Engineering Journal issn: 0125-8281

Article

Diabetic Retinopathy Classification: Performance Evaluation of Pre-trained Lightweight CNN using Imbalance Dataset

Pranajit Kumar Das[1,2,a] **and Suree Pumrin**[1,b,*]

1 Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Thailand
2 Department of Computer Science and Engineering, Sylhet Agricultural University, Bangladesh
E-mail: [a]pkd@sau.ac.bd, [b,*]suree.p@chula.ac.th (Corresponding author)

Abstract. Diabetic Retinopathy (DR) is an eye complication that arises from long-term diabetes and damages the retinal blood vessels. Various clinical studies claim that Diabetic retinopathy infects about eighty percent of patients who suffer from diabetes type 1 for the last 15 years and a hundred percent of patients with this disease for 20 years. The human evaluation method is challenging but useful because it can detect diseases by the presence of lesions associated with Diabetic Retinopathy in most cases, but it is also time-consuming, erroneous, and requires a sophisticated medical setup. An efficient and automatic Diabetic Retinopathy identification method is still a challenging task. The feature extraction part is a very significant part and plays a vital role in the automatic Diabetic Retinopathy identification system. CNN has demonstrated its efficiency in medical image classification tasks as compared to other neural networks and traditional image processing methods. In this study, two lightweight CNN models: MobileNet and MobileNetV2 are used via transfer learning for binary (2-class) and multiclass (5-class) Diabetic Retinopathy classification using the DDR dataset, which is highly imbalanced. The efficiency of the models is measured using accuracy, precision, recall, and F1-score values. The ROC curve is generated for both models in binary and multiclass classification. The MobileNet model performed slightly better than MobilenetV2 in Diabetic Retinopathy classification for binary and multiclass classification. MobileNet shows 80% and 71% accuracy whereas MobileNetV2 shows 79% and 69% in binary and multiclass classification, respectively.

Keywords: Diabetic Retinopathy, Lightweight CNN, MobileNet, MobileNetV2, Transfer Learning, Fundus Images.

ENGINEERING JOURNAL Volume 28 Issue 7
Received 8 August 2023
Accepted 17 July 2024
Published 31 July 2024
Online at https://engj.org/
DOI:10.4186/ej.2024.28.7.13

DOI:10.4186/ej.2024.28.7.13

1. Introduction

Diabetes is one of the most common disease that affects people all over the world, from children to old people, regardless of age. Diabetes means the human body has persistently high glucose levels in the bloodstream. Diabetes has detrimental effects on human health as it contributes to various cardiovascular diseases, eye disease, neuropathy, nephrology, kidney, and many other diseases [1]. Nowadays, Diabetic Retinopathy is a vision-threatening retina disease that is very common in most people who have suffered from diabetes for a long time [2]. According to the announcement of the World Health Organization (2021), 537 million adults have diabetes whose age range is 20 to 79 years and most of them are living in lower-income countries (the ratio is 3 out of 4). Alarmingly, the number of people with diabetes is projected to hit 643 million by 2030 and 783 million by 2045 [3]. The global DR prevalence is increasing day by day and the situation is expected to worsen[4]. Diabetic Retinopathy is a retina disease that damages blood vessels and their walls internally. These blood vessels can become swollen and leaky, cutting off blood flow in the long run [5]. At the beginning of diabetes, Diabetic Retinopathy does not show any physical symptoms but worsens over time and can eventually destroy the vision. So, Early diagnosis of Diabetic Retinopathy can prevent the vision of most DR patients as treatment may cure or sometimes slow or stop the damage of the retina. The manual DR screening process has some limitations, for example, medical instrumental setup, insufficient medical centers with expert ophthalmologists, skilled technologists, and misdiagnoses, that obstruct the timely DR diagnosis process. Moreover, the manual process is time-intensive and error-prone. On contrary, the automatic DR diagnosis process addressed and resolved most of the problems over the manual process with the advantages of high accuracy, low expenditure, and minimum dependencies on expert oculists and clinicians. So, The automatic process for DR classification and identification is cost and time effective and more accurate than the manual process [6]. Although fundus photography is technically challenging and expensive, retinal fundus images are popular among researchers in assessing Diabetic Retinopathy disease and other ophthalmic abnormalities [7] [2] . Primarily the DR severity classes are graded into five categories: No DR, Mild Non-Proliferative DR (Mild NPDR), Moderate Non-Proliferative DR (Moderate NPDR), Severe Non-Proliferative DR (severe NPDR), and Proliferative DR (PDR) [8]. But in a broad sense, the major stages in DR are Non-Proliferative DR (NPDR) and proliferative DR (PDR). The 4-2-1 rule is used in severe DR diagnosis in which IRMA (Intra Retinal Microvascular Abnormalities) plays a vital role. In the DR severity classification using fundus images, the severity stages are defined by various symptoms observed in the fundus images. The DR severity stages associated with rules and symptoms are listed in Table 1.

The considerable symptoms that are significant in NPDR grading are (i) Microaneurysms (MA's) (ii) Hemorrhages (HM), and (iii) Soft and Hard Exudates. The advanced stages with (i) Neovascularization, and (ii) Vitreous or Preretinal Hemorrhage are treated as Proliferative DR [8]. Microaneurysms (MA's) are the first ever diagnosable symptoms of DR that look like red-colored small dots with uneven or irregular shapes. The size of MA's is less than 125 μm or 1/12 of the optic disc diameter. Hemorrhages (HM) are also red-colored spots whose size is greater than 125 μm but symmetric in shape and occur because of leakage in the retina's vessel wall. Soft Exudates (Cotton Wools) can be seen as round or oval-shaped white or pale yellowish white or grayish-white color spots because of injury in the nerve fiber layer. Hard Exudates look like white or yellowish white-colored bright spots and occurred due to plasma leakage. They usually appear at the outer periphery of the retina [9] [10]. The characteristics of the normal retina and retina with those DR symptoms are displayed in Figure 1. Figure 1(a) illustrates the normal retina with no DR sign, whereas Figure 1(b) depicts the DR retina with several DR symptoms of varying severity levels.

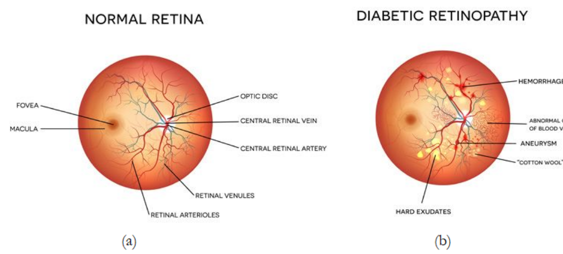

Fig. 1. Normal and DR diseased retina with significant symptoms [11].

In Deep Learning, CNN is a type of supervised machine-learning algorithm that requires a sufficient amount of data to train the system and the most prominent feature of CNN is learning features automatically. Because of the recent advancements in CNN in medical image analysis, a lot of research work has been carried out using the CNN method. Among deep learning algorithms, CNN has proved its domination in image classification, detection, segmentation, registration, and so on [12] [13]. Currently, CNN

Table 1. Diabetic Retinopathy symptoms accoding to severity stages and class level [8].

Class	Severity Stage	DR Symptoms
0	No DR	No Abnormality
1	Mild NPDR	Microaneurisms only
2	Moderate NPDR	More than Mild NPDR but less than Severe NPDR
3	Severe NPDR	Any of the following -Internal Hemorrhage in every 4 quadrants -Venous beading in 2+ quadrants -Prominent IRNA in 1+ quadrants -No apparent sign of PDR
4	Proliferative DR	One or more of the following -Neovascularization -Vitreous or preretinal hemorrhage

is universally used in the medical field with high accuracy and efficiency [14] [15]. CNN is built and runs on a powerful workstation with graphical processing units as they perform convolution operation which is a very expensive operation in terms of processing power and memory. Tremendous improvements in processing power and graphical processing units (GPU) have also improved the performance of CNN and consequently increased their widespread use in the medical field [16]. But Most of the CNNs have massive parameters as they have huge layers and complex structures. Even after GPU rendering takes a long time and consequently cannot be implemented on all platforms. Therefore, we studied lightweight CNN architectures, for example, MobileNet and MobileNetV2 with transfer learning in Diabetic Retinopathy multiclass classification to test their applicability on low-resource devices. This lightweight CNN architecture has a limited number of parameters and smaller structures. Their dependencies on GPU are not as substantial as other heavy architecture. The main goal of this research work is to study lightweight CNN architectures for DR multiclass classification. Many studies have been conducted by researchers worldwide; however, significant limitations remain, leaving room for improvement. The main contribution of this work is as follows:

1. Study the behavior, outcome and power of-transfer learning of lightweight CNN architectures with an imbalanced dataset (fundus images).

2. Evaluate and analyze the performance of lightweight CNN models in Diabetic Retinopathy severity (multi class and binary class) classification.

The structure of the remaining sections of the paper is as follows. Related research works are summarized precisely in section 2. Materials and Methodology are described briefly in section 3. In section 4, the Results and discussion are presented following the conclusion and future work.

2. Related Work

Automatic identification and detection of Diabetic Retinopathy with accuracy is a very challenging task. The feature extraction part is very important in the whole detection process and CNN shows supreme capabilities in disease classification compared to other traditional machine learning models. This in-depth literature analysis made us familiar with different prize-winning CNN architectures such as LeNet, AlexNet, OverFeat, ZFNet, VGG Net, GoogleNet, ResNet, SqeezeNet, DenseNet, ExceptionNet, MobileNet, YOLO, and U-Net. Recently, several deep-learning-based approaches have been developed and proposed for the classification and detection of Diabetic Retinopathy using fundus images whose specificity and sensitivity are satisfactory [17] [18] [19] . Especially, Convolutional Neural Networks (CNN) has been shown their supremacy in automatic feature extraction for various disease analyses in the medical domain [20], which has been applied in Diabetic Retinopathy classification and detection [21] [22]. The authors in [23] proposed a lightweight model for DR classification where CNN is used for feature extraction and various machine learning classifiers (SVM, AdaBoost, Naive Bayes, Random Forest, and J48) are used for classification using KAGGLE, IDRiD, and MESSIDOR datasets. Several evaluation metrics, e.g., accuracy, precision, recall, specificity, FPR, and Kappa Score are calculated and the J48 classifier outperforms others. In [24], the authors proposed three different Deep CNN models with varying numbers of epochs and batch sizes for 5-class DR classification. All three models are evaluated using red, green, and blue channel images and grayscale images, and the best result score was 97.07% for the blue channel of fundus images. Shufflenet [25] was originally designed for small power-constrained devices like mobile which was an efficient CNN model that performs better than the MobileNet [26] in classification tasks on ImageNet. Shufflenet shows 13 times higher accuracy than AlexNet on ARM-based mobile devices. In [27], the authors develop a lightweight CNN to classify DR and DME automatically. In hyper-parameters tuning, the authors examine four sets of combinations: Adam optimizer with a learning rate of 10-4, SGD optimizer with a learning rate of 10-4, Adam optimizer with the cosine decay learning rate, and SGD optimizer with the cosine decay learning rate where the model expresses best performance with SGD optimizer with the cosine decay learning rate. A transfer learning-based approach is proposed for DR screening using fundus images captured by smartphones [28]. They used NasNetMobile lightweight model as a feature descriptor where a multilayer perceptron was used as a classifier. A custom dataset was used for training and validating the model performances which was prepared by expert ophthalmologists. The achieved result was 94.44% sensitivity, 96.92% specificity, 95.91% accuracy, and 95.71% precision at the validation stage. Another transfer-learning based study demonstrate binary classification of diabetic retinopathy where VGG16, InceptionV3, and MobileNet pre-trained cnn models are applied. VGG16 model shows best classification result than others [29]. In [30], a novel AI-based approach has been proposed for Diabetic Retinopathy detection using public funds images datasets. Two different U-net architectures are used for Optic Disc (OD) and Blood Vessels (BV) segmentation, and CNN with Singular value decomposition (SVD) was used for appropriate feature selection following extraction. At last, transfer learning-based InceptionV3 was applied for DR diagnosis, and various famous metrics were calculated to compare the result. The authors in [31] present various preprocessing techniques for raw fundus images as those images are very difficult to analyze using a traditional ML system. Used preprocessing techniques are resizing, histogram equalization, image enhancement, and green channel extraction. They also perform microaneurysm and hemorrhage detection, and optic disc and exudate elimination. The seven most significant features out of 14 extracted features are used for classifying normal images from DR-diseased images. In [32], the authors proposed a study showing the importance of required image preprocessing in classifying diabetic eye diseases. The final classification task is achieved through a series of intermediate steps, e.g., enhancing the quality of images, segmenting the region of interest (ROI), augmenting the image dataset using geometric transformation, and finally classification. Their custom-built CNN architecture showed the optimized result such as accuracy, sensitivity, and specificity with traditional image processing techniques. A transfer learning based VGGNet CNN model was proposed for detecting diabetic retinopathy automatically [33]. Data augmentation technique is used to tackle limited and imbalanced datasets. Various preprocessing methods are applied to enhance the images quality. The transfer learning method outperforms other similar methods in terms of accuracy. In [34], the authors propose three different custom CNN architectures that are built using common CNN layers. The model was trained as a single model and as an ensemble of three models with image datasets provided by EyePACS. As a single model, the quadratic weighted kappa score is 0.386 and the ensemble of three models gives 0.3996.

3. Materials and Methodology

This study uses the power of transfer learning to classify the stages of Diabetic Retinopathy. The transfer learning approach enables CNN models to overcome the demand for a lot of data for training and saves computational resources along with memory. The pre-trained lightweight CNN architectures are applied to learn features from retina images for 2-stage (binary) and 5-stage (multi-class) classification. Small-sized lightweight models are investigated to test their applicability on low-resource devices such as mobile phones. CNN models with computationally inexpensive and higher performance are mostly expected in the current context. CNN is built with stacked layer architecture that allows the model to extract significant features layer by layer automatically without any human supervision. CNN is superior than its predecessor in terms of performance as it select and extract features accordingly. The extracted information is then flattened and fed into a fully connected layer for classification. Fully connected layers are final layers that are responsible for calculating class probabilities. The true class is identified depending on the class probability value. The overall methodology of the proposed work is presented in Figure 2. In the proposed work the model was trained and validated in 5 (five) steps with transfer learning of CNN (Convolutional Neural Network).

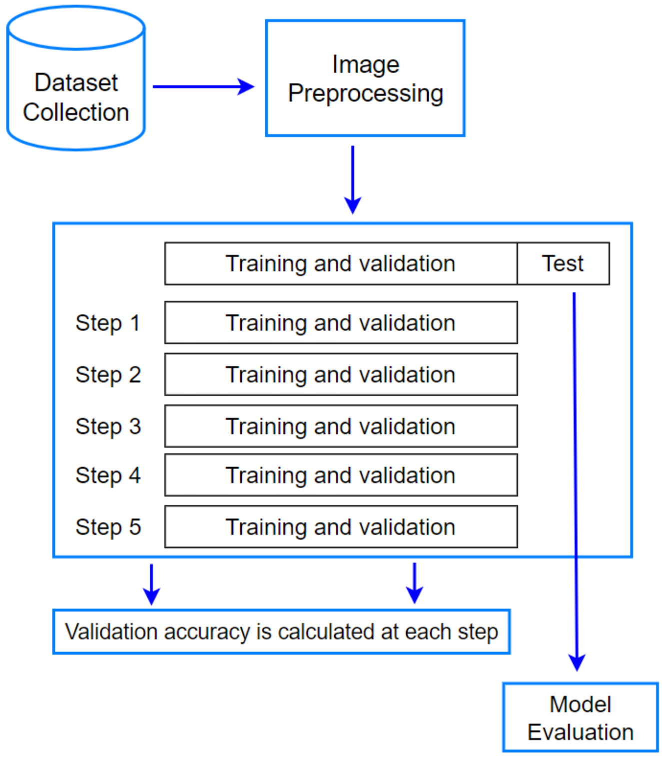

Fig. 2. Workflow in Diabetic Retinopathy Classification.

3.1. Dataset

The DDR dataset [35] is used for training, validation, and testing the performance of lightweight CNN architecture that consists of 13,673 images of 9,598 individuals where 48.23% are male and 51.77% are female patients covering ages from 1 to 100. The images were collected from 2016 to 2018 at 147 hospitals in 23 provinces in China and prepared for general purposes. This dataset supplied three different annotations: image level, pixel level, and bounding box annotations. The images in this dataset are graded by seven expert graders and a voting system is used to determine the final grading. All images in this dataset are grouped into six classes: No DR, Mild NPDR, Moderate NPDR, Severe NPDR, Proliferative DR, and Ungradable. The ungradable images are beyond classification by expert graders because of very low image quality and are usually ignored by the research community. The frequency distribution of each class in the DDR dataset is presented in Table 2. Since all the images in the ungradable class are excluded from this study, the class distribution of the DDR dataset for this paper is as in the Figure 4.

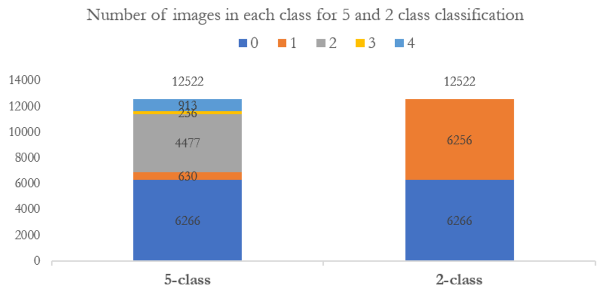

Fig. 4. Frequency distribution of each class in binary and multiclass classification.

3.1.1. Data Preprocessing

Data preprocessing is a mandatory step in medical image processing as it cleans, refine, enhance, remove noise, eliminate unwanted areas, prepare, and arrange dataset before fed into the deep learning models. The images dataset collected from Kaggle repository [36] are found cropped, that means unwanted black areas are removed. Only the retina encircled images are collected from Kaggle. Then the contrast limited adaptive histogram equalization (CLAHE) is applied for image enhancement. CLAHE is an image enhancement method used in contrast enhancement on low contrast images found in medical imaging such as retinal fundus images. CLAHE is the evolution of the probability theory-based Histogram Equalization (HE) and Adaptive Histogram Equalization (AHE) techniques that is used to enhance and resolve the comparability issue for digital images. [37] CLAHE carried out the procedure by increasing the Histogram Equalization (HE) Complications which makes the improvements in medical images even better by increasing the HE Compensation. Sample images from each class are shown in Figure 3 where NPDR stand for non proliferative diabetic retinopathy. All the images shown in Figure 3 are preprocessed images.

3.2. Used lightweight CNN Model

Currently, deep CNNs are used extensively in medical image analysis because of its unique features and unparallel benefit. Although CNN requires a lot of data to train the model, the research community is attracted by CNN's dynamic features learning and selection. There are two ways of using CNN effectively in image classification: designing the custom CNN model from scratch and transfer learning (TL) the pre trained CNN architectures to solve similar problem. Transfer learning is a popular method where the pre trained weights are used in case of fewer dataset and limited computational powers. In

DOI:10.4186/ej.2024.28.7.13

Table 2. Number of Images in each class in the DDR dataset.

Severity Stages	No DR	Mild	Moderate	Severe	Proliferative DR	Ungradable	Total
Class	**0**	**1**	**2**	**3**	**4**	**-**	**-**
No. of Images	6266	630	4477	236	916	1151	13673
Percentage(%)	45.83%	4.53%	32.74%	1.73%	6.68%	8.42%	100%

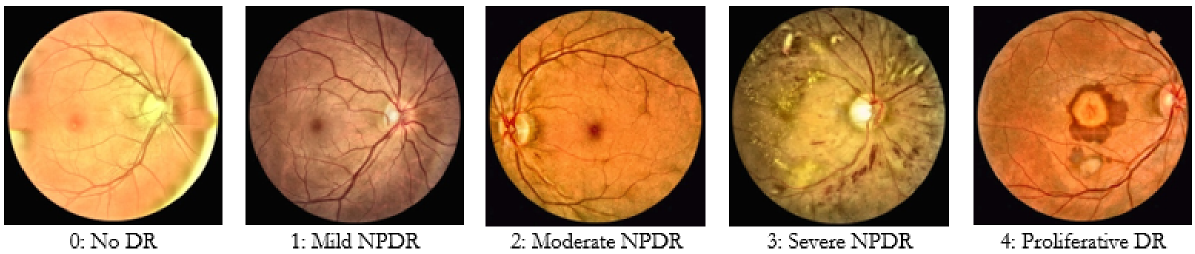

Fig. 3. Sample images from each class of the DDR dataset that used in this study .

this study, the lightweight CNN architectures are considered for DR classification because its depth is less, requires low resources for computation, and shorter training time. Considering those issues, MobileNet and MobileNetV2 are used via transfer learning. In transfer learning paradigm, the process are generally implemented through two steps: the top layer weights are reused and the parameter and weights of the classifier are updated at training time with the intended dataset.

MobileNet [26] is a first model designed for mobile and embedded devices. MobileNet is small, low depth, and has fewer number parameters to meet the requirements of low-resource devices. MobileNet uses depth wise separable convolution which decreases the number of parameters and results in lightweight. A depth wise separable convolution is combination of two distinct operations: depth wise convolution and pointwise convolution.

MobileNetV2 [38] was invented by Google in which inverted residual blocks are used for improving performance. In this model, two types of blocks are used: one is residual block with stride 1 and another block is stride 2. Each block has 3 layers. The first layer is 1 by 1 convolution layer with Relu6, depth wise convolution is second layer, and third layer is 1 by 1 convolutional layer with no non-linearity.

3.3. Framework and Experimental Setup

The proposed work was carried out using Python software package. Anaconda with Python default version 3.9 is used in Windows environment. Spyder is used as a development environment which is free IDE (Integrated Development Environment) included in Anaconda. A Windows 10 Desktop PC with 1T SSD, 16 GB RAM, and NVIDIA GEFORCE RTX 2080 (16 GB) was used for experimental work. The deep learning API Keras is used with Tensorflow as its back end because it runs smoothly on both CPU and GPU. Moreover, Keras supports almost all neural network models.

3.4. Model Training and Validation

All the training and validation works were performed using a Desktop PC whose specifications are described in above paragraph. Two lightweight Convolutional Neural Network (CNN) architectures: MobileNet and MobileNetV2 models are utilized to classify the severity stages of Diabetic Retinopathy through transfer learning. The trained weights of the CNN models are transferred and used in the training process where those models are originally trained on ImageNet dataset[39], a very large image dataset that contains more than 14 million images of thousand classes. All the upper layers remain frozen during training that means the weights are not being updated during training. The transfer learning includes fully connected layers at lower level which is called classifier according to the number of classes to be classified. Only the fully connected (FC) layers are being updated during the training period. Same hyperparameters values are utilized for both CNN models. Cross-entropy loss in the training period as well as validation period is calculated at each epoch for a total of 500 epochs using training and validation dataset, respectively. Adam optimizer with default learning rate 0.001 is used in training because it has the tendency to converge faster than other in many

Table 3. MobileNet performance on validation data for binary classification.

Model	Step 1	Step 2	Step 3	Step 4	Step 5
Accuracy	80%	87%	89%	92%	93%
Precision	81%	88%	89%	92%	93%
Recall	80%	87%	89%	92%	93%
F1-Score	79%	87%	89%	92%	93%

cases. In CNN, adam is works well and used popularly and hence considered default optiizer. A batch size of 32 is used to estimate the error before adjusting weights throughout the training period. At last in the fully connected layer, sigmoid and softmax activation function is used as a requirement for binary and multiclass classification, respectively. Accuracy and loss at each epoch for every validation step is calculated.

Before feeding the dataset into the model, the dataset is prepared according to the number of classes in classification. For 5-class classification, the dataset is utilized as it is. But in case of 2-class (binary) classification, we made the 2 class by mixing classes. Images in Class 1, Class 2, Class 3, and Class 4 are combined to form the diseased class where images in Class 0 are considered as healthy class as well. Class wise frequency distribution for 5-class and 2-class classification are depicted in Figure 6. Training and validation of the models are performed through five iteration with 100 epochs at each step, means 500 epoch in total. The training, validation and testing process graphically shown in Figure 2.

Repeated training and validation is typically effective to learn the model about dataset when there are class imbalance problems in the dataset. First the whole dataset is divided into two folders: Training and Testing. Before dataset splitting, shuffles it to avoid any ordering in dataset. The weights of each step is transfered to next step. So, the validation performances increased step by step. There a clear sign of overfitting at the beginning of validation because of class imbalance in the dataset. Classification report generated at validation time clearly indicate that the validation accuracy improves as itertion progresses. The cross validation result of the MobileNet and MobileNetV2 for binary and multiclass classification are presented in Table 3, 4, 5, and 6.

4. Testing Result and Discussion

After training and validation, the performance of the lightweight CNN models are evaluated using a test dataset. There are 1250 and 1251 images that were used for 5 class and 2 class classification, respectively. For both models, the testing performance was evaluated by calculating accuracy, precision, recall, F1-score values, and ROC curves.

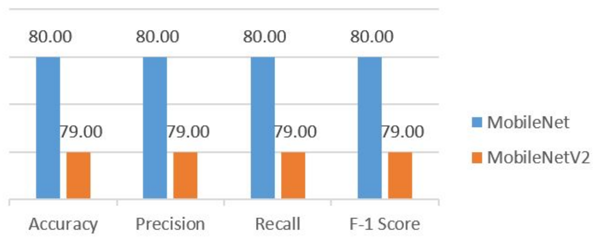

Fig. 5. Testing result for binary classification.

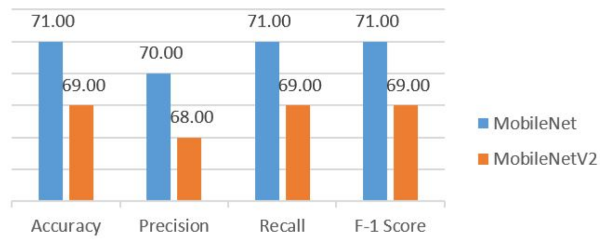

Fig. 6. Testing result for multiclass (5-class) classification.

The values of accuracy-score, f1-score, precision-score, recall-score, confusion-matrix were calculated by using builtin function of sklearn.metrics. The overall testing results for binary (2-class) and multiclass (5-class) classification are presented in the Figure 5 and Figure 6, respectively.

Mobilenet showed slightly better performance than MobileNetV2 in both modes of classification. In case of binary classification, the accuracy, precision, recall and F1-score of MobileNet model are 80%, 80%, 80% and 80%, respectively and the accuracy, precision, recall and F1-score of MobileNetV2 model are 79%, 79%, 79% and 79%, respectively. For the multiclass classification, the MobileNet model achieved 71% accuracy, 70% precision, 71% recall, and 71% F1-score. A 69% accuracy, 68% precision, 69% recall, and 69% F1-score achieved by MobileNetV2 architecture.

DOI:10.4186/ej.2024.28.7.13

Table 4. MobileNetV2 performance on validation data for binary classification.

Model	Step 1	Step 2	Step 3	Step 4	Step 5
Accuracy	78%	85%	87%	88%	90%
Precision	78%	85%	87%	88%	90%
Recall	78%	85%	87%	88%	90%
F1-Score	78%	85%	87%	88%	90%

Table 5. MobileNet performance on validation data for multiclass (5-class) classification.

Model	Step 1	Step 2	Step 3	Step 4	Step 5
Accuracy	71%	80%	87%	89%	90%
Precision	70%	79%	87%	89%	90%
Recall	71%	80%	87%	89%	90%
F1-Score	70%	79%	86%	89%	90%

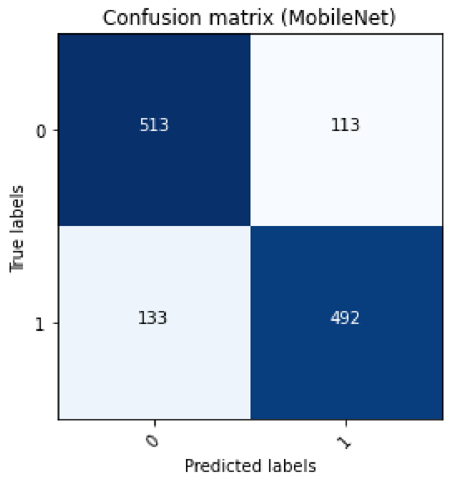

Fig. 7. Confusion Matrix for binary classification generated by MobileNet.

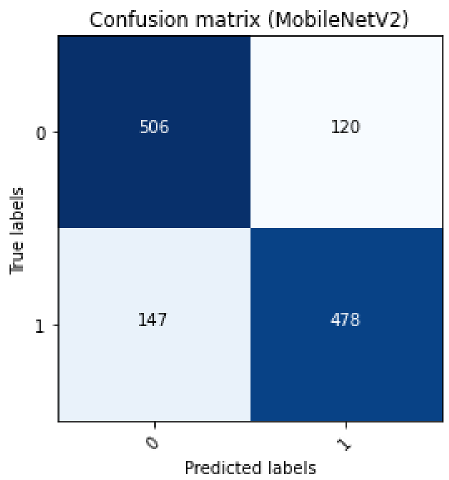

Fig. 8. Confusion Matrix for binary classification generated by MobileNetV2.

Figure 7 and 8 represents the confusion matrix results for binary classification by MobileNet and MobileNetV2, respectively. Figure 9 and 10 represents the confusion matrix results for multiclass (5-class) classification by MobileNet and MobileNetV2, respectively. In case of binary classification, MobileNet made 513 correct predictions and 113 misprediction among 626 images of (Class 0) healthy class and 492 correct prediction and 133 misprediction among 625 images of (Class 1) diseased class. Similarly, MobileNetV2 predict 506 images correctly and 120 misprediction among 626 images of (Class 0) healthy class whereas 478 correct prediction and 147 misprediction among 625 images of (Class 1) diseased class. MobileNet showed 81.94% and 78.72% accuracy for healthy and diseased class, respectively. MobileNetV2 showed an accuracy of 80.83% and 76.48% for healthy and diseased class, respectively.

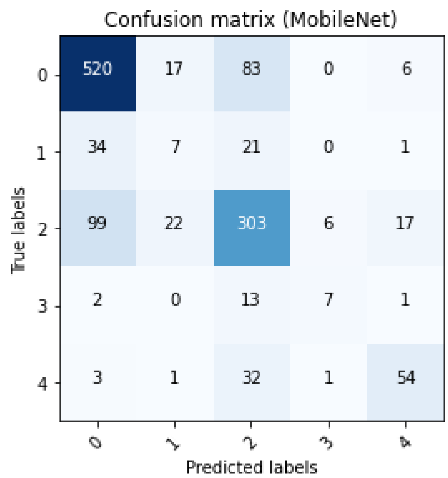

Fig. 9. Confusion Matrix for 5-class classification generated by MobileNet.

Table 6. MobileNetV2 performance on validation data for multiclass (5-class) classification.

Model	Step 1	Step 2	Step 3	Step 4	Step 5
Accuracy	67%	78%	82%	87%	88%
Precision	67%	77%	82%	87%	88%
Recall	67%	78%	82%	87%	88%
F1-Score	67%	77%	81%	87%	88%

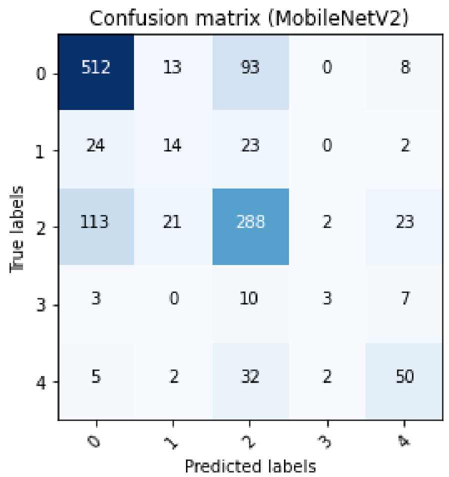

Fig. 10. Confusion Matrix for 5-class classification generated by MobileNetV2.

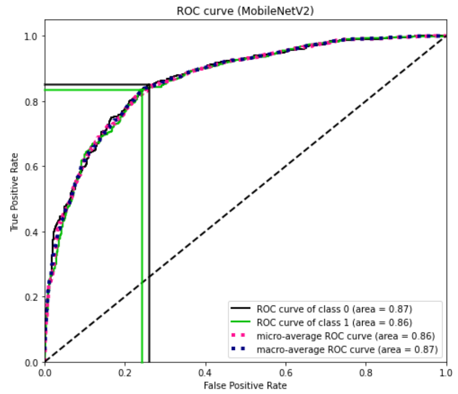

Fig. 12. Receiver Operating Characteristic (ROC) curve for binary classification by MobileNetV2.

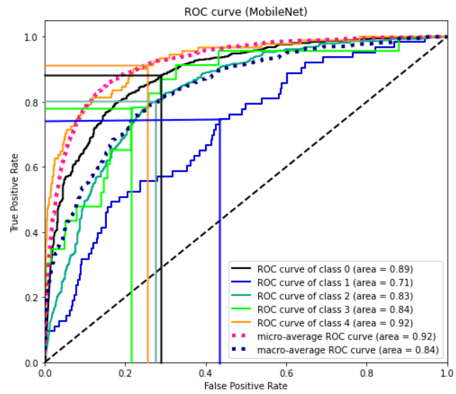

Fig. 13. Receiver Operating Characteristic (ROC) curve for 5-class classification by MobileNet.

It is evident from the above discussion that MobileNet performed well on both healthy and diseased class than MobileNetV2. In multiclass classification shown in Figure 10, both models comparatively performed well on Class 0 (No Disease) and class 2 classes (Moderate DR) but showed worst performance on class 1 (Moderate DR) and class 3 (Severe DR). In the case of class 4 (Proliferative DR), both models performed moderately.

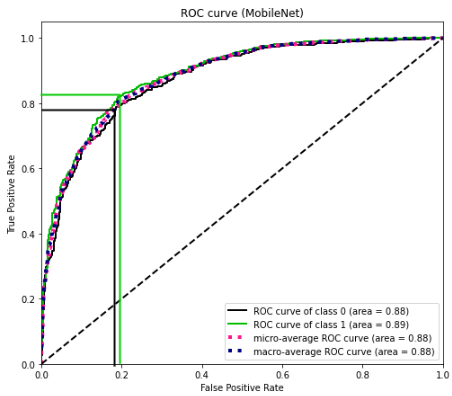

Fig. 11. Receiver Operating Characteristic (ROC) curve for binary classification by MobileNet.

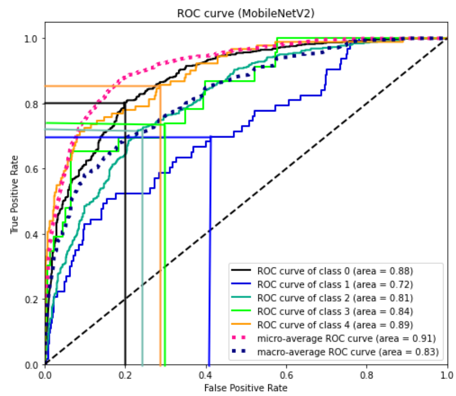

Fig. 14. Receiver Operating Characteristic (ROC) curve for 5-class classification by MobileNetV2.

Table 7. Performanes comparison with others state of art works.

Algoritms	Dataset Used	Results
Custom build CNN model by H. Pratt et al. [8]	Kaggle Dataset (80,000 images)	Accuracy: 75% Sensitivity: 30% Specificity: 95%
CNN Transfer Learning by Wang et al. [40]	EyePACS (35,126 images)	AlexNet: 37.43% VGG: 50.16% InceptionV3: 63.23%
MobileNet and MobileNetV2 via Transfer Learning (Our Work)	DDR Dataset (12,522 images)	Binary Classification: Accuracy: 80% Multiclass Classification: Accuracy: 71%

ROC (Receiver Operating Characteristics) curve is plotted using false positive rate (FPR) and true positive rate (TPR) values. Figure 11 and 12 shows the ROC curve for binary classification by MobileNet and MobileNetV2, respectively. In binary classification, MobileNet occupy the area under the ROC curves are 88% for healthy class and 89% for diseased class and MobileNetV2 occupy the area under the ROC curves are 87% for healthy class and 86% for diseased class. ROC curve for multiclass (5-class) classification by MobileNet and MobileNetV2 are shown in Figure 13 and 14, respectively. For multiclass classification, class 4 occupied the highest area 92% and 89% by MobileNet and MobileNetV2, respectively.

5. Conclusion and Future Research

Long-term diabetic retinopathy may damage the retina of the eye and lead to permanent vision loss. Nowadays, along with the increase in diabetic patients, the number of diabetic retinopathy patients is increasing alarmingly. So, detection and classification of diabetic retinopathy at an early stage can lead to early treatment which can save eye vision. Expert ophthalmologists are currently investigating diabetic retinopathy by scrutinizing the lesions and other symptoms associated with the disease. This clinical process is very effective for DR screening, but it is time-consuming, cost and resource-effective, and error-prone. The symptoms of DR are initially invisible which makes early DR detection a very challenging task using the traditional way. Therefore, the automatic DR detection method is a prime need, especially in the rural areas of lower-middle-income countries where expert doctors are very few in number compared to DR patients. CNN has demonstrated its efficacy in DR detection through automatic feature extraction and selection. In this paper, the performances of lightweight CNN architectures: MobileNet and MobileNetV2 are analyzed in DR severity classification via transfer learning. Although the DDR dataset has enough image data, it is a highly imbalanced dataset that has a maximum of 6266 images in class 0 (45.83% of the total) and a minimum of 236 images in class 3 (1.73% of the total). To improve the performances of CNN using an imbalanced dataset, the model has been trained and valilate in five (5) repeated steps, and therefore the validation result increased with each iteration. The highest 93% and 90% validation accuracy were achieved by MobileNet after the last iteration in binary and multiclass classification, respectively. The performances of the models are tested using a test set and the highest 80% and 71% accuracy were achieved during testing by MobileNet in binary and multiclass classification, respectively. In the future, the same work may be done with some modification and finetuning of the models to improve the performance.

Acknowledgment

We are deeply grateful to the scholarship authority of the Graduate Scholarship Program for ASEAN and NON-ASEAN countries for their generous support, which made our research work possible. We would also like to express our heartfelt appreciation to the Embedded System and IC Design Research Laboratory (ESID) and the Department of Electrical Engineering at Faculty of Engineering, Chulalongkorn University for their invaluable assistance throughout the research process.

References

[1] “What is diabetes?” https://www.cdc.gov/diabetes/basics/diabetes.html, (Accessed: 23 March, 2023).

[2] M. D. Abramoff, M. K. Garvin, and M. Sonka, “Retinal imaging and image analysis,” ***IEEE re-***

views in biomedical engineering, vol. 3, pp. 169–208, 2010.

[3] "Idf diabetes atlas: Tenth edition." https://diabetesatlas.org/, accessed: 15 March, 2023.

[4] Z. L. Teo, Y.-C. Tham, M. Yu, M. L. Chee, T. H. Rim, N. Cheung, M. M. Bikbov, Y. X. Wang, Y. Tang, Y. Lu *et al.*, "Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis," *Ophthalmology*, vol. 128, no. 11, pp. 1580–1591, 2021.

[5] "Diabetic retinopathy: Causes, symptoms, treatment." https://www.aao.org/eye-health/diseases/what-is-diabetic-retinopathy, (Accessed: 23 March, 2023).

[6] G. S. Scotland, P. McNamee, A. D. Fleming, K. A. Goatman, S. Philip, G. J. Prescott, P. F. Sharp, G. J. Williams, W. Wykes, G. P. Leese *et al.*, "Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy," *British Journal of Ophthalmology*, vol. 94, no. 6, pp. 712–719, 2010.

[7] G. A. Williams, I. U. Scott, J. A. Haller, A. M. Maguire, D. Marcus, and H. R. McDonald, "Single-field fundus photography for diabetic retinopathy screening: a report by the american academy of ophthalmology," *Ophthalmology*, vol. 111, no. 5, pp. 1055–1062, 2004.

[8] C. P. Wilkinson, F. L. Ferris III, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, J. T. Verdaguer *et al.*, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, 2003.

[9] E. T. D. R. S. R. Group *et al.*, "Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified airlie house classification: Etdrs report number 10," *Ophthalmology*, vol. 98, no. 5, pp. 786–806, 1991.

[10] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," *Informatics in Medicine Unlocked*, vol. 20, p. 100377, 2020.

[11] "Diabetic retinopathy- can it be reversed??" https://neoretina.com/blog/diabetic-retinopathy-can-it-be-reversed/, (Accessed: 23 March, 2023).

[12] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85–112, 2020.

[13] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[14] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[15] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 47, 2018.

[16] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *Ieee Access*, vol. 6, pp. 9375–9389, 2017.

[17] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, "Automatic detection of diabetic eye disease through deep learning using fundus images: a survey," *IEEE access*, vol. 8, pp. 151 133–151 149, 2020.

[18] N. Gour, M. Tanveer, and P. Khanna, "Challenges for ocular disease identification in the era of artificial intelligence," *Neural Computing and Applications*, pp. 1–23, 2022.

[19] Y. Jeong, Y.-J. Hong, and J.-H. Han, "Review of machine learning applications using retinal fundus images," *Diagnostics*, vol. 12, no. 1, p. 134, 2022.

[20] X. Li, C. Li, M. M. Rahaman, H. Sun, X. Li, J. Wu, Y. Yao, and M. Grzegorzek, "A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches," *Artificial Intelligence Review*, vol. 55, no. 6, pp. 4809–4878, 2022.

[21] S. M. S. Islam, M. M. Hasan, and S. Abdullah, "Deep learning based early detection and grading of diabetic retinopathy using retinal fundus images," *arXiv preprint arXiv:1812.10595*, 2018.

[22] P. Burlina, W. Paul, P. Mathew, N. Joshi, K. D. Pacheco, and N. M. Bressler, "Low-shot deep learning of diabetic retinopathy with potential applications to address artificial intelligence bias in retinal diagnostics and rare ophthalmic diseases," *JAMA ophthalmology*, vol. 138, no. 10, pp. 1070–1077, 2020.

[23] S. Gayathri, V. P. Gopi, and P. Palanisamy, "A lightweight cnn for diabetic retinopathy classification from fundus images," *Biomedical Signal Processing and Control*, vol. 62, p. 102115, 2020.

[24] M. M. Butt, G. Latif, D. A. Iskandar, J. Alghazo, and A. H. Khan, "Multi-channel convolutions

neural network based diabetic retinopathy detection from fundus images," *Procedia Computer Science*, vol. 163, pp. 283–291, 2019.

[25] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6848–6856.

[26] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[27] Z. Lu, J. Miao, J. Dong, S. Zhu, X. Wang, and J. Feng, "Automatic classification of retinal diseases with transfer learning-based lightweight convolutional neural network," *Biomedical Signal Processing and Control*, vol. 81, p. 104365, 2023.

[28] Y. Elloumi, N. Abroug, and M. H. Bedoui, "End-to-end mobile system for diabetic retinopathy screening based on lightweight deep neural network," in *Advances in Intelligent Data Analysis XX: 20th International Symposium on Intelligent Data Analysis, IDA 2022, Rennes, France, April 20–22, 2022, Proceedings*. Springer, 2022, pp. 66–77.

[29] P. K. Das and S. Pumrin, "Cnn transfer learning for two stage classification of diabetic retinopathy using fundus images," in *2023 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON)*. IEEE, 2023, pp. 443–447.

[30] A. Bilal, L. Zhu, A. Deng, H. Lu, and N. Wu, "Ai-based automatic detection and classification of diabetic retinopathy using u-net and deep learning," *Symmetry*, vol. 14, no. 7, p. 1427, 2022.

[31] D. S. Sisodia, S. Nair, and P. Khobragade, "Diabetic retinal fundus images: Preprocessing and feature extraction for early detection of diabetic retinopathy," *Biomedical and Pharmacology Journal*, vol. 10, no. 2, pp. 615–626, 2017.

[32] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, J. Ma, and K. Wang, "Image preprocessing in classification and identification of diabetic eye diseases," *Data Science and Engineering*, vol. 6, no. 4, pp. 455–471, 2021.

[33] M. K. Jabbar, J. Yan, H. Xu, Z. Ur Rehman, and A. Jabbar, "Transfer learning-based model for diabetic retinopathy diagnosis using retinal images," *Brain Sciences*, vol. 12, no. 5, p. 535, 2022.

[34] D. Doshi, A. Shenoy, D. Sidhpura, and P. Gharpure, "Diabetic retinopathy detection using deep convolutional neural networks," in *2016 international conference on computing, analytics and security trends (CAST)*. IEEE, 2016, pp. 261–266.

[35] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, pp. 511–522, 2019.

[36] "Ddr dataset." https://www.kaggle.com/datasets/mariaherrerot/ddrdataset, (Accessed: 5 January, 2023).

[37] R. Zulfahmi, D. S. Noviyanti, G. R. Utami, A. N. Harison, P. S. Agung *et al.*, "Improved image quality retinal fundus with contrast limited adaptive histogram equalization and filter variation," in *2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*. IEEE, 2019, pp. 49–54.

[38] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[39] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[40] X. Wang, Y. Lu, Y. Wang, and W.-B. Chen, "Diabetic retinopathy stage classification using convolutional neural networks," in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*. IEEE, 2018, pp. 465–471.

DOI:10.4186/ej.2024.28.7.13

Pranajit Kumar Das received bachelor of Engineering (B.Eng.) degree in Computer Science and Engineering from the Department of Computer Science and Engineering under the faculty of applied science and technology, Shahjalal University of Science and Technology, Sylhet, Bangladesh in 2010. Then he completed Master of Engineering (M. Eng.) degree at the Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand under the Graduate scholarship program for ASEAN and Non-ASEAN countries in 2023. He holds a faculty position at Department of Computer Science and Engineering, Sylhet Agricultural University, Bangladesh. He is passionate about doing cutting-adge research on application of AI in medical field. In particular, Deep Learning, Computer Vision, Data Analytics, and Medical Image Analysis.

Suree Pumrin received her B.Eng. degree in electronics engineering from King Mongkut's Institute of Technology Ladkrabang, Thailand, in 1990, followed by M.S.E. degree in electrical engineering from Arizona State University, USA., in 1992, and Ph.D. degree in electrical engineering from University of Washington, USA., in 2002. She is currently an assistant professor at the Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Thailand. Her research interests include image processing, computer vision, and embedded systems.