

Received 7 August 2022; revised 6 November 2022, 29 November 2022, and 12 December 2022; accepted 13 December 2022.  
Date of publication 15 December 2022; date of current version 23 December 2022.

Digital Object Identifier 10.1109/JTEHM.2022.3229561

# Classification of IHC Images of NATs With ResNet-FRP-LSTM for Predicting Survival Rates of Rectal Cancer Patients

TUAN D. PHAM<sup>1</sup>, (Senior Member, IEEE), VINAYAKUMAR RAVI<sup>1</sup>,  
CHUANWEN FAN<sup>2,3</sup>, BIN LUO<sup>2,3,4</sup>, AND XIAO-FENG SUN<sup>2,3</sup>

<sup>1</sup>Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar 31952, Saudi Arabia

<sup>2</sup>Department of Oncology, Linköping University, 58185 Linköping, Sweden

<sup>3</sup>Department of Biomedical and Clinical Sciences, Linköping University, 58185 Linköping, Sweden

<sup>4</sup>Department of Gastrointestinal Surgery, Sichuan Provincial People's Hospital, Chengdu 610032, China

CORRESPONDING AUTHOR: T. D. PHAM (tdpham123@gmail.com)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Ethics Committee of Linköping University under Application No. Dnr 2012-107-31 and No. Dnr 2014-79-31, and performed in line with the Declaration of Helsinki.

**ABSTRACT** Background: Over a decade, tissues dissected adjacent to primary tumors have been considered “normal” or healthy samples (NATs). However, NATs have recently been discovered to be distinct from both tumorous and normal tissues. The ability to predict the survival rate of cancer patients using NATs can open a new door to selecting optimal treatments for cancer and discovering biomarkers. Methods: This paper introduces an artificial intelligence (AI) approach that uses NATs for predicting the 5-year survival of pre-operative radiotherapy patients with rectal cancer. The new approach combines pre-trained deep learning, nonlinear dynamics, and long short-term memory to classify immunohistochemical images of RhoB protein expression on NATs. Results: Ten-fold cross-validation results show 88% accuracy of prediction obtained from the new approach, which is also higher than those provided from baseline methods. Conclusion: Preliminary results not only add objective evidence to recent findings of NATs’ molecular characteristics using state-of-the-art AI methods, but also contribute to the discovery of RhoB expression on NATs in rectal-cancer patients. Clinical impact: The ability to predict the survival rate of cancer patients is extremely important for clinical decision-making. The proposed AI tool is promising for assisting oncologists in their treatments of rectal cancer patients.

**INDEX TERMS** Rectal cancer, 5-year survival prediction, artificial intelligence, deep learning, fuzzy recurrence plots.

## I. INTRODUCTION

Immunohistochemistry (IHC) is an image-staining technique for visualizing antigens (proteins) in cells of tissue by employing fluorescence microscopy. IHC images are used for detecting abnormal cells such as those found in malignant tumors [1], discovering biomarkers of diseases, and validating candidate drug efficacy [2]. Techniques for image analysis of IHC staining have been well recognized as an important role in digital pathology [3], [4]. Furthermore, the combination of artificial intelligence and image analysis techniques have recently been reported to significantly enable new discoveries in oncology [5], including colorectal

cancer [6], [7], [8], and advance research in pathology and medicine [9], [10].

Rectal cancer is a disease in which cancerous cells uncontrollably develop in the tissues of the rectum. The overall 5-year survival rate for patients with rectal cancer is about 63% [11]. The American Cancer Society reported that 5-year (relative) survival rates for rectal cancer are 90%, 73%, and 17% for localized, regional, and distant SEER (Surveillance, Epidemiology, and End Results) staging, respectively; and 67% for the combination of all SEER stages [12].

Over the past decade, tissues dissected adjacent to tumors are considered as normal, and so-called normal tissues

adjacent to tumors (NATs) [13], [14]. Recent findings have shown that NATs' molecular characteristics are quite different from both healthy tissues and tumors [15]. NATs are essential for understanding recurrent tumors and selecting optimal surgical strategies [16]. However, it is suggested to use computational methods to independently confirm the distinct molecular patterns of NATs [15].

In an attempt to discover the prognosis of breast cancer, a study investigated the use of gene expression profiles in NATs for survival prediction of breast cancer patients [17]. The study was carried out by analyzing the genetic characteristics of NATs obtained from The Cancer Genome Atlas, and a particular gene expression profile in normal tissues around breast tumors was found to be associated with the 10-year survival rate for breast cancer women with estrogen receptor-positive cases, which accounts for about 70% among patients with breast cancer. It was reported in a recent study that NATs provide important molecular information, which can be used to differentiate tumors of being more and less aggressive [18]. The study investigated 52 NAT samples of prostate cancer patients obtained from the Cancer Genome Atlas to develop a genetic model for the prediction of recurrence-free survival. The survival-prediction power of NATs was confirmed with an independent cohort, suggesting distinct prognostic mechanisms between NATs and tumor tissues. Similarly, another study of NATs for survival prediction in hepatocellular carcinoma (HCC), which is the most frequent primary liver cancer, reported the critical function of Hippo signaling in normal tissues adjacent to HCC with respect to the cancer development [14]. The study constructed a model of Hippo-related gene expression profiles in NATs for predicting the prognosis of patients with HCC. The univariate Cox regression analysis identified 14 genes that can be used for the survival prediction in HCC with an area under the receiver operating characteristic curve of 0.75.

This paper reports the first attempt to investigate the power of RhoB expression in images of rectal-cancer NATs with deep learning and nonlinear time-series analysis. Using IHC images of RhoB protein expression on rectal-cancer NATs, this study combines the methods of pre-trained convolutional neural networks (CNNs), nonlinear dynamics, and long short-term memory (LSTM) for classifying IHC images of the NATs. Pretrained CNNs are utilized for extracting deep features of IHC images. The method of fuzzy recurrence plots (FRPs) [19] developed for studying nonlinear dynamics is then applied to transform long CNN-based feature vectors into FRPs, which are short multidimensional time series. Finally, an LSTM network is used to learn these FRPs for time-series classification to predict the life expectancy of two cohorts of rectal-cancer patients who took pre-operative radiotherapy and lived more or less than 5 years after surgery.

The motivation for extracting features of the IHC images with pretrained CNNs and constructing FRPs of these extracted features is based on the following aspects of data analysis for machine learning. Firstly, extracting deep features of new data from pre-trained CNNs has been reported

very useful in terms of robustness and computational advantage for the classification of complex biomedical images [20], [21], [22], [23]. Secondly, the transformation of flattened deep image features into FRPs is expected to enhance the power of machine learning as the spatial-temporal content of the original IHC data can be captured by this method of nonlinear dynamics. Finally, the transformed spatial-temporal features represented by the FRPs can be used to provide multiple feature dimensions in a much shorter time series for increasing the sequential learning power of the LSTM. Thus, the combination of two state-of-the-art AI approaches (CNN and LSTM models) by means of the FRP construction is unique for predicting survival rates in patients with rectal cancer, which is the major contribution of this study.

The rest of this paper is organized as follows. Section II describes methods employed for deep learning and classification of IHC-images to predict the 5-year survival of rectal cancer patients. Section III presents and discusses results. Finally, Section IV provides concluding remarks on the finding.

## II. METHODS

### A. FEATURE EXTRACTION FROM PRE-TRAINED CNNs

CNNs are deep neural networks, which consist of a series of three main types of layers known as convolutional, pooling, and fully-connected layers. A feature map  $G$  of an input image  $I$  can be produced using the convolutional process as follows.

$$\begin{aligned} G(x, y) &= W \otimes I(x, y) \\ &= \sum_{dx=-\alpha}^{\alpha} \sum_{dy=-\beta}^{\beta} W(dx, dy)I(x + dx, y + dy), \end{aligned} \quad (1)$$

where  $W$  is a filter kernel.

To avoid negative values of the feature map, the rectified linear unit (ReLU) [24], denoted as  $L$ , is then applied to  $G$ . This nonlinear function is defined as

$$L(\omega) = \begin{cases} 0, & \text{for } \omega < 0 \\ \omega, & \text{for } \omega \geq 0 \end{cases} \quad (2)$$

There is no single theoretical method for selecting an optimal activation function for hidden layers of a deep net to learn on a particular dataset. The ReLU is the most widely used activation function because it can effectively address the vanishing gradient problem when training a deep neural network [25]. However, the disadvantage of the ReLU is that, in some given datasets, it transforms many neurons with negative values into zeros, resulting in many dead neurons that will never be updated. To resolve the dead-neuron problem, several extended versions of ReLU were developed to avoid the nullification of negative-value neurons, including leaky ReLU (LReLU) [26], parametric ReLU (PReLU) [27], randomized leaky ReLU (RLReLU) [28], and S-shaped ReLU (SReLU) [29]. Mathematical descriptions and applications of these ReLU variants were reviewed in [25].

Next the pooling operator is applied to the rectified feature map in an attempt to produce down-sampling. The most widely used pooling for CNNs is the maximum operator, denoted as  $D_{\max}$ , which operates on a collection of  $M$  pooling regions  $O^p$ ,  $p = 1, \dots, M$ , as

$$D_{\max}(O^p) = \max_{1 \leq \kappa \leq m \times m} (o_{\kappa}^p). \quad (3)$$

where  $o_{\kappa}^p \in O^p = (o_1^p, \dots, o_{m \times m}^p)$ , and  $m \times m$  is the number of pixels in a pooling region.

After a series of operations on convolution and pooling, the final feature map is flattened and fed into the fully-connected (hidden) layer followed by the softmax layer for computing the classification probability distribution over a set of objects.

In this study, the IHC images were input into a pre-trained CNN, which went through the network learning. Finally, the flattened final feature map was then extracted for classification by other machine-learning models. In fact, pretrained CNNs that were trained on the ImageNet database [30] for learning and extracting complex features from different types of images have been often used to classify new objects, where the sample size in the new task is small [6], [20]. Such utilization of a pretrained network is known as transfer learning, which offers certain advantages for feature extraction of complex medical images for pattern classification having mentioned earlier. Furthermore, the extraction of IHC image features from pre-trained CNNs has been utilized for pattern classification, because it offers the fastest procedure to take advantage of the power of deep learning [31]. This type of feature extraction is time-saving, because it requires only a single pass through the data and avoids the need for network training.

Three pre-trained CNNs that were used in this study for extracting features of the IHC images are briefly described as follows.

- NASNet-Large is a variant of the Neural Architecture Search Network (NASNet) models [32]. This net was designed to consist of normal and reduction cells to carry out search space, search strategy, and performance estimation to identify the best algorithm in order to achieve the best optimal performance over a certain piece of work. NASNet-Large has the image input size of  $331 \times 331 \times 3$  pixels.
- DenseNet-201 is a pretrained CNN, which is a variant of the DenseNet [33]. DenseNet-201 has 201 layers. This network has an architecture that allows collective information received from the prior layers can reduce the number of channels, making the network become dense. The size of an input image to DenseNet-201 is of  $224 \times 224 \times 3$  pixels.
- ResNet-101 is one of the ResNet (Residual Networks) family [34]. It is a pretrained CNN. This network requires the input image size of  $224 \times 224 \times 3$  pixels. This network learns the residual functions with reference to the layer inputs instead of the signals and stacks residual blocks on top of each other to form a network

of 101 layers deep. Because of such architecture, it is relatively easier to optimize these networks, which are expected to increase classification accuracy by increasing the network depth.

As pre-trained CNNs built hierarchical representations of input images, where deeper CNN layers generated higher-level features, which were constructed based on lower-level features obtained from earlier layers. Feature extraction of the IHC images from pre-trained CNNs was taken from the fully-connected layers that combined features of the input images over all spatial locations. This layer selection for extracting features from pre-trained CNNs was reported as being effective for classification tasks by several previous studies [35], [36], [37], [38],

Because the above three pre-trained networks require input images of different sizes ( $331 \times 331 \times 3$  for NASNet-Large and  $224 \times 224 \times 3$  for DenseNet-201 and ResNet-101), all training and test IHC images were resized to the specified sizes before they were input to the pre-trained CNNs. NASNet-Large, DenseNet-201, and ResNet-101 are among popular pre-trained deep-learning models, and discussions on their computing designs and structures are widely available in AI literature reviews. Interested readers can refer to original developments reported in [32], [33], and [34] for detailed descriptions of the architectures of the pre-trained CNNs adopted in this study.

## B. FRPs OF PRETRAINED CNN FEATURES

Let  $g = (g_1, g_2, \dots, g_N)$  be a vector of extracted image features flattened by a pretrained CNN. To significantly reduce the computational complexity of constructing an FRP of a long time series,  $g$  can be represented with a sequence of its cluster centers  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , where  $n \ll N$  by using the fuzzy  $c$ -means (FCM) algorithm [39]. To compute an FRP of  $\mathbf{x}$ , first a phase-space reconstruction from  $\mathbf{x}$  is carried out using Takens' time-delay embedding theorem [40], yielding

$$\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_M), \quad (4)$$

where  $M = n - (d - 1)\tau$ ,  $d$  is an embedding dimension,  $\tau$  is a time delay, and

$$\mathbf{s}_i = (x_i, x_{i+\tau}, \dots, x_{i+(d-1)\tau}), \quad i = 1, \dots, n - (d - 1)\tau. \quad (5)$$

Given a number of clusters  $c$  and fuzzy exponent  $m$ , the FCM is then applied to divide  $\mathbf{S}$  into  $c$  groups, which are represented with a set of  $c$  cluster centers  $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c)$ . The FCM also yields a matrix of fuzzy membership grades  $\mu(\mathbf{s}_i, \mathbf{v}_k) \in [0, 1]$ ,  $i = 1, \dots, M$ ,  $k = 1, \dots, c$ , which express the similarity between  $\mathbf{s}_i$  and  $\mathbf{v}_k$ . Finally, using the FCM results, an FRP, denoted as  $\mathbf{F}$ , is constructed as an  $M \times M$  grayscale image that takes real values in  $[0, 1]$  as [19]

$$\mathbf{F}(i, j) = \mu(\mathbf{s}_i, \mathbf{s}_j), \quad i, j = 1, \dots, M, \quad (6)$$

where  $\mu(\mathbf{s}_i, \mathbf{s}_j)$ , which is a fuzzy membership of similarity between  $\mathbf{s}_i$  and  $\mathbf{s}_j$ , is determined using the following three properties.

1) Reflexivity:

$$\mu(\mathbf{s}_i, \mathbf{s}_i) = 1, \quad i = 1, \dots, M. \quad (7)$$

2) Symmetry:

$$\mu(\mathbf{s}_i, \mathbf{v}_k) = \mu(\mathbf{v}_k, \mathbf{s}_i), \quad i = 1, \dots, M, k = 1, \dots, c. \quad (8)$$

3) Transitivity:

$$\mu(\mathbf{s}_i, \mathbf{s}_j) = \max [\min \{\mu(\mathbf{s}_i, \mathbf{v}_k), \mu(\mathbf{v}_k, \mathbf{s}_j)\}], \quad i \neq j; k = 1, \dots, c. \quad (9)$$

The determination of  $\mu(\mathbf{s}_i, \mathbf{v}_k)$ ,  $i = 1, \dots, M$ ,  $k = 1, \dots, c$ , can be obtained by the FCM that tries to minimize the following objective function [39]:

$$F_m = \sum_{i=1}^M \sum_{k=1}^c [\mu(\mathbf{s}_i, \mathbf{v}_k)]^m \|\mathbf{s}_i - \mathbf{v}_k\|^2, \quad (10)$$

where  $m \in [1, \infty)$  is the weighting exponent, and  $F_m$  is subject to

$$\sum_{k=1}^c \mu(\mathbf{s}_i, \mathbf{v}_k) = 1; \quad i = 1, \dots, M. \quad (11)$$

The objective function of the FCM is minimized by a numerical scheme, which iterates the update the fuzzy membership grades and cluster centers until the values converge. Given initialized fuzzy memberships, both  $\mathbf{v}_k$  and  $\mu(\mathbf{s}_i, \mathbf{v}_k)$  are iteratively updated as

$$\mathbf{v}_k = \frac{\sum_{i=1}^N (\mu(\mathbf{s}_i, \mathbf{v}_k))^m \mathbf{s}_i}{\sum_{i=1}^N (\mu(\mathbf{s}_i, \mathbf{v}_k))^m}. \quad (12)$$

$$\mu(\mathbf{s}_i, \mathbf{v}_k) = \frac{1}{\sum_{j=1}^c \left( \frac{\|\mathbf{s}_i - \mathbf{v}_k\|}{\|\mathbf{s}_i - \mathbf{v}_j\|} \right)^{2/(m-1)}}; \quad (13)$$

The procedure for carrying out the FCM can be outlined as follows.

#### FCM Algorithm

- 1) Given data  $\mathbf{S}$ , number of clusters  $c > 1$ , weighting exponent  $m$ , threshold  $\delta$ , and maximum number of iterations  $Q$ .
- 2) Set index  $q = 1$ .
- 3) Initialize fuzzy-membership matrix  $\mathbf{U}_q = [\mu_q(\mathbf{s}_i, \mathbf{v}_k)]$ ,  $i = 1, \dots, M; k = 1, \dots, c$ .
- 4) Compute  $\mathbf{v}_k$ ,  $k = 1, \dots, c$  using Eq. (12).
- 5) Compute  $\mu_{q+1}(\mathbf{s}_i, \mathbf{v}_k)$  using Eq. (13).
- 6) If  $|\mathbf{U}_q - \mathbf{U}_{q+1}| \leq \delta$  or  $q = Q$ , stop.
- 7) Otherwise, set  $q = q + 1$  and return to Step 4.

#### C. TRAINING LSTM NETWORKS WITH FRPs

LSTM networks [41], [42] have the capability of handling the long-term dependency problem encountered by other recurrent neural networks. The long-term dependency induces a vanishing gradient, which becomes negligible to allow the

updating of the network weights. Applications of LSTM networks have reportedly been useful for classifying physiological signals [43], [44], [45] and histopathological images [46].

Basically, the repeating block in an LSTM regulates how data at each time step are processed using four interacting components:

- Input gate, denoted as  $u$ , which controls the level of cell state update.
- Forget gate, denoted as  $f$ , which controls the level of cell state reset or forgetting.
- Cell candidate, denoted as  $w$ , which keeps and adds useful information to a cell state.
- Output gate, denoted as  $e$ , which controls the level of a cell state added to a hidden state.

LSTM learnable parameters of a layer consist of input weights ( $\mathbf{I}$ ), recurrent weights ( $\mathbf{R}$ ), and bias ( $\mathbf{b}$ ), which are mathematically expressed as

$$\mathbf{I} = [\mathbf{i}_u, \mathbf{i}_f, \mathbf{i}_w, \mathbf{i}_e]^T \quad (14)$$

$$\mathbf{R} = [\mathbf{r}_u, \mathbf{r}_f, \mathbf{r}_w, \mathbf{r}_e]^T \quad (15)$$

$$\mathbf{b} = [b_u, b_f, b_w, b_e]^T \quad (16)$$

The cell and hidden states at time step  $t$  are denoted as  $\mathbf{c}_t$  and  $\mathbf{h}_t$ , respectively, which are defined as

$$\mathbf{c}_t = f_t \circ \mathbf{c}_{t-1} + u_t \circ w_t, \quad (17)$$

where  $\circ$  denotes the Hadamard product, and

$$\mathbf{h}_t = e_t \circ \sigma_c(\mathbf{c}_t). \quad (18)$$

At time  $t$ , input gate ( $u_t$ ), forget gate ( $f_t$ ), cell candidate ( $w_t$ ), and output gate ( $e_t$ ) are defined as

$$u_t = \sigma_g(\mathbf{i}_u \mathbf{z}_t + \mathbf{r}_u \mathbf{h}_{t-1} + b_u), \quad (19)$$

$$f_t = \sigma_g(\mathbf{i}_f \mathbf{z}_t + \mathbf{r}_f \mathbf{h}_{t-1} + b_f), \quad (20)$$

$$w_t = \sigma_c(\mathbf{i}_w \mathbf{z}_t + \mathbf{r}_w \mathbf{h}_{t-1} + b_w), \quad (21)$$

$$e_t = \sigma_g(\mathbf{i}_e \mathbf{z}_t + \mathbf{r}_e \mathbf{h}_{t-1} + b_e), \quad (22)$$

in which  $\mathbf{z}_t$  is the input feature vector at time  $t$ ,  $\sigma_c$  and  $\sigma_g$  are state and gate activations usually expressed as the hyperbolic tangent and sigmoid functions, respectively.

In this study, an FRP, which is the final transformation of the associated IHC image described previously, is used as an input into an LSTM for the network training and classification. To be more explicit,  $F(i, j)$ ,  $i = 1, 2, \dots, M$ , where  $M$  is the number of features, is a sequence input into an LSTM, and  $F(i, j)$ ,  $j = 1, 2, \dots, M$ , where  $M$  is the number of time steps, is the input sequence of  $M$  features at each time step  $j$ . More explicitly,  $F(i, j)$ ,  $i = 1, 2, \dots, M$ , are the fuzzy membership grades of similarity computed to construct the FRP of an IHC image described in Section II-B, and used as a time series of multiple features for LSTM learning and prediction task.

Figure 1 shows the proposed procedure for extracting deep-learning features of IHC-NAT images from a pre-trained CNN, transforming the extracted deep-learning features into FRPs that are used for training and testing an LSTM network for the survival prediction.

Figure 2 shows how the flow of an FRP  $\mathbf{F}$ , which is presented as a time series of length  $M$  with  $M$  features, is processed by an LSTM layer. The first LSTM block takes information from the initial state of the network and the first time-step of the FRP to determine the first hidden and cell states. At time step  $i$ , the LSTM block uses the previous state pair of the network ( $\mathbf{h}_{i-1}$  and  $\mathbf{c}_{i-1}$ ) and  $M$  features of the time series (FRP) at time  $i$  to compute  $\mathbf{h}_i$  and  $\mathbf{c}_i$ . As a result, the current cell state contains information obtained from previous time steps, and at each time step, the LSTM layer adds important information to or removes redundant memory from the cell state during training. The network layer regulates these updates using the gate operators defined earlier.

#### D. PERFORMANCE MEASURES

To evaluate the performance of the proposed approach for survival prediction using IHC images of RhoB expression in rectal cancer biopsy,  $> 5$ -year and  $< 5$ -year survival times are considered as positive and negative conditions, respectively. The numbers of samples of  $>$  and  $<$  5 years of survival are denoted as  $P$  and  $N$ , respectively. True positive ( $TP$ ) is the number of samples of  $>$  5 years, which are correctly predicted as  $>$  5 years; whereas true negative ( $TN$ ) is the number of samples of  $<$  5 years being correctly predicted as  $<$  5 years. False positive ( $FP$ ) is the number of samples of  $<$  5 years, which are misclassified as  $>$  5 years. False negative ( $FN$ ) is the number of samples of  $>$  5 years, which are misclassified as  $<$  5 years.

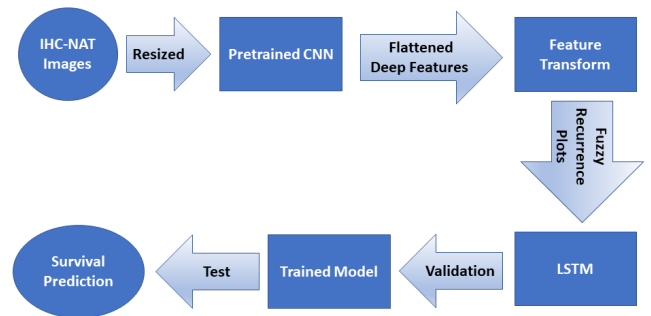
Equations used for calculating prediction/classification accuracy ( $ACC$ ), sensitivity or true positive rate ( $SEN$ ), specificity or true negative rate ( $SPE$ ), precision or positive predictive value ( $PRE$ ), and  $F_1$  score are expressed in Table 1. For all measures, a higher value indicates better performance of the model in terms of the defined measure.

**TABLE 1.** Performance measures of prediction models.

$ACC$	$SEN$	$SPE$	$PRE$	$F_1$
$\frac{TP + TN}{P + N}$	$\frac{TP}{P}$	$\frac{TN}{N}$	$\frac{TP}{TP + FP}$	$\frac{2TP}{2TP + FP + FN}$

### III. RESULTS

The proposed approach was tested using a subset of IHC images obtained from rectal-cancer patients. This subset consists of samples of RhoB-protein expression by IHC staining on tissues that were removed adjacent to primary tumors. The data were collected from two groups of rectal-cancer patients who had survival rates of less or more than 5 years. The study was conducted in accordance with the Declaration of Helsinki, and protocols approved by the Institutional Ethics Committee of Linköping University (Dnr 2012-107-31 and Dnr 2014-79-31). These patients were participants included



**FIGURE 1.** CNN-FRP-LSTM model for survival prediction in rectal cancer using IHC-NAT images.

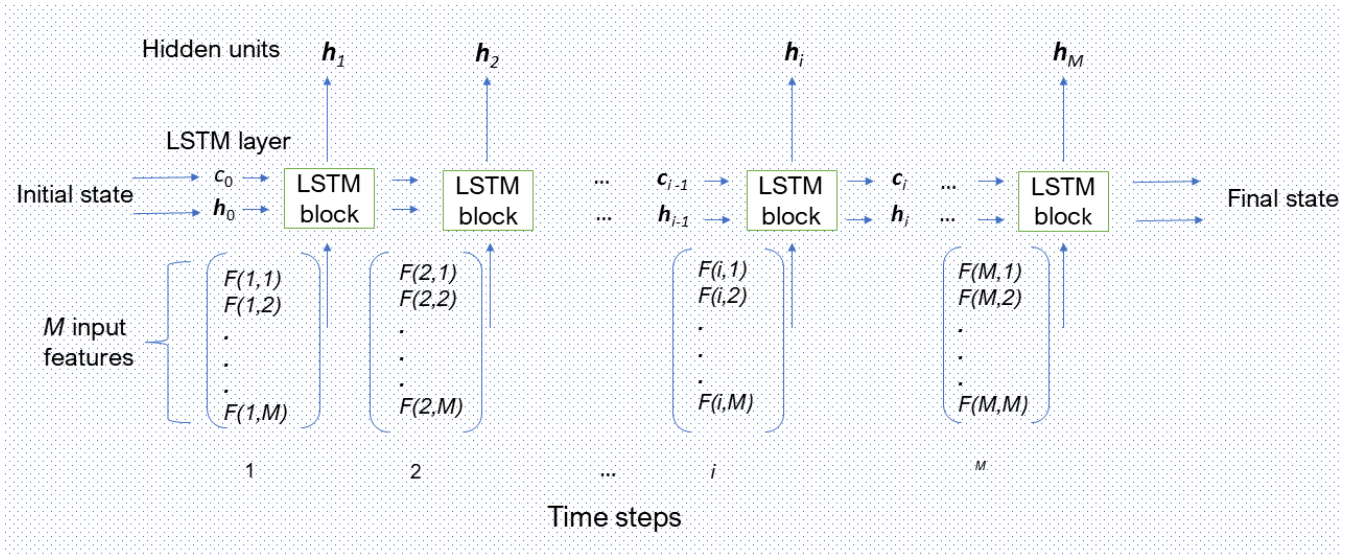
**TABLE 2.** Performance measures for 10-fold cross-validation among different classification models, where NASNet, DenseNet, and ResNet are short names for NASNet-Large, DenseNet-201, and ResNet-101, respectively.

Model	%ACC	%SEN	%SPE	%PRE	$F_1$
GLCM-SVM	75.00	100.00	33.33	71.43	0.83
GoogLeNet	62.50	100.00	0.00	62.50	0.77
AlexNet	62.50	100.00	0.00	62.50	0.77
NASNet	75.00	80.00	66.67	80.00	0.80
NASNet-SVM	37.50	60	0.00	50.00	0.55
NASNet-FRP-LSTM	75.00	100	33.33	71.43	0.83
DenseNet	75.00	100.00	33.33	71.43	0.83
DenseNet-SVM	75.00	100.00	33.33	71.43	0.83
DenseNet-FRP-LSTM	75.00	100.00	33.33	71.43	0.83
ResNet	75.00	100.00	33.33	71.43	0.83
ResNet-SVM	50.00	80.00	0.00	57.14	0.67
ResNet-FRP-LSTM	<b>87.50</b>	100.00	66.67	<b>83.33</b>	<b>0.91</b>

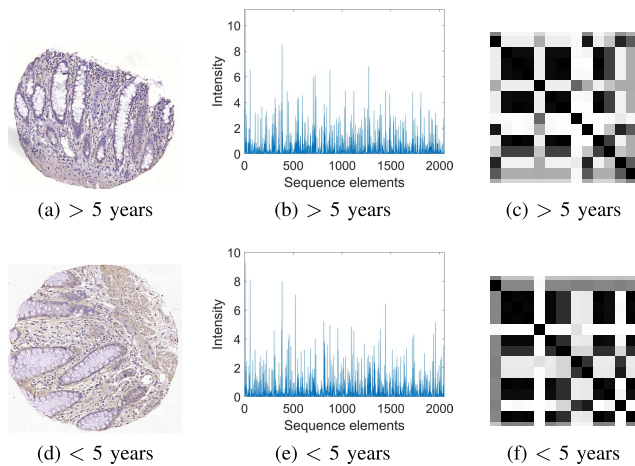
**TABLE 3.** Time complexity among different classification models for 10-fold cross-validation on a single CPU, where NASNet, DenseNet, and ResNet are short names for NASNet-Large, DenseNet-201, and ResNet-101, respectively. TT = training time in seconds, TRT = total running time in seconds, and #P = number of parameters in millions.

Model	TT (sec)	TRT (sec)	#P (MM)
GLCM-SVM	4	91	
GoogLeNet	540	589	7.0
AlexNet	360	399	61.0
NASNet	23571	23583	88.9
NASNet-SVM	4	183	88.9
NASNet-FRP-LSTM	21	265	88.9
DenseNet	1626	1632	20.0
DenseNet-SVM	3	51	20.0
DenseNet-FRP-LSTM	30	125	20.0
ResNet	1401	1407	44.6
ResNet-SVM	3	43	44.6
ResNet-FRP-LSTM	70	141	44.6

in the randomized Swedish Rectal Cancer Trial of preoperative radiotherapy between 1987 and 1990 [48]. There are 80 IHC samples of NATs, where 29 and 51 images are of less and more than 5-year disease-free survival, respectively.



**FIGURE 2.** LSTM layer architecture with an FRP as input time series of multiple features.



**FIGURE 3.** RhoB-expressed IHC images of tissues dissected adjacent to primary tumors of rectal cancer (a) & (d), associated ResNet-101 flattened features (b) & (e), and FRPs (c) & (f).

These samples were collected from rectal cancer patients who took pre-operative radiotherapy.

The layer named “global\_average\_pooling2d\_2” (global average pooling features), which is before the fully connected layer, was used for feature extraction with NASNet-Large. Similarly, for DenseNet-201, the layer named “avg\_pool” (global average pooling features), which is before the fully connected layer, was used for feature extraction. Likewise, for ResNet-101, the layer named “pool5” (global average pooling features), which is before the fully connected layer, was used for feature extraction. For classification using the three pre-trained CNNs, parameter specifications for the training were as follows. Minimum batch size = maximum number of epochs = 20, initial learning rate = 0.0003, number of epochs for dropping the learning rate = 10, factor for dropping the learning rate = 0.1, factor for  $L_2$  regularization (weight

decay) = 0.0001, gradient threshold method =  $L_2$  norm, gradient threshold =  $\infty$ , training and validation data were shuffled once before training, and networks were trained using stochastic gradient descent with momentum. Parameter fine-tuning was performed in this study for the transfer learning of the IHC data, because the three CNNs were pre-trained on different types of images [32], [33], [34]. Such parameter fine-tuning was necessary and found to be effective in several similar applications [6], [31], [49], [50].

Transfer learning of the resized IHC images resulted in flattened feature lengths of 4032, 1920, and 2048 obtained from NASNet-Large, DenseNet-201, and ResNet-101, respectively, at the selected deep layers described earlier. These flattened features were then used as the input for constructing the FRPs.

For model comparison, the same deep-learning parameters given above were used for the survival prediction performed directly by the three stand-alone pre trained CNNs (NASNet-Large, DenseNet-201, and ResNet-101). For another purpose of comparison, the linear support vector machine (SVM) model was adopted for the survival prediction using the same pretrained CNN-based features (NASNet-Large-SVM, DenseNet-201-SVM, and ResNet-101-SVM).

To further compare with other classifiers, 14 features from the gray-level co-occurrence matrices (GLCM) described in [51] were extracted from the grayscale IHC images and then used for the survival prediction by the linear SVM algorithm. In addition, the GoLeNet [52] and AlexNet [53], which are two widely used pre-trained CNNs, were also applied for classifying the IHC images for the survival prediction.

To illustrate the performance of the proposed approach, the same deep-learning features of the IHC images extracted from the pre-trained CNNs were adopted for constructing the FRPs that were modeled as time series of multiple features

for the survival prediction by the LSTM-based classifier. For the use of LSTM, the bidirectional LSTM model was used for sequence classification, and its other parameter specifications are: maximum number of epochs = 300, minimum batch size = 150, initial learning rate = 0.01, and gradient threshold = 1.

For the classification by LSTM with the input as FRPs of features extracted from pre-trained CNN models, the number of clusters  $c = 13$  was used to compress the original CNN-based feature vectors into much shorter sequences of length  $c$  by the FCM. To construct FRPs of the compressed sequences,  $c = 3$ , fuzzy exponent  $m = 2$ , embedding dimension  $d = 1$ , and time delay  $\tau = 1$  were adopted in this study. In two cases for computing the FCM, threshold  $\delta = 0.00001$ , and the maximum number of iterations  $Q = 100$ . The construction of FRPs using the flattened features produced by each of the three pretrained CNNs and the specified FRP parameters resulted in output  $13 \times 13$  matrices. These  $13 \times 13$  FRPs were then used as the input into the LSTM to produce two outputs of scores for predicting the two survival times of  $<$  and  $>$  5 years.

Figure 3 shows IHC images of RhoB expression on NATs of rectal-cancer patients who had  $<$  or  $>$  5 years of survival after surgery, their feature vectors extracted from the ResNet101, and corresponding FRPs, respectively.

Table 2 shows the performance metrics based on 10-fold cross-validation obtained from different classification models, where NASNet-Large indicates the classification was entirely carried out by the network, NASNet-Large-SVM means the image features were extracted by the pre-trained network and used for training the SVM to carry out the classification task, NASNet-Large-FRP-LSTM is the proposed model that uses the pre-trained network for feature extraction, constructs the FRPs of the flattened pre-trained CNN-based features, and adopts the LSTM for classifying the FRPs as multifeatured time series. Similar meanings apply for DenseNet-201, DenseNet-201-SVM, and DenseNet-201-FRP-LSTM; and ResNet-101, ResNet-101-SVM, and ResNet-101-FRP-LSTM. The results obtained from the GLCM-SVM are the same as those from DenseNet and ResNet. GooLeNet and AlexNet yielded the same values of the performance measures, where the accuracy (62.50%) is lower than the GLCM-SVM (75%). While the classification accuracy obtained from NASNet-Large and NASNet-Large-FRP-LSTM as well as DenseNet-201 and DenseNet-201-FRP-LSTM are the same (75%), ResNet-101-FRP-LSTM achieved higher accuracy (87.50%), which is also the best rate among other 11 models. The ResNet-101-FRP-LSTM model also provided the best results for precision (83%) and  $F_1$  score (0.91). The SVM-based classification performed poorly with features extracted from the NASNet-Large (accuracy = 37.50%) and ResNet-101 (accuracy = 50.00%).

Table 3 shows the time complexity of the 12 classification models, which produced the results shown in Table 2, required by a CPU@2.50GHz. Because of its large size,

the training of NASNet-Large took relatively much more computational time, which is 5890 and 1120 times longer than the training of a linear SVM and bidirectional LSTM, respectively. Training the proposed approach (NASNet-Large-FRP-LSTM) is only about 5 times longer than a linear SVM. Similar comparisons of time complexity can be drawn for the other two groups of classification methods: DenseNet-201, DenseNet-201-SVM, and DenseNet-201-FRP-LSTM; and ResNet-101, ResNet-101-SVM, and ResNet-101-FRP-LSTM. Because of its shallower depth, ResNet-101 training required the least computational time. However, both the training time and total running time for ResNet-101-FRP-LSTM were longer than those for DenseNet-201-FRP-LSTM. The training time taken by either GooLeNet or AlexNet was a few times less than the other three pre-trained CNNs. The GLCM-SVM required a longer running time than the ResNet-101-SVM mainly due to the extraction of the GLCM features. Such time saving suggests an advantage of feature extraction from the ResNet-101.

Table 3 also shows the numbers of parameters of the pretrained CNNs (the lower the better). In scientific computing, the counting of floating point operations (FLOPs) has been used to compare the computational complexity between algorithms (the lower the better). Both FLOPs and FLOPS (floating point operations per second) are adopted for different purposes. While FLOPs are used to quantify how many operations are needed to execute a computer algorithm, FLOPS (the higher the better or faster) are used for measuring the computational power of given hardware. FLOPS vary substantially between different microprocessor architectures [54].

It has been reported that the measure of FLOPs is no longer the governing factor in execution speed on modern computers [55] and becomes obsolete [56] as openly discussed in the community of numerical analysis [57]. In fact, the function for counting FLOPs is no longer available in updated versions of MATLAB. The AI community is seeking more accurate models for comparing algorithm efficiency [54]. It should be pointed out that total running times required by a pretrained CNN are different for using data of various sizes. However, measures of computer performance using ResNet-101, DenseNet-201, and NASNet-Large by means of FLOPs were reported in literature. For example, using the ImageNet database [30], total FLOPs for ResNet-101, DenseNet-201, and NASNet-Large are  $8 \times 10^9$  [58],  $4 \times 10^9$  [58], and  $30 \times 10^9$  [59], respectively. These FLOPs show the most and least favorable nets among the three are DenseNet-201 and NASNet-Large, respectively.

The above analyses suggest the usefulness of the proposed approach for classifying IHC images of RhoB expression on NATs of rectal-cancer patients who took pre-operative radiotherapy. Such classification results obtained from the proposed approach not only indicate RhoB protein is a potential prognosis biomarker of rectal cancer, but are also helpful for clinical decision-making if a patient with rectal cancer should be recommended for post-operative radiotherapy to reduce

local recurrence or other alternative treatments, depending on the predicted survival of the patient.

#### IV. CONCLUSION

The foregoing sections have presented the proposed CNN-FRP-LSTM approach that extracts features of IHC images from a pre-trained CNN, transforms them into FRPs, and then uses the transformed data as multifeatured time series for classification by an LSTM network. The results illustrate that the CNN-FRP-LSTM models are more favorable in terms of the combination of accuracy and computational complexity than other classification methods for predicting the 5-year survival of rectal-cancer patients using IHC images of RhoB expression on NATs. Such promising results encourage further study of the proposed approach by exploring other pre-trained CNN models and optimal parameters for constructing FRPs of the extracted deep-learning features.

A future investigation of the proposed AI approach is to use RhoB-expression tissues of healthy subjects, tumors, and NATs to discover the predictive power of the three tissue types and molecular changes associated with cancer. Such applications of AI and machine-learning methods can be helpful for posing new questions and validating hypotheses.

#### DATA AND CODE AVAILABILITY

IHC data and Matlab codes are available at the Tuan D. Pham's personal website: <https://sites.google.com/view/tuan-d-pham/codes>, under the name "Rectal-cancer NATs".

#### AUTHOR CONTRIBUTIONS

Tuan D. Pham: Conceptualized the study, developed the AI approach, and wrote the article; Vinayakumar Ravi: Assisted in the AI implementation; Chuanwen Fan and Bin Luo: organized the database; and XFS: Conceptualized the study. All authors analyzed the results and approved the submission.

#### REFERENCES

- [1] J. A. Ramos-Vara and M. A. Miller, "When tissue antigens and antibodies get along: Revisiting the technical aspects of immunohistochemistry—The red, brown, and blue technique," *Veterinary Pathol.*, vol. 51, no. 1, pp. 42–87, Jan. 2014.
- [2] W. J. Howat et al., "Antibody validation of immunohistochemistry for biomarker discovery: Recommendations of a consortium of academic and pharmaceutical based histopathology researchers," *Methods*, vol. 70, no. 1, pp. 34–38, Nov. 2014.
- [3] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.
- [4] P. W. Hamilton et al., "Digital pathology and image analysis in tissue biomarker research," *Methods*, vol. 70, no. 1, pp. 59–73, Nov. 2014.
- [5] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial intelligence in digital pathology—New tools for diagnosis and precision oncology," *Nature Rev. Clin. Oncol.*, vol. 16, no. 11, pp. 703–715, Nov. 2019.
- [6] T. D. Pham, C. Fan, H. Zhang, and X.-F. Sun, "Artificial intelligence-based 5-year-survival prediction and prognosis of DNP73 expression in rectal cancer patients," *Clin. Transl. Med.*, vol. 10, no. 4, p. e159, 2020.
- [7] S. H. Lee, I. H. Song, and H.-J. Jang, "Feasibility of deep learning-based fully automated classification of microsatellite instability in tissue slides of colorectal cancer," *Int. J. Cancer*, vol. 149, no. 3, pp. 728–740, 2021.
- [8] M. S. Kwak et al., "Deep convolutional neural network-based lymph node metastasis prediction for colon cancer using histopathological images," *Frontiers Oncol.*, vol. 10, Jan. 2021, Art. no. 619803.
- [9] J. H. Harrison et al., "Introduction to artificial intelligence and machine learning for pathology," *Arch. Pathol. Lab. Med.*, vol. 145, pp. 1228–1254, Oct. 2021.
- [10] V. Baxi, R. Edwards, M. Montalto, and S. Saha, "Digital pathology and artificial intelligence in translational medicine and clinical practice," *Modern Pathol.*, vol. 35, no. 1, pp. 23–32, Jan. 2022.
- [11] *Rectal Cancer, Cleveland Clinic*. Accessed: Mar. 4, 2022. [Online]. Available: <https://my.clevelandclinic.org/health/diseases/21733-rectal-cancer>
- [12] *Survival Rates for Colorectal Cancer, The American Cancer Society*. Accessed: Mar. 4, 2022. [Online]. Available: <https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/survival-rates.html>
- [13] V. Trevino et al., "Analysis of normal-tumour tissue interaction in tumours: Prediction of prostate cancer features from the molecular profile of adjacent normal cells," *PLoS ONE*, vol. 6, no. 3, Mar. 2011, Art. no. e16492.
- [14] Q. Pan et al., "Normal tissue adjacent to tumor expression profile analysis developed and validated a prognostic model based on Hippo-related genes in hepatocellular carcinoma," *Cancer Med.*, vol. 10, no. 9, pp. 3139–3152, May 2021.
- [15] D. Aran et al., "Comprehensive analysis of normal adjacent to tumor transcriptomes," *Nature Commun.*, vol. 8, no. 1, p. 1077, Dec. 2017.
- [16] K. Graham, X. Ge, A. De Las Morenas, A. Tripathi, and C. L. Rosenberg, "Gene expression profiles of estrogen receptor-positive and estrogen receptor-negative breast cancers are detectable in histologically normal breast epithelium," *Clin. Cancer Res.*, vol. 17, no. 2, pp. 236–246, Jan. 2011.
- [17] M. A. Troester et al., "DNA defects, epigenetics, and gene expression in cancer-adjacent breast: A study from the cancer genome atlas," *Npj Breast Cancer*, vol. 2, no. 1, pp. 1–7, Dec. 2016.
- [18] R. Zhou et al., "Prediction of biochemical recurrence-free survival of prostate cancer patients leveraging multiple gene expression profiles in tumor microenvironment," *Frontiers Oncol.*, vol. 11, Sep. 2021, Art. no. 632571.
- [19] T. D. Pham, "Fuzzy recurrence plots," *EPL (Europhysics Letters)*, vol. 116, no. 5, p. 50008, Dec. 2016.
- [20] U. K. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Comput. Biol. Med.*, vol. 89, pp. 135–143, Oct. 2017.
- [21] S. Saxena, S. Shukla, and M. Gyanchandani, "Pre-trained convolutional neural networks as feature extractors for diagnosis of breast cancer using histopathology," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 3, pp. 577–591, Sep. 2020.
- [22] A. A. M. Teodoro et al., "An analysis of image features extracted by CNNs to design classification models for COVID-19 and non-COVID-19," *J. Signal Process. Syst.*, pp. 1–13, Nov. 2021.
- [23] A. Aljuaid and M. Anwar, "Survey of supervised learning for medical image processing," *Social Netw. Comput. Sci.*, vol. 3, no. 4, p. 292, May 2022.
- [24] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [25] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," 2018, *arXiv:1811.03378*.
- [26] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. 30th Int. Conf. Mach. Learn.*, vol. 28, 2013, p. 3.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," 2015, *arXiv:1502.01852*.
- [28] B. Xu, N. Wang, H. Kong, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolution network," 2015, *arXiv:1505.00853*.
- [29] X. Jin et al., "Deep learning with S-shaped rectified linear activation units," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1737–1743.
- [30] *ImageNet*. Accessed: Jul. 1, 2022. [Online]. Available: <http://www.image-net.org>
- [31] T. D. Pham, C. Fan, H. Zhang, and X.-F. Sun, "Deep learning of P73 biomarker expression in rectal cancer patients," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.
- [32] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," 2017, *arXiv:1707.07012*.



- [33] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [35] R. Paul et al., "Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma," *Tomography*, vol. 2, no. 4, pp. 388–395, 2016.
- [36] M. Awais et al., "Can pre-trained convolutional neural networks be directly used as a feature extractor for video-based neonatal sleep and wake classification?" *BMC Res. Notes*, vol. 13, no. 1, p. 507, Dec. 2020.
- [37] B. Petrovska, E. Zdravetski, P. Lameski, R. Corizzo, I. Stajduhar, and J. Lerga, "Deep learning for feature extraction in remote sensing: A case-study of aerial scene classification," *Sensors*, vol. 20, no. 14, p. 3906, Jul. 2020.
- [38] S. Barburiceanu, S. Meza, B. Orza, R. Malutan, and R. Terebes, "Convolutional neural networks for texture feature extraction. Applications to leaf disease classification in precision agriculture," *IEEE Access*, vol. 9, pp. 160085–160103, 2021.
- [39] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York, NY, USA: Plenum Press, 1981.
- [40] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick (Lecture Notes in Mathematics)*, vol. 898. Berlin, Germany: Springer, 1981, pp. 366–381.
- [41] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [42] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
- [43] Q.-J. Lv et al., "A multi-task group Bi-LSTM networks application on electrocardiogram classification," *IEEE J. Transl. Eng. Health Med.*, vol. 8, pp. 1–11, 2020.
- [44] J. Cheng, Q. Zou, and Y. Zhao, "ECG signal classification based on deep CNN and BiLSTM," *BMC Med. Informat. Decis. Making*, vol. 21, no. 1, p. 365, Dec. 2021.
- [45] T. D. Pham, "Time-frequency time-space LSTM for robust classification of physiological signals," *Sci. Rep.*, vol. 11, p. 6936, Mar. 2021.
- [46] T. D. Pham, "Time-frequency time-space long short-term memory networks for image classification of histopathological tissue," *Sci. Rep.*, vol. 11, no. 1, p. 13703, Jul. 2021.
- [47] L. A. Zadeh, "Similarity relations and fuzzy orderings," *Inf. Sci.*, vol. 3, no. 2, pp. 177–200, Apr. 1971.
- [48] S. R. C. Trial, "Improved survival with preoperative radiotherapy in resectable rectal cancer," *New England J. Med.*, vol. 8, pp. 980–987, Apr. 1997.
- [49] N. Becherer, J. Pecarina, S. Nykl, and K. Hopkinson, "Improving optimization of convolutional neural networks through parameter fine-tuning," *Neural Comput. Appl.*, vol. 31, no. 8, pp. 3469–3479, Aug. 2019.
- [50] T. D. Pham, "Classification of COVID-19 chest X-rays with deep learning: New models or fine tuning?" *Health Inf. Sci. Syst.*, vol. 9, no. 1, pp. 1–11, Dec. 2021.
- [51] T. D. Pham, Y. Watanabe, M. Higuchi, and H. Suzuki, "Texture analysis and synthesis of malignant and benign mediastinal lymph nodes in patients with lung cancer on computed tomography," *Sci. Rep.*, vol. 7, no. 1, p. 43209, Mar. 2017.
- [52] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [53] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.
- [54] A. Geifman. (Apr. 16, 2020). *Are All FLOPs Created Equal? A Comparison of FLOPs Vs Runtime*. [Online]. Available: <https://deci.ai/blog/flops-vs-run-time-comparison/>
- [55] C. Moler. *MATLAB Incorporates LAPACK*. Accessed: Nov. 5, 2022. [Online]. Available: <https://www.mathworks.com/company/newsletters/articles/MATLAB-incorporates-lapack.html>
- [56] *Counting Floating Point Operations in MATLAB*. Accessed: Dec. 8, 2022. [Online]. Available: <http://www.stat.uchicago.edu/lekheng/courses/309f14/flops/>
- [57] (Dec. 3, 2000). *Exchanges in NA Digest Regarding Flops*. [Online]. Available: <http://www.stat.uchicago.edu/lekheng/courses/309f14/flops/vmdd.html>
- [58] *GitHub*. Accessed: Dec. 1, 2022. [Online]. Available: <https://github.com/albanie/convnet-burden>
- [59] *NASNet*. Accessed: Dec. 1, 2022. [Online]. Available: <https://paperswithcode.com/model/nasnet?variant=nasnetlarge>

• • •