques in medical imaging are gaining more and more interests from health science researchers and practitioners. The high interest is due to efficiency of modern algorithms such as convolutional neural networks (CNN) and other types of deep learning. CNN is the most popular deep learning algorithm because of its prominent capability on learning key features from images that help capturing the correct class of images. Moreover, several sophisticated CNN architectures with many learning layers are available in the cloud computing environment. In this study, we are interested in performing empirical research work to compare performance of CNNs when they are dealing with noisy medical images. We design a comparative study to observe performance of the AlexNet CNN model on classifying diseases from medical images of two types: images with noise and images without noise. For the case of noisy images, the data had been further separated into two groups: a group of images that noises harmoniously cover the area of the disease symptoms (NIH) and a group of images that noises do not harmoniously cover the area of the disease symptoms (NNIH). The experimental results reveal that NNIH has insignificant effect toward the performance of CNN. For the group of NIH, we notice some effect of noise on CNN learning performance. In NIH group of images, the data preparation process before learning can improve the efficiency of CNN.**

Index Terms—**Deep learning, convolution neural network, AlexNet, medical images, noise.**

I. INTRODUCTION

Machine learning (ML) is an automatic learning process in which computers seek from data some key patterns of features (called a model) that can be used for classifying data into a proper category, predicting values on some specific features, forming a group of data based on the key features, and many other kinds of learning task. Researchers in the ML community have introduced hundreds of algorithms that are efficient and effective on doing some specific learning tasks. Generally, these algorithms can be categorized into three major types based on their learning assumption. These three main learning types are supervised learning (i.e., the algorithms are guided by the labeled target field), unsupervised learning (i.e., the guiding field is unavailable), and reinforcement learning (i.e., the algorithm learns to achieve some task through the reward/penalty system).

There is no single ML algorithm that is good in all tasks. Instead, each algorithm is invented to be good at some specific type of data or works well in some environment. The success of applying ML thus depends on the type and quality of data as well as the selection of appropriate algorithms to be used with such specific data. In this research, we focus on the image medical data with the assumption that these images are not perfect in the sense that some of them may contain noise. Such noisy images can occur in a daily practice of medical imaging.

Nowadays, the development of computer models that use machine learning for medical diagnosis is gaining popularity, in particular, the use of deep learning (DL). For example, DL has been used for creating a disease classification model from medical images. From the study of related work, we found that some researchers applied a series of image processing for preparing proper pictures before learning, while some researchers ignored the image preprocessing. For instance, the work of Elhoseny & Shankar [1] applied Bilateral Filter (BF) for filtering, then used Dragonfly (DF) and Modified Firefly (MFF) algorithm to optimize the control parameters in medical imaging (MI) denoising process. After extracting the denoised images, they used it as training data to create classification model with CNN, Support vector machine (SVM), Neural Network (NN) and Navies Bayes (NB). The results show that CNN is the best method as it can correctly classify the denoised image as either normal or abnormal with a high classification rate. The work of Liu *et al.* [2] proposed a Genetic algorithm (GA)-based method to construct CNN structures, named EvoNets, for medical image denoising. The results from comparing with state-of-the-art deep learning methods in medical image denoising with BM3D [3] and DnCNN [4] show that EvoNets outperform others consistently at various noise levels. Other research team [5] also proposed a magnetic resonance imaging (MRI) denoising method by applying the residual encoder–decoder Wasserstein generative adversarial network (RED-WGAN). They use both clinical and simulated datasets to compare with three methods: CNN3D (RED-WGAN with only the generator part and the MSE loss), BM4D [6] and PRI-NLM3D [7] to validate the performance of their proposed RED-WGAN. The results show that the RED-WGAN achieves superior performance as compared to other several state-of-the-art methods in both two types of data. Their method shows powerful abilities in both noise suppression and structure preservation.

In this work, we study CNN modeling for disease

Kittipat Sriwong, Kittisak Kerdprasop, and Nittaya Kerdprasop are with the School of Computer Engineering, SUT, 111 University Avenue, Muang, Nakhon Ratchasima 30000, Thailand (e-mail: kittipat.re@gmail.com, kerdpras@sut.ac.th, nittaya@sut.ac.th).

classification from medical images with the main focus of observing performance of CNN when noises appear in the images. We thus prepare two kinds of image data, that are, images with noise (images with adding noise) and images without noise (original image). This is to test the hypothesis that whether image preprocessing (to remove noise prior to the learning process) is necessary or not for the CNN algorithm. The CNN architecture used in this work is AlexNet.

II. PRELIMINARY AND LITERATURE REVIEW

A. Different Types of Noise

Noise is something that is created in the image. During image acquisition or image transmission such as sending images through channels that are scrambled. Noise affects the quality of images differently depending on the type of interference. Noise can be classified as follows:

• Impulse Noise (Salt and Pepper Noise)

This type of noise results in black and white dots in the image [8], so it is called Salt and Pepper Noise. This noise occurs in the image due to the sudden changes of image signal. The white color is caused by changing the color value of the image pixel to the highest value, and black is caused by changing the color value of the image pixel to the lowest value.

• Gaussian Noise (Amplifier Noise)

This type of noise can be found in nature [9]. It is characterized following Gaussian distribution. This means that each pixel in image is caused by the sum between the true pixel point value and the random value of the noise with the Gaussian distribution.

• Poisson Noise (Photon Noise)

This type of noise occurs when the number of photons that the sensor senses is not sufficient to provide detectable statistical information [9]. It is characterized following Poisson distribution.

• Speckle Noise

This type of noise is caused by the random values multiplications with pixel values of the image. It occurs in interconnected imaging systems such as synthetic aperture radar (SAR) and medical ultrasonic images. This noise reduces the quality of active radar and SAR images [9].

B. Deep Learning Method

DL method is based on the artificial neural network (ANN) concept. DL uses many processing layers more than ANN. Therefore, DL takes longer time of training process than ANN. The time-consuming tradeoff is that the accuracy of DL is normally higher than ANN.

For example, Zhang *et al.* [10] present the application of DL to build the land cover classification model. They report in their research results that Joint Deep Learning Land Cover (JDL-LC) model has an overall accuracy as high as 89.64 % and 90.72 %. These accuracy rates are higher than the ANN method that can classify land cover types over Southampton and Manchester areas in the U.K. with 81.29 % and 82.22 %, respectively.

In general, DL method can be classified into four major types based on the network architecture [11]. These DL types are Unsupervised Pretrained Networks (UPNs), Recurrent Neural Networks, Recursive Neural Networks, and Convolutional Neural Networks (CNNs). In this paper, we focus in CNNs because they are state-of-the-art DL that are appropriate for learning from image data.

C. Convolutional Neural Networks (CNNs)

CNNs are one of most famous DL method [12]. The main goal of CNNs is to learn image patterns for the purpose of recognition and classification such as recognition the symbols on the street, face recognition, and many other object recognition tasks. At present, there are many public pre-trained CNN models available for adopting to a specific tasks. These most popular models include AlexNet [12], GoogLeNet [13], VGGNet [14], and ResNet [15]

Fig. 1. AlexNet architecture.

We used the pre-trained CNN model of AlexNet in this study. AlexNet has an architectural as demonstrated in Fig. 1. There are five main layers in AlexNet architecture.

1) *Input Layer*. This layer is for inputting image data of size $227 \times 227 \times 3$.

2) *Convolutional Layer*. This layer is used for extracting important features from images.

3) *Pooling Layer*. This layer obtains data from the previous convolutional layer. The main advantage of this layer is to reduce the spatial dimension (width and height) of the input data that will sent forward to the next layer (may be convolutional layer or fully-connected layer).

4) *Fully-Connected Layer*. The nodes in this layer are fully connected to the output from the previous layer (convolutional layer or fully-connected layer).

5) *Output Layer*. This layer receives data from the previous layer for making classification decision on an image.

The ReLu and Softmax in Fig. 1 are activation function.

We employ CNN architecture of AlexNet with transfer learning. The transfer learning method is the parameter learning of network for a new problem. With transfer learning, ones can apply the pre-train CNN to quickly learn suitable parameters for a new set of images [16]. The advantages of this method are two folds. First, this method spends shorter time in training as compared to the train-from-scratch method that users have to start the learning process from identifying

number of layers, assigning parameter weights, and iteratively learning from sample images to search for the optimal set of parameters. Second, this method requires number of images for training significantly less than the train-from-scratch method. However, this method has some limitation, that is, size of images (pixel x pixel) depending on the specification identified by the pre-trained model. Therefore, we are require to resize the images before staring the modeling process.

D. Literature Review

There are many research works in the literature that have applied ML in the development of models for disease classification by using patient data for training. The data may be in the form of texts, numbers, images, and others. Amoroso *et al.* [17] use MRI images for the classification of Parkinson's disease. They use random forest algorithm in the feature extraction steps and then apply support vector machine in the classification steps. They report classification result at the accuracy of $93 \pm 4\%$.

Nowadays, researchers who focus their interest on disease classification from medical images have increasingly applied a more sophisticated technique such as DL. For example, Ting *et al.* [18] classify breast cancer from mammography images that are obtained from the mammographic image analysis society [19] using the CNN model to yield classification accuracy as high as 90.50 %.

Biswas *et al.* [20] classify Fatty Liver Disease (FLD) from ultrasound images using three algorithms: support vector machine, extreme learning machine, and CNN. They report the model accuracy as 82.08 %, 92.22 %, and 100 %, respectively. It is noticeable from the work of Biswas *et al.* [20] that among the three learning algorithms, CNN can achieve the best performance with significant accuracy rate at 100%.

Based on the success of CNN as found in the literature, we thus decide to further study performance of CNN in a different aspect from other researchers in that we focus on performance evaluation of CNN regarding the quality of images used in the training process of CNN. Our research is a comparison of the effectiveness of the CNN medical image classification with noise and no noise in training data for creating models.

III. MODEL CREATION METHOD

In this research, we design the process for CNN modeling from medical images in two groups: images with noise and images without noise. We employ CNN architecture of AlexNet with transfer learning scenario as the starting point. Images without noise are original data. Images with noise are those that we intentionally add noise at different levels. We create four schemes of modeling as shown in Fig. 2.

A. Data

This study uses two sets of data: Chest X-Ray Images (Pneumonia) and Retinal OCT Images (optical coherence tomography) [21]. We have added speckle noise with 3 different levels of noise to the original image. The result of the adding speckle noise is that the data are separated into two groups: images with noises harmoniously cover the area of the disease symptoms (NIH) and images with noises not harmoniously cover the area of the disease symptoms (NNIH).

Each group of images (i.e., NIH and NNIH) has 4 datasets of images according to the noise level, that is, level 0, 1, 2 and 3. Level 0 means images without noise (or the original images), whereas level 3 is images with the maximum level of noises added. Detail of noise level according to standard deviation (SD) is as follows:

- Level 0: No noise added
- Level 1: Adding speckle noise with 1×SD
- Level 2: Adding speckle noise with 2×SD
- Level 3: Adding speckle noise with 3×SD

Image data in the NIH group are chest X-ray images [21] containing two classes of patients: pneumonia and normal. Sample pictures of each class at different level of noises are shown in Fig. 3. Statistics summary of this dataset is presented in Table I. In total, a dataset has 5,856 images. Most images are in a class of pneumonia, which consists of 4,273 images.
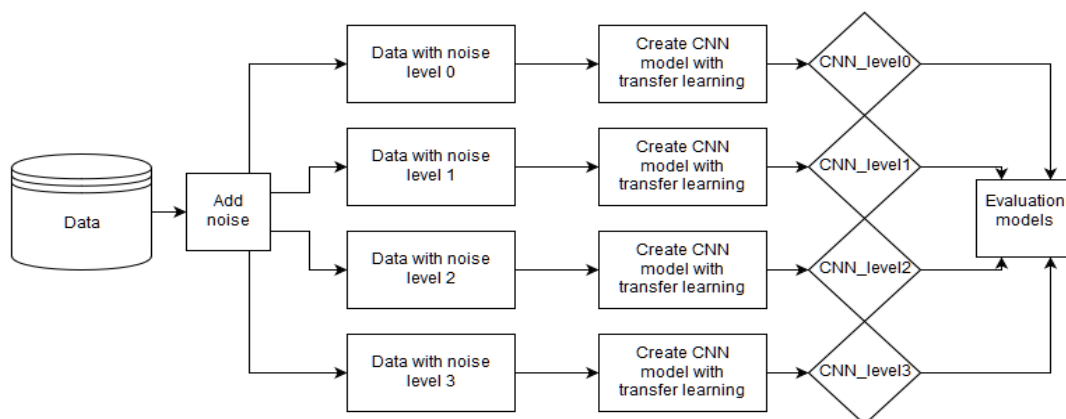

Fig. 2. The research framework.

The image data in NNIH group are Retinal OCT Images (optical coherence tomography) [21]. There are four classes in this dataset: CNV disease, DME disease, Drusen disease, and normal case. Sample pictures from each class are shown in Fig. 4. Statistics summary is presented in Table II. In total, a dataset contains 84,484 images. Most images are in a class of choroidal neovascularization (CNV).

B. Image Classification Modeling

In this study, the data set is divided into 3 subsets as train

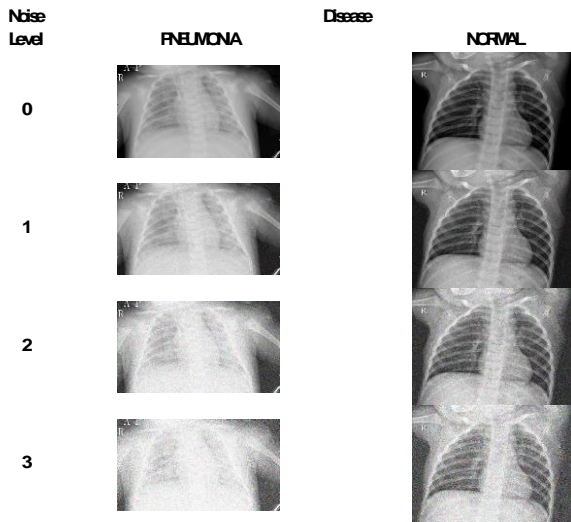data, validation data, and test data. The details of each set are summarized in Table III.

Fig. 3. Sample pictures from the two classes of chest X-ray images with different level of noises added.

TABLE I: STATISTICS OF CHEST X-RAY IMAGES DATASET

Disease	Number
PNEUMONIA	4,273
NORMAL	1,583
Summation	5,856

TABLE II: SUMMARY OF RETINAL OCT IMAGES DATASET

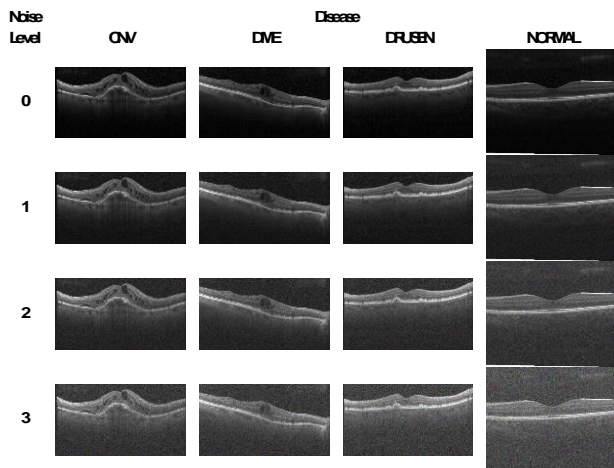| Disease	Number
Choroidal neovascularization (CNV)	37,455
Diabetic macular edema (DME)	11,598
DRUSEN	8,866
NORMAL	26,565
Summation	84,484

Fig. 4. Sample pictures of data from the four classes of retinal OCT images with different level of noises added.

TABLE III: THE NUMBER OF IMAGES IN EACH DATA SUBSET

Data set	Train	Validation	Test
Chest X-Ray Images (Pneumonia)	5,216	16	624
Retinal OCT Images (optical coherence tomography)	83,484	32	968

On creating the CNNs models, we use the available pre-trained CNN architecture named Alexnet [12] with the transfer learning scheme. For each dataset, we create four models according to the four different degrees of noises added

to the images (level 0, 1, 2, and 3 in increasing amount of noises). With the two datasets (retinal OCT images and chest X-ray images) and four levels of noises, there are totally 8 models in our experiments. The parameter setting in each modeling experiment is the same in every experiment. This is for controlling the experimentation environment. We assign name for each model as shown in Table IV.

TABLE IV: NAME OF THE MODELS FROM EACH DATASET

Data set	Chest X-Ray Images (Pneumonia)	Retinal OCT Images (optical coherence tomography)
Model using images with noise level 0 (no noise added)	Pneumonia_CNN_level0	OCT_CNN_level0
Model using images with noise level 1	Pneumonia_CNN_level1	OCT_CNN_level1
Model using images with noise level 2	Pneumonia_CNN_level2	OCT_CNN_level2
Model using images with noise level 3	Pneumonia_CNN_level3	OCT_CNN_level3

IV. NOISE EFFECT EVALUATION RESULTS

We used overall accuracy (i.e., the average of accuracy from all classes of the dataset) as a metric for evaluating the performance of the models to assess the correctness of disease classification. The modeling experimentation is observed based on the noise distribution strategies namely NIH and NNIH.

Pneumonia_CNN models are the CNN models built from noise distribution in the NIH group, in which noises harmoniously cover the area of disease symptoms. Model performance when noises are added increasingly from level 0 up to 3 are shown in Table V. It can be noticed from the accuracy metric that the more noises added to the lung images, the higher the classification accuracy. This phenomenon is due to the fact that the speckle noises we intentionally add to the images are random values that multiply with pixel values of the image. Such speckle noises and the distribution form of NIH help CNN models differentiate easier the pneumonia case images from images in the normal cases.

Fig. 5 depicts the change in accuracy of the CNN models as more NIH noises have been added. The percentage of change is computed by comparing against the model performance when no noise has been added to the images, that is, level 0 of noise. The changes in accuracy are around one percent at noise levels 1 and 2. The increase in accuracy is more than three percents at noise level 3.

TABLE V: MODEL PERFORMANCES IN THE NIH GROUP OF NOISES

Model	Accuracy
Pneumonia_CNN_level0	79.81 %
Pneumonia_CNN_level1	80.61 %
Pneumonia_CNN_level2	80.45 %
Pneumonia_CNN_level3	82.53 %

For the NNIH group of images that noises do not harmoniously cover the area of disease symptom, when the

training images used for creating the CNN model have more noises, the classification accuracy of the models decrease. The accuracy performances are summarized in Table VI. From the table, we can notice that the accuracy of models trained by images in the NNIH group decreases with the pattern according to equation 1.

$$\text{Accuracy} = \text{Accuracy}[0] - (N*2.3916 - 0.0647) \qquad (1)$$

Accuracy [0] means accuracy of the CNN model level 0 in which no noise has been added, and N is the noise level. The percentages of accuracy decrease at each noise level are also graphically displayed in Fig. 6.
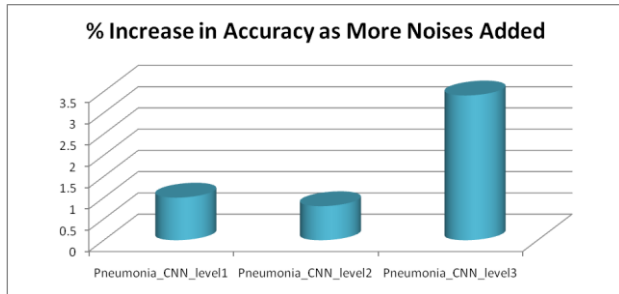

Fig. 5. Accuracy changes at each level of noises in the NIH group of images using noise level 0 as a base case of comparison.

TABLE VI: Model Performances in the NNIH Group of Noises

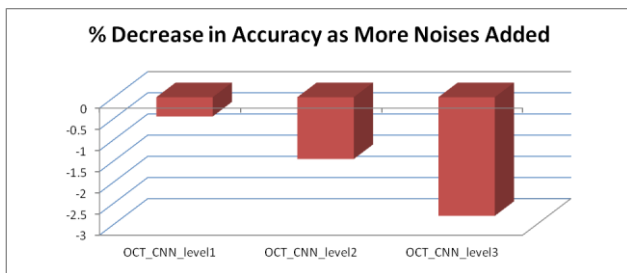| Model	Accuracy
OCT_CNN_level0	92.36 %
OCT_CNN_level1	91.94 %
OCT_CNN_level2	91.01 %
OCT_CNN_level3	89.77 %

Fig. 6. Accuracy changes at each level of noises in the NNIH group of images when compared to learning from images without noise.

V. Conclusion

Convolution neural network, or CNN, is a deep learning algorithm that has been widely accepted as the most accurate algorithm suitable for learning patterns from images. Many research laboratories and renowned organization such as Google provide the pre-trained CNN models for further deployment. In this research work, we adopt the AlexNet architecture to build the CNN models.

The focus of our research is to observe the performance of CNN models when the trained images contain noises at various levels. Medical images are the scope of our observation. We categorized medical noisy images into two groups, namely NIH and NNIH groups. The NIH group is a group of medical images that noises have been added to cover harmoniously the area of the disease symptoms, whereas the NNIH group is the group of images that noises do not

harmoniously cover the area of the disease symptoms.

The experimental results show that noises in the NIH group have positive effect to the CNN modeling process. But noises in the NNIH group show negative effect. We therefore suggest that for the case of NNIH noises, the preprocessing steps to reduce noises prior to CNN modeling are necessary for the learning performance improvement. In the future, we will try to develop model for medical image denoising for the case of NNIH with different noise levels.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The first author is responsible for designing the research framework, organizing the experimentation steps and preparing the manuscript. The second author helps confirming the experimental results and discussing the future research direction. The last author contributes editing the manuscript and validating the research steps.

References

[1] M. Elhoseny and K. Shankar, "Optimal bilateral filter and convolutional neural network based denoising method of medical image measurements," *Measurement*, vol. 143, pp. 125-135, September 2019.

[2] P. Liu, M. D. El Basha, Y. Li, Y. Xiao, P. C. Sanelli, and R. Fang, "Deep evolutionary networks with expedited genetic algorithms for medical image denoising," *Medical Image Analysis*, vol. 54, pp. 306-315, May 2019.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "BM3D image denoising with shape-adaptive principal component analysis," *Signal Processing with Adaptive Sparse Structured Representations*, April 2009.

[4] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142-3155, July 2017.

[5] M. Ran, J. Hu, Y. Chen, H. Chen, H. Sun, J. Zhou, and Y. Zhang, "Denoising of 3D magnetic resonance images using a residual encoder–decoder Wasserstein generative adversarial network," *Medical Image Analysis*, vol. 55, pp. 165-180, July 2019.

[6] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 119-133, January 2013.

[7] J. V. Manjón, P. Coupé, A. Buades, D. L. Collins, and M. Robles, "New methods for MRI denoising based on sparseness and self-similarity," *Medical Image Analysis*, vol. 16, no. 1, pp. 18-27, January 2012.

[8] C. Boncelet and A. C. Bovik, "Image noise models," *Handbook of Image and Video Processing*, pp. 325-335, 2005.

[9] P. Patidar, M. Gupta, S. Srivastava, and A. K. Nagawat, "Image de-noising by various filters for different noise," *International Journal of Computer Applications*, vol. 9, no. 4, pp. 45-50, November 2010.

[10] C. Zhang, I. Sargent, X. Pan, H. Li, A. Gardiner, J. Hare, and P. M. Atkinson, "Joint deep learning for land cover and land use classification," *Remote Sensing of Environment*, vol. 221, pp. 173-187, February 2019.

[11] A. Gibson, J. Patterson, *Deep Learning*, O'Reilly Media, Inc., 2017, ch. 4.

[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, November 1998.

[13] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1-9.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv: 1409.1556, April 2015.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[16] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, October 2010.

[17] N. Amoroso, M. La Rocca, A. Monaco, R. Bellotti, and S. Tangaro, "Complex networks reveal early MRI markers of Parkinson's disease," *Medical Image Analysis*, vol. 48, pp. 12-24, August 2018.

[18] F. F. Ting, Y. J. Tan, and K. S. Sim, "Convolutional neural network improvement for breast cancer classification," *Expert Systems with Applications*, vol. 120, pp. 103-115, April 2019.

[19] Mammographic Image Analysis Homepage. (2018). Mammographic Image Analysis Society (MIAS). [Online]. Available: http://www.mammoimage.org/databases/

[20] M. Biswas *et al.*, "Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm," *Computer Methods and Programs in Biomedicine*, vol. 155, pp. 165-177, March 2018.

[21] D. K. Kermany and M. Goldbaum, "Labeled optical coherence tomography (OCT) and Chest X-Ray images for classification," *Mendeley Data*, vol. 2, January 2018.

Kittipat Sriwong is currently a master student with the School of Computer Engineering, Institute of Engineering, Suranaree University of Technology, Thailand. He has been fully financial support by grant from Suranaree University of Technology throughout his bachelor and master study. He is a research assistant in the Data and Knowledge Engineering Research Unit, SUT. His current research of interest includes data mining, support vector machine and deep learning techniques, statistical data mining, and data mining applications in medical science.

Kittisak Kerdprasop is an associate professor and chair of the Computer Engineering School, Suranaree University of Technology, Thailand. He received his B.Ed. degree in mathematics from Srinakarinwirot University, Thailand, in 1986, M.S. in computer science from the Prince of Songkla University, Thailand, in 1991 and Ph.D. in computer science from Nova Southeastern University, U.S.A., in 1999. His current research includes machine learning, statistical data modeling, and artificial intelligence.

Nittaya Kerdprasop is an associate professor and head of the Data and Knowledge Engineering Research Unit, School of Computer Engineering, Suranaree University of Technology, Thailand. She received her B.S. degree in radiation techniques from Mahidol University, Thailand, in 1985, M.S. in computer science from the Prince of Songkla University, Thailand, in 1991 and Ph.D. in computer science from Nova Southeastern University, U.S.A., in 1999. Her current research includes data mining, business intelligence, computational modeling in health and environmental sciences.