This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3168887, IEEE Transactions on Industrial Informatics

Hercules: Deep Hierarchical Attentive Multi-Level Fusion Model with Uncertainty Quantification for Medical Image Classification

Moloud Abdar, Mohammad Amin Fahami, Leonardo Rundo, Petia Radeva, Alejandro F. Frangi, *Fellow, IEEE*, U. Rajendra Acharya, *Senior Member, IEEE*, Abbas Khosravi, *Senior Member, IEEE*, Hak-Keung Lam, *Fellow, IEEE*, Alexander Jung, *Member, IEEE*, Saeid Nahavandi, *Fellow, IEEE*

***Abstract*—The automatic and accurate analysis of medical images (e.g., segmentation, detection, classification) prerequisites for modern disease diagnosis and prognosis. Computer-aided diagnosis (CAD) systems empower accurate and effective detection of various diseases and timely treatment decisions. The past decade witnessed a spur in deep learning (DL)-based CADs showing outstanding performance across many healthcare applications. Medical imaging is hindered by multiple sources of uncertainty ranging from measurement (aleatoric) errors, physiological variability, and limited medical knowledge (epistemic errors). However, uncertainty quantification (UQ) in most existing DL methods is insufficiently investigated, particularly in medical image analysis. Therefore, to address this gap, we propose a simple yet novel hierarchical attentive multi-level feature fusion model with an uncertainty-aware module for medical image classification coined *Hercules*. This approach is tested on several real medical image classification challenges. The proposed *Hercules* model consists of two main feature fusion blocks, where the former concentrates on attention-based fusion with uncertainty quantification module and the latter uses the raw features. *Hercules* was evaluated across three medical imaging datasets, i.e., retinal OCT, lung CT, and chest X-ray. *Hercules* produced the best classification accuracy in retinal OCT (94.21%), lung CT (99.59%), and chest X-ray (96.50%) datasets, respectively, against other state-of-the-art medical image classification methods.**

***Index Terms*—Deep learning, Medical image classification, Early fusion, Feature fusion, Uncertainty quantification, Attention mechanisms.**

M. Abdar, A. Khosravi,and S. Nahavandi are with the Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia. S. Nahavandi is also with the Harvard Paulson School of Engineering and Applied Sciences, Harvard University, Allston, MA 02134 USA (e-mails: mabdar@deakin.edu.au & abbas.khosravi@deakin.edu.au & saeid.nahavandi@deakin.edu.au).

M.A. Fahami is with the Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran (e-mail: mafahami@gmail.com).

L. Rundo is with the Department of Radiology, University of Cambridge, Cambridge, United Kingdom, and the Cancer Research UK Cambridge Centre, University of Cambridge, Cambridge, United Kingdom (e-mail: lr495@cam.ac.uk).

P. Radeva is with Mathematics and Computer Science Department, University of Barcelona, Barcelona, Spain (e-mail: radevap@gmail.com).

A. F. Frangi is with the CISTIB Centre for Computational Imaging & Simulation Technologies in Biomedicine, School of Computing and School of Medicine, University of Leeds, Leeds, U.K., and also with the Departments of Cardiovascular Sciences and Electrical Engineering, KU Leuven, 3000 Leuven, Belgium. (e-mail: a.frangi@leeds.ac.uk).

U. R. Acharya is with the Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Clementi, Singapore, and also with the Department of Biomedical Engineering, School of Science and Technology, SUSS University, Singapore, and also with the Department of Biomedical Informatics and Medical Engineering, Asia University, Taichung, Taiwan (e-mail: aru@np.edu.sg).

H.K. Lam is with the Centre for Robotics Research, Department of Engineering, King's College London, London, United Kingdom (e-mail: hak-keung.lam@kcl.ac.uk).

A. Jung is with the Department of Computer Science, Aalto University, Finland (e-mail: alex.jung@aalto.fi).

* Corresponding author: Moloud Abdar, m.abdar1987@gmail.com

I. Introduction

MACHINE learning (ML) and deep learning (DL) techniques have proven effective across many problems and diverse benchmark datasets. ML and DL met-hods extract hidden information from raw data and make predictions utilizing these models [1]. The performance of predictive models can be hindered by the uncertainty in input data and modelling priors. Imprecise or noisy data and limiting or wrong model assumptions are sources of uncertainty. Handling uncertainties effectively is crucial for trustworthy machine learning, particularly in safety-critical applications like healthcare. Uncertainty quantification plays thus an important role in ML [2], [3].

Identifying the sources of uncertainty most affecting the predictions in our estimation problem is essential to tackle them [4]. There are two major sources of uncertainty in predictive modelling. First, irreducible uncertainty in data gives rise to uncertainty in predictions, also known as aleatoric (or data uncertainty). The second source of uncertainty is knowledge or epistemic uncertainty. Epistemic uncertainty can arise due to wrong assumptions on the model input variables (e.g., their distribution) or the structure of the model itself (e.g., an incomplete mechanistic understanding of the underlying system). Here, the model can produce erroneous predictions even with perfect measurements.

Dealing with uncertainty in medical image analysis pipelines is critical as errors propagate through subsequent image analysis tasks and ultimately can mislead diagnosis. In this study, we propose a novel uncertainty quantification (UQ) method for medical image analysis. Deep neural networks (DNNs) have demonstrated their potential in medical image analysis and computer-aided diagnosis. For example, DNNs exhibited superhuman or comparable performance against clinicians on diabetic retinopathy detection [5], skin cancer classification [6], [7], and many more. Different evaluation metrics such as the Receiver

Operating Characteristic (ROC) Area under the Curve (AUC), F1-score, specificity, sensitivity or accuracy mainly deal with the discriminative power of the predictive models assuming deterministic DNN outputs. Conventional DNNs do not produce well-calibrated, reliable uncertainty estimates in their predictions [8]. Uncertainty estimates are, however, critical in medical image analysis [9]–[11]. Modern DL models still are insufficiently robust to be deployed in real-world clinical scenarios [10].

Aleatoric and epistemic uncertainties can thwart fully automated analysis and diagnosis systems in a clinical practice where critical life decisions are made. Uncertainty estimates can enhance the transparency and trustworthiness of ML methods and assist in promoting their uptake.

A. Research Gaps

Feature fusion is a combination of various features obtained from different branches, layers or networks that can be considered as an omnipresent part of new modern developed networks. Therefore, a comprehensive literature review [9], [10], [12] on previous studies reveal several research gaps in medical image classification:

- Very few studies assess the robustness (robust decision making) during the testing or consider the influence of noisy inputs and their uncertainty.
- Multi-view feature fusion has obtained less attention in medical image classification. Various fusion techniques such as feature fusion (also called early fusion) and decision fusion (also called late fusion) improve DL performance. Ensemble learning can be used to quantify ML and DL predictions uncertainty.

B. Main Contributions

This work proposes a new, simple, yet effective DL model considering its uncertainty for medical image classification. The major contributions of this study are:

- A new hierarchical fusion model coined *Hercules* for accurate classification of medical images with two fusion blocks (Fig. 1 in the supplementary material).
- A modified channel attention (CA) module combining dropout and Monte Carlo (MC) dropout [8] during fusion of features obtained from the CA module.
- A novel hierarchical attention fusion block further enhances the *Hercules* to process information. Inspired by [13], a modified multi-view feature fusion model is proposed using rich feature extractors directly by pre-trained models.
- Last but not least, the proposed Hercules model not only considers the important features coming from attention mechanisms, but it also benefits from richer features.

The remainder of this manuscript is organized as follows. Section II summarizes a few relevant studies. Section III formulates the proposed methodology. The main experiment results of this study are discussed in Section IV. Section V presents the results and comprehensively compares them against previous literature. Finally, Section VI concludes the paper.

II. Literature Review

In this section, a brief review of a few recent studies conducted on deep learning-based medical data analysis, a wide range of fusion approaches used for disease identification, and, finally the importance of using UQ methods in medical data (image) analysis.

A. Deep learning-based medical data analysis

Recent progress in image classification due to DL [14]–[16] is of great assistance to not only the health sector specifically for disease diagnosis using medical imaging, but also many other applications [10]. A wide variety of DL models have shown promising outcomes in complicated diagnostic areas across radiology, pathology, ophthalmology, dermatology and so on. Incorporation of DL techniques in image-based diagnosis have yielded good results as the models achieved strikingly human like performance. Also, automated feature learning capability of DL models make them adaptable and flexible in learning characteristics features which help them in providing better classification results using medical images. Researchers have used different DL techniques in medical image analysis pertaining to different areas – detecting carcinogenic lesions in organs and tissues, understanding pulmonary changes, brain tumor segmentation, diabetic retinopathy, and so on [17]–[19]. Lakshmanaprabu et al. [17] in their work have analyzed CT (computed tomography) scan of lungs to identify the location and staging of oncogenic tissue using linear discriminate analysis (LDA) and optimal DNN (ODNN). The dimension of extracted deep features is reduced using LDA to identify the nodules present in lungs while ODNN is used to classify the lung carcinoma with the help of gravitational search algorithm optimizer. Their proposed method obtained an accuracy of 94.56%, sensitivity of 96.2% and specificity of 94.2%. Generally, pathologists perform visual examination of histopathology slides to evaluate the staging, nature and subtypes of different lesions and tumors. This is similar for lung cancers where adenocarcinoma and squamous cell carcinoma are the most common types and hence necessitates experienced review by pathologists. To address such necessity, Coudray et al. [18] in their work used a deep CNN (DCNN) to automatically classify lung carcinoma. The proposed model used many independent datasets and evaluated the model performance using area under the curve.

B. Information Fusion in Medical Systems

In this sub-section, we briefly reviewed few recent published studies on fusion-based medical data analysis.
Combining multiple images from different imaging modalities into a single fused image to obtain more defined information is the central idea behind fusion-based image analysis. As medical imaging plays an imperative role in the diagnosis and therapy, detailed and accurate images are necessary. Fusion technique is one of the solutions for achieving high spatial and spectral information from a single image. The issue of image quality and heterogeneity

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3168887, IEEE Transactions on Industrial Informatics

can be handled by fusion-based processing of image as fused image has both high quality as well as intensity [20]. Therefore, inclusion of image fusion in medical image analysis is an important facet for accurate diagnosis and prediction. Several studies have been conducted in recent years with the pivotal concept of image fusion and authors have analyzed the idea from a different perspective. Among the various aspects and approaches used for image fusion, level of fusion, when and how to fuse, what to fuse and methods required for rule-based fusion [21] are important. The use of different ways of fusion of one single model or multiple fusion of more than one deep model has been employed to analyze the effect of fusion-based image analysis.

For example, in the area of brain tumor detection fusion-based image analysis has been proposed by Sharif et al. [22]. Brain surface extraction (BSE) has been used initially for skull removal followed by an optimization technique (particle swarm optimization) for tumor segmentation. To extract the deep and inherent features of tumor, local binary pattern (LBP) and genetic algorithm (GA) have been used. The obtained results showed a clear advantage over existing methods for brain tumor detection. For precise segmentation of tumor lesion, Grab cut method has been used in [23]. Serial-based technique has been applied to concatenate the features obtained through transfer learning. The fused feature vector when used for classification achieved a high dice similarity coefficient for brain tumor segmentation. It is well-known that computed tomography (CT) scan images used for tumor detection are accompanied by different challenges such as low distinguishability of the affected region, negative rates and so on.

C. Uncertainty Quantification in Medical Data Analysis

In this section, we summarise some more studies on UQ techniques used in ML and DL for medical data analysis. For example, Wang el al. in [24] have suggested a double-uncertainty weighted method that is loosely based on teacher-student model for semi-supervised segmentation. This method helped in addressing features as well as segmentation uncertainty. Also, for unsupervised learning process, a learnable uncertainty loss has been proposed such that balance can be maintained between supervised and unsupervised training processes. The current methods applied for the purpose of UQs are principally based on Bayesian networks even though they have their shortcomings. Therefore, researchers have used modified methods to address the problems such as approximate posterior inference [25] and frequentist coverage [26]. Uncertainty has also been addressed using deep ensemble approach (collection of a broad range of DNNs) [27], [28] as they provide an advanced approach for handling uncertainty estimation. Standard deviation of relevance score across each model is taken into consideration which helps in providing more accurate and reliable clarifications. This provides more trustworthy and dependable systems for healthcare sectors. Probabilistic DL techniques have been used in building more generalized and effective models to address the intrinsic and parameter uncertainty and enumerate predictive confidence. Extensive studies related to UQ and estimation in DL are being done such that it helps in developing more reliable models as DL plays a pivotal role in medical image analysis. Finally, the combination of various classifiers (as a kind of late fusion or ensemble) has been significantly studied for measuring and quantifying uncertainties in the literature (for example, please see [29]–[31]).

III. Proposed Methodology: *Hercules*

In this section, we explain in detail the proposed model coined *Hercules* and shown in 1. *Hercules* model includes two main blocks:

1) *Block 1*: Hierarchical multi-view fusion of CA and SA modules.
2) *Block 2*: Multi-view feature fusion.

1) Feature Extraction Module: Let us assume a medical image classification problem where each sample comprises one image (x_{img}) and the associated label $y \in \{1, 2, \ldots, N_{\text{cla}}\}$. Here, N_{cla} represents the number of classes (labels). We define the feature extractor ψ_r, where $r \in \{1,2,\ldots,N\}$ is the number of pre-trained feature extractors.

We considered a medical image classification problem that each medical sample is composed of one image (x_{img}), and a label $y \in \{1, 2, \ldots, N_{\text{cla}}\}$, where N_{cla} represents the total number of classes (labels). Thereafter, we define the feature extractor ψ_r where $r \in \{1,2,\ldots,N\}$ is the number of pre-trained feature extractors. In this study, N is equal to 4 as four well-known pre-trained models are used. Therefore, let us define four sets of features (for each pre-trained feature extractor) as:

$$\widetilde{\mathbf{X}}^{\text{r}}_{\text{img}} = \psi_{\text{img}}(\mathbf{X}_{\text{img}}), \tag{1}$$

where $\widetilde{\mathbf{X}}_{\text{img}} \in \mathbb{R}^{i_{\text{img}} \times j_{\text{img}} \times k_{\text{img}}}$ in which i_{img} and $j_{\text{img}} \times k_{\text{img}}$ are the number and order of the feature maps, respectively. As stated earlier, in Eq. 1, four well-known pre-trained models are used in this study, i.e., $\widetilde{\mathbf{X}}^{\text{Dens}}_{\text{img}}$, $\widetilde{\mathbf{X}}^{\text{VGG}}_{\text{img}}$, $\widetilde{\mathbf{X}}^{\text{Effi}}_{\text{img}}$, and $\widetilde{\mathbf{X}}^{Res}_{\text{img}}$. Finally, we propose a new fusion deep learning model that estimates the probability of y by assuming a class $c \in \{1, 2, \ldots, N_{\text{cla}}\}$ given an image:

$$\hat{y} = p(y = c|\widetilde{\mathbf{X}}^{r}_{\text{img}}) \tag{2}$$

2) Attention Mechanism: In this study, the modified Convolutional Block Attention Module (CBAM) [32] was adopted to focus and improve the representation of interest in the input images. In other words, the concept of attention in deep learning (computer vision) can lead to focusing on the most important part of input images. We, therefore, employed CA (Fig. 2) and spatial attention (SA) (Fig. 3 modules) as the main modules of CBAM [32]. The main motivation behind the combination of channel CA and SA modules is that each of these branches can significantly learn ‘what’ and ‘where’ to consider in the channel and spatial axes, respectively. Unlike most previous studies, we

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3168887, IEEE Transactions on Industrial Informatics

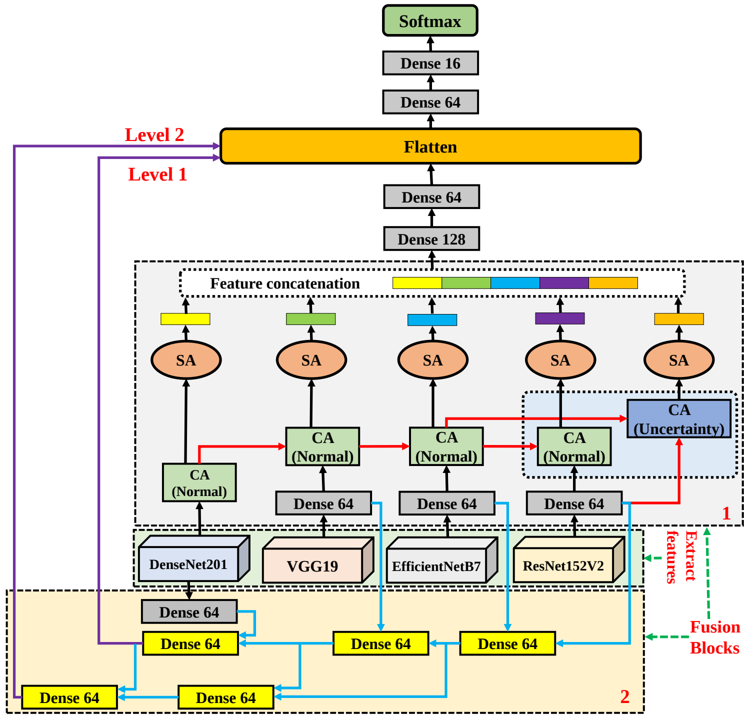


Fig. 1: A detailed overview of the proposed *Hercules* model. As stated earlier, the proposed *Hercules* model includes two main fusion blocks. *CA (Normal)* is the channel attention module with normal dropout, *CA (Uncertainty)* is the CA module with Monte Carlo (MC) dropout as our uncertainty quantification module, and *SA* is the spatial attention module. We adopted numerous pre-trained models as feature extractors in the proposed *Hercules* model, including DenseNet201, VGG19, EfficientNetB7 and ResNet152V2. It should be noted that all the pre-trained models are hierarchically connected.

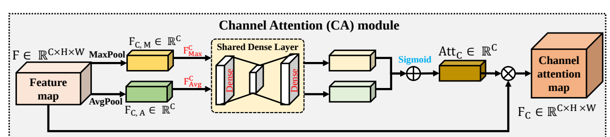
used a modified version of the CA module with the SA module as a part of the proposed *Hercules* (please see block 1 in Fig. 1). As stated in Fig. 1, we proposed hierarchical multi-view fusion of CA and SA modules. The CA mechanism is widely employed in different architectures of CNNs, which uses a scalar to evaluate and represent the importance of each channel. Let $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$ be the image feature tensor in a network, C represents the number of channels, H represents the height of the obtained feature tensor, and finally W represents the width of the feature tensor. Therefore, the attention mechanism can be formulated as [33]:

$$\mathbf{Att} = \sigma(\mathrm{FC}(\mathrm{Compr}(\mathbf{X}))), \tag{3}$$

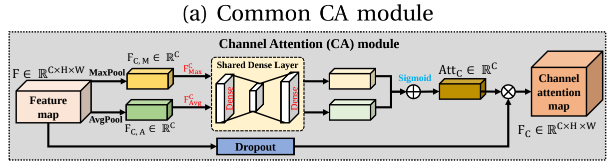
where $\mathbf{Att} \in \mathbb{R}^{C}$ represents the attention vector, σ is the Sigmoid activation function, and FC is the mapping functions (also called a convolution operation), such as fully connected (FC) layer or one-dimensional (1D) convolution, and finally, $\mathrm{Compr}: \mathbb{R}^{C\times H\times W} \to \mathbb{R}^{C}$ represents a compression method. Using this procedure, we can obtain the related attention vector of all C channels. We can scale each channel of input $\mathbf{X}_{\mathrm{img}}$ using the value of corresponding attention as follows [33]:

$$\widetilde{O}_{(:,i,:,:)} = \mathbf{Att}_i \mathbf{X}_{(:,i,:,:)}, \quad \text{s.t.} \quad i \in \{0,1,2,\ldots,C-1\}, \tag{4}$$

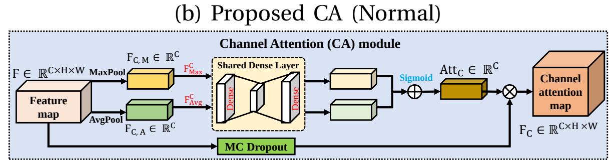
where $\widetilde{O}$ is the output of the applied attention mechanism, $\mathbf{Att}_i$ represents the i-th component of the attention vector, and finally $\mathbf{X}_{(:,i,:,:)}$ is related to the i-th channel of input.



(a) Common CA module



(b) Proposed CA (Normal)



(c) Proposed CA (Uncertainty)

Fig. 2: Main CA modules used in this study and difference with common CA module presented in [32]. Unlike the common CA module (i.e., Fig. 2a), our proposed CA modules (i.e., Figs. 2b and 15c) benefit from applying dropout and MC dropout to prevent over-fitting and quantify uncertainties.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3168887, IEEE Transactions on Industrial Informatics

Fig. 3 shows the general view of the spatial attention (SA) module which can be computed as:

$$F_S = \sigma(FC([\text{AvgPool}(\mathbf{F};\text{MaxPool}(\mathbf{F})])) = \sigma(FC([\mathbf{F}^s_{\text{avg}};\mathbf{F}^s_{\text{max}}])), \tag{5}$$

where σ is the Sigmoid activation function, and FC is the convolution operation, such as fully connected (FC) layer or one-dimensional (1D) convolution.

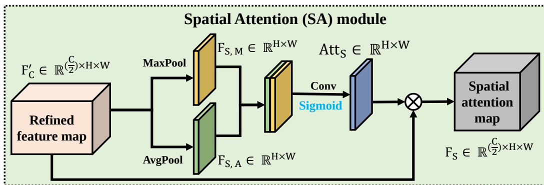


Fig. 3: A general view of the applied SA module as a part of the CBAM presented in [32].

Therefore, the attention process for both CA (Normal) and CA (Uncertainty) can be formulated in Eqs. (6) and (7) as:

$$\mathbf{F}^{\text{Nor}}_C = \mathbf{M}^{\text{Drop}}_c(\mathbf{F}) \otimes \mathbf{F}, \tag{6}$$

$$\mathbf{F}^{\text{Uncer}}_C = \mathbf{M}^{\text{MCD}}_c(\mathbf{F}) \otimes \mathbf{F}, \tag{7}$$

where $\mathbf{F}^{\text{Nor}}_C$, $\mathbf{M}^{\text{Drop}}_c$, $\mathbf{F}^{\text{Uncer}}_C$, and $\mathbf{M}^{\text{MCD}}_c$ are regular CA features, CA maps with classical dropout, uncertainty CA features, and CA with MC dropout, respectively. According to Fig. 1, the final fusion block of the attention mechanism adopted in our proposed *Hercules* model can be formulated as:

$$\begin{aligned} F_{\text{Att}} = \text{Concat}\big[&\text{SA}^{\text{Dens}}_{\text{img}}, \text{SA}^{\text{VGG}}_{\text{img}}, \text{SA}^{\text{Effi}}_{\text{img}}, \\ &\text{SA}^{\text{Res}}_{\text{img}}, \text{SA}^{\text{Uncertainty}}_{\text{CA}}\big]. \end{aligned} \tag{8}$$

According to Fig. 1, the second fusion block is a multi-level feature fusion, a modified version of bottom-top feature fusion proposed by Sindagi and Patel [13]. Thus, the final feature fusion F_{final} of the two main blocks used in the proposed model can be formulated as:

$$F_{\text{final}} = \text{Flatten}[F_{\text{Att}}, F^{\text{Level1}}_{\text{img}}, F^{\text{Level2}}_{\text{img}}], \tag{9}$$

where $F^{\text{Level1}}_{\text{img}}$ and $F^{\text{Level2}}_{\text{img}}$ are the feature fusion obtained at the first and second levels (i.e., Level1 and Level2).

Hercules totals 162,074,532 parameters, of which 1,298,829 are trainable, and 160,775,703 are non-trainable parameters. Moreover, the learning rate and batch size are 0.000001 and 32, respectively. In addition, the dropout rate of the proposed model is 0.3.

3) Model Uncertainty Calculation: The Monte Carlo (MC) dropout proposed by Gal and Ghahramani [8] is a simple, yet efficient, UQ approach used for performing variational inference (VI) on Bayesian Neural Networks (BNNs). The detailed information on MC dropout can be found in [8], [34]. The normal dropout simply switches off some random neurons of the model at each training step, whereas, MC dropout relies on the repeated random sampling procedure to obtain a distribution of input samples. Gal and Ghahramani [8] showed that, the normal dropout approach can be interpreted as a Bayesian approximation which is a well-known probabilistic model. In other words, various networks with different dropped out neurons can be treated as MC samples from the space-related to all available models. Therefore, this can provide mathematical grounds to have a precise reason regarding the method's uncertainty.

TABLE I: Details of each class of the datasets analyzed in this study.

Dataset	Disease	Class	# Samples
Retinal OCT	Retinal disease	CNV	37455
		DME	11598
		DRUSEN	8866
		Normal	51390
		Total	109309
Lung CT	COVID-19	NiCT	5705
		nCT	9979
		pCT	4001
		Total	19685
Chest X-ray	Pneumonia	Normal	1583
		Pneumonia	4273
		Total	5856

IV. Experiments

In this section, we first briefly describe the datasets used in our study. We present the experimental results obtained by our proposed fusion model.

A. Datasets

In this research, we evaluated the proposed *Hercules* model using three publicly available medical imaging datasets [11], [12]: optical coherence tomography (OCT), COVID-19 lung CT scans and pneumonia chest X-ray images. Table I explains the datasets used in our study. Fig. 4 shows some randomly selected samples of the retinal OCT, lung CT, and chest X-ray image datasets, respectively.

(a) Retinal OCT (b) Lung CT (c) Chest X-ray

Fig. 4: Some randomly selected samples of retinal OCT, lung CT, and chest X-ray datasets.

We randomly split the data into two main sets which 80% of data was used for training and the rest (20%) used as our test set. There are two important points regarding the studied datasets in this work. First, we include both binary and multi-class medical image classification tasks. Furthermore, we used different attained images of body parts, i.e., retinal OCT, lung CT, and chest X-ray (Table I).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3168887, IEEE Transactions on Industrial Informatics

TABLE II: Performance comparison of various deep learning models and our proposed *Hercules* fusion model applied to the retinal OCT dataset at the validation stage.

Method	Class	Performance (%)			
		Precision	Recall	F1-score	Accuracy
BARF (Direct) [12]	CNV	80.39	100	89.12	-
	DME	97.14	95.20	96.16	-
	DRUSEN	99.45	72.80	84.06	-
	Normal	93.85	97.60	95.68	-
	Average	92.70	91.40	91.25	91.40
BARF (Cross) [12]	CNV	81.70	100	89.92	-
	DME	99.15	93.20	96.08	-
	DRUSEN	99.48	77.20	86.93	-
	Normal	93.94	99.20	96.59	-
	Average	93.56	92.40	92.38	92.50
Hercules (Ours)	CNV	95.35	95.70	95.48	-
	DME	91.63	87.37	89.48	-
	DRUSEN	80.58	72.77	76.51	-
	Normal	96.64	97.86	97.25	-
	Average	94.50	93.97	94.23	94.21

TABLE III: Performance comparison of various deep learning models and our proposed *Hercules* fusion model applied to the lung CT dataset at the validation stage.

Method	Class	Performance (%)			
		Precision	Recall	F1-score	Accuracy
BARF (Direct) [12]	nCT	100	98.49	99.23	-
	NiCT	97.10	100	98.52	-
	pCT	99.87	99.37	99.61	-
	Average	98.99	99.28	98.12	99.11
BARF (Cross) [12]	nCT	99.40	99.74	99.56	-
	NiCT	99.55	98.94	99.24	-
	pCT	99.62	99.62	99.62	-
	Average	99.52	99.43	99.47	99.49
Hercules (Ours)	nCT	99.75	99.65	99.69	-
	NiCT	99.39	99.47	99.42	-
	pCT	99.50	99.62	99.55	-
	Average	99.59	99.59	99.56	99.59

B. Experimental Results

To evaluate the effectiveness of the proposed Hercules fusion model from different perspectives, we experimented on three real medical image datasets as listed in the previous sub-section (see sub-section IV-A). The obtained results for the retinal OCT, lung CT, and chest X-Ray datasets are shown in Tables II, III, and IV, respectively. The recently proposed fusion models [12], i.e., direct-based BARF and Cross-based BARF, are applied to all three medical datasets. In all tables, the best and second-best (accuracy) obtained results are shown in red and blue colours, respectively. In this study, we reported the weighted average while Abdar et al. [12] used the macro average in their study for the the retinal OCT and chest X-Ray image datasets. Therefore, we also reported the macro average for the CT dataset after applying BARF models. The number of epochs used in both Direct-based BARF and Cross-based BARF models tested on the lung CT dataset is 20 epochs as increasing the number of the epochs led to an increase in the validation loss.

As shown in Tables II, III, and IV, our proposed *Hercules* model outperformed the other two fusion models, i.e., Direct-based BARF and Cross-based BARF. Our proposed fusion model achieved the accuracy of 94.21%, 99.59%, and 96.50% for the retinal OCT, lung CT, and chest X-ray datasets, respectively. The training curves of the proposed *Hercules* model are shown in Figs. 5, 6, and 7, for the retinal OCT, lung CT, and chest X-ray datasets, respectively.

TABLE IV: Performance comparison of various deep learning models and our proposed *Hercules* fusion model applied to the chest X-Ray dataset at the validation stage.

Method	Class	Performance (%)			
		Precision	Recall	F1-score	Accuracy
BARF (Direct) [12]	Normal	100	72.22	83.86	-
	Pneumonia	85.71	100	92.30	-
	Average	92.85	86.11	88.08	89.58
BARF (Cross) [12]	Normal	100	73.93	85.01	-
	Pneumonia	86.47	100	92.74	-
	Average	93.23	86.96	88.87	90.22
Hercules (Ours)	Normal	93.94	93.05	93.49	-
	Pneumonia	97.43	97.77	97.59	-
	Average	96.50	96.50	96.50	96.50

In addition, the validation curves of the proposed model are shown in Figs. 8, 9, and 10, for the retinal OCT, lung CT, and chest X-ray datasets, respectively.

Finally, we provided the accuracy and loss curves versus epochs of our proposed fusion model during the training and validation stage in Fig. 11. As stated in Fig. 11, our proposed fusion model can converge at certain local optima and achieve better local optima. It is illustrated by the smaller losses during training and validation stages with all three medical image datasets used in this study. Our results demonstrate that the proposed fusion model can achieve better sub-optimal performance.

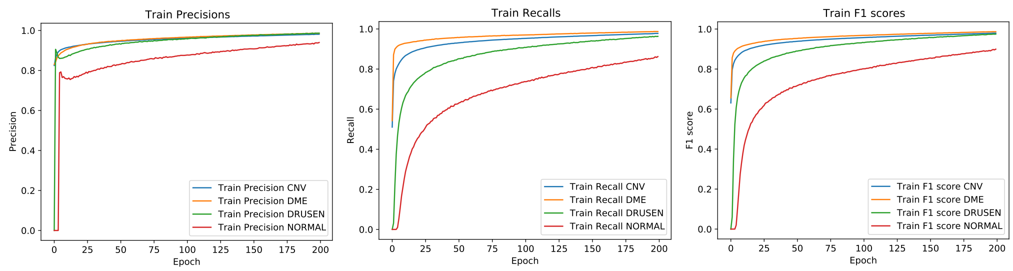
V. Discussion

Medical data analysis is one of the significant applications of artificial intelligence (AI) technologies. In the last few decades, many ML and DL methods have been broadly applied to various applications such as engineering, computer vision and image processing, natural language processing (NLP), weather forecasting, healthcare, etc. [10].

The impressive performance of these DL methods is a reason to use the methods to analyze medical data. In this study, we, therefo-re, apply different well-known DL methods for the classification of various medical image datasets. However, most DL methods rarely achieve the desired performance due to the lack of labelled medical datasets. Different fusion approaches, therefore, have been proposed to deal with this issue [11], [12], [35]–[38]. Inspired by previous studies, we propose a novel fusion model for efficient medical image classification named the *Hercules* model. Unlike most previous fusion models, the proposed *Hercules* fusion model takes advantage of a new approach called *deep hierarchical attentive multi-view fusion* to overcome uncertainties and over-fitting. Meanwhile, the *Hercules* fusion model still shows the early raw valuable features with more useful information.

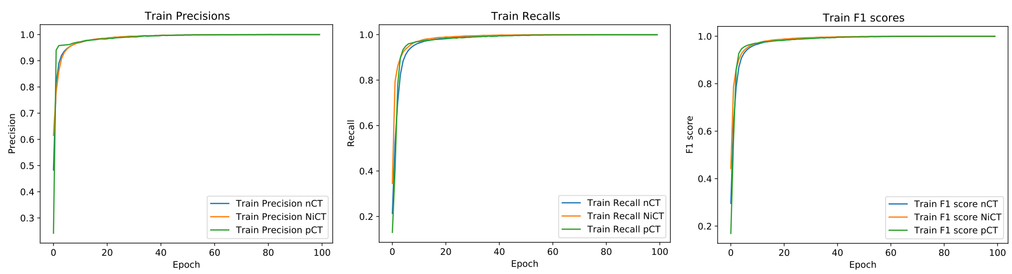
We developed a novel hierarchical multi-view feature fusion model in two main levels, as stated in Fig. 1. As we indicated in Figure 1, in the first fusion block, we hierarchically obtain the features from the first feature extractor (*i.e.* DenseNet201) and we followed this strategy in all four pre-trained models. Unlike the previous multi-view feature fusion, we did not use original features extracted by models (here pre-trained models), but we proposed a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3168887, IEEE Transactions on Industrial Informatics



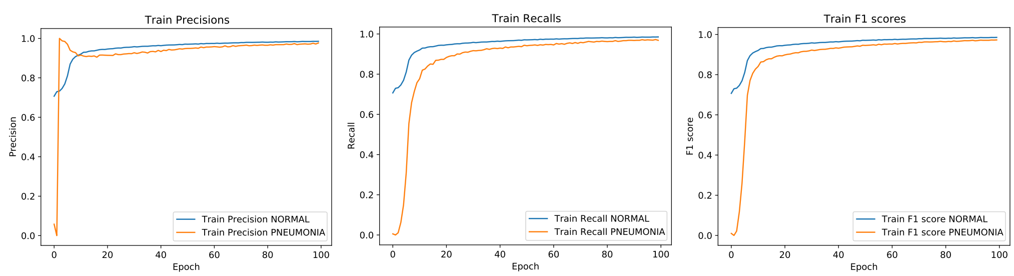
(a) Precision (b) Recall (c) F1-score

Fig. 5: Precision, Recall and F1-score per epoch of the proposed *Hercules* model for the retinal OCT dataset at the training stage.



(a) Precision (b) Recall (c) F1-score

Fig. 6: Precision, Recall and F1-score per epoch of the proposed *Hercules* model for the lung CT dataset at the training stage.



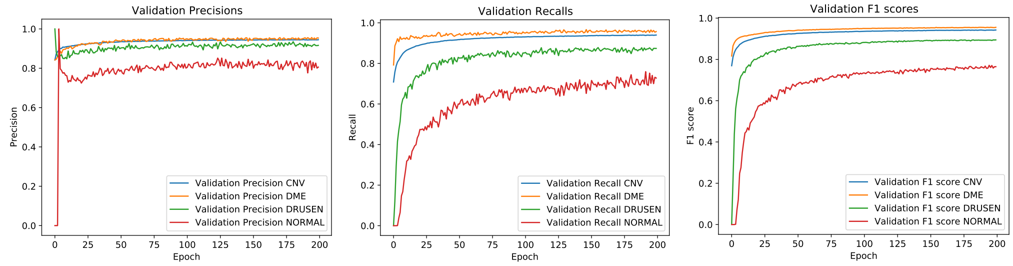
(a) Precision (b) Recall (c) F1-score

Fig. 7: Precision, Recall and F1-score per epoch of the proposed *Hercules* model for the X-Ray dataset at the training stage.

feature fusion based on the attention mechanism. The use of an attention mechanism can be used to detect the most important features in each stage and then combine them. In addition, as most of the previous studies did not consider the uncertainty of their model, we tried to include an uncertainty quantification module in the CA (see the block CA (Uncertainty) block in Fig. 1). Unlike previous studies on those applied CA mechanisms [39]–[41], the applied CA blocks include either simple dropout or MC dropout to deal with over-fitting and uncertainty in the attention modules (see Fig. 2). We quantify uncertainties while applying the attention mechanism. This is the first study on medical image classification considering uncertainties in the attention mechanism merged with a bottom-up multi-view feature fusion to the best of our knowledge. The main cause of over-fitting in different deep learning methods is the lack of sufficient training data. These experimental outcomes indicate that our new fusion model can prevent over-fitting problems caused due to limited medical training data.

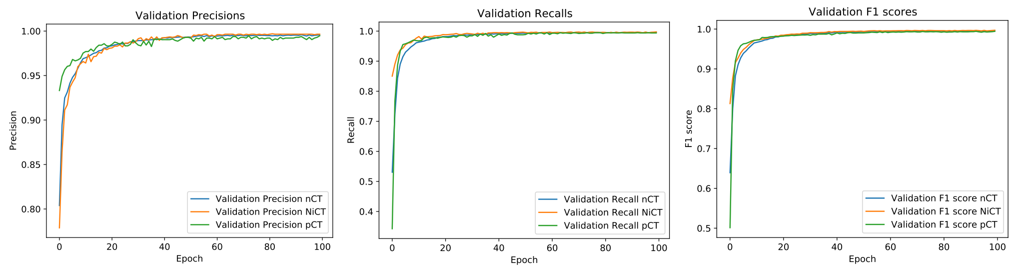
Another strength of the proposed fusion model is to consider its uncertainty during predictions. Like our previous study [12], good performance of the fusion model is considered and quantifying of uncertainty is considered. Thus, to better reveal the importance and impact of our proposed Hercules fusion model, its performance is compared with other methods in Table V. Our new fusion model outperformed the other state-of-the-art medical image classification methods. Table V also reveals that very few studies considered UQ methods, whereas most of the previous studies did not quantify uncertainties. The visual explanation of the proposed *Hercules* fusion model using Gradient-weighted Class Activation Mapping (Grad-CAM) and Grad-CAM++ techniques is presented in Fig. 12 for the retinal OCT, the lung CT and chest X-Ray datasets, respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3168887, IEEE Transactions on Industrial Informatics



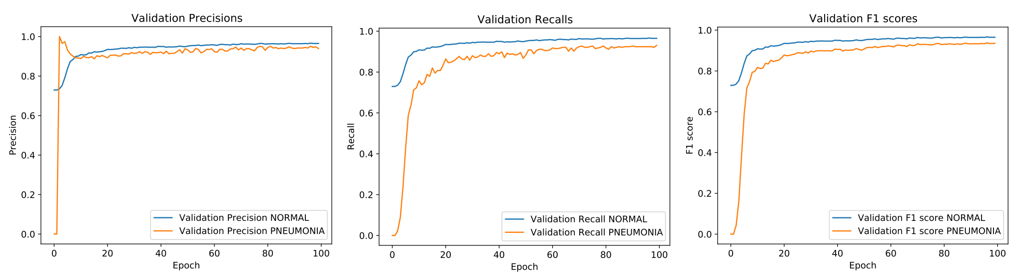
(a) Precision (b) Recall (c) F1-score

Fig. 8: Precision, Recall and F1-score per epoch of the proposed *Hercules* model for the retinal OCT dataset at the validation stage.



(a) Precision (b) Recall (c) F1-score

Fig. 9: Precision, Recall and F1-score per epoch of the proposed *Hercules* model for the lung CT dataset at the validation stage.



(a) Precision (b) Recall (c) F1-score

Fig. 10: Precision, Recall and F1-score per epoch of the proposed *Hercules* model for the chest X-Ray dataset at the validation stage.

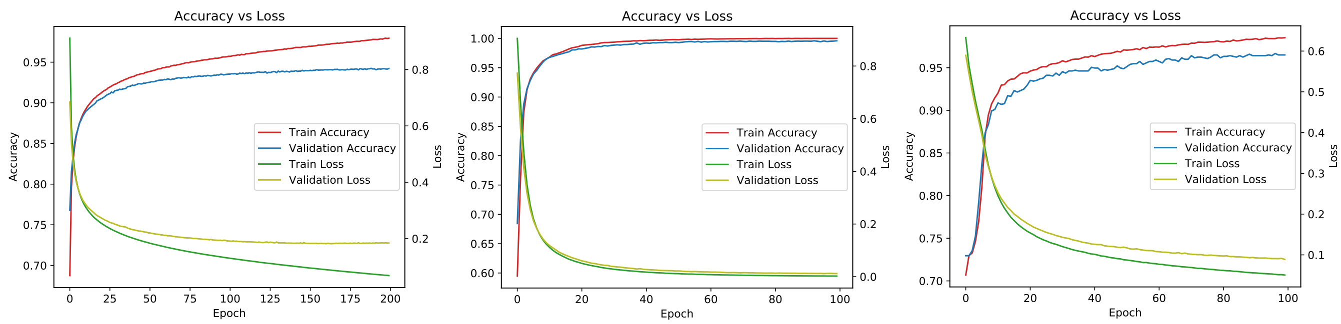
Finally, the probability distribution of the predictions for each class obtained by the proposed *Hercules* fusion model for the retinal OCT, lung CT and chest X-Ray datasets are presented in Figs. 13, 14, and 15, respectively. In Figs. 13, 14, and 15, we show misclassified samples in *orange* color and the correctly classified samples in *blue* color. Note that in Figs. 13, 14, and 15, *Mis* means the misclassified samples, and *Cor* means the correctly classified samples.

It should be noted that since Ning et al. [45] just reported the outcomes per each class, we report the average of the obtained results in [45].

According to Figs. 13, 14, and 15, we make 500 predictions for one sample test and plot its distribution for each dataset. As stated in Figs. 13, 14, and 15, we can see that the proposed *Hercules* is correct and also fairly certain about its predictions. These figures show that there is not overlap between true class and the other classes in each dataset. In other words, having more overlap between the distributions shows that the model is fairly uncertain about its predictions. For example, in Fig. 13, the prediction distributions of misclassified samples belonging to CNV, DME, and NORMAL classes (orange color) are concentrated close to zero whereas the prediction distribution of DRUSEN class (blue color) is concentrated close to one. It is worth noting that there may exist some hard samples to classify and we therefore refer them to medical experts to decide what to do with those examples. An important point in Figs. 13, 14, and 15 is that the overall distribution represents the model's uncertainty (the x-axis) for 500 predictions per each sample test. However, each bar chart shows the model's prediction.

Altogether, the key advantages of using the proposed *Hercules* are as follows: 1) It considers a multi-view approach of feature extraction plus a multi-level feature fusion, 2) The *Hercules* model also uses an UQ method, i.e., MC

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3168887, IEEE Transactions on Industrial Informatics



(a) Retinal OCT (b) Lung CT (c) Chest X-ray

Fig. 11: Accuracy vs Loss per epoch curves obtained for our proposed *Hercules* model during training and validation stages using retinal OCT, lung CT and chest X-ray datasets.

TABLE V: Comprehensive comparison of our results with existing techniques on automated medical image classification.

Dataset	Study	Year	Method	# of Samples	Performance (%)				UQ Method
					Precision	Recall	F1-score	Accuracy	
	Abdar et al. [12]	2021	BARF (Cross)	109309	93.56	92.40	92.38	92.50	Yes
Retinal OCT	Zhang et al. [42]	2021	TS-SSL[1] (1%)	109309	N/A	N/A	N/A	82.60	No
	Zhang et al. [42]	2021	TS-SSL (10%)	109309	N/A	N/A	N/A	93.60	No
	Wang et al. [43]	2021	DVAS[2]	109309	89.74	93.30	91.36	95.13	No
	Yang et al. [44]	2021	ResNet-50 (224)	109309	N/A	N/A	N/A	77.60	No
	Ours	**2021**	*Hercules*	**109309**	**94.50**	**93.97**	**94.23**	**94.21**	**Yes**
	Abdar et al. [11]	2021	UncertaintyFuseNet	19685	99.08	99.08	99.08	99.08	Yes
Lung CT	Abdar et al. [12]	2021	BARF (Cross)	19685	99.52	99.43	99.47	99.49	Yes
	Ning et al. [45]	2020	Deep learning (CNN)	19685	90.28	93.62	N/A	94.59	No
	Javidi et al. [46]	2021	RegCS.CapsDenseNet[3] (IR = 5)	19685	100	99.70	N/A	99.94	No
	Rahman et al. [47]	2021	DenseNet201	18479	94.55	94.56	94.53	95.11	No
	Ours	**2021**	*Hercules*	**19685**	**99.59**	**99.59**	**99.56**	**99.59**	**Yes**
	Abdar et al. [12]	2021	BARF (Cross)	5856	93.23	86.96	88.87	90.22	Yes
Chest X-Ray	Liang and Zheng [48]	2020	CNN	5856	89.10	96.70	92.70	90.50	No
	Lujan-Garcia et al. [49]	2020	Xception-Network	5232	84.30	99.20	91.20	87.98	No
	Chhikara et al. [50]	2020	Deep CNN	5856	90.70	95.70	93.10	90.10	No
	Ours	**2021**	*Hercules*	**5856**	**96.50**	**96.50**	**96.50**	**96.50**	**Yes**

[1] Twin self-supervision based semi-supervised learning; [2] Deep virtual adversarial self-training with consistency regularization; [3] Regularized cost-sensitive CapsNet.

dropout, to quantify uncertainty inside attention mechanism, 3) In general, the proposed *Hercules* model quantifies uncertainties at an early stage of model development, means that during feature extraction from raw images, 4) and last but not least, the proposed *Hercules* model can be considered as an ensemble model. As stated in [10], ensemble techniques are very useful methods which help to quantify uncertainties effectively. Moreover, the radar charts of all average evaluation metrics (i.e., accuracy, F1-score, precision and recall) of the applied methods in terms of the all predicted classes.

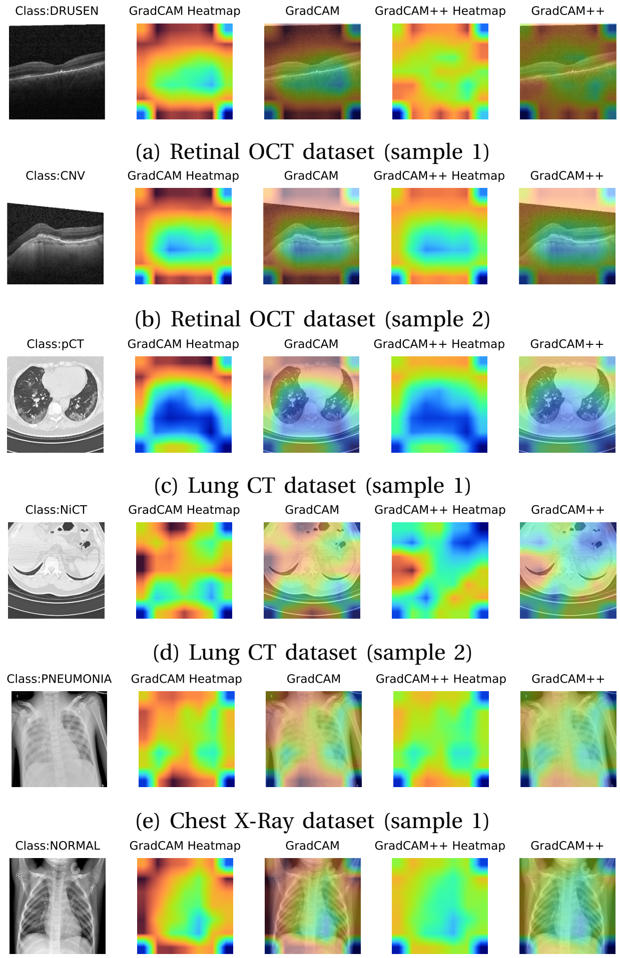
According to Figure 16, we can clearly observe that the proposed *Hercules* fusion model achieves the best classification performance on all datasets and the all evaluation metrics. This indicates that our new fusion model is entirely promising to distinguish and classify the characteristics of different classes in various medical image datasets. Finally, the computation time obtained for the applied methods using Core i7-9700KF@3.60 GHz, 64 GB RAM, and NVIDIA RTX 2080 GPU is shown in Table VI.

TABLE VI: The computation time for the applied methods.

Study	Method	Time		
		Hours	Minutes	Seconds
	BARF (Direct, 20 Epoch) [12]	13	36	48
Retinal OCT	BARF (Cross, 20 Epoch) [12]	13	42	59
	Hercules (Ours, 200 Epoch)	105	56	40
	BARF (Direct, 30 Epoch) [12]	3	22	3
Lung CT	BARF (Cross, , 30 Epoch) [12]	3	22	54
	Hercules (Ours, 100 Epoch)	8	35	10
	BARF (Direct, 40 Epoch) [12]	1	32	36
Chest X-Ray	BARF (Cross, 40 Epoch) [12]	1	32	1
	Hercules (Ours, 100 Epoch)	3	2	1

VI. Conclusion

This study proposed a novel, simple, yet adequately efficient, feature fusion model with a UQ module based on a hierarchical attentive multi-level approach (named *Hercules*). The proposed fusion model is validated using three well-known medical image datasets. Our proposed model has successfully captured the spatial relationships between features extracted by different pre-trained DL models and obtained high classification performance using a new feature fusion approach. The *Hercules* model generalizes the attention mechanism concept as a selective type of feature fusion into the main learning framework. The *Hercules* model yielded the highest classification performance owing to the multi-level feature fusion strategy. *Hercules* outperformed the other related DL methods, thus showing its impact on industrial applications by testing on the medical image classification datasets. We plan to



(a) Retinal OCT dataset (sample 1)
(b) Retinal OCT dataset (sample 2)
(c) Lung CT dataset (sample 1)
(d) Lung CT dataset (sample 2)
(e) Chest X-Ray dataset (sample 1)
(f) Chest X-Ray dataset (sample 2)

Fig. 12: Visual explanation of the proposed *Hercules* fusion model using Grad-CAM and Grad-CAM++ techniques for the retinal OCT, lung CT, and chest X-ray datasets.

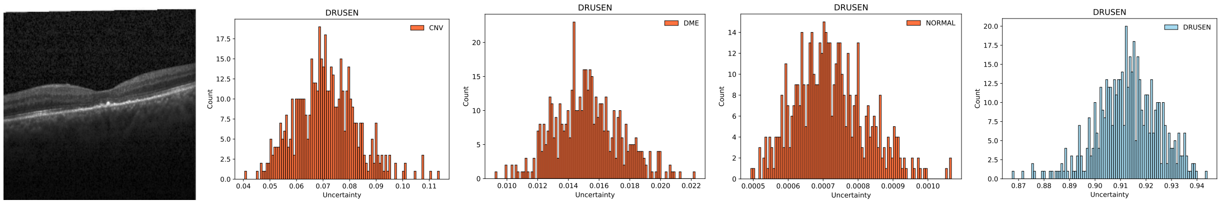
extend our proposed fusion model to perform medical image segmentation and study the impact of multi-level feature fusion on the robustness of the segmentation. Also, this work can be further expanded by employing Bayesian-based ensemble methods to our proposed method.

VII. Acknowledgment

This research was partially supported by the Australian Research Council's Discovery Projects funding scheme (project DP190102181) and The Mark Foundation for Cancer Research and Cancer Research UK Cambridge Centre [C9685/A25177].

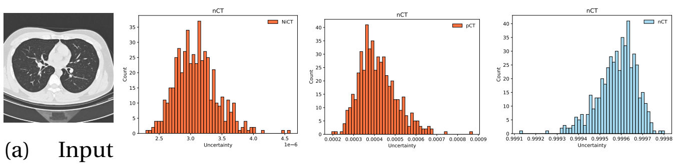
References

[1] A. Malinin, "Uncertainty estimation in deep learning with application to spoken language assessment," Ph.D. dissertation, University of Cambridge, UK, 2019.
[2] T. M. Mitchell, *The need for biases in learning generalizations*. Department of Computer Science, Laboratory for Computer Science Research, Rutgers University, NJ, USA, 1980.
[3] C. C. Aggarwal, X. Kong, Q. Gu, J. Han, and S. Y. Philip, "Active learning: A survey," in *Data Classification: Algorithms and Applications*. CRC Press, 2014, pp. 571–605.
[4] B. T. Phan, "Bayesian deep learning and uncertainty in computer vision," Master's thesis, University of Waterloo, On, Canada, 2019.
[5] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
[6] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
[7] M. Abdar, M. Samami, S. D. Mahmoodabad, T. Doan, B. Mazoure, R. Hashemifesharaki, L. Liu, A. Khosravi, U. R. Acharya, V. Makarenkov *et al.*, "Uncertainty quantification in skin cancer classification using three-way decision-based Bayesian deep learning," *Comput. Biol. Med.*, p. 104418, 2021.
[8] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. International Conference on Machine Learning (ICML)*, 2016, pp. 1050–1059.
[9] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, and S. Wahl, "Leveraging uncertainty information from deep neural networks for disease detection," *Sci. Rep.*, vol. 7, no. 1, pp. 1–14, 2017.
[10] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inform. Fusion*, 2021.
[11] M. Abdar, S. Salari, S. Qahremani, H.-K. Lam, F. Karray, S. Hussain, A. Khosravi, U. R. Acharya, and S. Nahavandi, "UncertaintyFuseNet: Robust uncertainty-aware hierarchical feature fusion with ensemble Monte Carlo dropout for COVID-19 detection," *arXiv preprint arXiv:2105.08590*, 2021.
[12] M. Abdar, M. A. Fahami, S. Chakrabarti, A. Khosravi, P. Pławiak, U. R. Acharya, R. Tadeusiewicz, and S. Nahavandi, "BARF: A new direct and cross-based binary residual feature fusion with uncertainty-aware module for medical image classification," *Inform. Sci.*, vol. 577, pp. 353–378, 2021.
[13] V. A. Sindagi and V. M. Patel, "Multi-level bottom-top and top-bottom feature fusion for crowd counting," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 1002–1012.
[14] X. Wang, Y. Zhao, and F. Pourpanah, "Recent advances in deep learning," 2020.
[15] F. Pourpanah, M. Abdar, Y. Luo, X. Zhou, R. Wang, C. P. Lim, and X.-Z. Wang, "A review of generalized zero-shot learning methods," *arXiv preprint arXiv:2011.08641*, 2020.
[16] Y. Luo, X. Wang, and F. Pourpanah, "Dual VAEGAN: A generative model for generalized zero-shot learning," *Appl. Soft Comput.*, p. 107352, 2021.
[17] S. Lakshmanaprabu, S. N. Mohanty, K. Shankar, N. Arunkumar, and G. Ramirez, "Optimal deep learning model for classification of lung cancer on CT images," *Fut. Gener. Comput. Syst.*, vol. 92, pp. 374–382, 2019.
[18] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nat. Med.*, vol. 24, no. 10, pp. 1559–1567, 2018.
[19] T. Y. Tan, L. Zhang, and C. P. Lim, "Intelligent skin cancer diagnosis using improved particle swarm optimization and deep learning models," *Appl. Soft Comput.*, vol. 84, p. 105725, 2019.
[20] F. E.-Z. A. El-Gamal, M. Elmogy, and A. Atwan, "Current trends in medical image registration and fusion," *Egypt. Inform. J.*, vol. 17, no. 1, pp. 99–124, 2016.
[21] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method based on convolutional neural networks," in *Proc. 20th International Conference on Information Fusion (Fusion)*. IEEE, 2017, pp. 1–7.
[22] M. Sharif, J. Amin, M. Raza, M. Yasmin, and S. C. Satapathy, "An integrated design of particle swarm optimization (pso) with fusion of features for detection of brain tumor," *Pattern Recognition Letters*, vol. 129, pp. 150–157, 2020.
[23] T. Saba, A. S. Mohamed, M. El-Affendi, J. Amin, and M. Sharif, "Brain tumor detection using fusion of hand crafted and deep learning features," *Cogn. Syst. Res.*, vol. 59, pp. 221–230, 2020.
[24] Y. Wang, Y. Zhang, J. Tian, C. Zhong, Z. Shi, Y. Zhang, and Z. He, "Double-uncertainty weighted method for semi-supervised learning," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2020, pp. 542–551.



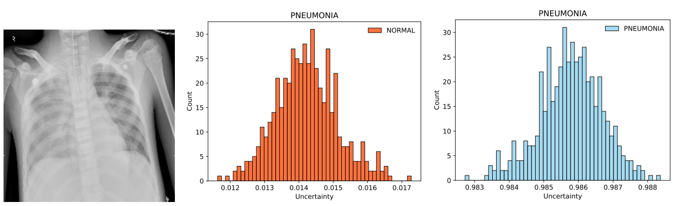
(a) Input image (b) Mis (CNV) (c) Mis (DME) (d) Mis (NORMAL) (e) Cor (DRUSEN)

Fig. 13: Visualization of the probability distribution of the predictions for each class obtained by the proposed *Hercules* fusion model for the retinal OCT dataset.



(a) Input image (b) Mis (NiCT) (c) Mis (pCT) (d) Cor (nCT)

Fig. 14: Visualization of the probability distribution of the predictions for each class obtained by the proposed *Hercules* fusion model for the lung CT dataset.



(a) Input image (b) Mis (NORMAL) (c) Cor (PNEUMONIA)

Fig. 15: Visualization of the probability distribution of the predictions for each class obtained by the proposed *Hercules* fusion model for the chest X-ray dataset.

[25] K. Posch and J. Pilz, "Correlated parameters to accurately measure uncertainty in deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, 2020.

[26] A. Alaa and M. Van Der Schaar, "Discriminative jackknife: Quantifying uncertainty in deep learning via higher-order influence functions," in *Proc. International Conference on Machine Learning (ICML)*. PMLR, 2020, pp. 165–174.

[27] K. K. Wickstrøm, K. ØyvindMikalsen, M. Kampffmeyer, A. Revhaug, and R. Jenssen, "Uncertainty-aware deep ensembles for reliable and explainable predictions of clinical time series," *IEEE J. Biomed. Health Inform.*, 2020.

[28] A. Ashukha, A. Lyzhov, D. Molchanov, and D. Vetrov, "Pitfalls of in-domain uncertainty estimation and ensembling in deep learning," *arXiv preprint arXiv:2002.06470*, 2020.

[29] Z. Yu, L. Li, J. Liu, and G. Han, "Hybrid adaptive classifier ensemble," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 177–190, 2014.

[30] Z.-G. Liu, L.-Q. Huang, K. Zhou, and T. Denoeux, "Combination of transferable classification with multisource domain adaptation based on evidential reasoning," *IEEE Trans. Neural Netw. Learn. Syst.*, 2020.

[31] Z. Liu, X. Zhang, J. Niu, and J. Dezert, "Combination of classifiers with different frames of discernment based on belief functions," *IEEE Trans. Fuzzy Syst.*, 2020.

[32] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[33] Z. Qin, P. Zhang, F. Wu, and X. Li, "FcaNet: Frequency channel attention networks," *arXiv preprint arXiv:2012.11879*, 2020.

[34] Y. Gal and Z. Ghahramani, "Bayesian convolutional neural networks with Bernoulli approximate variational inference," *arXiv preprint arXiv:1506.02158*, 2015.

[35] S. Nemati, R. Rohani, M. E. Basiri, M. Abdar, N. Y. Yen, and V. Makarenkov, "A hybrid latent space data fusion method for multimodal emotion recognition," *IEEE Access*, vol. 7, pp. 172 948–172 964, 2019.

[36] M. E. Basiri, S. Nemati, M. Abdar, S. Asadi, and U. R. Acharrya, "A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets," *Knowl.-Based Syst.*, vol. 228, p. 107242, 2021.

[37] P. Wang, P. Li, Y. Li, J. Wang, and J. Xu, "Histopathological image classification based on cross-domain deep transferred feature fusion," *Biomed. Signal Process. Control*, vol. 68, p. 102705, 2021.

[38] M. R. Hassan, S. Huda, M. M. Hassan, J. Abawajy, A. Alsanad, and G. Fortino, "Early detection of cardiovascular autonomic neuropathy: A multi-class classification model based on feature selection and deep learning feature fusion," *Inform. Fusion*, 2021.

[39] N. Guo, K. Gu, J. Qiao, and J. Bi, "Improved deep CNNs based on nonlinear hybrid attention module for image classification," *Neural Netw.*, vol. 140, pp. 158–166, 2021.

[40] Y. Li, K. Guo, Y. Lu, and L. Liu, "Cropping and attention based approach for masked face recognition," *Appl. Intell.*, vol. 51, no. 5, pp. 3012–3025, 2021.

[41] J. Chen, Y. Chen, W. Li, G. Ning, M. Tong, and A. Hilton, "Channel and spatial attention based deep object co-segmentation," *Knowl.-Based Syst.*, vol. 211, p. 106550, 2021.

[42] Y. Zhang, M. Li, Z. Ji, W. Fan, S. Yuan, Q. Liu, and Q. Chen, "Twin self-supervision based semi-supervised learning (ts-ssl): Retinal anomaly classification in sd-oct images," *Neurocomputing*, vol. 462, pp. 491–505, 2021.

[43] X. Wang, H. Chen, H. Xiang, H. Lin, X. Lin, and P.-A. Heng, "Deep virtual adversarial self-training with consistency regularization for semi-supervised medical image classification," *Medical image analysis*, vol. 70, p. 102010, 2021.

[44] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, and B. Ni, "Medmnist v2: A large-scale lightweight benchmark for 2d and 3d biomedical image classification," *arXiv preprint arXiv:2110.14795*, 2021.

[45] W. Ning, S. Lei, J. Yang, Y. Cao, P. Jiang, Q. Yang, J. Zhang, X. Wang, F. Chen, Z. Geng *et al.*, "Open resource of clinical data from patients with pneumonia for the prediction of covid-19 outcomes via deep learning," *Nat. Biomed. Eng.*, vol. 4, no. 12, pp. 1197–1207, 2020.

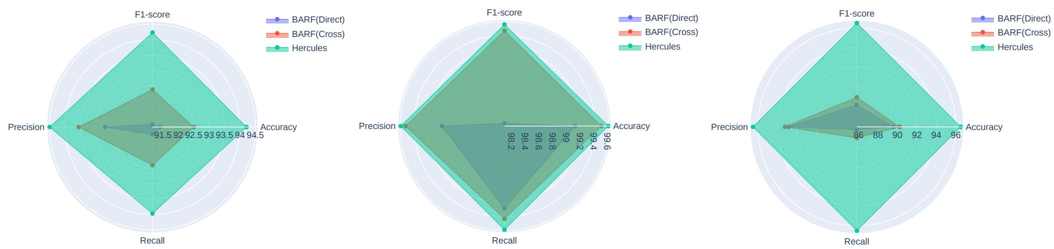
[46] M. Javidi, S. Abbaasi, S. Naybandi Atashi, and M. Jampour, "Covid-19 early detection for imbalanced or low number of data using a regularized cost-sensitive capsnet," *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021.

[47] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaier, M. S. Khan *et al.*, "Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images," *Computers in biology and medicine*, vol. 132, p. 104319, 2021.

[48] G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Comput. Methods Programs Biomed.*, vol. 187, p. 104964, 2020.

[49] J. E. Luján-García, C. Yáñez-Márquez, Y. Villuendas-Rey, and O. Camacho-Nieto, "A transfer learning method for pneumonia classification and visualization," *Appl. Sci.*, vol. 10, no. 8, p. 2908, 2020.

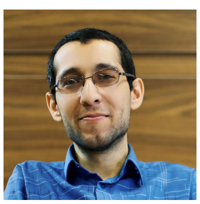
[50] P. Chhikara, P. Singh, P. Gupta, and T. Bhatia, "Deep convolutional neural network with transfer learning for detecting pneumonia on chest X-rays," in *Advances in Bioinformatics, Multimedia, and Electronics Circuits and Signals*. Springer, 2020, pp. 155–168.



(a) Retinal OCT (b) Lung CT (c) Chest X-Ray

Fig. 16: The radar charts of all average evaluation metrics of BARF (Direct), BARF (Cross) and *Hercules* fusion models for the retinal OCT, lung CT and chest X-Ray datasets at the validation stage, respectively.

Moloud Abdar is currently pursuing the Ph.D. degree with the Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Geelong, VIC, Australia. He was a recipient of the Fonds de Recherche du Quebec—Nature et Technologies Award (ranked 5th among 20 candidates in the second round of selection process), in 2019. His research interests include data mining, machine learning, deep learning, computer vision, sentiment analysis, and medical image analysis.



Mohammad Amin Fahami is a Data Scientist at FANAP Company in Tehran, Iran. He recognized as a distinguished researcher at FANAP Company in December 2021. His main research interests lie in the areas of Computer Vision, Deep Learning, Machine Learning and their applications, especially in Robotics, Autonomous Vehicles and Medical Image Analysis. He graduated with a B.Sc. degree in Computer Engineering (Software) at Isfahan University of Technology in September 2017 and ranked 1st among undergraduate students studying the same major as him and graduated with him. He is an expert in Algorithms and Complexity Theory and accordingly he was the teaching assistant for the course "Design and Analysis of Algorithms" at IUT for several semesters. Furthermore, he Ranked 3rd in the Soccer 3D Simulation League at the 12th international RoboCup Iran Open competitions in Tehran, Iran.

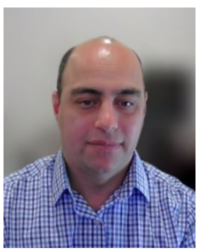
Leonardo Rundo received his Bachelor's and Master's Degrees in Computer Science Engineering from the University of Palermo, Palermo, Italy, in 2010 and 2013, respectively. His Master thesis area was on adaptive data-driven Graphical User Interfaces (GUIs) based on DICOM for medical imaging software. Since December 2013, he was a Research Fellow at the Institute of Molecular Bioimaging and Physiology, National Research Council of Italy (IBFM-CNR), Cefalù (PA), Italy. In February 2019, he obtained his Ph.D. in Computer Science at the University of Milano-Bicocca, Milan, Italy. From November 2018, Leonardo was a Research Associate at the Department of Radiology, University of Cambridge, Cambridge, UK, in collaboration with the Cancer Research UK Cambridge Centre. Since October 2021, he is a Tenure-Track Assistant Professor at the Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, Fisciano (SA), Italy. His research activities are mostly focused on computer vision, medical

Petia Radeva is a Full professor at the Universitat de Barcelona (UB), Head of the Consolidated Research Group "Computer Vision and Machine Learning" at the University of Barcelona (CVMLUB) at UB (www.ub.edu/cvmlub) and Senior researcher in Computer Vision Center (www.cvc.uab.es). She is an Associate editor of Pattern Recognition journal and International Journal of Visual Communication and Image Representation. She is a Research Manager of the State Agency of Research (Agencia Estatal de Investigación, AEI) of the Ministry of Science and Innovation of Spain. Petia Radeva belongs to the top 2% of the World ranking of scientists with the major impact in the field of TIC according to the citations indicators of the popular ranking of Stanford. Moreover, she was awarded IAPR Fellow since 2015, ICREA Academia assigned to the 30 best scientists in Catalonia for her scientific merits since 2014, "Aurora Pons Porrata" of CIARP, Prize "Antonio Caparrós", etc.

Alejandro F. Frangi (Fellow, IEEE) received the undergraduate degree in telecommunications engineering from the Technical University of Catalonia, Barcelona, in 1996, and the Ph.D. degree in medicine from the Image Sciences Institute, University Medical Centre Utrecht on model-based cardiovascular image analysis. He is a Diamond Jubilee Chair of Computational Medicine with the University of Leeds, Leeds, U.K., with joint appointments at the School of Computing and the School of Medicine. He holds an honorary joint position with the Department of Electrical Engineering and Department of Cardiovascular Science, KU Leuven, Belgium. He leads the Centre for Computational Imaging and Simulation Technologies in Biomedicine. He is also a Royal Academy of Engineering Chair of Emerging Technologies. His main research interests lie at the crossroad of medical image analysis and modeling with emphasis on machine learning (phenomenological models) and computational physiology (mechanistic models). He has a particular interest in statistical methods applied to population imaging phenomics and in silico clinical trials. His highly interdisciplinary work has been translated to the areas of cardiovascular, musculoskeletal, and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2022.3168887, IEEE Transactions on Industrial Informatics

U. Rajendra Acharya (Senior Member, IEEE) received the M.Tech. degree from Mangalore University, Mangalore, India, in 1998, the Ph.D. degree from the National Institute of Technology Karnataka, Surathkal, India, in 2002, the D.Eng. degree from Chiba University, Chiba, Japan, in 2008, and the D.Sc. degree from the AGH University of Science and Technology, Kraków, Poland, in 2020.He is currently a Senior Faculty Member with Ngee Ann Polytechnic, Singapore. He is also an Adjunct Professor at the University of Malaya, Kuala Lumpur, Malaysia, and Asia University, Taichung, Taiwan, an Associate Faculty Member at the Singapore University of Social Sciences, Singapore, and an Adjunct Professor at the University of Southern Queensland, Toowoomba, QLD, Australia. He has published more than 500 papers, in refereed international SCI-IF journals (345), international conference proceedings (42), and books (17) with more than 39672 citations in Google Scholar (with H-index of103). He has worked on various funded projects, with grants worth more than5.5 million SGD. He is ranked in the top 1% of the highly cited researchers for the last five consecutive years (2016–2020) in computer science according to the Essential Science Indicators of Thomson. Dr. Acharya is on the editorial board of many journals and has served as a guest editor for many journals.



Abbas Khosravi (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2002, the M.Sc. degree in electrical engineering from the Amirkabir University of Technology, Tehran, in 2005, and the Ph.D. degree from Deakin University, Geelong, VIC, Australia, in 2010. He is currently an Associate Professor with the Institute for Intelligent systems Research and Innovation (IISRI), Deakin University. His research interests include artificial intelligence, data mining, and optimization.

Hak-Keung Lam (M'98-SM'12-F'20) received the B.Eng. (Hons.) and Ph.D. degrees from the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, in 1995 and 2000, respectively. During the period of 2000 and 2005, he worked with the Department of Electronic and Information Engineering at The Hong Kong Polytechnic University as Post-Doctoral Fellow and Research Fellow, respectively. He joined as a Lecturer at King's College London in 2005 and is currently a Reader. His current research interests include intelligent control, computational intelligence and machine learning. He has served as a program committee member, international advisory board member, invited session chair and publication chair for various international conferences and a reviewer for various books, international journals and international conferences. He was an associate editor for IEEE Transactions on Circuits and Systems II: Express Briefs and is an associate editor for IEEE Transactions on Fuzzy Systems, IET Control Theory and Applications, International Journal of Fuzzy Systems, Neurocomputing and Nonlinear Dynamics; and guest editor for a number of international journals. He is on the editorial board of Journal of Intelligent Learning Systems and Applications, Journal of Applied Mathematics, Mathematical Problems in Engineering, Modelling and Simulation in Engineering, Annual Review of Chaos Theory, Bifurcations and Dynamical System, The Open Cybernetics and Systemics Journal, Cogent Engineering and International Journal of Sensors, Wireless Communications and Control. He was named as a highly cited researcher and is an IEEE Fellow. He is a coeditor of two edited volumes: Control of Chaotic Nonlinear Circuits (World Scientific, 2009) and Computational Intelligence and Its Applications (World Scientific, 2012), and author/-

Alexander Jung obtained a Ph.D. (with sub auspiciis) in 2012 from Technical University Vienna. After Post-Doc periods at TU Vienna and ETH Zurich, he joined Aalto University as an Assistant Professor for Machine Learning in 2015. He leads the group "Machine Learning for Big Data," which studies explainable machine learning in network structured data. Alex first-authored a paper that won a Best Student Paper Award at IEEE ICASSP 2011. He received an AWS Machine Learning Research Award and was the "Computer Science Teacher of the Year" at Aalto University in 2018. Currently, he serves as an associate editor for the IEEE Signal Processing Letters and as the chair of the IEEE Finland Joint Chapter on Signal Processing and Circuits and Systems (SP/CAS). His textbook "Machine Learning: The Basics" was published in 2022 by Springer, Singapore.

Saeid Nahavandi received a Ph.D. from Durham University, U.K. in 1991. He is an Alfred Deakin Professor, Pro Vice-Chancellor, Chair of Engineering, and the Founding Director of the Institute for Intelligent Systems Research and Innovation at Deakin University. His research interests include modelling of complex systems, robotics and haptics. He has published over 900 scientific papers in various international journals and conferences. Professor Nahavandi is Editor-In -Chief: IEEE SMC Magazine, the Senior Associate Editor: IEEE Systems Journal, Associate Editor of IEEE Transactions on Systems, Man and Cybernetics: Systems, and IEEE Press Editorial Board member. Professor Nahavandi is a Fellow of IEEE (FIEEE), Engineers Australia (FIEAust), the Institution of Engineering and Technology (FIET). Saeid is a Fellow of the Australian Academy of Technology and Engineering (ATSE).