Diabetic Retinopathy Detection Using Convolutional Neural Network: A Comparative Study on Different Architectures

Tassanee Hattiya[1], Kwankamon Dittakan[1,*] and Salang Musikasuwan[2]

[1] College of Computing, Prince of Songkla University, Phuket Campus
80 M.1 Vichitsongkram Road, Kathu, Phuket, Thailand

[2] Department of Mathematics and Computer Science, Faculty of Science and Technology,
Prince of Songkla University, Pattani Campus,
181 Charoenpradit Road, Rusamilae, Muang, Pattani, Thailand

s6130621001@phuket.psu.ac.th, kwankamon.d@phuket.psu.ac.th*, and salang.m@psu.ac.th

Abstract. *Diabetic retinopathy (DR) is a diabetes complication affects the eyes. The patients who lack treatments may be affected to the visual such as no clear vision, bleeding, or blindness. The problem of diabetic patients is the difficulty in the detection of DR until the symptom has happened. Early diagnosis is typically made using retina imagery obtained from the fundus camera. In this paper, an automated mechanism for DR screening is proposed. The idea is to construct a classifier that can be used to distinguish between DR or non-DR retina images. The fundamental idea is to segment retina images for obtaining the region of interest (ROI), while remaining compatible with the classification process. The ROI is then transformed into an appropriate format. It is suggested that the convolutional neural network (CNN) is the most classifier learning mechanism to be considered. With respect the work presented in this paper, seven convolutional neural network architectures have been applied to compare the classification performance: (i) AlexNet, (ii)ResNet50, (iii) DenseNet201, (iv) InceptionV3, (v) MobileNet, (vi) MnasNet and (vii) NASNetMobile. The process is fully described and evaluated. The data used for the evaluation was obtained from the Kaggle of 23,513 images. The best results were obtained from AlexNet learning mechanism with accuracy values of 98.42% and 81.32% for training and testing sets, respectively.*

Received by	15 November 2020
Revised by	14 December 2020
Accepted by	9 January 2021

Keywords:

Diabetic Retinopathy detection, Image Analysis, Image Classification, Neural network, Convolutional neural network

1. Introduction

Diabetes mellitus (DM) is the health condition in the human body which happen when the body has a high sugar level in the blood [1]. The main cause of DM is the body could not produce enough of the insulin hormone. Once the lack of insulin, the body is unable to transform the sugar into energy. Thus, a large amount of sugar is accumulated in the bloodstream. The information from the International Diabetes Federation or IDF reported that 425 million people all over the world are patients with diabetes in 2017 and expected that it will rise to 629 million people by 2045. Early stages of diabetes are difficult to diagnose because there is no sign of any symptoms. However, the patient with DR might face several symptoms such as (i) frequent hunger, (ii) very thirsty, (iii) weight loss, (iv) frequent urination, (v) dry mouth, (vi) the wound is difficult to heal, (vii) blurred vision, and (viii) numbness especially the hands and legs.

There are three different types of diabetes mellitus [2] [3]: (i) Type 1 diabetes is a condition that happens when the immune system destroys the pancreas cells. Therefore, the body is incapable of producing insulin; (ii) Type 2 diabetes is a condition in which the body unable to produce enough insulin or unable to utilize insulin efficiently. This type of DR is the most common and usually found in middle-aged to old age; and (iii) Gestational diabetes is a type of diabetes that can occur at any stage of pregnancy. In general, most often diagnosed during the 24-28 weeks of pregnancy. It is caused by high blood sugar levels because the body is unable to produce enough insulin. This type of DR can adversely affect mothers and babies.

Long-term complications of diabetes mellitus are categorized into two main groups: (i) microvascular, and (ii) macrovascular. The microvascular complication is the disease or symptom that small blood vessels are damaged, the example of microvascular complications includes: (i) eye damage (retinopathy), (ii) nerve damage (neuropathy), (iii) kidney damage (nephropathy), (v) feet damage. On the other hand, the macrovascular complication is the disease or symptom concerned with larger blood vessels include (iv) cardiovascular disease such as strokes, heart attacks, and narrowing of arteries (atherosclerosis).

Diabetic Retinopathy is a complication of diabetes and is the number one cause of blindness of the patient with DM

[4]. The main cause of DR because the high blood sugar levels of diabetes damage the blood vessels of the retina. More specifically, when too much sugar in the blood, the small and tiny blood vessels are blocked. Thus, no blood supply to nourish the retina. Therefore, the body is trying to grow new blood vessels. However, the new blood vessels are delicate causing broken blood vessels. However, the new blood vessels are not strong enough causing broken blood vessels. The patient with diabetic retinopathy probably no sign of the symptoms in the early stages. However, the progression of the symptoms of DR might include: (i) blurred vision, (ii) spot or strip floating in the vision, (iii) the vision is getting worse, (iv) eye pain, (v) impaired colour vision, and (vi) blindness [5].

There are four stages of diabetic retinopathy [6]: (i) mild non-proliferative retinopathy is the first stage of DR. There are microaneurysms in the retina blood vessels; (ii) moderate non-proliferative retinopathy is the second stage of DR. The blood vessels of the retina are blocked and swell. This stage may have a fluid in the macula region of the retina; (iii) severe non-proliferative retinopathy, the third stage or DR. The large number of blood vessels in the retina become blocked, thus the new small blood vessels are grown, and (iv) proliferative diabetic retinopathy (PDR) is the final stage of DR. The tiny and small new blood vessels are proliferated. The blood leak and bleed has occurred. In general, diabetic retinopathy could be diagnosed by using a comprehensive dilated eye exam. Before an ophthalmologist or expert doctor could be commenced, the doctor has to place drops to the patient's eyes in order to dilate the pupils which allow the doctor easy to look through a special lens to see pathology, and irregular inside the eyes include: (i) abnormal blood vessels, (ii) swelling in the retina, (iii) the growth of new blood vessels, (iv) scar tissue, (v) bleeding, and (vi) abnormal in the optic nerve. There are some medical imaging could be applied in order to help the doctor in diagnosis process: (i) ophthalmoscopy or fundus photography, (ii) fundus fluorescein angiography (FFA), and (iii) optical coherence tomography (OCT). The example image of retina from the fundus camera is illustrated as shown in Fig. 1.

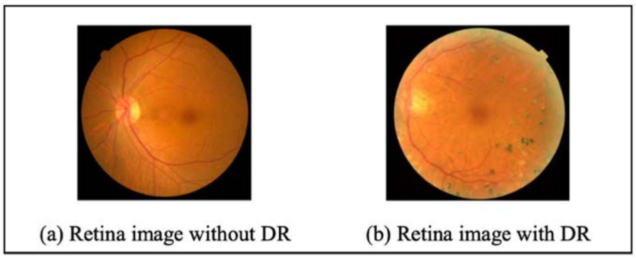

Fig. 1: Example of the retina image with and without DR

From the Fig.1, an example fundus image of the retina without DR is presented in Fig. 1(a), while the retina image with DR is shown in Fig. 1(b). There are many problems associated with the traditional diabetic retinopathy diagnosis: (i) the requirement of domain expert; (ii) the accuracy of diagnosis because it is depended on domain expert experience, (iii) the lack of specialized equipment, and (iv) time-consuming in data collection and analysis processes.

The artificial intelligence (AI) [7], [8] is the simulation or imitation of the human control system for the machine or computer system. These processes include: (i) learning, (ii) reasoning, (iii) visual perception, (iv) speech recognition, and (v) decision-making. Machine learning is a subset of AI that allows the machine to learn and improve their ability using its experience. The learning process initiates with sample data (also known as the training set) is applied with a machine-learning algorithm to find interesting patterns and can be used to make a better decision in the future. The machine learning method is divided into four different mechanisms: (i) Supervised machine learning algorithms, (ii) Unsupervised machine learning algorithms, (iii) Semi-supervised machine learning algorithms, and (iv) Reinforcement machine learning algorithms. The new trend of machine learning is deep learning. In the public, when the term of deep learning is used, it usually refers to a deep artificial neural network. The deep artificial neural network has been applied to many applications such as natural language processing, pattern recognition, and image detection. Deep learning has been also successfully applied in medical, especially in the diagnosis process. The examples include deep learning was applied to determine metastatic breast cancer. The deep learning was used to classify the medical image quality for DR screening.

The solution proposed in this paper is found on the idea of generating the classifiers that can be used to detect diabetic retinopathy and non-diabetic retinopathy from retinal imagery. The retina image is obtained from the fundus camera. The fundamental idea of the proposed research begins with a collection of labelled retinal image data. A set of training data is then pre-processed in order to facilitate the learning process. The desired classifiers have been generated using the convolutional neural network (CNN) which could be applied in the future. The proposed approach offers the following advantages: (i) low cost, in term of human resource and financial, (ii) automatic processing speed up the waiting time for the patient. In the context of the work presented in this paper, the comparative study on the classification performance of seven architectures of the convolutional neural network has been compared, include (i) AlexNet, (ii) ResNet50, (iii) DenseNet201, (iv) InceptionV3, (v) MobileNet, (vi) MnasNet and (vii) NASNetMobile.

The rest of this paper is organised as follows: Section 2 explains the overview of some related works. The proposed framework of diabetic retinopathy detection using convolutional neural network is presented in Section 3. The detail of image pre-processing processes is described in Section 4. The brief description of seven architectures of CNN used with respect to the work presented in this paper is described in Section 5. Section 6 explains the results and evaluation of the proposed framework. And Section 7 presents the discussion and conclusion.

2. Related Work

Medical imaging is the technique, process, and technology which can be used to create various type of image of the human body. The aims of medical imaging are (i) diagnosis, (ii) monitor, and (iii) treatment. Different type of technology gives different information and can be used in different situations. The example of existing medical imaging includes [9]: (i) magnetic resonance imaging (MRI) uses the powerful magnetic fields and radio waves to generate the cross-sections images of the body; (ii) ultrasound used high frequencies of sound waves to create an organ image inside the body; (iii) X-rays stand for X-radiation uses the high energy electromagnetic radiation to obtain the structure inside the body; (iv) computed tomography (CT) uses a beam of X-rays spins around the body to generate the internal structure images; and (v) positron emission tomography (PET) is a technique of medical imaging that uses nuclear medicine to construct the image of metabolism in the body. Medical imaging had been applied in many applications which reported in the literature. The MRI brain scan was used in order to detect the area and number of cerebral micro bleed in the brain [10]. MRI also used in [11] for edge detection, in [12], represent about to predicting CT Image using MRI, and in paper [13], proposed the three-dimensional surface reconstruction using optical flow.

As noted in Section I, diabetic retinopathy can be diagnosed using various medical imaging. There are five different kinds of medical imaging used for an eye examination. The first one is optical coherence tomography (OCT) uses the light wave to obtain the shape of the retina and optic nerve. This is commonly used to observe abnormal of the optic nerve and macula lutea area. The second is the standard automated perimetry (SAP), which is another technique of the eye examination similar to OCT but it is used to identify glaucoma. The other technology is fundus fluorescein angiography (FFA). It uses the laser to obtain the image. In general, it is used to observe the retina and blood vessels. The binocular slit-lamp is the major tool for an eye examination. This technology uses the high-intensity light source that can be used to examine both anterior and posterior segment of the eye: retina, conjunctiva, cornea, eyelid, natural crystalline lens, iris, and sclera. The last one is the fundus photography, it uses an intricate microscope to observe the retina, optic disc, and macula area in the eye. The fundus photography is commonly used to diagnose diverse disorders such as diabetic retinopathy, and macular degeneration.

From the literature, it was reported that the progression of diabetic retinopathy could be identified by a complication of diabetes. The abnormal of diabetic retinopathy could be diagnosed using microaneurysm, haemorrhage, exudate, venous change, neovascularization, and retinal thickening. Automated detection of diabetic retinopathy has been found in many works. The keys technology used to solve this problem were (i) image processing and (ii) machine learning.

Image processing is a process to perform some operations on an image such as (i) extract some useful information from the image, (ii) enhance the image quality, and (iii) segment an interesting area from the whole image. Image processing is a subset of signal processing. The image is used as the input and output could be the image or others. Image processing has been applied to analyse of various diseases. In [14] reported that like nuclear medicine diagnosis, logic programming to find a pathology [2], create a model for distributed medical image [15], create a model used to x-ray image, axial CT, MRI, fluoroscopy, and ultrasound [16]. In the context of diabetic retinopathy, image processing was used in [17] detection of diabetic retinopathy, develop automated software for screening and diagnosing diabetic retinopathy [18], feature extraction for early detection diabetic retinopathy [19]. Machine learning (ML) is the applications of artificial intelligence (AI) that allows the machine or system has the capability to automatically learn and be better from experience. With respect to diabetic retinopathy, machine learning was also applied for the automated detection system. In paper [20] reported that create computer-aided screening systems that analysis fundus images, create the model an automated method for image-based classification [21], and detection early diabetic retinopathy using fundus image [22].

The new technology that is dramatically developed for machine learning is deep learning. Deep learning has been developed from the neural networks which allow the system to learn. The learning in deep learning could be (i) supervised learning, (ii) semi-supervised learning, and (iii) unsupervised learning. Deep learning is the method that trying to imitate the complex system in the human central system. From the literature, deep learning has been applied in many applications. In the context of diabetic retinopathy disease, image classification, pattern recognition and machine leaning were used to generate the classification models from the colour retina images [23]. The generated models were then used to detect diabetic retinopathy. In [24], used a regular image processing and machine learning techniques was applied to identify the hard exudates which were later used in diabetic retinopathy.

3. DR Detection Mining Framework

The framework of proposed diabetic retinopathy detection using the convolutional neural network is presented in this section. The schematic of the proposed framework is illustrated in Fig. 2.

From Fig. 2, it can be observed that the framework consists of three processes: (i) image segmentation, (ii) image transformation, and (iii) classifier generation.

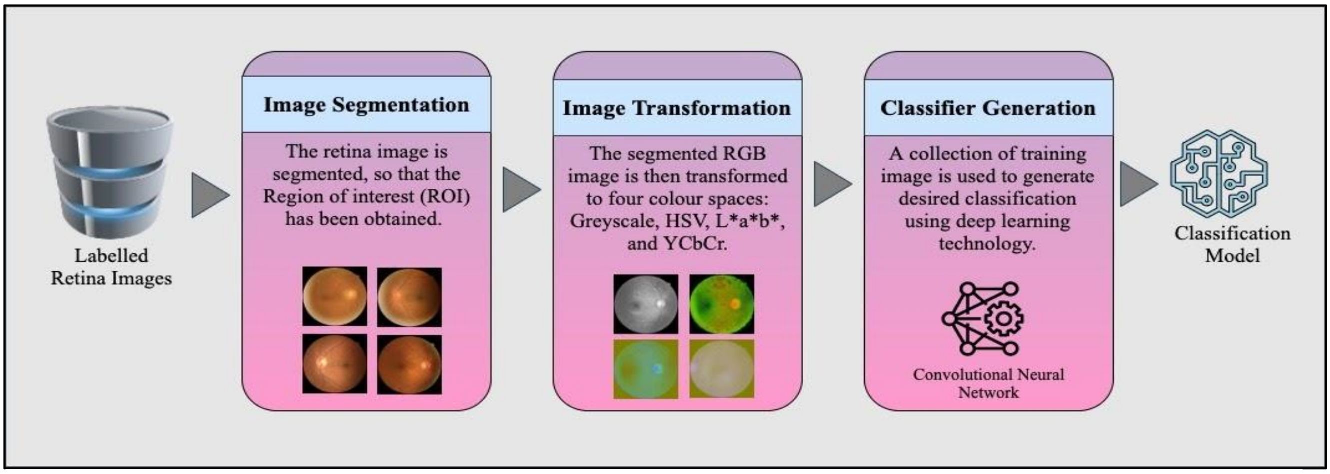

Fig. 2: Schematic illustration the proposed DR detection mining from retina images framework

The work starts with a collection of labelled retina images (each image was labelled by the experts) has been used as a training set. The first process is image segmentation. Each retina image is segmented in order to obtain a region of interest (ROI). The aim of this process is to isolate the area of the retina from the fundus camera image. The second process is image transformation. The objective of this process is to convert RGB image data set to four different colour spaces: (i) Greyscale colour space, (ii) HSV (Hue, Saturation, and Value) colour space, (iii) L*a*b* (CIE L*a*b*, lightness, green–red component, and blue-yellow component) colour space, and (iv) YCbCr (Luma, Blue difference, and Red difference) colour space. The aim of converting the RGB image to other colour spaces is to evaluate what is the most appropriate colour space for diabetic retinopathy classification. The detail of image segmentation and image transformation is later presented in Section 4.

Once each set of colour space images has been transformed, a collection of image data set is then used to generate the desired classifiers. The deep learning mechanism is suggested. More specifically, the convolutional neural network (CNN) is applied to construct the classification models. With respect to the work described in this paper, seven convolutional neural network architectures are suggested: (i) AlexNet, (ii) ResNet50, (iii) DenseNet201, (iv) InceptionV3, (v) MobileNet, (vi) MnasNet and (vii) NASNetMobile. The brief description of CNN and the seven-architectures are described in Section 5.

4. Image Pre-processing

This section presents the image pre-processing process. As mentioned in Section 3 the proposed framework, before the classification generation process can be commenced, the image must be segmented and transformed. Thus, we have combined image segmentation and image transformation processes as image pre-processing process. The processes are illustrated in Fig. 3.

The first process in the proposed framework is image segmentation. The aim of this process is to segment an individual image into an appropriate region of interest (ROI). The original RGB image (Fig. 3(a)) converted into HSV colour space as shown in Fig. 3(b). Only the third layer which is the value channel (from Hue, Saturation and Value) is then isolated in order to fascinate for the next step. Fig. 3(c) shows an example of the value channel image. The threshold value of 0.30 is used to convert the image to a binary image (as presented in Fig. 3(d)). After that Canny edge detection mechanism is applied to detect the border of the eye as given in Fig. 3(e). Once the bounding line of the eye has been detected. The 0.02% of the eye's width and 0.02% of the eye's height are used to make the eye's image located in the centre. Then the ROI of retina image has been obtained as illustrated in Fig. 3(f).

Once the ROI image has been obtained, next process is to align all the retina images to be the same direction. The optic disc of an eye must be located in the left of the target image. To make it happen, the obtained ROI image is then converted to a greyscale image (as shown in Fig. 3(g). Then the threshold value of 230 is used to transform it into a binary image as given in Fig. 3(h). The white area of the binary image presents the eye's optic disc (as the brightest area of the eye). Then the white area of the binary image has been used to examine the location of the optic disc is in the left side or right side. In case the optic disc is in the left just do nothing. On the other hand, if the optic disc is on the right side, the image is then flip to the left side. The example is illustrated in Fig. 3(i). The image resizing process is then applied in order to make the appropriate image size before the classifier generation process could be applied. A collection of obtained ROI images is then resized to be 244x244 pixels (Fig. 3(j)).

When a collection of ROI images is transformed into the same direction, the colour space transformation is then applied in order to investigate what is the most appropriate colour space for constructing the desired classifier. The RGB colour space is then converted into four different colour spaces: (i) YCbCr colour space, (ii) L*a*b* colour space, (iii) HSV colour space, and (iv) Greyscale colour space as presented the examples in Fig. 3(k), Fig. 3(l), Fig. 3(m), and Fig. 3(n), respectively.

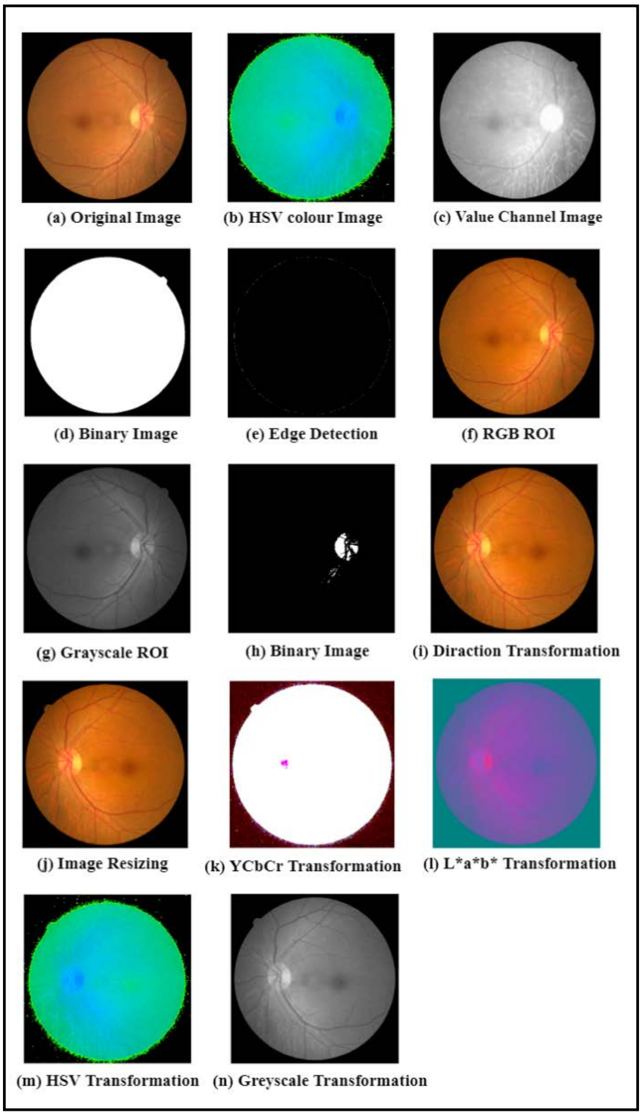

Fig. 3: The image pre-processing process

5. Convolutional Neural Network

This section describes the details of the convolutional neural networks refer to CNN which is used in the classifier generation process mentioned in Section 3. CNN is a type of deep learning algorithms. In general, CNN is applied to analysing the visual imagery such as image classification, image clustering and object recognition. The example of the CNN applications includes face detection, face recognition, street signs detection, tumour detection and many other aspects of visual data. The CNN process is complicated because it comprises of multiple complex mathematical operations. The calculation begins with the configuration of a filter or kernel. The filter is used to extract image properties such as line detection or image sharpening.

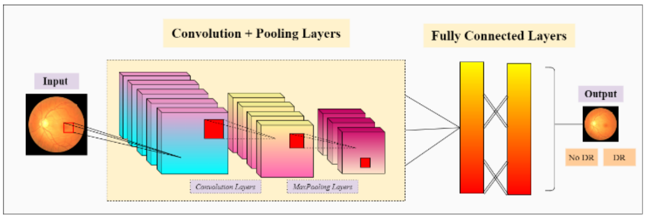

Fig. 4: Convolutional Neural Network architecture

CNN is a three-layers architecture as shown in Fig. 4. From the Fig, it can be observed that the CNN is comprised of three main layers: (i) convolutional layer, (ii) pooling layer, and (iii) fully-connected layer.

Architecture	Year	Size (MB)	Parameter (M)	Layer	Model Description
AlexNet	2012	238	62.3	8	– Convolution Layers – Pooling Layers – Fully Connected Layers
ResNet50	2015	98	25.6	50	– Convolution Layers – Residual Block – Fully Connected Layers
DenseNet201	2017	33	20.2	201	– Convolution Layers – Dense Block – Fully Connected Layers
InceptionV3	2014	92	23.8	42	– Convolution Layers – Inception Block – Fully Connected Layers
MobileNet	2017	16	2.2	-	– Convolution Layers – Depthwise separable Blocks – Fully Connected Layers
MnasNet	2019	18	3.1	-	– Convolution Layers – Depthwise separable Blocks – MBConvolution Block – Fully Connected Layers
NASNet Mobile	2018	23	5.3	-	– Convolution Layers – Reduction Block – Normal Block – Fully Connected Layers

Table 1 The comparison of different CNN architectures

The first layer consists of a set of the learnable kernels which is the process to extract the features from an input image. With respect to the work presented in this paper, the input image is a collection of pre-processed retina images. The second layer is the pooling layer. The objective of this layer is to reduce the resolution of the feature map. Pooling splits, the inputs into the disjoint areas in order to produce one output from each area. There are two pooling operations: (i) maximum pooling, and (ii) average pooling. The final layer is a fully-connected layer which connects every unit (neuron) in one layer to every unit (neuron) in another layer.

The flattened matrix goes through a fully-connected layer to perform image analysis.

In the context of the work presented in this paper, seven different CNN architectures are applied to generate the desired classifiers which can be used to identify the DR retina image in the future. There are seven CNN architectures include (i) AlexNet, (ii) ResNet50, (iii) DenseNet201, (iv) InceptionV3, (v) MobileNet, (vi) MnasNet, and (vii) NASNetMobile. The conclusion of the difference of seven CNN architecture is given in Table 1.

The detail of seven CNN architectures are presented in Sub-section 5.1.1 to 5.1.7, respectively.

5.1.1 AlexNet

AlexNet was firstly introduced by Alex Krizhevsky in 2012 [25] and popular model of CNNs. The architecture of AlexNet consists of three different layers: (i) five convolutional layers, (ii) three pooling layers, and (iii) three fully connected layers. The architecture of Alexnet is shown in Fig. 5.

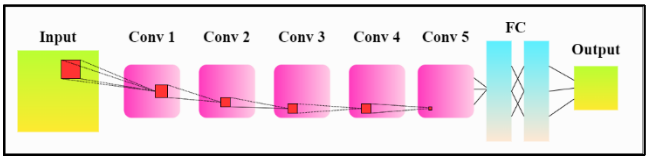

Fig. 5: AlexNet deep convolutional architecture

5.1.2 RestNet50

ResNet50 refers to Residual Neural Network was proposed by Microsoft in 2015 by Kaiming He [26]. ResNet consists of the gated units (gated recurrent units or "skip connections") and features heavy batch normalisation. The architecture of ResNet50 provides a new layer which is 'Residual Block'. ResNet50 is a 50-layers architecture includes (i) 47 convolutional layers, and (ii) three fully connected layers. The architecture of ResNet50 is shown in Fig. 6.

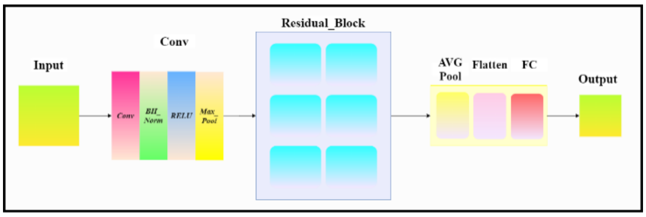

Fig. 6: RestNet50 deep convolutional architecture

5.1.3 DenseNet201

DenseNet201 (Densely Connected Convolutional Networks) is one of the CNN architecture. It was introduced by Gao Huang and member in 2017 [27]. DenseNet201 is a 201-layers architecture. The DenseNet201 builds from the dense block and pooling. In each block, it has the layers densely connected. The architecture of DenseNet201 shown in Fig. 7.

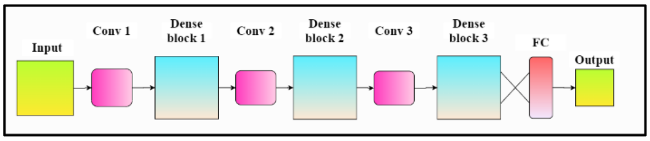

Fig. 7: DenseNet201 deep convolutional architecture

5.1.4 InceptionV3

InceptionV3 was developed from GoogleNet in 2014 and was published in 2016 [28]. Google has already released the Inception four versions. Each version has something different in order to solve a particular problem. InceptionV3 contains 42 deep learning network layers [28]. In this version, it was designed to decrease the number of parameters in order to reduce the complexity of the architecture. The architecture of InceptionV3 is illustrated in Fig. 8.

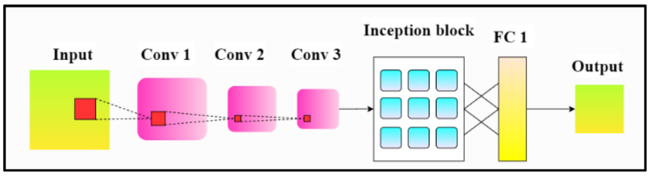

Fig. 8: InceptionV3 deep convolutional architecture

5.1.5 MobileNet

MobileNet is also initially proposed by Google in 2017 [30]. MobileNet is a CNN architecture which developed in order to implement an application on the mobile device. MobileNet is designed to construct a small neural network. The new layer of MobileNet is a depthwise separable convolution. The depthwise separable convolution can be separated into two different layers: (i) depthwise convolution layer and (ii) pointwise convolution layer. The architecture of MobileNet is given in Fig. 9.

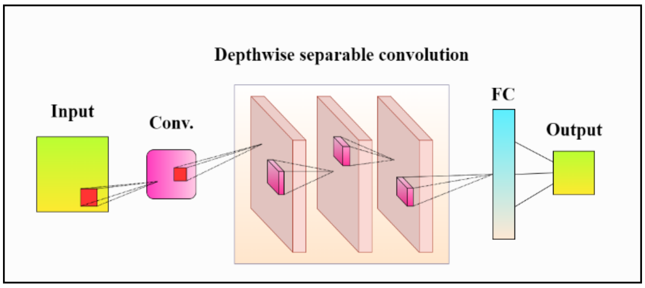

Fig. 9: MobileNet deep convolutional architecture

5.1.6 MnasNet

MnasNet refers to Mobile Neural Architecture Search (MNAS). It was proposed by Google Brain and Google Inc in 2019 [18]. MnasNet has developed from MobileNet in order to improve the performance in NASNet model problem. MnasNet provides a new block which is called MB_Convolution blocks' or MBConv. The MBConv contains convolution layer and depthwise convolution layer as shown in Fig. 10.

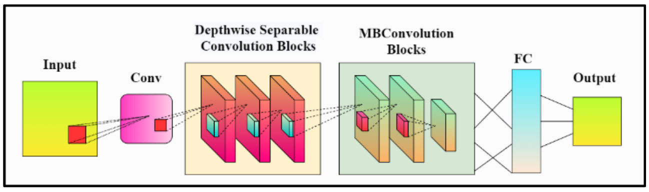

Fig. 10: MnasNet deep convolutional architecture

5.1.7 NASNetMobile

NASNet was firstly introduced by Google Brain in 2018 [31]. NASNetMobile is a version for the mobile device. NAS refers to Neural Architecture Search (NAS). This model was specifically designed. It has specifically two blocks or cells include (i) Reduction block, and (ii) Normal block. The reduction block returns the feature values the height and width and reduced by a factor. The normal block returns the feature values of the same dimension. The architecture of NASNet is shown in Fig. 11.

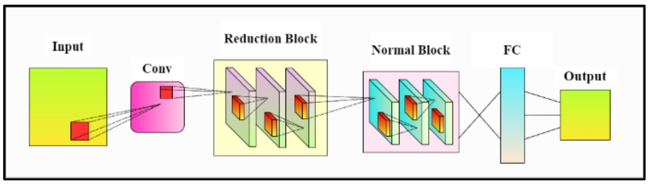

Fig. 11: NASNet deep convolutional architecture

6. Evaluation

The evaluation of the proposed mechanism to DR screening using CNN is presented in this section. The evaluation was conducted by considering a specific case study directed at retina images taken from the fundus camera obtained from open data, Sub-section 6.1 is presented the detail of the data used in this research. A collection of the obtained images was then processed using various image processing techniques in order to make an appropriate image for the classifier construction process. The results of pre-processing are presented in Sub-section 6.2. Once the images were ready, the classifier generation process was then applied. The detail of the model construction process is given in Sub-section 6.3.

6.1 Data Set

This sub-section describes the detail of a set of image data used in this research. A collection of retina images was acquired from the Kaggle competition[1]. Retinal images from Kaggle were provided by EyePACS[2] which is a free platform for retinopathy screening. A collection of obtained images contains 23,513 retina images. The image data set was separated into two different classes by the domain experts provided by the data source. There are two categories include: (i) 11,447 retina images without DR, and (ii) 12,066 retina images with DR. The image was represented in RGB colour space in JPEG (Joint Photographic Experts Group) format. The example of a set of collection of retina images is presented in Fig. 12.

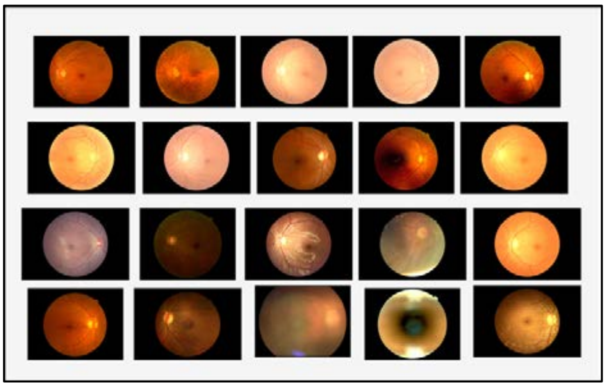

Fig. 12: The example of the image set from Kaggle

6.2 Image pre-processing

The detail of image pre-processing process is given in this Sub-section. As noted in Section 3, prior process before the classifier generation process could be commenced is the image pre-processing mechanism. The pre-processing processes were already described in Section 4. The objective of this process is to make a training set to be ready before the classification learning method could be done. The summary of the image pre-processing methods is presented in Fig. 13.

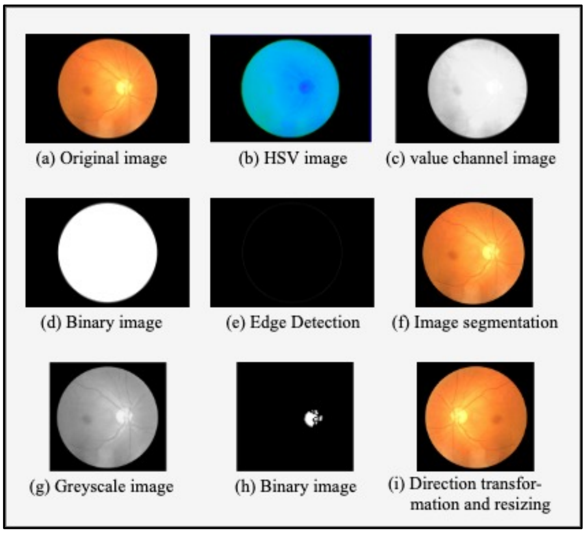

Fig. 13: Image pre-processing process

From the Fig, the example of the original image is given in Fig. 13(a). The RGB image was converted into HSV colour space as seen in Fig. 13(b). The value channel from HSV image was selected as presented in Fig. 13(c). The thresholding value of 0.3 was used to convert the value channel to a binary image as illustrated in Fig. 13(d). The Canny edge detection method was applied to detect the

[1] : https://www.kaggle.com/c/diabetic-retinopathy-detection
[2] : http://www.eyepacs.com/

border of the retina image is presented in Fig. 13(e). The border of the retina was used to calculate the image size before the image could be segmented as the sample in Fig. 13(f). Once the segmented image was obtained, the image was then transformed into the greyscale image as shown in Fig. 13(g). The thresholding value of 230 was then used to detect optic disc from a retina image as the binary image as presented in Fig. 13(h). The optic disc was used to determine that the retina was aligned in the right direction or not, if not transformed the image and finally resized into a size of 244x244 pixels as shown in Fig. 13(i).

The example images after the pre-processing process are given in Fig. 14, and Fig. 15.

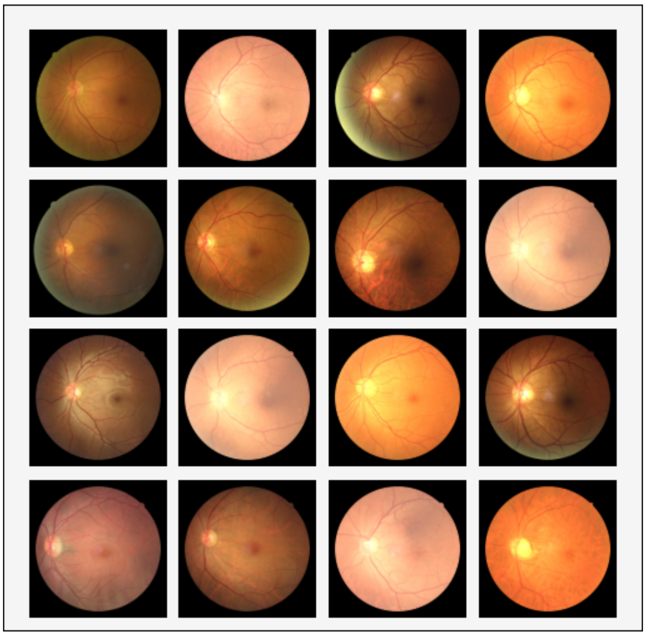

Fig. 14: The processed no-DR retina images

Fig. 14 and Fig. 15 show the example of retina images without and with DR after applied the pre-processing process.

Once the retina image was processed into the appropriate size and direction, the image was then transformed into four different colour spaces: (i) YCbCr colour space, (ii) L*a*b* colour space, (iii) HSV colour space, and (iv) Greyscale colour space.

6.3 Classifier Generation

When the image pre-processing processes have completed, then the classifier construction could be commenced.

The classifier generation using the CNN methods are presented in this sub-section. To this end, two sets of experiments were conducted according to two objectives:

- Objective 1 colour space reports in Sub-section 6.3.1: a set of experiments to identify the most appropriate colour spaces used in the classifier generation process.
- Objective 2 convolutional neural network reports in Sub-section 6.3.2: a set of experiments to determine the most appropriate CNN architectures for DR detection.

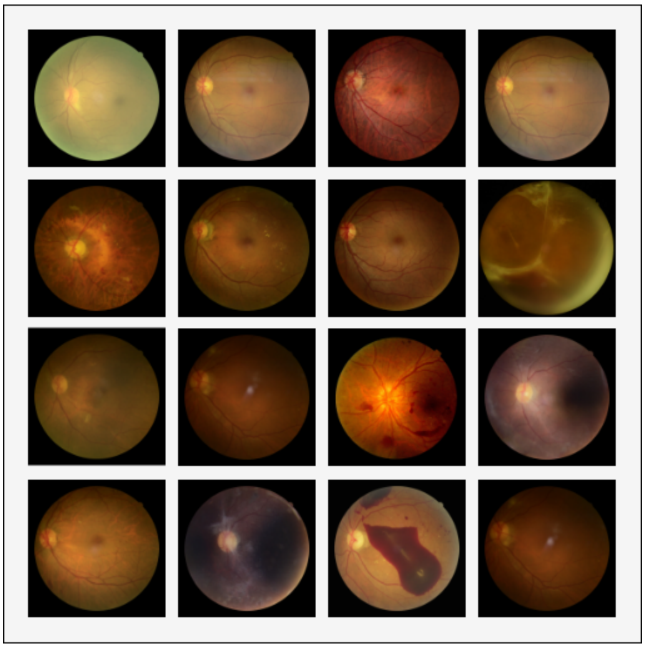

Fig. 15: The processed DR retina images

The experiments conducted with respect to the work presented in this research have been implemented using AMD FX(TM)-8350 8-Core processor computer includes 8 GB RAM and GeForce GTX 1080/PCis/SSE2 GPU 8 GB. The operating system is Ubuntu 18.04.2 LTS. Softwares includes (i) Python 3.7.3, (ii) Miniconda 3.4, (iii)Tensorflow-GPU 1.13.1, (iv) Keras 2.3.1, (v) OpenCV-Python 4.1.1.26, and (vi) Numpy 1.16.4.

The classification performances were recorded in term of accuracy. Eighty percent (80%) of the sample was used as the training set and 20% was used as the test set. The accuracy in the parentheses refers to accuracy values using test set data.

6.3.1 Colour Space

This sub-section presents the results from the first of the evaluation objectives. Thus, a set of experiments were conducted in order to identify the most appropriate colour spaces in term of classification performance.

There were five different colour spaces: (i) RGB colour space, (ii) Greyscale colour space, (iii) HSV colour space, (iv) L*a*b* colour space, and (v) YCbCr colour space. For the experiments conducted in this objective, the AlexNet learning method was applied because it was found these tended to generate the best performance classifier (as will be seen from the following reported results). The results obtained are presented in Table 2.

Colour Space	Train		Test		Time (s/epoch)
	Acc	Loss	Acc	Loss	
RGB	**98.42**	0.0496	81.32	0.7150	298
Greyscale	96.65	0.0971	64.84	1.2130	753
HSV	98.11	0.0563	78.47	0.8079	603
L*a*b*	98.02	0.0606	74.86	0.8472	302
YCbCr	98.06	0.0617	78.62	0.7469	342

Table 2 The results obtained from the five colour spaces

From Table 2, it can be observed that the best result was obtained by using the RGB colour space image data set with the best recorded accuracy value of 98.42% (81.32%) with a loss rate of 0.0496. RGB shows that the running time 298 seconds per epoch. While HSV, L*a*b*, and YCbCr image sets produced the classification performances with 98.11% (78.47%), 98.02% (74.86%) and 98.06% (78.62%), respectively. On another hand, the greyscale colour space image data produced the worst results with an accuracy value of 96.65% (64.84%).

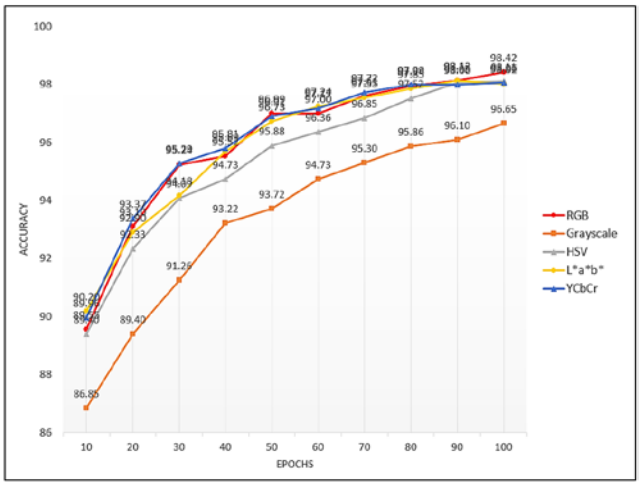

Fig. 16: The classification performance results obtained from using different colour spaces

The graph given in Fig. 16 gives a different perspective on the recorded results in Table 2. From the graph, it can be observed that: (i) the highest classification performances were obtained from RGB data set, (ii) the lowest classification performance obtained from greyscale colour space data set, and (iii) the experiment with higher epochs produced the better classification performance on every data sets.

It should be noted that the doctors or clinicians who analyse retina images for DR detection typically consider RGB colour space image.

6.3.2 Convolutional Neural Network

This sub-section reports on the evaluation conducted from the second objective. Thus, a set of experiments were conducted to identify the most appropriate CNN architectures in term of classification performance.

There were seven different CNN architectures: (i) AlexNet, (ii) ResNet50, (iii) DenseNet201, (iv) InceptionV3, (v) MobileNet, (vi) MnasNet and (vii) NASNetMobile. For the experiments, the RGB colour space image data set was used (because experiments reported above have shown that the RGB colour space produced the best result). The results obtained are presented in Table 3.

CNNs	Train		Test		Time (s/epoch)
	Acc	Loss	Acc	Loss	
AlexNet	**98.42**	0.0496	81.32	0.7150	298
ResNet50	76.09	0.5181	67.04	0.6112	809
DenseNet201	93.02	0.1784	50.01	4.9271	328
InceptionV3	59.74	0.6550	58.69	0.6605	331
MobileNet	75.39	0.5058	66.48	0.6809	831
MnasNet	95.37	0.1775	71.18	1.3828	288
NASNetMobile	71.19	0.5778	70.70	0.6907	361

Table 3. The results obtained from the five colour spaces

From Table 3 indicates that the best result was obtained from the AlexNet learning mechanism with the best-recorded accuracy value of 98.42% (81.32%) and the running time of 298 seconds per epoch. While InceptionV3 learner produced the worst result with an accuracy value of 59.74% (58.69%) running time 331 seconds per epoch.

The obtained results can be illustrated on a different perspective and presented in Fig. 17. From the graph, it can be seen that: (i) the best classifier in term of classification performance was recorded from AlexNet learner with an accuracy value of 98.42% (81.32%), (ii the lowest accuracy comes from the InceptionV3 with an accuracy value of 59.74% (58.69%), however if when the accuracy of test data was considered, DenseNet201 also produced the worst performance with an accuracy value 93.02% (50.01%) with very high loss value on the test set at 4.9271, and (iii) using multiple epochs experiments produced the higher classification performance.

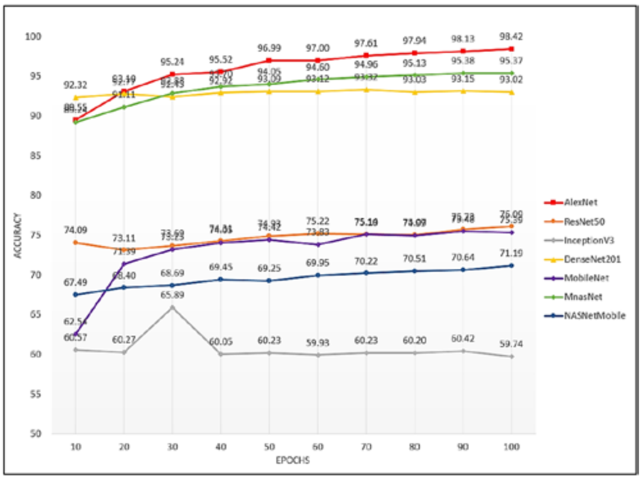

Fig. 17: The classification performance results obtained from using different CNN learning architectures

The AlexNet learning method produced better results, in terms of classification performance than other CNN architectures. It was conjectured that this was because AlexNet has 62 million trainable parameters. Also, each layer of the architecture of AlexNet has deep technical and math to calculate. So it gave higher accuracy than other architectures.

It can be concluded that AlexNet deep learning mechanism is considered as the most appropriate CNN architecture in term of DR detection application.

7. Conclusion

A framework for the diabetic retinopathy (DR) detection from retina images taken from the fundus camera has been presented. The main idea presented in this paper is to generate a classifier using the convolutional neural networks (CNN) for future DR detection. To evaluate the proposed classifiers, experiments were conducted using 23,513 retina test images. The main findings of this study can be presented as follows:

- A mechanism for retina image pre-processing process in order to facilitate in the classifier generation process.
- In order to identify the most appropriate colour spaces to be used with respect to the AlexNet architecture for DR screening, five different colour spaces were considered: (i) RGB, (ii) Greyscale, (iii) HSV, (iv) L*a*b*, and (v) YCbCr. The best result obtained was found from the RGB colour space image data set with a recorded accuracy value of 98.42% and 81.32% for training and test sets, respectively.
- In the context of identifying the most appropriate CNN architecture to be adopted for the DR detection, seven CNN architectures were considered: (i) AlexNet, (ii) ResNet50, (iii) DenseNet201, (iv) InceptionV3, (v) MobileNet, (vi) MnasNet and (vii) NASNetMobile. The best result obtained from AlexNet method with a recorded accuracy value of 98.42% and 81.32% for training and testing sets, respectively. It was conjectured that this was because AlexNet has a deeper and complex architecture than other CNN architectures. Therefore, it could extract the salient features which can be used to distinguish between DR and Non-DR images.

Thus, in conclusion, the proposed framework produced the good results indicating a good way to the DR screening. However, in the future work, more CNN techniques may be carried out, AlexNet tuning could be worked through and the larger training data sets should be conducted.

References

[1] Habtamu Wondifraw Baynes. Classification, Pathophysiology, Diagnosis and Management of Diabetes Mellitus. Journals of Diabetes and Metabolism, 2015.

[2] Association, A. D., et al. Diagnosis and classification of diabetes mellitus. Representation Diabetes Care, January 2013.

[3] Avula Benjamin and Chandan Chakraborty. Detection of Hard Exudates in Retinal Fundus Images Using Deep Learning. In Proceedings of the 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pages 465-469, Kitakyushu, Ja-pan, February 2019.

[4] Oliver Faust, Rajendra Acharya U., E. Y. K. Ng, Kwan-Hoong Ng and Jasjit S. Suri. Journal of Medical Systems, February 2012.

[5] Victor Alves, Jose Neves, Moreira Maia and Luis Nelas. Medical image processing using ex-tended logic programming and artificial neural networks. In Proceedings of the International Symposium on VIPromCom Video/Image Processing and Multimedia Communications, pages 39-42, Zadar, Croatia, June 2002.

[6] Akara Thammastitkul, and Bunyarit Uyyanonvara. Diabetic Retinopathy Stages Identification Using Retinal Images, In Proceedings of the Conference on Intelligent Computing, Computer Science & Information Systems (ICCSIS-16), Pattaya, Thailand, April 2016.

[7] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad and Andrew H. Beck. Deep Learning for Identifying Metastatic Breast Cancer, representation arXiv preprint arXiv:1606.05718, 2016.

[8] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad and Andrew H. Beck. Deep Learning for Identifying Metastatic Breast Cancer, representation arXiv preprint arXiv:1606.05718, 2016.

[9] Hany Kasban, Mohsen El-boundary and Dina Salama. A Comparative Study of Medical Imaging Techniques. International Journal of Information Science and Intelligent System. 4, April 2015

[10] Sujittra Sa-ngiem, Kwankamon Dittakan, Kanya Temkiatvises, Sirisak Yaisoongnern and Kongkiat Kespechara, Cerebral Microbleed Detection by Extracting Area and Number from Susceptibility Weighted Imagery using Convolutional Neural Network, May 2019, Journal of Physics: Conference Series, Volume 1229, pp 012038.

[11] Mark us Gudmundsson, Es-sam A. El-Kwae and Mansur R. Kabuka. Edge detec-tion in medical images using a genetic algorithm. IEEE Transactions on Medical Imaging, pages 469-474, June 1998.

[12] FengLi Yu, Jing Sun, Annan Li, Jun Cheng, Cheng Wan and Jiang Liu. Image Quality Classification for DR Screening Using Deep Learning, In Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 664-667, Seogwipo, South Korea, September 2017.

[13] Wei Yang, Liming Zhong, Yang Chen, Liyan Lin, Zhentai Lu, Shupeng Liu, Yao Wu, Qianjin Feng and Wufan Chen. Predicting CT Image From MRI Data Through Feature Matching With Learned Nonlinear Local Descriptors, In IEEE Transactions on Medical Imaging, pages 977-987, January 2018.

[14] Piyamas Suapang, Kobchai Dejhan and Surapun Yimmun. Medical Image Processing and Analysis for Nuclear Medicine Diagnosis, In International Conference on Control, Automation and Systems 2010, pages 2448-2451, Gyeonggi-do, South Korea, December 2010.

[15] Euclides de Moraes Barros Jr, Michel Sehn, Aldo von Wangenheim and Dirk Krechel. In Proceedings of the IEEE Symposium on Computer-Based Medical Systems, pages 189-194, USA, August 2002

[16] Asif Khan, Jian-Ping Lii and Riaz Ah-med Shaikw. Medical Image Processing using Fuzzy Logic In Proceedings of the 12th International Com-puter Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), pages 163-167, Chengdu, China, June 216

[17] Enrique V. Carrera, Andres Gonzalez and Ricardo Carrera. Automated detection of diabetic retinopathy using SVM. In Proceedings of the IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), pages 1-4, Cusco, Peru, October 2017.

[18] Tanapat Ratanapakorn, Athi-wath Daengphoonphol, Nawapak Eua-Anant and Yosanan Yospaiboon. Digital image processing software for diagnosing diabetic retinopathy from fundus photographs. Journal of Clinical Ophthalmology, Auckland, New Zealand, 2019

[19] Dilip Singh Sisodia, Shruti Nair and Pooja Khobragade. Diaebtic Retinal Fundus Images: Preprocessing and Feature Extraction for Early Detection of Diabetic Retinopathy, Biomedical and Pharmacology Journal 10, pages 615-626, June 2017.

[20] Sohini Roychowdhury, Stu-dent Member, Dara D. Koozekanani, and Keshab K. Parhi. DREAM: Diabetic Retinopathy Analysis Using Machine Learning, In IEEE journal of biomedical and health informatics 18, pages 1717-1728, December 2013.

[21] Pilar Perez Conde, Jorge de la Calleja, Antonio Benitez and Ma. Auxilio Medina. Image-Based Classification of Diabetic Retinopathy using Machine Learning. In Proceedings of the 12th International Conference on Intelligent Systems Design and Applications (ISDA), pages 826-830, Kochi, India, January 2013.

[22] Nikita Gurudath, Mehmet Celenk, and H. Bryan Riley. Machine Learning Identification of Diabetic Retinopathy from Fundus Images. In Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium (SPMB). pages 1-7, Cusco, Peru, October 2017

[23] Kanungo, Y.S., Srinivasan, B., & Choudhary, S. (2017). Detecting diabetic retinopathy using deep learning. 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 801-804.

[24] Benzamin, A., & Chakraborty, C. (2018, June). Detection of Hard Exudates in Retinal Fundus Images Using Deep Learning. In 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (pp. 465-469). IEEE.

[25] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks, In Advances in Neural Information Processing Systems, pages 1097-1105, January 2012.

[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep Residual Learning for Image Recognition, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770-778, Las Vegas, NV, USA, December 2016.

[27] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q.Densely connected convolutional networks. InProceedings of the IEEE conference on computer vision and pattern recognition (2017), pp. 4700–4708.

[28] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. (2015). Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition pp. 1—9.

[29] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z.Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition(2016),pp. 2818–2826.

[30] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto and Hartwig Adam. MobileNet: Efficient Convolutional Neural Network for Mobile Vision Application, representation arXiv:1704.04861, April 2017.

[31] Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. V.Learning transfer able architectures for scalable image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (2018), pp. 8697–8710.

Biographies

Tassanee Hattiya is a mater student in the information of technology major at College of Computing, Prince of Songkla University, Phuket Campus, Thailand.

Kwankamon Dittakan is a lecturer at Prince of Songkla University, Thailand since 2005. Her research interests are focused on image understand, signal analysis and others related to the data science. She has been conducting research in the field of data science since 2011.The initial work was image classification as her PhD thesis. She obtained her PhD from the Department of Computer Science, University of Liverpool, UK.

Salang Musikasuwan is currently an Assistant Professor in Computer Science at Faculty of Science and Technology, Prince of Songkla University, Thailand. He received his PhD in Computer Science from The University of Nottingham, United Kingdom. His research interests are machine learning techniques, artificial intelligence, data analytics, and fuzzy set and system applications.