

## RESEARCH ARTICLE

# Adenoma Dysplasia Grading of Colorectal Polyps Using Fast Fourier Convolutional ResNet (FFC-ResNet)

**MAY PHU PAING**  **AND CHUCHART PINTAVIROOJ**

Department of Biomedical Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

Corresponding authors: May Phu Paing (may.pa@kmitl.ac.th) and Chuchart Pintavirooj (chuchart.pi@kmitl.ac.th)

This work was supported by the King Mongkut's Institute of Technology Ladkrabang (KMITL) Research and Innovation Services (KRIS) under Grant KREF186505.

**ABSTRACT** Colorectal polyps are precursor lesions of colorectal cancer; hence, early detection and dysplasia grading of polyps are essential for determining cancer risk, the possibility of developing subsequent polyps, and follow-up recommendations. The significant contribution of this study is the development of an enhanced deep-learning model called Fast Fourier Convolutional ResNet (FFC-ResNet) to classify dysplasia grades of polyps. It is based on the ResNet-50 architecture and uses cross-feature fusion, which combines local features extracted by traditional spatial convolution with global features extracted by Fourier convolution. Due to the compensatory effect between local and global features, the learnability and performance of FFC-ResNet have increased. The proposed FFC-ResNet was developed and tested using UniToPatho, a dataset containing 7000  $\mu\text{m}$  and 800  $\mu\text{m}$  hematoxylin-and-eosin (H&E)-stained colorectal images. And a favorable performance of sensitivity 0.95, specificity 0.93, balance accuracy 0.94, precision 0.95, F1 score 0.95, and AUC 0.99 was obtained using 800  $\mu\text{m}$  polyp patches.

**INDEX TERMS** Colorectal cancer, computer-aided diagnosis, deep learning, frequency domain, principal component analysis.


## I. INTRODUCTION

Colorectal cancer (CRC), commonly known as colon and rectal cancer, is the third most common and second most fatal cancer worldwide [1]. Most CRCs arise from precursor lesions called colorectal polyps, which are aberrant growths of the mucosal layer of the large intestine [2]. Typically, these precancerous lesions take several years to transform into CRCs [3]. Hence, removing them before they transform into CRCs is the most crucial step in preventing CRC. This reason strongly motivates the necessity for early diagnosis. In current clinical practice, a variety of screening tests are available to detect polyps, such as the fecal occult blood test (FOBT), fecal immunochemical test (FIT), stool DNA testing, X-ray with barium enema, sigmoidoscopy, colonoscopy, and CT colonography [3], [4]. However, according to population-based studies, colonoscopy is the most common and preferred

test [4]. In addition, it is recommended as a gold-standard test by international guidelines.

Through colonoscopy, practitioners can detect colorectal polyps and analyze their physical indications such as size, pit pattern, shape (sessile or pedunculated), and elevation (protruded, superficially elevated, flat, or depressed) [5]. These indicators help in classifying different types of polyps and predicting cancer risks. Nevertheless, evaluating merely the physical features is inadequate to identify the specific dysplasia grades of polyps. Cell replication inside the polyp tissue varies across polyp types, and thus analysis of cell manners is imperative for grading and determining the probability of polyps transforming into CRC. For this reason, the histopathological examination of polyp tissues is usually performed after a colonoscopy. Pathologists then analyze the characteristics of the glands and cells inside the polyp tissue samples to investigate the manner and rate of cell replication.

Histopathological images of polyps are high-resolution images of tissue samples, typically stored at different

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico .

magnifications [6]. Each magnification scale can provide different visualizations: glands (coarser structures) require low magnification, whereas individual cells (finer structures) require higher magnification. Pathologists can analyze the characteristics of glands and cells from histopathological images and determine the degree of malignancy, the possibility of developing subsequent polyps, and follow-up recommendations. According to the US Multi-Society Task Force guidelines on colorectal cancer, the histopathological characterization of colorectal polyps is critical in determining the risk of CRCs and future rates of patient surveillance [7].

Colorectal polyps can be pathologically classified into three main groups: normal, hyperplastic (HP), and adenoma. Both the normal and hyperplastic polyps showed no signs of malignancy. However, normal cells tend to become cancerous or malignant following unusual changes. More specifically, cells with abnormal changes but not cancer can be defined as hyperplastic. Unlike normal and hyperplastic cells, adenoma cells exhibit a higher malignancy. Therefore, polyps containing these cells have a higher potential for CRC development. Adenoma cells are of two structures: tubular (branching tubules) and tubulovillous (tubules and leaf- or finger-like villi). Based on these structures, adenoma cells can be further divided into four types: (i) tubular adenoma high-grade dysplasia (TA.HG), (ii) tubular adenoma low-grade dysplasia (TA.LG), (iii) tubulovillous adenoma high-grade dysplasia (TVA.HG), and (iv) tubulovillous adenoma low-grade dysplasia (TVA.LG) [8].

The existing burden in histopathological analysis of colorectal polyps is a manual method that usually takes time and sometimes leads to inter-observer variation in the diagnosis results. Furthermore, although the manual classification of binary classes between non-malignant (normal, hyperplastic) and malignant (adenoma) is generally reliable, the multi-classification of different polyp types and specific dysplasia grades remains a significant challenge, even for expert pathologists. In an effort to overcome these difficulties, the development of more reliable, rapid, and automated diagnostic methods has recently attracted considerable interest and demand [9].

## II. RELATED WORKS

Several studies have focused on the development of automated diagnostic methods for colorectal polyp classification. Owing to their enhanced performance and efficacy, the most recent polyp diagnosis methods in the literature have applied different deep-learning approaches. For example, [7] used state-of-the-art deep learning models, including AlexNet, VGG, GoogleNet, and a variation of ResNets, to identify five different colorectal polyp structures: hyperplastic, sessile serrated, traditional serrated, tubular, and tubulovillous/villous. They obtained the best classification performance using ResNet-D, showing 93.00% accuracy, 89.7% precision, 88.3% recall, and 88.8% F1 score. Comparably, [12] also used ResNet to classify four polyp types namely tubular adenoma (TA), tubulovillous or villous adenoma

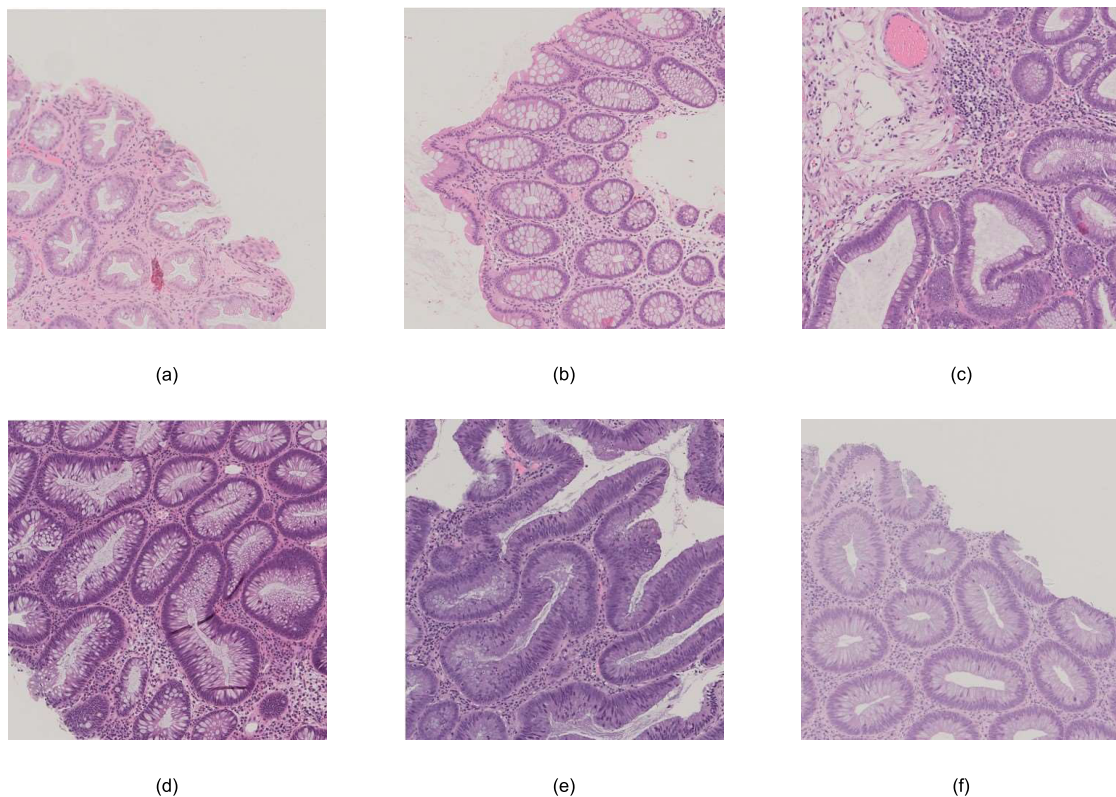
(TVA), hyperplastic polyp (HP), and sessile serrated adenoma (SSA). They compared ResNet with the manual prognostication of local pathologists and observed that ResNet outperformed the manual method by an accuracy of 2.1%. Alternatively, in the studies of [10] and [11], the use of Inception-v3 was found. As the dimensions of the whole slide image (WSI) are extremely large, they initially divided the WSI into tiles or patches and then conducted two classification levels. Reference [10] used Inception-v3 as a tile-level feature extractor, followed by a recurrent neural network (RNN) for slice-level classification. However, in [11], slice-level information was aggregated by the topological connection of tile clusters, and the final cancer status on the WSI was determined. The reported performance of [10] was 0.99 AUC for classifying three classes: adenocarcinoma, adenoma, and non-neoplastic, while [11] achieved an AUC of 0.988 for the binary classification of CRCs and non-CRCs. Although the performances of these cited related works [7], [10], [11], [12] were remarkable, they focused only on polyp type or CRC classification. Adenoma dysplasia grading of polyps that can indicate information on low- or high-risk CRC cannot be obtained from their studies. Indeed, the grading information is more critical and necessary for treatment and follow-up suggestions to reduce colorectal cancer.

Only a limited number of computer-aided schemes [6], [8], [13], [14], [15] for dysplasia grading have been reported in the literature. Reference [6] classified polyp tissues into five different types: cancer, high-grade dysplasia, low-grade dysplasia, hyperplastic, and normal. They exploited multi-scale task multiple instance learning (MuSTMIL), which combines contextual and detailed pathological information from multi-scale patches with different magnifications.

Similarly, [8] also conducted a dysplasia grading using multi-resolution hematoxylin-and-eosin (H&E)-stained colorectal images. The main difference between [6] and [8] is that the model in [8] was based on a three-cascaded style of ResNet and able to predict six different classes of polyps including grades: normal (NORM), hyperplastic (HP), low-grade tubular adenomas (TA.L), high-grade tubular adenomas (TA.H), low-grade tubulovillous adenomas (TVA.L), and high-grade tubulovillous adenomas (TVA.H). First, [8] hypothesized that hyperplastic (HP) polyps are related to the more minor details of tiny gland edges; thus, they were segregated using a finer resolution of  $800 \mu\text{m}$  ( $1812 \times 1812$  pixel dimension) with the help of a binary ResNet classifier. Subsequently, the second classifier on a coarser scale of  $7000 \mu\text{m}$  ( $15855 \times 15855$  pixel dimension) was used to classify the remaining types of polyps, that is, normal and adenomas (TA.L, TA.H, TVA.L, and TVA.H), as their best distinguishability strongly relied on large-scale macrostructures and entire glands. Finally, the last classifier was used for low-and high-grade TA and TVA. Their method thoroughly considered the pathological nature of polyps and used different magnifications in a cascaded style of classifiers. Nevertheless, the method in [8] achieved an accuracy of

**TABLE 1.** Some computer-aided colorectal polyp grading schemes in literature.

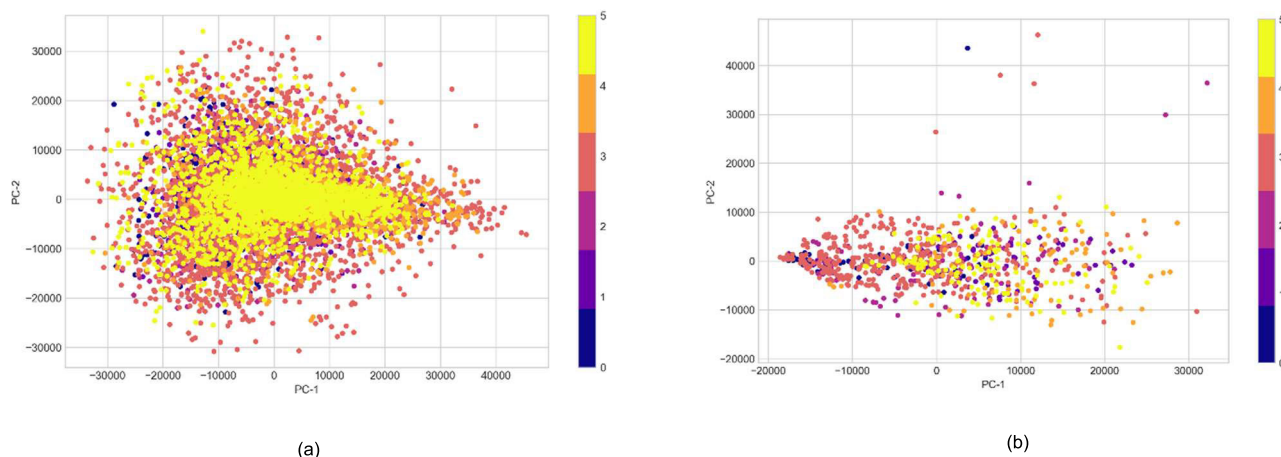
Ref (Year)	Dataset	Number of classes: Names	Based Deep-learning Model	Performance (Accuracy)
[6] (2021)	Private dataset	<b>5 Classes:</b> Cancer, High-grade dysplasia, Low-grade dysplasia, Hyperplastic (HP), Normal	ResNet-34	87%
[8] (2021)	UniToPatho	<b>6 Classes:</b> Normal (NORM), Hyperplastic (HP), Tubular adenomas low grade (TA.L), Tubular adenomas high grade (TA.H), Tubulovillous adenomas low grade (TVA.L), Tubulovillous adenomas high grade (TVA.H)	ResNet-18	67%
[13] (2020)	Private dataset	<b>6 Classes:</b> Low grade tubular, High grade tubular, Low grade villous, High grade villous, Low grade tubulovillous and High grade tubulovillous	DeepLabv2 with ResNet-34	90% (WSI)
[14] (2022)	DeepHealth	<b>6 Classes:</b> Normal (NORM), Hyperplastic (HP), Tubular adenoma high-grade dysplasia (TA.HG), Tubular adenoma low-grade dysplasia (TA.LG), Tubulo-villous adenoma high-grade dysplasia (TVA.HG), Tubulovillous adenoma low-grade dysplasia (TVA.LG)	ResNet-18	70% (WSI)
[15] 2022	CRC	<b>3 Classes:</b> Normal, Low grade, High grade	ResNet-34	90.19% (WSI)

**FIGURE 1.** Example of polyp patches in UniToPatho (a) HP (b) NORM (c) TA.HG (d) TA.LG (e) TVA.HG and (f) TVA.LG.

67% for the patch-level classification, which requires further improvement.

Moreover, [13] developed an automated polyp classification and grading system using an improved deep neural network based on DeepLabv2 with ResNet-34. They classified polyps into three main types: tubular (T), villous (V) and

tubulovillous (TV). Each type was then divided into two grades (high and low); hence, their study had six different grades of polyps. As well, a recent study [14] proposed polyp grading using different resolutions. They used eight possible magnifications ranging from 300 to 1000  $\mu\text{m}$  with a step size of 100  $\mu\text{m}$ . They classified polyps into six different types



**FIGURE 2.** Visualization of class separability by conducting principal component analysis (PCA), (a) 800  $\mu\text{m}$  patches and (b) 7000  $\mu\text{m}$  patches.

using a simple ResNet-18 classifier rather than a cascaded or ensemble style.

Furthermore, another study [15] proposed colorectal polyp grading (three classes: normal, low-grade, and high-grade) using a semi-supervised approach containing three-stepped classifiers. Initially, they trained a ResNet-34 as a supervised tile classifier on a smaller data set. Subsequently, the trained weights were applied as the initial weights for an additional supervised classifier trained on an extensive dataset. These two classifiers produced tile-level classification results. For slide-level classification, the feature vectors of each tile were extracted by removing the fully connected layer of the second classifier and were fed into six different models. As a slide-level result, they reported 90.19% accuracy, 98.8% sensitivity, and 85.7% specificity.

Table 1 summarizes the aforementioned colorectal polyp grading schemes in the literature and their performances. From this table, we can see that different studies used different datasets and grading scales. Among these, the dysplasia grades used by [8] and [14] were more specific than those used by others. Therefore, we chose to utilize UniToPatho [8] as the input dataset for this study and to grade according to their grading scales.

The ultimate goal of this study was to develop a deep-learning-based colorectal polyp grading system that can provide better performance and more reliable results. Although some methods in the literature were performed at two levels (patch-level and whole slide image (WSI)-level), this study primarily focuses on patch-level classification for the following reasons.

- 1) First, we intend to classify polyps into more detailed grades, that is, six different grades (NORM, HP, TA.HG, TA.LG, TVA.HG, and TVA.LG) because knowing specific grades can indicate better treatment decisions. As shown in Table 1, the UniToPatho and DeepHealth datasets provide annotations for these particular grades. However, only UniToPatho

is publicly available but offers only the patch-level annotations.

- 2) Second, patch-level classification is fundamental and crucial to the final WSI-level results. If the patch-level prediction is inaccurate, the WSI-level prediction will be incorrect. Therefore, this study focuses mainly on patch-level classification using an enhanced deep-learning model. Subsequently, WSI-level classification will be conducted as a future plan after ensuring patch-level classification performance.

To obtain more precise and reliable patch-level grading results, this study proposes an enhanced deep-learning model called FFC-ResNet as a significant contribution. The proposed FFC-ResNet takes advantage of the ease and effectiveness of convolution in the frequency/Fourier domain. In the frequency domain, the kernels are the same size as the input feature map and non-local respective fields; thus, convolution can extract global features, ensuring sufficient information about polyp grades. Moreover, the proposed FFC-ResNet fuses local and global features to obtain a compensatory effect and improve learnability.

### III. MATERIALS AND METHODS

#### A. MATERIALS

UniToPatho, the dataset provided in [8], was used as the input material for this study. It is downloadable from the IEEE data port [16] and contains 9536 annotated hematoxylin and eosin (H&E)-stained colorectal images, which are the most relevant patches extracted from 292 whole-slide images. There were six types of patches: normal, HP, TA.HG, TA.LG, TVA.HG, and TVA.LG, with two magnifications of 800 and 7000  $\mu\text{m}$ , respectively. We divide the dataset into five folds in order to conduct 5-fold cross validation. Four training folds contain 80% of the total patches, and one validation fold contains 20%. Figure 1 demonstrates some examples of polyp patches showing different grades of adenomatous dysplasia.

## B. METHODS

### 1) DATA ANALYSIS AND PREPROCESSING

The performance of a deep-learning model is strongly correlated with the quality of the input data. Thus, before we propose any method and start any experiments, we initially conducted a data analysis to gain insight and enhance the classification power of the input data. Principal Component Analysis (PCA) is a standard tool for exploratory data analysis and dimensionality reduction. It transforms the pixels of images into a new set of features called principal components (PCs). Simply put, PC is a series of unit vectors; that is,

$$PC = \{PC_1, PC_2, PC_3, \dots, PC_i, \dots, PC_n\}$$

where  $i^{th}$  vector is the direction of the line that best fits the data and is orthogonal to the  $i - 1$  vectors. This transformation effectively compresses a massive amount of information across the images in the dataset into fewer and more salient features.

For visual analysis, we extracted PCs from our polyp patches and plotted  $PC_1, PC_2$  that capture the most and second-most variation in polyp features onto a two-dimensional coordinate, as illustrated in Figure 2. From this plotting, we can visualize the separability between polyp classes. As shown in Figure 2, the scatter plots of  $PC_1, PC_2$  for different classes (0=HP, 1=Normal, 2=TA.HG, 3=TA.LG, 4=TVA.HG, and 5=TVA.LG) are overlapping and highly correlated with each other. This indicates that the complexity between the features in the polyp patches is very high, resulting in low separability. Compared with the  $800 \mu m$  patches (Figure 2a), the  $7000 \mu m$  patches (Figure 2b) seem more separable because  $7000 \mu m$  focuses on coarser structures, similar to the zoomed-in view, thus producing fewer patches.

PCA analysis revealed that reducing the complexity between pixels or features is essential to increasing separability. In general, most deep learning models conduct data preprocessing before model training. Such preprocessing is a simple pixel-wise or feature-wise normalization and standardization that subtracts the mean of the image pixels or features and divides them by their standard deviations. These preprocessing techniques help the pixel values to be centered around zero and have a unit standard deviation [17]. However, although they help increase the model's stability and performance, they cannot effectively compress input data.

Therefore, in our study, we conducted normalization and standardization as well as PCA compression for data preprocessing. We apply PCA to compress and reduce the dimensions of the input images by removing irrelevant and redundant pixels. With the help of PCA, both the dimensions and storage of the input patches can be reduced before feeding them into the deep learning models. This is crucial because training deep learning models using larger images requires extensive computational resources and time. Especially in our research, input polyp patches are very high in image resolution ( $1812 \times 1812$  pixel dimension for  $800 \mu m$  patches and  $15855 \times 15855$  pixel dimension for  $7000 \mu m$  patches) and large in file size, nearly 6.5 MB per image. For this reason,

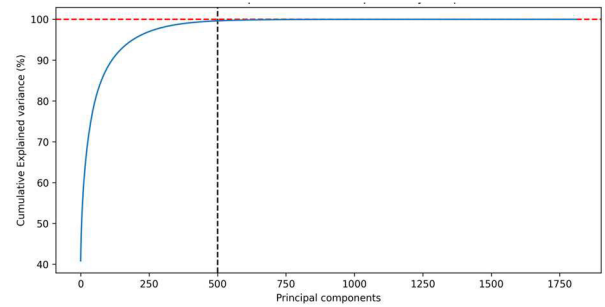


FIGURE 3. Cumulative explained variance by 500 PCs.

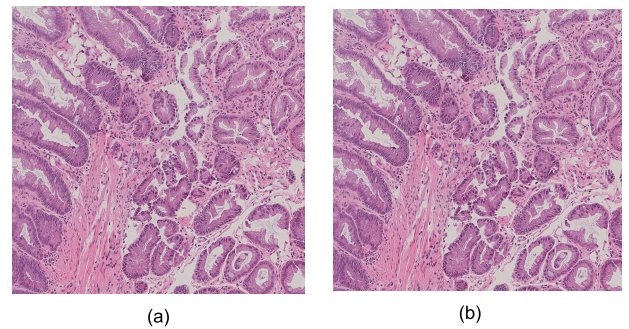
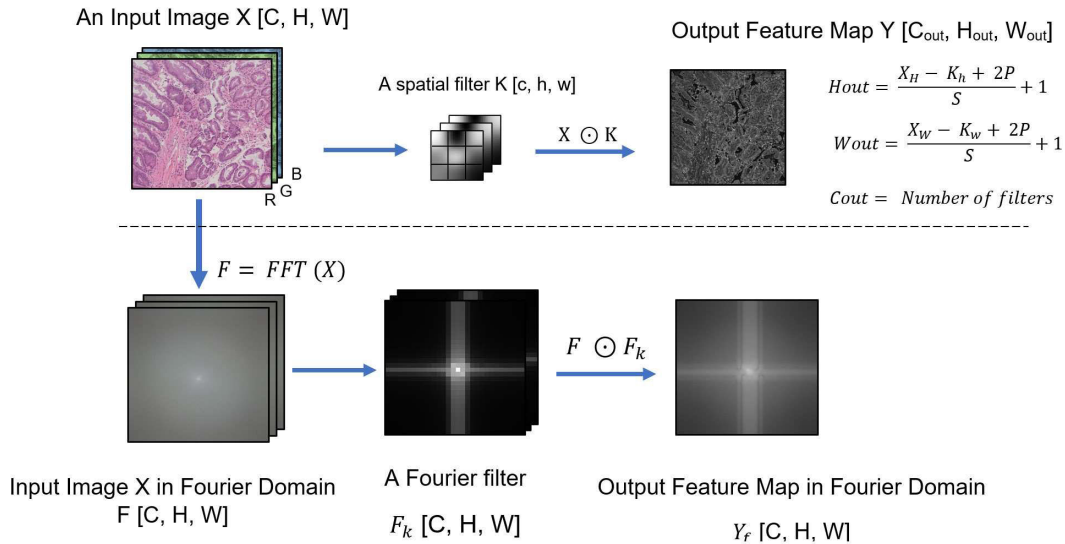


FIGURE 4. Preprocessing of polyp patches using Principal Component Analysis (PCA) compression (a) Original patch and (b) Pre-processed patch.

we tried to downgrade them using PCA at the preprocessing stage. Instead of directly downsampling the input images, we use PCA because it retains only the pixels which have salient features and contribute most to the variance. However, as the PCA is a lossy compression, it removes the details in the input image. Hence, we maintain as much of the quality of the image as possible by controlling the number of principal components (PCs) selected.

We empirically selected 500 PCs and deleted the rest because they had a high possibility of being indistinct or potentially redundant features. Figure 3 demonstrates the plot of the cumulative percentage of variance explained by the PCs. For example, if an original input image has  $1812 \times 1812$  dimensions, there will be 1812 PCs. Among them, we selected only 500 PCs because they can explain nearly 99.99% of the input image's information, as can be seen in Figure 3. 500 PCs will be able to guarantee the minimum loss of data and help to get better predictions while downgrading the image. As a noticeable effect of using PCA, the file size of input images can be reduced from ( $\approx 6.5$  MB) to 4 MB. Moreover, it can also help the transmission and training processes become more manageable and flexible. Figure 4 compares the original polyp patch (Figure 4a) and the output of PCA compression (Figure 4b). We can see no significant visual differences between the before and after PCA compression in these figures, but unnecessary pixel information had been discharged. The other benefit of PCA on model performance will be highlighted in the experimental results and discussion



**FIGURE 5.** Comparison of convolution operations in spatial domain and frequency domain. Spatial domain convolution uses a small filter/kernel that has local respective field. Frequency domain convolution uses a Fourier kernel having the same size as the input image and global respective field.

section (Section IV) by comparing the performance of deep-learning models using original images to those using PCA compression.

## 2) FAST FOURIER CONVOLUTIONAL NETWORK (FFC-NET)

The Fast Fourier convolutional (FFC-Net) network [19] is an enhanced deep-learning model that exploits the efficiency of learning in the frequency domain. One of the existing challenges of conventional CNNs is the high computational effort and time required for the training process [19]. This is because they operate in the spatial domain, and the convolution operation in the spatial domain is expensive. For example, a spatial domain convolution of a two-dimensional (2D) image  $I$  using a kernel  $k$ , requires a sliding window operation, where  $k$  is moved across  $I$  and convolution is computed by the summation of the element-wise product  $\odot$  of kernel value with the corresponding image pixel value.

$$y[i, j] = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} k[m, n] \odot I[i - m, j - n]$$

where  $y$  is the convoluted image,  $i$  and  $j$  are the indices of image  $I$ ;  $m$  and  $n$  are the indices of kernel  $k$ .

According to the Fourier theorem [20], convolution in the spectral or frequency domain can be simply conducted using a single element-wise product, without requiring a sliding window and summation.

$$F(y) = F(k) \odot F(I)$$

where  $F$  is a discrete Fourier transform function,  $F(I)$  is the Fourier transform of image  $I$ ,  $F(k)$  is the Fourier transform of kernel  $k$ , and  $F(y)$  is the output of the convolution in the frequency domain. Figure 5 illustrates the difference between

the spatial- and frequency-domain convolutions of a three-color channel (RGB) polyp image. An inverse Fourier transform (IF) is necessary to transform the Fourier image back into the spatial domain.

Another advantage of using Fourier convolution is that it can alleviate the local respective field problems that commonly occur in spatial convolutions. A respective field is a region of an image or input feature map that can be accessed by one filter [18]. Conventional spatial convolutions frequently use small filters with small receptive fields. In general, the most widely used and acceptable filter size for spatial convolution layers in modern deep neural networks is  $3 \times 3$  or  $5 \times 5$ . Nevertheless, these small filters have a drawback in that they look for very few pixels under their small receptive fields and learn only local features. For this reason, most deep neural networks follow architectures that stack many convolutions deeply with small receptive fields, with the aim of increasing the receptive field linearly or exponentially. However, as a negative result, stacking multiple convolutional layers can increase the effort and complexity of the network. In addition, it is ineffective for some problems, particularly context-sensitive tasks. Convolution in the frequency domain can be implemented by using a non-local respective field. As described in the previous equation, Fourier convolution requires no sliding window operation or summation of element-wise products. Instead, it uses simple element-wise multiplication of the Fourier input and Fourier kernel. Thus, the kernel size of the Fourier convolution is the same as that of the input.

Although Fourier convolution offers such advantages, it also provides tradeoffs. The effort and time required to conduct the DFT and IFT are the most important factors to consider when implementing Fourier CNNs. Unlike other Fourier

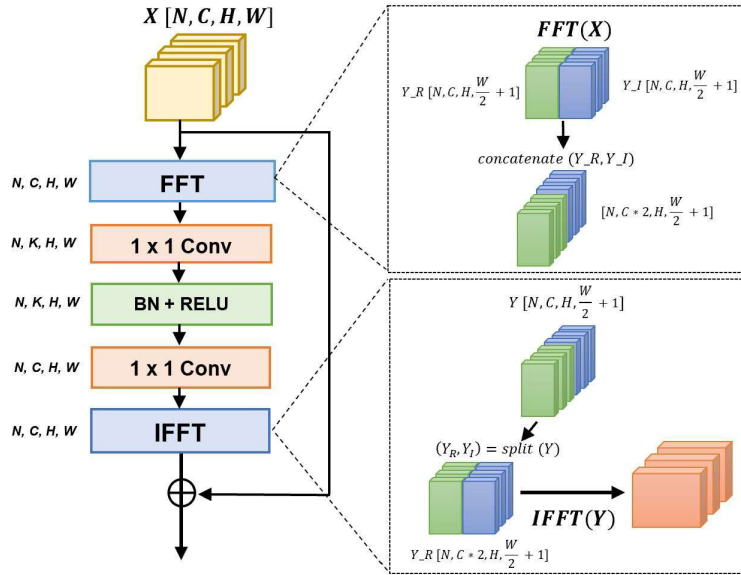


FIGURE 6. Fourier unit of FFC Network, which uses a bottleneck style.

convolution-based networks [21], [22], the FFC proposed in this study applies a Fast Fourier Transform (FFT) with the Cooley-Tukey algorithm and uses a bottleneck style. Figure 6 illustrates the Fourier convolution unit of the proposed FFC. As illustrated in that figure, it was constructed in a bottleneck style, where  $X$  is an input feature map with  $N$ ,  $C$ ,  $H$ , and  $W$  dimensions, which represent the batch size, color channel, height, and width, respectively. FFT and IFFT represent the fast Fourier transform and inverse Fourier transform, respectively. The major purpose of FFT is to transform the input feature map from the spatial domain to the frequency domain so that frequency-domain convolution can be performed on that input feature map. The FFT converts the values of the input feature maps into a Hermitian matrix of complex numbers that contain real and imaginary values. Owing to the conjugate symmetric property of the Hermitian matrix, the feature map in the frequency domain can be transformed back into the spatial domain using an inverse Fourier transform (IFFT).

As shown in the diagram of the Fourier unit in Figure 6, the input frequency-domain feature map can be split into two parts: the real part  $Y_R(N, C, H, \frac{W}{2} + 1)$  and the imaginary part  $Y_I(N, C, H, \frac{W}{2} + 1)$ . Then,  $Y_R$  and  $Y_I$  are concatenated as  $Y$  along dimension  $C$  for an easy convolution process. Subsequently, a pair of  $1 \times 1$  convolutions were conducted to create a bottleneck. The use of a bottleneck in a Fourier unit has two main advantages. First, it can reduce the number of parameters and computational burden. Second, it ensures global feature updates, because a  $1 \times 1$  convolution is conducted on the frequency-domain. It does not need to create Fourier kernels/filters of the same size as the input frequency-domain feature map, as shown in Figure 5. The

**Algorithm 1** Comparison of Time Elapsed

**Input:**  $X(N, C, H, W)$ , an input spatial feature map having  $N$  batch sizes,  $C$  channels and  $H, W$  dimensions

**Output:**  $tf, t1$  where  $tf$  is time elapsed for Fourier convolution using element wises product and  $t1$  is time elapsed for  $1 \times 1$  convolution

```
# Fourier Convolution
start_tf = Time.time () # record current time before convolution
ffted_X = FFT (X) # Convert input spatial feature map into Fourier domain
using FFT
s_k = Random (X.size()) # Initilze a random spatial kernel, the same size as
X
ffted_s_k = FFT (s_k) # Convert s_k as a Fourier kernel
convld_Xf = ffted_X * s_k # Convolution by element-wise multiplication
end_tf = Time.time() # record current time after convolution
tf = end_tf - start_tf
```

```
# 1 x 1 Convolution in Fourier Unit of FFC.
start_t1 = Time.time () # record current time before convolution
Y_R, Y_I = FFT (X) # Convert input spatial feature map into Fourier domain
using FFT where Y_R is the real part, and Y_I is the imaginary part
ffted_X1 = Concatenate ([Y_R, Y_I], dim=1)
s_k = Random (1,1) # Initilze a random spatial kernel having the same size
as X
convld_X1 = ffted_X1 ⊙ s_k # Convolution by element-wise multiplication
end_t1 = Time.time() # record current time after convolution
t1 = end_t1 - start_t1
```

kernel size can be fixed at  $1 \times 1$ , because any operation in the spectral or frequency domain has a global receptive field. Simply put, it performs a spatial  $1 \times 1$  convolution on a frequency or spectral domain feature map.

**Algorithm 2** Cross-Feature Fusion

**Input:**  $X(C, H, W)$ , an input spatial feature map having  $C, H, W$  channels, height, and width

**Output:**  $Y(C, H, W)$ , an output feature map that combines non-local and local features

# Cross-feature fusion

$X = \{x^l, x^g\}$  # input feature map is divided into two parts to extract local and non-local features

$\alpha_{in} \in [0, 1]$  # User-defined parameter to split the input feature map

$x^l \in \mathbb{R}^{H, W, (1-\alpha_{in})C}$  # To traditional convolution units to extract local features

$x^g \in \mathbb{R}^{H, W, \alpha_{in}C}$  # To Fourier Unit to extract non-local features

$Y = \{y^l, y^g\}$  # where  $y^l$  is the local feature map and  $y^g$  is the non-local feature map

**TABLE 2.** Comparisons of elapsed time for convolution operations.

	Spatial Domain	Frequency Domain	
		Fourier Kernels same size as input	Fourier Unit in FFC
Input size	(1, 3, 224, 224)	(1, 3, 224, 224)	(1, 3, 224, 224)
Kernels	1 × 1, 64 3 × 3, 64 1 × 1, 64	224 × 224, 64	1 × 1, 64 3 × 3, 64 1 × 1, 64
FFT ( $X$ )	-	0.003 s	0.003 s
Convolution	0.1208 s	0.113 s	0.041 s
IFFT ( $X$ )	-	0.035 s	0.035 s
Total Time	0.1208 s	0.178 s	<b>0.079 s</b>
Output size	(1, 64, 224, 224)	(1, 64, 224, 224)	(1, 64, 224, 224)

Algorithm 1 can be used to compare the time elapsed for a  $1 \times 1$  convolution and a Fourier (elementwise) convolution that uses a Fourier kernel of the same size as the input feature map. Using a polyp patch as an input, that is, batch size = 1, where  $X$  (1, 3, 224, 224), we can compare the time elapsed of three different convolutions, namely (i) spatial convolution (ii) Fourier convolution using Fourier kernels having the same sizes as input feature map [21], [22], and (iii) Fourier convolution using proposed Fourier unit which uses  $1 \times 1$  convolution. Table 2 summarizes the comparative results of these three convolutions. Based on those results, it is obvious that the proposed Fourier convolution in FFC-Net is approximately two times faster than its counterparts. Indeed, although the FFCNet requires FFT and IFFT to convert the original image to frequency domain (DFT) and its inverse (IFFT) processes, these two operations have to be performed only at the very first and last layers of FFCNet.

Another important improvement of the FFC is that it can apply both spatial and Fourier convolutions in an add-on style. To do so, the FFC uses a cross-feature fusion method that combines the non-local features extracted by the Fourier unit with the local features extracted by spatial convolutions. Algorithm 2 describes the step-by-step procedure

**TABLE 3.** Hyperparameter values for all models.

Hyperparameters	Values
Batch size	32
Epochs	100
The split ratio for local and global ( $\alpha_{in}$ )	0.5
Learning rate	0.1
Learning rate steps	[30, 60, 80]
Weight decay	1e-4
Momentum	0.9
Optimizer	SGD
Loss	Cross Entropy

for cross-feature fusion. We can define a parameter  $\alpha_{in} \in [0, 1]$  to set the ratio of local and global features in fusion. Cross-feature fusion provides a compensatory effect between the local and global features. Thus, the learnability and performance of the proposed model could be improved. The main concept behind the proposed FFC was adopted from the original study in [18]. Its performance had been validated for three representative computer vision tasks such as image classification, video classification and human key point detection using ImageNet, Kinetics-400, and COCO datasets, respectively. It achieved a promising performance of accuracy of 77.8 on ImageNet classification, 76.1 on video detection and 79.4 on keypoint detection. Therefore, we adopted the idea behind FFC [18] and developed an FFC-ResNet using 50 layers and cross-feature fusion. The model summary of FFC-ResNet-50 used for adenoma-dysplasia grading of colorectal polyps is shown in Figure 7. We set  $\alpha_{in} = 0.5$ , thus the local and global features were mixed equally in half.

**IV. RESULTS AND DISCUSSION**

As described in the Materials (Section III-A), the UnitoPatho dataset contains H&E-stained images of polyp patches at two magnifications: (i) 800  $\mu\text{m}$  (1812 × 1812 dimensions) and (ii) 7000  $\mu\text{m}$  (15855 × 15855 dimensions). The total number of patches in the dataset is 9,536 (8,669 patches for 800  $\mu\text{m}$  and 8,67 patches for 7000  $\mu\text{m}$ ). For each magnification, 4 folds (80% of total patches) for training and 1 fold (20% of total patches) for validation were used and a 5-fold cross validation was conducted. Before developing the models and training them, we performed data analysis as an initial step and found that the separability between the different polyp classes was unclear. Therefore, we preprocessed the input polyp patches using PCA feature reduction to reduce the undesirable correlations between pixels. After that, to ensure the effectiveness of PCA feature reduction, we conducted an ablation study that compared the performances of models without PCA and those with PCA.

Since the performance of ResNet for polyp classification has been approved in many studies, we selected it as the base model architecture. Then, we created eight different deep-learning models: (i) ResNet50 using original 7000  $\mu\text{m}$  patches, (ii) ResNet50 using PCA



ResNet-50		FFC ResNet-50		
Conv1	7x7, 64, s=2	Conv1	7x7, 64, s=2	
	3 x3 MaxPool, s=2		3 x3 MaxPool, s=2	
		Cross-feature fusion	Local Features (Spatial Convolution)	Global Features (Fourier Convolution)
Conv2_x	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	Conv2_x	$\begin{bmatrix} 1 \times 1, 32 \\ 3 \times 3, 32 \\ 1 \times 1, 128 \end{bmatrix}$	$\begin{bmatrix} 1 \times 1, 32 \\ 3 \times 3, 32 \\ 1 \times 1, 128 \end{bmatrix} \times 3$
Conv3_x	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	Conv3_x	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix}$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 4$
Conv4_x	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	Conv4_x	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix}$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 6$
Conv5_x	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2084 \end{bmatrix} \times 3$	Conv5_x	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1042 \end{bmatrix}$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1042 \end{bmatrix} \times 3$
Average pool, 6-d FC, SoftMax		Average pool, 6-d FC, SoftMax		

FIGURE 7. The architecture of the conventional ResNet-50 and proposed FFC-ResNet50, which uses local and global feature fusion.

TABLE 4. Comparisons of performance measurements.

Models	No. of Parameters	Training Time (Hours)	Sensitivity (Recall)	Specificity	Balance Accuracy	Precision	F1 Score	AUC
ResNet50-Original 7000 μm	25,557,032	1.11	0.61	0.52	0.56	0.52	0.55	0.82
ResNet50-PCA 7000 μm		1.00	0.66	0.53	0.59	0.67	0.61	0.85
FFC-ResNet50-Original 7000 μm	28,142,054	1.68	0.77	0.62	0.69	0.78	0.76	0.95
FFC-ResNet50-PCA 7000 μm		1.41	0.82	0.80	0.78	0.85	0.83	0.97
ResNet50-Original 800 μm	25,557,032	3.25	0.91	0.88	0.90	0.91	0.91	0.99
ResNet50-PCA 800 μm		3.14	0.93	0.92	0.92	0.94	0.94	0.99
FFC-ResNet50-Original 800 μm	28,142,054	3.44	0.93	0.91	0.92	0.93	0.93	0.99
FFC-ResNet50-PCA 800 μm		3.33	<b>0.95</b>	<b>0.93</b>	<b>0.94</b>	<b>0.95</b>	<b>0.95</b>	<b>0.95</b>

TABLE 5. Performance measurements of each class using FFC-ResNet50-PCA 800 μm.

Classes	Sensitivity (Recall)	Specificity	Balance Accuracy	Precision	F1 Score	AUC
HP	0.97	0.99	0.98	0.97	0.97	1.00
NORM	0.97	0.99	0.98	0.96	0.97	1.00
TA.HG	0.86	0.99	0.93	0.95	0.90	1.00
TA.LG	0.95	0.96	0.96	0.95	0.95	1.00
TVA.HG	0.95	0.99	0.97	0.96	0.95	1.00
TVA.LG	0.94	0.97	0.96	0.93	0.94	1.00

reduced 7000 μm patches, (iii) FFC-ResNet50 using original 7000 μm patches, (iv) FFC-ResNet50 using PCA reduced 7000 μm patches; (v) ResNet50 using original 800 μm

patches; (vi) ResNet50 using PCA reduced 800 μm patches, (vii) FFC-ResNet50 using original 800 μm patches, and (viii) FFC-ResNet50 using PCA reduced 800 μm patches.

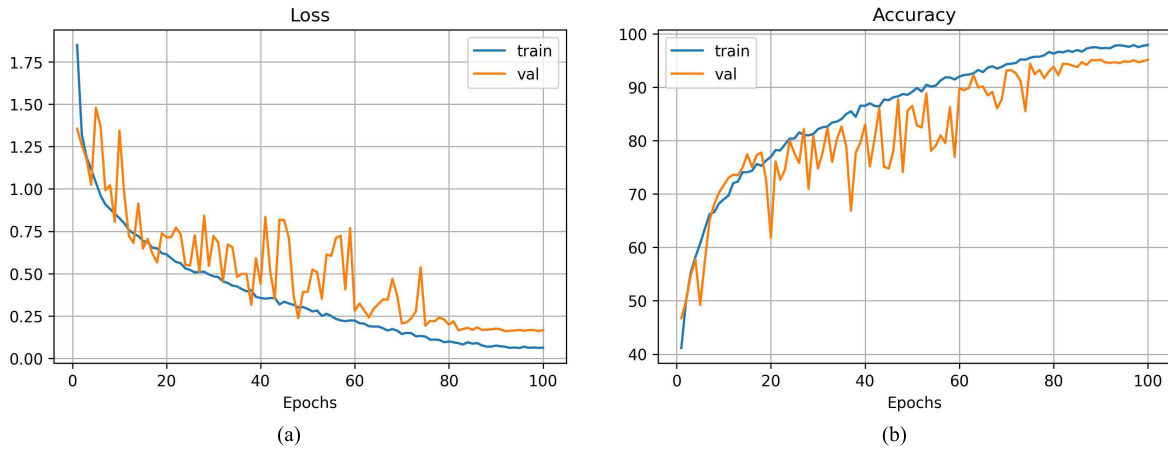


FIGURE 8. Learning curves of the FFC-RESNET50-PCA 800  $\mu$ M (a) Loss and (b) Accuracy.

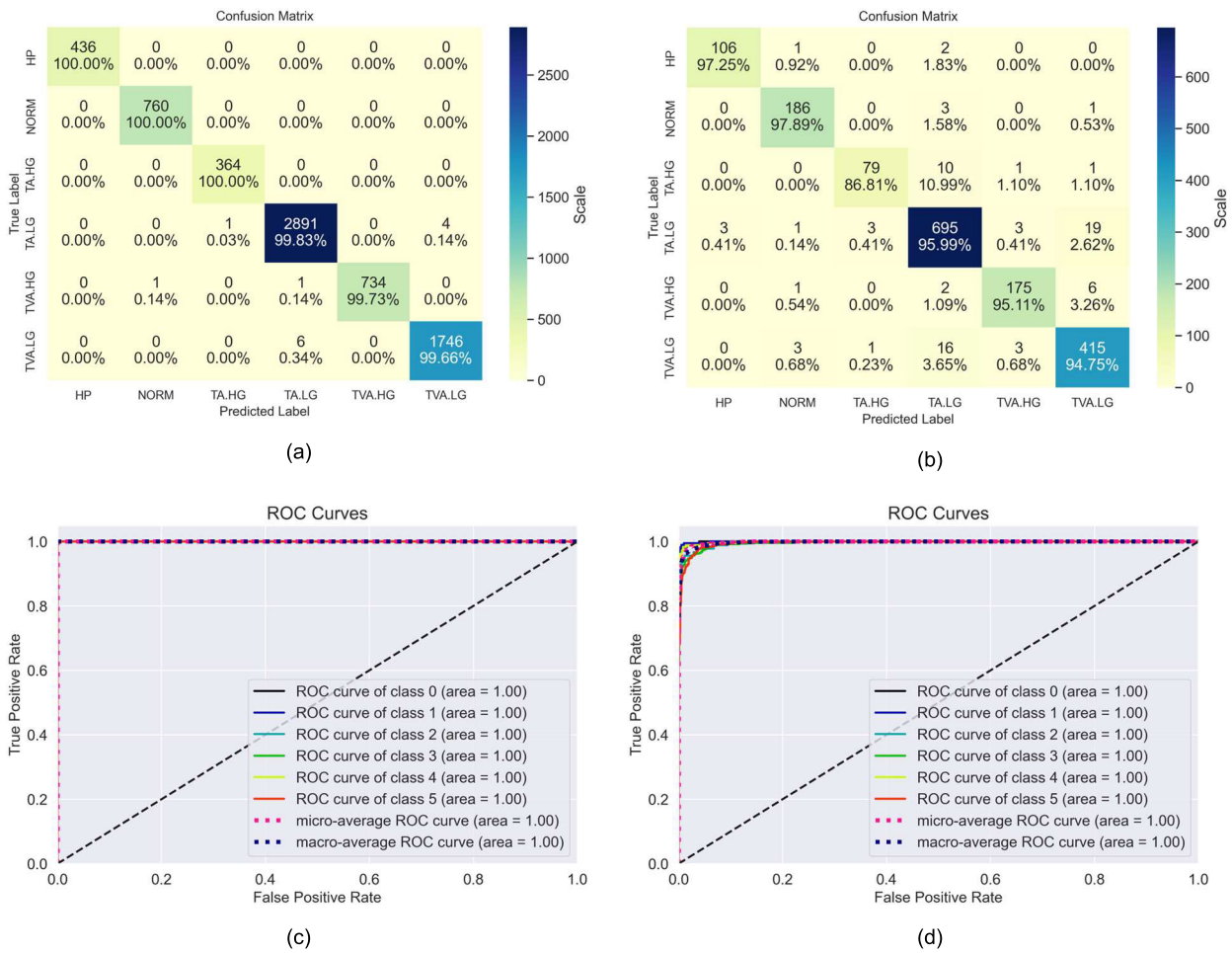
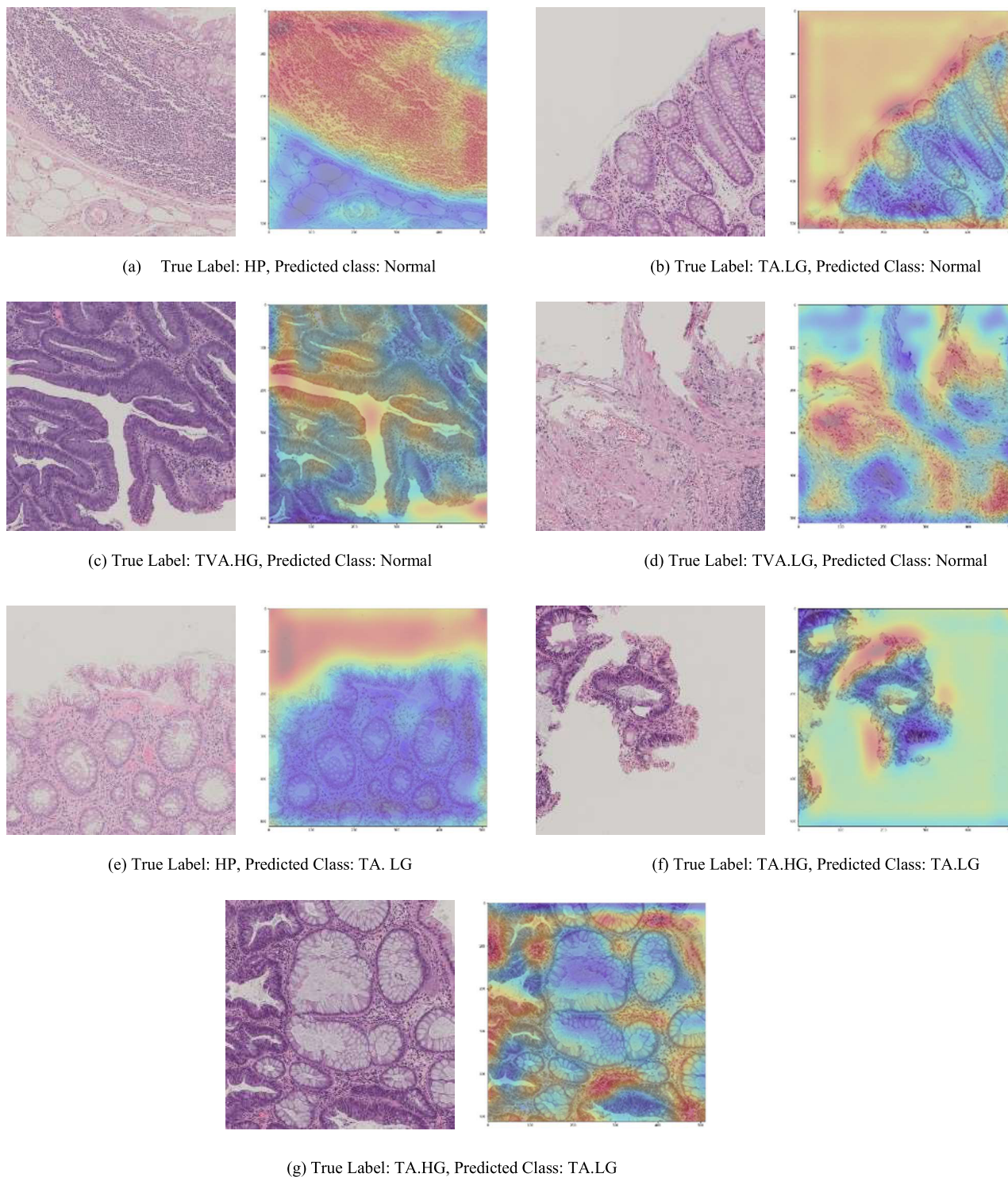


FIGURE 9. Classification results of proposed FFC-ResNet50 (a) Confusion Matrix of training data (b) Confusion Matrix of validation data (c) Receiver operating characteristics (ROC) curve of training data and (d) Receiver operating characteristics (ROC) curve of validation data.

Here, original patches mean raw input images having  $1812 \times 1812$  or  $15855 \times 15855$  dimensions, and they are directly down-sampled into  $224 \times 224$  once they are inputted into the models. Unlikely, PCA-reduced patches mean patches

that are compressed un-salient features and reconstructed as  $224 \times 224$  using the PCA method.

To ensure a fair comparison, we empirically selected the same hyperparameter values for all models. Table 3



**FIGURE 10.** Some examples of false predictions showing input patches and heatmaps of model weights.

summarizes the hyperparameters and their corresponding values. For the performance comparison of the models, we calculated a number of assessment measurements, such as sensitivity, specificity, balance accuracy, precision, F1 score, and AUC. Moreover, we also measured the total number of

parameters and training time to determine the efficiency of the models. Table 4 summarizes the performance assessments of the models. These results were collected and calculated by finding the average of 5-fold cross validation outputs. Based on these assessment values, it was found that PCA reduction

**TABLE 6.** Comparisons of FFC-ResNet50 with other state of the art models using 800  $\mu\text{m}$  pca feature reduced patches.

Models	Sensitivity (Recall)	Specificity	Balance Accuracy	Precision	F1 Score	AUC
VGG [7]	0.84	0.83	0.84	0.86	0.85	0.97
GoogleNet [7]	0.70	0.60	0.65	0.67	0.68	0.92
Xception	0.79	0.62	0.71	0.80	0.78	0.95
Inception V3 [10] [11]	0.86	0.80	0.83	0.86	0.85	0.97
DenseNet101	0.88	0.83	0.85	0.88	0.88	0.98
<b>Proposed FFC-ResNet50</b>	<b>0.95</b>	<b>0.93</b>	<b>0.94</b>	<b>0.95</b>	<b>0.95</b>	<b>0.99</b>

could improve the performance of the models. Models using PCA patches provided more accurate outputs, regardless of the magnification (7000  $\mu\text{m}$  or 800  $\mu\text{m}$ ) and model architecture (ResNet or FFC-ResNet). However, it was found that the results using 7000  $\mu\text{m}$  were very poor compared with those using 800  $\mu\text{m}$ . This could be attributed to two factors. First, 7000  $\mu\text{m}$  magnification focuses on coarser structures, such as glands; therefore, it cannot provide detailed information about the cells. Second, as 7000  $\mu\text{m}$  appears as a zoomed-in structure, it does not produce many patches. The total number of polyp patches with 7000  $\mu\text{m}$  was 8,67 patches while there were 8,669 patches for 800  $\mu\text{m}$ . Deep-learning models perform better if sufficient input images are provided. Therefore, the outputs of the models using 7000  $\mu\text{m}$  are less effective than those using 800  $\mu\text{m}$ .

After analyzing the effectiveness of PCA, we analyzed the advantages of using FFC-ResNet. As we can see in Table 4, experimental results reveal that the performance of the models improved if we used Fourier convolution and global feature learning. The balance accuracy of ResNet-50 using original 7000  $\mu\text{m}$  patches increased from 0.56 to 0.69 when we used FFC. As well, the balance accuracy of ResNet-50 using PCA 7000  $\mu\text{m}$  patches increased from 0.59 to 0.78. Moreover, in the cases of 800  $\mu\text{m}$  patches, FFC helped to increase balance accuracies from 0.90 to 0.92 for original patches and from 0.92 to 0.94 for PCA patches.

The total number of parameters in FFC-ResNet (28,142,054) was higher than that in the original ResNet (25,557,032) because the FFC used cross-feature fusion to obtain the compensative effect among the local and global features. However, the training time of FFC-ResNet can be reduced using the PCA compression of the input images. In particular, if we use FFC-ResNet50 on 800  $\mu\text{m}$  PCA compressed patches, we could obtain superior performance. As a result, we selected FFC-ResNet50 using 800  $\mu\text{m}$  PCA patches as the most appropriate classifier to do adenoma dysplasia grading of colorectal polyps. Table 5 shows the performance measurements on each class of polyps using FFC-ResNet50-PCA 800  $\mu\text{m}$ . From this table, it is evident that the proposed model performed well for every class of polyp. Besides, in order to access the performance of the proposed model, the loss and accuracy curves of the training and validation processes are also stated in Figure 8. Furthermore, the details of classification performance can also be seen in

the confusion matrices and ROC curves of Figure 9. Using the test data, our proposed FFC-ResNet50 correctly classified 97.25%, 97.89%, 86.81%, 95.99%, 95.11%, and 94.75% of the HP, NORM, TA.HG, TA.LG, TVA.HG, and TVA.LG, respectively. It significantly outperformed the original Uni-Patho [8] study that used the multiple-class ensemble method and obtained 86%, 79%, 60%, 50%, 78%, and 52% of the HP, NORM, TA.HG, TA.LG, TVA.HG, and TVA.LG, respectively.

However, as stated in confusion matrix of the validation data (Figure 9 b), our proposed FFC-ResNet produces some false predictions. For example, Figure 10 illustrates some of polyp patches which are wrongly predicted by our proposed method. Along with those patches, we also plotted the heatmaps of model weights in order to gain insights about the false predicts. Based on the visual analysis of the heatmaps, most of these wrong gradings happened because the model gave more weights (denoted by the red color) to background or minute histopathological structures.

Finally, we compared the performance of our proposed FFC-ResNet with that of other state-of-the-art models, such as VGG [7], GoogleNet [7], Inception V3 [10], [11], and DenseNet as shown in Table 6. Here, it should be noted that all of the models in Table 6 were redeveloped and tested on the same dataset in order to provide a fair comparison, as performance in the literature was assessed using different datasets. According to the comparative results, it is evident that the FFC-ResNet can produce superior results than methods in literatures.

## V. CONCLUSION

This paper presents a computer-aided scheme for dysplasia grading of colorectal polyps using hematoxylin and eosin (H&E)-stained patches. An enhanced convolutional neural network called Fast Fourier convolutional ResNet (FFC-ResNet) was developed for better predictions. As the name implies, the proposed FFC-ResNet is based on the ResNet50 architecture and exploits the advantages of convolution in the frequency domain. Because convolution in the frequency domain uses kernels with non-local respective fields, it can provide global features. In addition to global features, FFC enhances learning ability by fusing local features learned by spatial convolutions. The proposed model was evaluated on a publicly available, labelled histopathological

dataset called UniToPatho, and achieved 95% sensitivity, 93% specificity, 94% balance accuracy, 95% precision, 95% F1 score, and 99% AUC. The results produced by the proposed method are promising and superior to those of state-of-the-art methods. However, because the proposed method focuses only on the precise prediction of polyp patches, there can still be further improvements in WSI-level predictions and magnification invariants.

## ACKNOWLEDGMENT

The authors would like to thank the owners of the UNITOPATHO dataset for publicly sharing it for research purposes.

## REFERENCES

- [1] F. Baidoun, K. Elshawy, Y. Elkeraia, Z. Merjaneh, G. Khoudari, M. T. Sarmini, M. Gad, M. Al-Husseini, and A. Saad, "Colorectal cancer epidemiology: Recent trends and impact on outcomes," *Current Drug Targets*, vol. 22, no. 9, pp. 998–1009, Jul. 2021, doi: [10.2174/1389450121999201117115717](https://doi.org/10.2174/1389450121999201117115717).
- [2] J. A. Sninsky, B. M. Shore, G. V. Lupu, and S. D. Crockett, "Risk factors for colorectal polyps and cancer," *Gastrointestinal Endoscopy Clinics North Amer.*, vol. 32, no. 2, pp. 195–213, Apr. 2022, doi: [10.1016/j.giec.2021.12.008](https://doi.org/10.1016/j.giec.2021.12.008).
- [3] M. Mahanty, D. Bhattacharyya, D. Midhunchakkaravarthy, and T.-H. Kim, "Detection of colorectal cancer by deep learning: An extensive review," *Int. J. Current Res. Rev.*, vol. 12, no. 22, pp. 150–157, 2020, doi: [10.31782/IJCRR.2020.122234](https://doi.org/10.31782/IJCRR.2020.122234).
- [4] T. N. Tran, A. Ferrari, S. Hoec, M. Peeters, and G. Van Hal, "Colorectal cancer screening: Have we addressed concerns and needs of the target population?" *Gastrointestinal Disorders*, vol. 3, no. 4, pp. 173–203, Oct. 2021, doi: [10.3390/gdisord3040018](https://doi.org/10.3390/gdisord3040018).
- [5] C. Walsh, "Colorectal polyps: Cancer risk and classification," *Gastrointestinal Nursing*, vol. 15, no. 5, pp. 26–32, Jun. 2017, doi: [10.12968/gasn.2017.15.5.26](https://doi.org/10.12968/gasn.2017.15.5.26).
- [6] N. Marini, S. Otálora, F. Ciompi, G. Silvello, S. Marchesin, S. Vatrano, G. Buttafuoco, M. Atzori, and H. Müller, "Multi-scale task multiple instance learning for the classification of digital pathology images with global annotations," in *Proc. MICCAI Workshop Comput. Pathol.*, vol. 12, 2021, pp. 170–181.
- [7] B. Korbar, A. M. Olofson, A. P. Mirafior, C. M. Nicka, M. A. Suriawinata, L. Torresani, A. A. Suriawinata, and S. Hassanpour, "Deep learning for classification of colorectal polyps on whole-slide images," *J. Pathol. Informat.*, vol. 8, no. 1, p. 30, Jan. 2017, doi: [10.4103/jpi.jpi\\_34\\_17](https://doi.org/10.4103/jpi.jpi_34_17).
- [8] C. A. Barbano, D. Perlo, E. Tartaglione, A. Fiandrotti, L. Bertero, P. Cassoni, and M. Grangetto, "Unitopatho, a labeled histopathological dataset for colorectal polyps classification and adenoma dysplasia grading," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Anchorage, AK, USA, Sep. 2021, pp. 76–80, doi: [10.1109/ICIP42928.2021.9506198](https://doi.org/10.1109/ICIP42928.2021.9506198).
- [9] L. D. Tamang and B. W. Kim, "Deep learning approaches to colorectal cancer diagnosis: A review," *Appl. Sci.*, vol. 11, no. 22, p. 10982, Nov. 2021, doi: [10.3390/app112210982](https://doi.org/10.3390/app112210982).
- [10] O. Iizuka, F. Kanavati, K. Kato, M. Rambeau, K. Arihiro, and M. Tsuneki, "Deep learning models for histopathological classification of gastric and colonic epithelial tumours," *Sci. Rep.*, vol. 10, no. 1, p. 1504, Jan. 2020, doi: [10.1038/s41598-020-58467-9](https://doi.org/10.1038/s41598-020-58467-9).
- [11] K. S. Wang et al., "Accurate diagnosis of colorectal cancer based on histopathology images using artificial intelligence," *BMC Med.*, vol. 19, no. 1, p. 76, Dec. 2021, doi: [10.1186/s12916-021-01942-5](https://doi.org/10.1186/s12916-021-01942-5).
- [12] J. W. Wei, A. A. Suriawinata, L. J. Vaickus, B. Ren, X. Liu, M. Lisovsky, N. Tomita, B. Abdollahi, A. S. Kim, D. C. Snover, J. A. Baron, E. L. Barry, and S. Hassanpour, "Evaluation of a deep neural network for automated classification of colorectal polyps on histopathologic slides," *JAMA Netw. Open*, vol. 3, no. 4, Apr. 2020, Art. no. e203398, doi: [10.1001/jamanetworkopen.2020.3398](https://doi.org/10.1001/jamanetworkopen.2020.3398).
- [13] Z. Song et al., "Automatic deep learning-based colorectal adenoma detection system and its similarities with pathologists," *BMJ Open*, vol. 10, no. 9, Sep. 2020, Art. no. e036423, doi: [10.1136/bmjopen-2019-036423](https://doi.org/10.1136/bmjopen-2019-036423).
- [14] D. Perlo, E. Tartaglione, L. Bertero, P. Cassoni, and M. Grangetto, "Dysplasia grading of colorectal polyps through convolutional neural network analysis of whole slide images," in *Proc. Int. Conf. Med. Imag. Comput.-Aided Diagnosis (MICAD)*, vol. 784, R. Su, Y.-D. Zhang, and H. Liu, Eds. Singapore: Springer, 2022, pp. 325–334, doi: [10.1007/978-981-16-3880-0\\_34](https://doi.org/10.1007/978-981-16-3880-0_34).
- [15] P. C. Neto, S. P. Oliveira, D. Montezuma, J. Fraga, A. Monteiro, L. Ribeiro, S. Gonçalves, I. M. Pinto, and J. S. Cardoso, "iMIL4PATH: A semi-supervised interpretable approach for colorectal whole-slide images," *Cancers*, vol. 14, no. 10, p. 2489, May 2022, doi: [10.3390/cancers14102489](https://doi.org/10.3390/cancers14102489).
- [16] L. Bertero, D. Perlo, A. Gambella, L. Cavallo, and A. Fiandrotti. (2021). *UNITOPATHO*. Accessed: Apr. 28, 2022. [Online]. Available: <https://iee-dataport.org/open-access/unitopatho>
- [17] F. Chollet, *Deep Learning With Python*. Shelter Island, NY, US: Manning, 2018.
- [18] L. Chi, B. Jiang, and Y. Mu, "Fast Fourier convolution," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 4479–4488. [Online]. Available: <https://proceedings.neurips.cc/paper/2020/file/2fd5d41ec6cfab47e32164d5624269b1-Paper.pdf>
- [19] O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for convolutional neural networks," 2015, *arXiv:1506.03767*.
- [20] R. N. R. N. Bracewell, *The Fourier Transform and Its Applications*. 2nd ed. New York, NY, USA: McGraw-Hill, 1978. [Online]. Available: <https://search.library.wisc.edu/catalog/999495438302121>
- [21] M. Mathieu, M. Henaff, and Y. LeCun, "Fast training of convolutional networks through FFTs," 2013, *arXiv:1312.5851*.
- [22] H. Pratt, B. Williams, F. Coenen, and Y. Zheng, "FCNN: Fourier Convolutional Neural Networks," in *Machine Learning and Knowledge Discovery in Databases*, vol. 10534, M. Ceci, J. Hollmén, L. Todorovski, C. Vens, and S. Džeroski, Eds. Cham, Switzerland: Springer, 2017, pp. 786–798, doi: [10.1007/978-3-319-71249-9\\_47](https://doi.org/10.1007/978-3-319-71249-9_47).



**MAY PHU PAING** received the Ph.D. degree in electrical engineering from the King Mongkut's Institute of Technology Ladkrabang, Thailand, in 2020. Since 2021, she has been a Lecturer with the Department of Biomedical Engineering, King Mongkut's Institute of Technology Ladkrabang. Her research interests include image processing, data science, and artificial intelligence.



**CHUCHART PINTAVIROOJ** received the Ph.D. degree in biomedical engineering, from Drexel University, Philadelphia, PA, USA, in 2000. In 2000, he joined the Department of Electronics, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. He is currently an Associate Professor at the Department of Biomedical Engineering, King Mongkut's Institute of Technology Ladkrabang. His research interests include healthcare information technology, biomedical instrumentation, and biomedical image/signal processing. He is an acting Chairperson of the Biomedical Engineering Society of Thailand affiliated with the IEEE EMBS (Thailand Section).

• • •