IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 September 2023, accepted 22 November 2023, date of publication 30 November 2023,
date of current version 6 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3338379

RESEARCH ARTICLE

3D Efficient Multi-Task Neural Network for Knee Osteoarthritis Diagnosis Using MRI Scans: Data From the Osteoarthritis Initiative

PAULINE SHAN QING YEOH[1], SIEW LI GOH[2], KHAIRUNNISA HASIKIN[1], XIANG WU[3], (Member, IEEE), AND KHIN WEE LAI[1], (Senior Member, IEEE)

[1]Department of Biomedical Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[2]Faculty of Medicine, Universiti Malaya, Kuala Lumpur 50603, Malaysia
[3]School of Medical Information and Engineering, Xuzhou Medical University, Xuzhou, Jiangsu 221110, China

Corresponding authors: Khin Wee Lai (lai.khinwee@um.edu.my) and Xiang Wu (wuxiang@xzhmu.edu.cn)

This work was supported in part by the Ministry of Higher Education Malaysia and Universiti Malaya under Fundamental Research Grant Scheme (FRGS) Project under Grant FRGS/1/2023/SKK05/UM/02/2.

ABSTRACT Deep learning, particularly Convolutional Neural Networks, has demonstrated effectiveness in computer-aided diagnosis applications, including knee osteoarthritis analysis. Two of the most common tasks done in medical imaging are segmentation and classification tasks. This research investigates the feasibility of multi-task models for volumetric analysis using Magnetic Resonance Imaging scans in knee osteoarthritis diagnosis, while considering computational efficiency. In order to leverage the correlation between segmentation and classification tasks, two 3D multi-task models, OA_MTL (Osteoarthritis_Multi-Task Learning) and RES_MTL (Residual_Multi-Task Learning) models are developed to simultaneously segment knee structures and classify knee osteoarthritis incidence. The performance of the multi-task models is evaluated against single-task baseline models and other existing convolutional neural network models using a total of eight different performance metrics, while comparing the computational complexity among the models. Experimental results demonstrate that multi-task model leverages the information of segmentation task to improve the classification performance. OA_MTL is a multi-task model that incorporates an encoder-decoder architecture, residual modules, and depthwise separable convolutions for enhanced performance. OA_MTL achieves superior performance for classification tasks with an accuracy score of 0.825, and a comparable segmentation DSC score of 0.915. OA_MTL achieves a favorable trade-off between computational complexity and model performance. The contribution of this work includes an approach that simultaneously performs knee structure segmentation and osteoarthritis classification in 3D MRI, which addresses the need for efficient models in the field of medical imaging, specifically on computationally challenging 3D medical imaging applications.

INDEX TERMS Convolutional neural network, deep learning, knee osteoarthritis, magnetic resonance imaging, multi-task model.

I. INTRODUCTION

In recent years, artificial intelligence, particularly in the area of deep learning, has revolutionized various domains, especially in medical image analysis. In the field of computer-aided medical diagnosis, the application of deep convolutional neural networks (CNN) has greatly benefited the field where medical imaging plays a crucial role as the input for accurate diagnosis [1], [2]. Moreover, CNNs are well suited in medical diagnosis due to its ability to extract complex patterns and relationships from high-dimensional medical imaging data [3]. Previous literature has shown that CNNs achieved remarkable performance in medical diagnostic tasks such as segmentation and classification through

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

medical images and have become the mainstream approach for disease analysis [1].

Over the past years, developing such methods has been a significant research focus with a notable shift from 2D CNNs to 3D CNNs [3]. Previous research employs 2D CNNs on volumetric medical images by a slice-by-slice manner [4], ignoring the relationship among image slices within the volume. With the increased availability of high-end computing resources, researchers have utilized 3D models to process volumetric medical data, employing 3D convolutions to capture the correlation between adjacent slices to extract the three-dimensional spatial context features of medical images. Unlike 2D CNN, this allows a more comprehensive analysis of the anatomical structures or abnormalities which potentially improves the automated tasks' performance. The robustness of this transition provides new research directions for future research [5]. Despite the advancement of 3D CNN in computer-aided diagnosis, there is still a gap of knowledge in 3D CNN applications. In addition, the utilization of 3D CNN architectures incurs significant computational overheads such as large number of parameters and computational costs, due to multiple layers of 3D convolutions, making them impractical for real-world deployment on standard hardware. Given the constraint of computing hardware resources in real-world scenarios, there is a need for efficient and lightweight models where the models achieve a good trade-off between computational efficiency and model accuracy, to allow a more feasible practical application. The development of efficient neural networks that can achieve high performance while utilizing fewer parameters and computing resources is an ongoing research area in the medical imaging field, especially on 3D medical imaging tasks that are computationally expensive due to the nature of the 3D medical images involved [1].

The multi-task technique serves as an efficient option to reduce the high computational costs associated with 3D neural networks [6]. Majority of the previous literature on medical diagnosis focuses on single-task learning models where the classification of disease and segmentation of anatomical structures were considered as two independent tasks, ignoring the correlation between the tasks [2]. Medical image classification is the process of assigning the medical scans into predefined categories based on extracted features. Classification can be applied to serve different purposes, such as differentiating between normal and abnormal images to detect the presence of a disease, or prediction of a certain condition either in terms of disease progression or severity staging. On the other hand, medical image segmentation involves partitioning medical images into multiple segments based on the region of interest (ROI) according to visual characteristics. Basically, it is the identification of ROI pixels or voxels from the background of medical scans to provide clearer visualization by either delineating or localizing specific regions of interest or anatomical structures. Moreover, both tasks have their respective limitations and may yield mutual benefits when performed together [7]. For example, the tissue type and structure information from segmentation tasks might be crucial for disease stage classification [2]. Multi-task learning emerges as one of the deep learning approaches to address the aforementioned challenges by performing multiple tasks simultaneously in a single model to improve diagnosis accuracy [7]. This technique enables effective information sharing among tasks, facilitating the learning of interrelated concepts and improves the general performance of the tasks, by optimizing several learning tasks simultaneously [8]. It has been proven that this can ultimately lead to improved performance compared to single-task learning [8], [9].

This research leverages the potential of multi-task learning and 3D CNNs in the domain of knee osteoarthritis diagnosis. As the ageing population continues to grow, the prevalence of this disease is expected to increase significantly in the coming years. Hence, it is crucial to develop a computer-aided diagnosis tool that performs a comprehensive analysis for this disease using 3D MRI scans. There are numerous deep learning applications in the field of knee osteoarthritis [1], [10]. While previous studies have shown promising results on osteoarthritis diagnosis using 2D CNN technique on plain radiography [11], knee OA is a complex whole joint disease [12], making MRI the preferred imaging tool for examination [4]. Moreover, due to the nature of 3D medical images which contain rich information, it allows intrinsic connections between multiple tasks. Therefore, this work employs 3D CNN to make full use of the 3D volumetric information of the knee condition. Besides, existing OA deep learning studies are restricted to a single task [13]. In this research, two of the important tasks that will be focused on are: knee joint structures segmentation and OA staging classification. The segmentation task aims to extract tissue structures such as bones and cartilages, while the classification task aims to detect the presence of osteoarthritis. Unlike existing methods that perform classification and segmentation separately, this research aims to investigate the feasibility of multi-task 3D CNN approach for knee OA diagnosis. Therefore, a research question arises: Can a multi-task model be developed to perform volumetric analysis for knee osteoarthritis diagnosis and achieve a better trade-off between efficiency and accuracy compared to single-task models?

To answer the question above, in this work, we address the aforementioned challenges by investigating the potential of 3D multi-task model of two different architectures in osteoarthritis diagnosis. It is worth noting that previous multi-task methods have not been applied to exploit the inherent relationship between knee joint segmentation and osteoarthritis detection tasks. To mitigate the aforementioned high computational complexity of 3D CNN models, besides using multi-task techniques as described previously, we incorporated depthwise separable convolutional layers [14] in both of the multi-task models for a more lightweight structure. Furthermore, the two common approaches for multi-task learning are the hard or soft parameter sharing

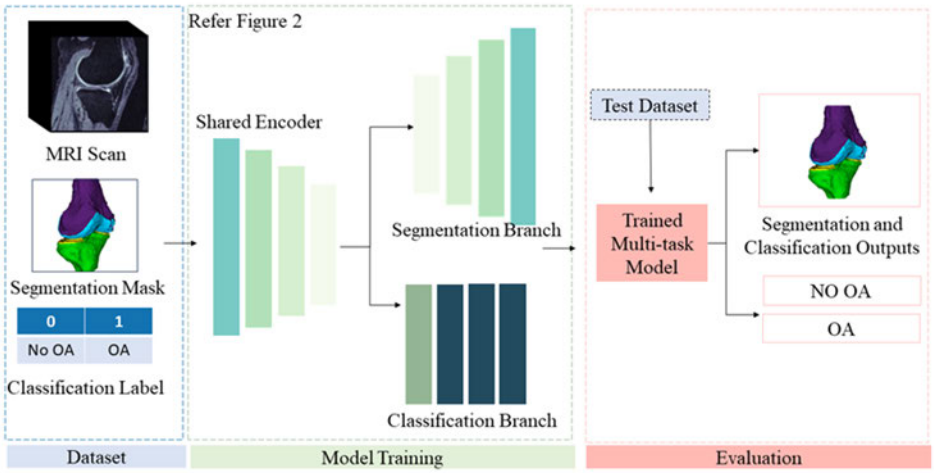

FIGURE 1. Research framework.

of hidden layers. In our work, the multi-task technique is applied through hard-parameter sharing, which involved sharing certain network layers among tasks while keeping other layers specific to each task, as it significantly reduces the number of parameters and computational costs associated with the model [9]. In this paper, we propose 3D multi-task neural network for joint segmentation of knee structures and classification of knee osteoarthritis. The key contributions of this paper are as follows:

1. We propose an end-to-end multi-task approach for jointly segmenting knee structures and classifying knee osteoarthritis using 3D MRI images. To our best knowledge, this is one of the pioneering works that performs knee osteoarthritis segmentation and classification in 3D MRI simultaneously.
2. We explore and compare two different network architectures for conducting multi-task learning, by modifying the encoder that serves as a parameter sharing between two tasks. For the first architecture, we incorporate several residual blocks in the encoder of the encoder-decoder network. In the second architecture, the encoder of the network is replaced with 3D ResNet-18. Both architectures employ shared encoders with separate task-specific decoders.
3. We extensively compare the proposed architectures with other popular CNN architectures and single-task baseline models to demonstrate the superior performance of multi-task networks.

II. METHODOLOGY

A. GENERAL OVERVIEW

The proposed multi-task neural networks are designed such that it integrates segmentation and classification in a single-stage end-to-end CNN architecture. Two architectures are proposed with different shared encoders for feature extraction, with the same decoder segmentation branch and classification branch. The shared encoder extracts relevant features from the 3D MRI volume to be utilized by both tasks. The models take a 3D MRI scan into the neural network and produce two different outputs simultaneously, which are the incidence of knee osteoarthritis and segmentation masks of four different knee structures: femur bone (FB), femoral cartilage (FC), tibia bone (TB) and tibial cartilage (TC).

To further validate the performance of the proposed multi-task, single-task networks are extracted from the proposed network as baseline models to compare the performances. Here, the baseline segmentation-task network consists of the shared encoder and the segmentation branch, which is the decoder that is symmetrical to the encoder. Whereas for the baseline classification-task network, it includes the shared encoder and the classification branch. The general overview of the research framework is presented in Fig. 1

B. DATA ACQUISITION AND PRE-PROCESSING

The dataset used in this work is obtained from the publically available knee MRI scans from the Osteoarthritis Initiative (https://nda.nih.gov/oai/). 400 3D sagittal double-echo steady-state (DESS) MRI scans from different subjects from the baseline time point are involved. These scans were acquired using Siemens 3T Trio systems. For segmentation tasks, the respective segmentation masks are acquired from Zuse Institute Berlin (ZIB) where each volume consists of background (BG), FB, FC, TB and TC [15]. For classification tasks, the presence of knee osteoarthritis is determined based on Kellgren-Lawrence (KL) grading, a common OA severity grading scale [12]. Similar to previous studies [16], [17], KL grades of 0 and 1 were categorized into ''No OA'' or class 0 while KL grades of 2, 3, and 4 were categorized into ''OA'' or class 1.

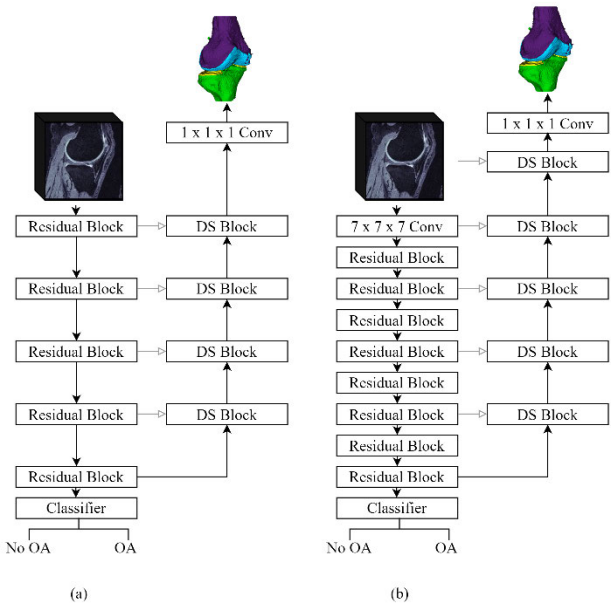

FIGURE 2. Architecture of multi-task networks. (a) OA_MTL (b) RES_MTL.

In this study, datasets from the Osteoarthritis Initiative were initially provided in Digital Imaging and Communications in Medicine (DICOM) format. To ensure consistency, all data used in the study underwent format conversion to the Neuroimaging Informatics Technology Initiative (NIfTI) standard. This preprocessing workflow was implemented using the SimpleITK library. All of the MRI volumes were resized into $160 \times 160 \times 160$ due to Graphics Processing Unit (GPU) memory limitations. Then, the scans are Z-normalized and standardized and are split into train, validation, and test sets. The training set contains 280 MRI scans (87 for No OA, 193 for OA), the validation set has 80 MRI scans (32 for No OA, 48 for OA), and the test set consists of 40 MRI scans (16 for No OA, 24 for OA).

C. MULTI-TASK NETWORKS ARCHITECTURES

Two different multi-task network architectures are explored where the base architectures and modules of the networks are inspired by the U-Net, residual modules from ResNet [18] and the depthwise separable convolutions. We utilize residual blocks [18] in our work as it has shown superior ability to extract features of knee osteoarthritis in our previous work [16], with a reduced computation complexity. The two models are presented in Fig. 2 and are named as (a) OA_MTL (Osteoarthritis_Multi-Task Learning) and (b) RES_MTL (Residual_Multi-Task Learning), respectively. Both networks have the two tasks share the same encoding path and then split into two task-specific paths.

1) ENCODER PATH

As described above, the two models differ in terms of their encoder where OA-MTL employs residual modules (shown in Fig. 3(a)) and the RES-MTL adopts the entire feature extractor of the ResNet-18 as the encoder. For OA-MTL, the encoder path consists of four residual blocks and four $2 \times 2 \times 2$ max-pooling layers with stride of 2. One residual block makes up the bottleneck. The proposed architecture implements 32, 64, 128, 256 and 512 filters at each stage of the encoder-decoder paths. The shared encoder fed the obtained representations into segmentation and classification branches, allowing a smaller computational cost while benefiting from both tasks. The segmentation branch is used to output the segmentation mask whereas the classification branch is used to detect knee OA incidence.

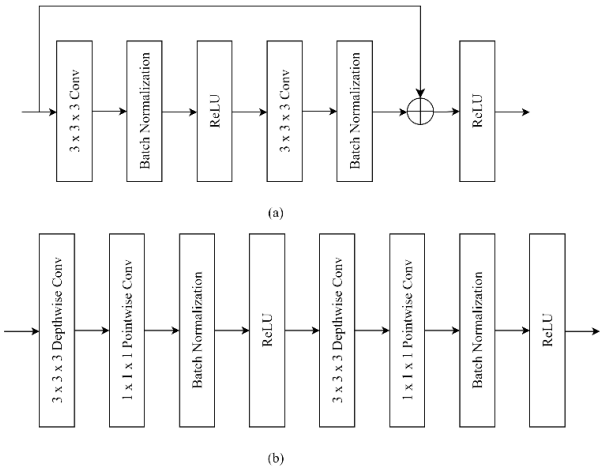

FIGURE 3. (a) Residual Block and (b) Depthwise Separable Convolutional Block.

2) SEGMENTATION BRANCH

The segmentation decoder path is symmetrical to the shared encoder. It consists of four transposed convolutions, where the output of the transposed convolution is then concatenated with the output of the corresponding encoding layer of the neural network using skip connections. Each is then followed by depthwise separable convolutional blocks (DS Block), as shown in Fig. 3(b). A $1 \times 1 \times 1$ standard convolution is implemented on the last layer of the network to reduce the number of output channels from 32 to 5, which is the number of classes in the segmentation mask All layers in the encoder-decoder paths except the $1 \times 1 \times 1$ convolutional layer will be followed by batch normalization and Rectified Linear Unit (ReLU) nonlinear activation function.

3) CLASSIFICATION BRANCH

The classification branch is built by extending the bottleneck of the encoder path with a classifier. The classifier is comprised of an average pooling layer followed by three fully connected layers with layer outputs of 128 and 32 with the last final output based on number of classes, which in this case is two. The first two fully connected layers are accompanied by a ReLU layer and a dropout layer with a rate of 0.5 is applied.

D. TRAINING SPECIFICATIONS

All experiments were conducted using Python using PyTorch deep learning framework, on a workstation equipped with Xeon W-2225 Central Processing Unit Intel and NVIDIA RTX A6000 Graphics Processing Unit (GPU) with Random Access Memory (RAM) of 32.0 GB. All the models are optimized via ADAM optimizer, using a learning rate of 1e-4 and a batch size of 2. All models were trained with a maximum epoch of 100 with an early stop patience of 10. Weighted Random Sampler, an oversampling strategy implemented by PyTorch [19], was used in the model training to address the class imbalance issue for classification tasks.

For the baseline networks and single-task networks, the loss function used was based on the tasks where segmentation tasks use the dice loss function whereas classification tasks use cross-entropy loss function. For the multi-task networks, we combined dice loss function from segmentation task and cross-entropy loss function from classification task to a multi-task loss function such that it can optimize two tasks simultaneously. The multi-task loss function (L_{TOTAL}) is formulated as follows:

$$\mathrm{L_{TOTAL}} = \alpha\,(\mathrm{L_{SEG}}) + (1-\alpha)(\mathrm{L_{CLS}}) \tag{1}$$

where α is the weight that balances both loss functions, and it is set to 0.7. $\mathrm{L_{SEG}}$ denotes the segmentation loss function whereas $\mathrm{L_{CLS}}$ denotes the classification loss function.

The formula of both loss functions can be expressed as follows:

$$\mathrm{L_{CLS}}\,(p, q) = -\sum_{i=1}^{N} p_i \log(q_i) \tag{2}$$

$$\mathrm{L_{SEG}}\,(p, q) = 1 - \frac{2\sum_{i=1}^{N} p_i q_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} q_i^2} \tag{3}$$

where p is the ground truth value of either 0 or 1, q is the predicted probability for i^{th} class and N is total number of classes.

E. EVALUATION STRATEGY AND METRICS

The performance evaluation of the models is conducted by assessing the performance of segmentation and classification tasks separately. It should be noted that although some metrics for the tasks are the same, they account for different definitions based on the nature of the task.

For classification tasks, we adopt accuracy (ACC), precision (PRE), recall (REC) and F1-Score to validate the classification performance of the model. All the classification metrics aforementioned above were computed according to the formulas below:

$$\mathrm{ACC} = \frac{\mathrm{TP_c} + \mathrm{TN_c}}{\mathrm{TP_c} + \mathrm{FP_c} + \mathrm{FN_c} + \mathrm{TN_c}} \tag{4}$$

$$\mathrm{PRE} = \frac{\mathrm{TP_c}}{\mathrm{TP_c} + \mathrm{FP_c}} \tag{5}$$

$$\mathrm{REC} = \frac{\mathrm{TP_c}}{\mathrm{TP_c} + \mathrm{FN_c}} \tag{6}$$

$$\text{F1 score} = \frac{2\mathrm{TP_c}}{2\mathrm{TP_c} + \mathrm{FP_c} + \mathrm{FN_c}} \tag{7}$$

where $\mathrm{TP_c}$ is the number of correctly classified samples as OA, $\mathrm{FP_c}$ is the number of incorrectly classified samples as OA, $\mathrm{TN_c}$ is the number of correctly classified samples as healthy (No OA) and $\mathrm{FN_c}$ is the number of incorrectly classified samples as healthy (No OA).

Evaluation metrics for the segmentation tasks includes precision (PRE), recall (REC), Dice Similarity Coefficient (DSC) and Jaccard Similarity Coefficient (JSC).The metrics are as follows:

$$\mathrm{PRE} = \frac{\mathrm{TP_s}}{\mathrm{TP_s} + \mathrm{FP_s}} \tag{8}$$

$$\mathrm{REC} = \frac{\mathrm{TP_s}}{\mathrm{TP_s} + \mathrm{FN_s}} \tag{9}$$

$$\mathrm{DSC} = \frac{2\mathrm{TP_s}}{2\mathrm{TP_s} + \mathrm{FP_s} + \mathrm{FN_s}} \tag{10}$$

$$\mathrm{JSC} = \frac{\mathrm{TP_s}}{\mathrm{TP_s} + \mathrm{FP_s} + \mathrm{FN_s}} \tag{11}$$

where $\mathrm{TP_s}$ is the true positive of specific class of interest, $\mathrm{FP_s}$ is the false positive of specific class of interest, $\mathrm{TN_s}$ is the true negative of specific class of interest and $\mathrm{FN_s}$ is the false negative of specific class of interest. For segmentation tasks, since it is a multi-class segmentation task, the average of the scores (FB, FC, TB, and TC) were obtained and reported to evaluate the overall performance of the model in segmentation task.

III. RESULTS

A. PERFORMANCE OF MULTI-TASK MODELS

Table 1 shows the comparison results between the two multi-task neural network models, OA_MTL and RES_MTL on segmentation and classification tasks. OA_MTL displays a superior overall performance compared to RES_MTL, in both classification and segmentation tasks. Without utilizing the whole ResNet-18 as the encoder, OA_MTL surpassed REC,_MTL by 6%, 16%, 2% in terms of ACC, PRE and F1-Score for classification task and by approximately 1% in terms of PRE, REC and DSC, and 2% in terms of JSC for segmentation task. The segmentation results in DSC, reported by classes are presented in Fig. 4. It is observed that both models display similar segmentation performance in all classes, with a lower performance in segmenting knee cartilages than that of segmenting knee bones.

B. COMPARISON WITH SINGLE-TASK CNN MODELS

We further verify the effectiveness of multi-task models compared to single-tasks models, which includes the baseline models and existing state of art single-task models that perform single segmentation or classification task independently. Here, baseline models are defined as model that take the shared encoder and a specified branch (segmentation or classification) of the multi-task model based on the specific task. To get the individual baseline results, the segmentation and classification only models are trained separately. We refer classification baseline model for OA_MTL and RES_MTL as OA_MTL_C and RES_MTL_C respectively, while OA_MTL_S and RES_MTL_S for their respective

TABLE 1. Comparison between OA_MTL and RES_MTL.

	Classification				Segmentation			
	ACC	PRE	REC	F1-Score	PRE	REC	DSC	JSC
OA_MTL	0.825	0.905	0.792	0.844	0.911	0.921	0.915	0.851
RES_MTL	0.775	0.778	0.875	0.824	0.902	0.909	0.904	0.833

TABLE 2. Comparison between multi-task models, baseline models and existing CNN models.

Models	Classification				Segmentation			
	ACC	PRE	REC	F1-Score	PRE	REC	DSC	JSC
OA_MTL	0.825	0.905	0.792	0.844	0.911	0.921	0.915	0.851
RES_MTL	0.775	0.778	0.875	0.824	0.902	0.909	0.904	0.833
OA_MTL_C	0.600	0.600	1.000	0.750				
RES_MTL_C	0.550	0.583	0.875	0.700				
DenseNet121	0.775	0.800	0.833	0.816				
ResNet50	0.700	0.700	0.875	0.778				
ResNeXT50	0.725	0.810	0.708	0.756				
ShuffleNet	0.575	0.630	0.708	0.667				
OA_MTL_S					0.913	0.922	0.917	0.853
RES_MTL_S					0.907	0.925	0.915	0.850
UNet					0.925	0.918	0.921	0.860
VNet					0.875	0.938	0.903	0.832
VoxResNet					0.928	0.514	0.621	0.484

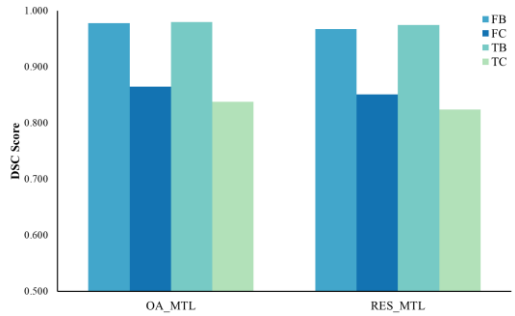

FIGURE 4. Detailed segmentation results by classes for OA_MTL and RES_MTL.

segmentation baseline models. For other existing models, the architectures are consistent with their respective papers with adjustment of the final output layer and use the same parameter as implementation details in Methodology section. We summarize the comparison of all the models in terms of segmentation and classification performances in Table 2.

1) COMPARISON WITH BASELINE MODELS

The effectiveness of multi-task model is demonstrated by OA_MTL and RES_MTL, where the classification accuracy improved by 38% and 41% as compared to OA_MTL_C and RES_MTL_C. Although there is no improvement in terms of segmentation performance when compared to OA_MTL_S and RES_MTL_S, the slight drop in segmentation tasks is very small and not significant. The multi-task models showed improvement in classification accuracy, precision and F1-score while maintaining the comparable segmentation results, when compared to the baseline models. Although OA_MTL_C achieved recall of 1, it is misleading to say that its performance is better because from other metrics, it shows that the model is making a lot of false positive predictions. Hence, metrics like F1-score play an important role in determining the model's overall performance by considering the harmonic mean of both precision and recall. Results demonstrated that OA_MTL and RES_MTL is able to segment knee structures accurately while classifying the knee conditions more effectively than the baseline models (as shown in Fig. 5), suggesting that the segmentation information may contribute additional context about the knee images and are helpful for classification tasks.

2) COMPARISON WITH OTHER CLASSIFICATION CNNS

We included four other classification CNNs to compare the classification performance with our multi-task model. The 3D classification models involved are 3D DenseNet121 [20], 3D ResNet-50 [18], 3D ResNeXT-50 [21] and 3D ShuffleNet [22]. All the classification models involved are evaluated with their complete architecture without extra modifications, only where models that are previously proposed in 2D are modified into 3D by replacing 2D operations with their 3D counterparts. From Table 2, it is observed that OA_MTL outperforms all the other models with the highest classification accuracy score of 0.825 and F1-score of 0.844. Following closely are RES_MTL and 3D DenseNet121, both achieved a good accuracy of 0.775. The performance of 3D DenseNet121 suggests that dense blocks might be suitable for

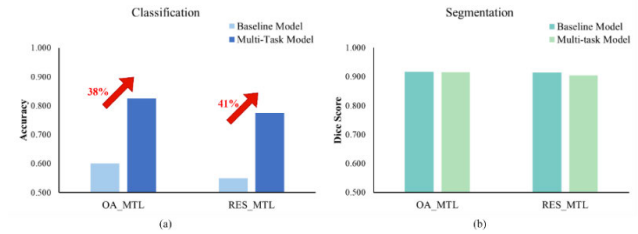

FIGURE 5. Performance change between baseline and multi-task models.

classification of knee osteoarthritis as well. 3D ResNeXT-50 and 3D ResNet-50, both utilizing residual skip connections displayed a moderate accuracy of 0.725 and 0.70 respectively, with 3D ResNet-50 achieved a high recall of 0.875. Meanwhile, ShuffleNet, recognized for its computational efficiency, appears to be the least effective in classification task, with an accuracy of 0.575. Hence, these findings highlight the importance of integrating different techniques in model design to improve classification performance.

3) COMPARISON WITH OTHER SEGMENTATION CNNs

The segmentation performance of the hybrid models are further compared with the most commonly used 3D volumetric architectures, UNet [23], VNet [24] and VoxResNet [25]. The performance of OA_MTL, RES_MTL, 3D UNet and 3D VNet are fairly similar to each other, whereas 3D VoxResNet achieves the lowest performance in terms of recall, DSC, and JSC. However, it is important to note the imbalanced score in precision and recall metrics of 3D VoxResNet, where the model tends to misclassify the pixels associated with knee structures as background, instead of any other knee structures of interest. In other words, in this work, 3D VoxResNet is found to be not well-suited for the knee MRI segmentation task, reflected by the lowest overall DSC and JSC. 3D UNet achieved the best DSC and JSC of 0.921 and 0.860 respectively whereas 3D VNet obtained the best recall score of 0.938. As shown in Fig. 6, 3D UNet and 3D VNet showed a similar trend of lower segmentation performance in the cartilages, which might be due to the imbalance volume of each class in one knee MRI.

C. COMPUTATIONAL COMPLEXITY ANALYSIS

We reported the total number of parameters and size in memory used of the models in Table 3. Fig. 7 displays the comparison between all the models compared in this paper in terms of segmentation performance, classification performance and computational complexity.

Compared to RES_MTL, OA_MTL achieved 1% and 6% better DSC in segmentation and accuracy in classification performance while requiring 2x less parameters and memory. From the two multi-task models proposed in this paper, OA_MTL is indeed more efficient than RES_MTL, by achieving better performances in both tasks while requiring a smaller number of parameters and memory, indicating OA_MTL is indeed excellent in performing 2-class

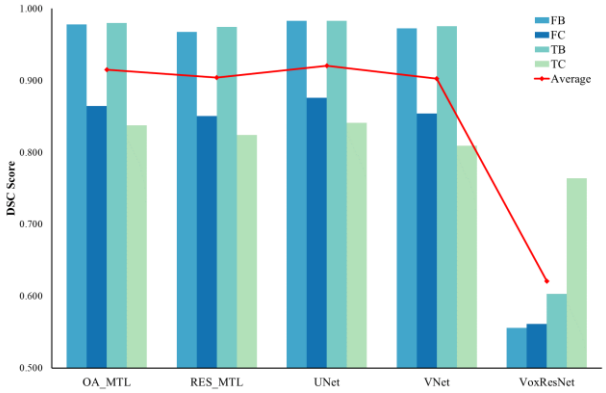

FIGURE 6. Detailed segmentation results by classes for each model.

TABLE 3. Number of parameters and memory used for all models.

	Model	Parameter (M)	Memory (MB)
Multi-Task Models	OA_MTL	16.074	64.296
	RES_MTL	34.986	139.946
Baseline Models	OA_MTL_C	14.379	57.516
	RES_MTL_C	33.200	132.919
	OA_MTL_S	16.004	64.017
	RES_MTL_S	34.917	139.666
Classification Models	DenseNet121	11.245	44.979
	ResNet50	46.159	184.637
	ResNeXT50	25.826	103.303
	ShuffleNet	0.945	3.782
Segmentation Models	UNet	19.074	76.427
	VNet	45.611	182.443
	VoxResNet	1.981	7.926

classification of OA diagnosis. Hence, we focus more on OA_MTL in the following results discussions.

From Table 3, it is observed that the parameter and memory required of the baseline models are smaller than that of multi-task models, especially for the classification model. This might be due to the complexity of the decoder structure required in the segmentation branch. However, by combining two tasks in one model, the number of parameters and memory used is definitely lesser than having two independent models for separate tasks. Not only multi-task models can use the number of parameters and memory more efficiently, but the time also used to train one model for two tasks is much more efficient than training two independent models, as the computational cost and resources required will be approximately double that of a multi-task model.

It is worth noting that OA_MTL requires less computational cost when compared to other existing segmentation and classification models. OA_MTL shows better efficiency than ResNet50 and ResNeXT50 by achieving better classification results with lower computational costs. OA_MTL requires only 16.074 M parameters where the model size is smaller than 3D ResNet-50 and 3D ResNeXt-50 by 65%

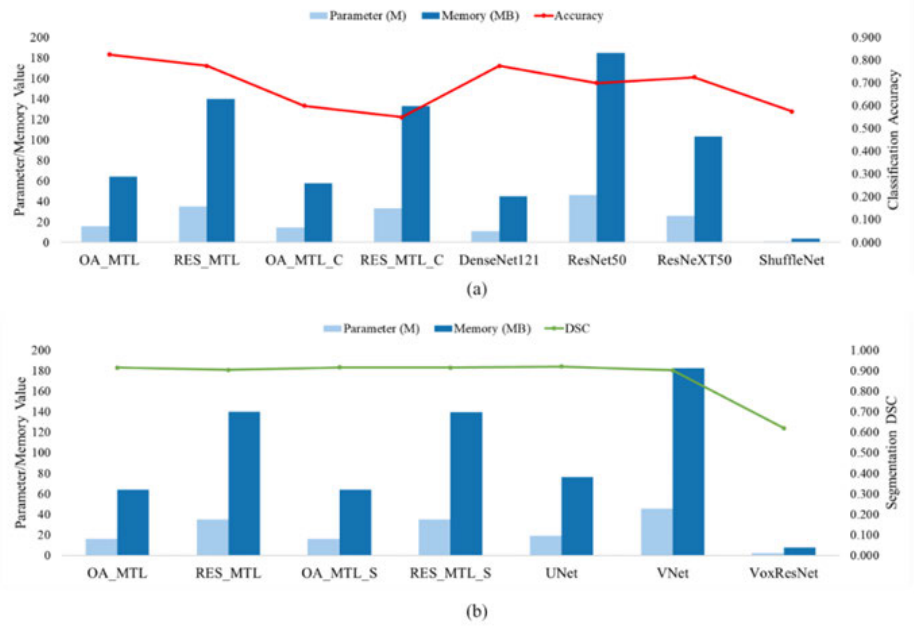

FIGURE 7. Comparison of models in terms of performance and computational complexity.

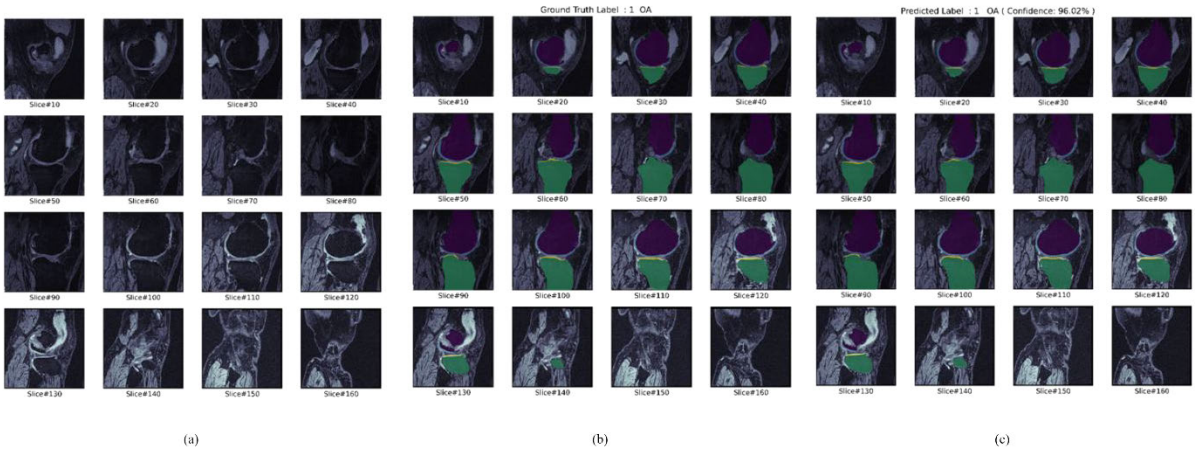

FIGURE 8. Example output of OA_MTL in 2D slices: (a) original image, (b) ground truth label and masks, (c) predicted label and masks.

and 38% respectively while surpassing their classification accuracy by 18% and 14% respectively. The results suggest that smaller models might be sufficient to achieve good classification performance. Besides, for models smaller than OA_MTL, 3D DenseNet121 and 3D ShuffleNet, their performance is worse than OA_MTL. Hence, it is important to achieve a good tradeoff between model size and performance. Moreover, it is notable that OA_MTL delivers a comparable segmentation performance as that of 3D V-Net and 3D U-Net while learning 2× and 1× fewer parameters. This makes our model preferable in medical imaging tasks [26]. Although VoxResNet requires the minimal computational cost, the segmentation results are very poor in terms of recall, DSC and JSC.

IV. DISCUSSION

In this paper, we established the effectiveness of multi-task approach in addressing the knee osteoarthritis diagnosis by using a 3D MRI as an input. We combined the segmentation of multiple knee structures and the classification of OA incidence as a multi-task learning strategy. The performance of OA_MTL is dominant in this study, where it jointly learns the segmentation and classification tasks, and leverages the information from segmentation tasks to improve the classification performance. The difference between the two architectures proposed in this work is that RES_MTL takes the entire ResNet-18 feature extractor architecture as the encoder and is larger in model size when compared to OA_MTL. OA_MTL is more lightweight and has

shown better ability in performing osteoarthritis diagnosis. OA_MTL is indeed efficient as it considers both parameter and computational efficiency, by utilizing residual blocks to enhance performance and depthwise separable convolutional blocks to improve computational efficiency.

Since the multi-task models can realize segmentation and classification outputs, the effectiveness of the multi-task models is also validated through comparison with single-task baseline models and existing CNN models that perform the tasks separately. From the findings above, it suggested that the segmentation of knee structures and osteoarthritis detection are highly related tasks, hence multi-task approach can make use of the inherent relationship between the two tasks to enhance the performance of either of the tasks, in this case, the classification task. OA_MTL has outperformed the existing classification models by 6% to 43% in classification accuracy.

One of the key challenges in implementing a 3D CNN model is the high computational complexity and expensive computational cost that causes impracticality in real-world applications. Although the practical aspects such as computational complexity and model size are important, it is usually neglected in previous works on artificial intelligence models in the healthcare domain [26]. We introduce an efficient lightweight 3D CNN model by incorporating a multi-task strategy where the encoder of the U-shaped architecture is shared and implemented depthwise separable convolutions in the decoder. These strategies result in a lightweight model that requires a lower number of parameter and memory, especially when compared to other single-task models as presented in the Results section above. OA_MTL strikes a balance between model performance and computational complexity where it achieves competitive segmentation performance (as shown in Fig. 8), and improved classification accuracy with a lower computational complexity. The U-shaped architecture in our work is adapted from UNet and our multi-task models are further developed with a mix-and-match strategy, where the architecture is modified based on our specific goals. This approach draws inspiration from the fact that several top- performing models in the field adopt similar architectures while incorporating different design choices [13], [27].

Using the computational setup described in the Methodology section, OA_MTL consists of 16.074 million parameters, requiring 64.296 MB of memory. OA_MTL required 22 hours and 9 minutes to train on a 2-class classification diagnosis. The inference time on one sample of knee MRI scan by OA_MTL takes around 3 seconds only, which is shorter than other reported studies [13]. Besides, the training and inference time of our model includes performing two tasks simultaneously, which is more efficient than the sum of resources needed for performing the tasks by two separate models independently. These make OA_MTL desirable in practical applications.

Currently, there are limited 3D deep learning applications in MRI for knee osteoarthritis diagnosis [1], [28], [29]. We contribute to deep learning in the OA field by considering several aspects in the development of the model such as practicality in terms of computational complexity and memory requirement, as well as on model performance. Overall, OA_MTL serves as a practical tool by combining two tasks together. Not only can the performance of the model be enhanced, but a more comprehensive output can also be provided to the clinicians for a more accurate diagnosis. However, this study presents some limitations that should be addressed in future works.

One limitation we observed is that there is a certain bias in the performance of segmentation tasks, particularly towards the femur bone and tibia bone, due to the nature of knee MRI scan. The imbalanced data of knee structures within one MRI scan might have led to bias in our study. One useful technique to address this limitation is to adjust the loss function by using weighted loss functions, which can be further investigated in future works. By enabling the model to pay more attention to the minority classes, which are the cartilages in this case, the model can extract more meaningful OA features from the cartilages, which might contribute to a more accurate diagnosis. Next, we evaluate the practicality of the model by comparing the model size in terms of parameters and memory used only. While these metrics provide meaningful insights, it is unfair to solely compare the models just from these aspects. To assess the overall computational efficiency, other aspects such as Floating-point operations per second utilization, inference time per scan and the training time of the models should be considered to provide better understanding on the model's complexity and efficiency. Moreover, this study does not employ k-fold cross-validation for model training. Although it is normally used in classical machine learning practices, it is not efficient to be used in deep learning research due to the high computing cost [28]. This is because deep learning models, particularly 3D CNNs that analyze 3D MRI input, are very large, resulting in longer training durations. Cross-validation involves repeated training and evaluating a model, it is computationally expensive compared to a straightforward train-validation-test split, making it impractical for this work. However, this can be further utilized in future work to obtain a more comprehensive and reliable estimate of the performance of our model.

V. CONCLUSION

In this work, we investigated the effectiveness of multi-task models by exploring two different architectures, focusing on the application of knee osteoarthritis diagnosis in the context of 3D CNNs. We introduce an efficient multi-task model, OA_MTL which simultaneously performs the segmentation of knee structures and classification of knee osteoarthritis incidence, using 3D MRI scans. The experimental results present that through multi-task learning, OA_MTL achieved a significant improvement of 38% in classification accuracy, reaching an accuracy of 0.825, while maintaining good segmentation DSC score of 0.915, which is

on par with other segmentation models. Besides, OA_MTL is computationally efficient, where the computational complexity is lower than that of employing two independent models for segmentation and classification tasks separately. OA_MTL demonstrates the potential of multi-task learning in improving knee osteoarthritis diagnosis while considering computational efficiency. In conclusion, a 3D efficient multi-task model, OA_MTL is proposed to perform knee osteoarthritis diagnosis by providing two types of output pertinent to the knee osteoarthritis condition using a single-stage, end-to-end model.

ACKNOWLEDGMENT

Dataset used in the preparation of this article were retrieved from the publicly available Osteoarthritis Initiative (OAI).

REFERENCES

[1] P. S. Q. Yeoh, K. W. Lai, S. L. Goh, K. Hasikin, Y. C. Hum, Y. K. Tee, and S. Dhanalakshmi, "Emergence of deep learning in knee osteoarthritis diagnosis," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–20, Nov. 2021, doi: 10.1155/2021/4931437.

[2] M. Li, X. Li, Y. Jiang, J. Zhang, H. Luo, and S. Yin, "Explainable multi-instance and multi-task learning for COVID-19 diagnosis and lesion segmentation in CT images," *Knowledge-Based Syst.*, vol. 252, Sep. 2022, Art. no. 109278, doi: 10.1016/j.knosys.2022.109278.

[3] Y. Zhou, H. Chen, Y. Li, Q. Liu, X. Xu, S. Wang, P.-T. Yap, and D. Shen, "Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101918, doi: 10.1016/j.media.2020.101918.

[4] C. W. Yong, K. W. Lai, B. P. Murphy, and Y. C. Hum, "Comparative study of encoder–decoder-based convolutional neural networks in cartilage delineation from knee magnetic resonance images," *Current Med. Imag. Formerly Current Med. Imag. Rev.*, vol. 17, no. 8, pp. 981–987, Aug. 2021, doi: 10.2174/1573405616666201214122409.

[5] H. Yuan, Y. Wu, J. Cheng, Z. Fan, and Z. Zeng, "Pulmonary nodule detection using 3-D residual U-Net oriented context-guided attention and multi-branch classification network," *IEEE Access*, vol. 10, pp. 82–98, 2022, doi: 10.1109/ACCESS.2021.3137317.

[6] Z. Yang, L. Leng, M. Li, and J. Chu, "A computer-aid multi-task lightweight network for macroscopic feces diagnosis," *Multimedia Tools Appl.*, vol. 81, no. 11, pp. 15671–15686, May 2022, doi: 10.1007/s11042-022-12565-0.

[7] H. Ryu, S. Y. Shin, J. Y. Lee, K. M. Lee, H.-J. Kang, and J. Yi, "Joint segmentation and classification of hepatic lesions in ultrasound images using deep learning," *Eur. Radiol.*, vol. 31, no. 11, pp. 8733–8742, Nov. 2021, doi: 10.1007/s00330-021-07850-9.

[8] K. Tan, W. Huang, X. Liu, J. Hu, and S. Dong, "A multi-modal fusion framework based on multi-task correlation learning for cancer prognosis prediction," *Artif. Intell. Med.*, vol. 126, Apr. 2022, Art. no. 102260, doi: 10.1016/j.artmed.2022.102260.

[9] S. Irshad, D. P. S. Gomes, and S. T. Kim, "Improved abdominal multi-organ segmentation via 3D boundary-constrained deep neural networks," *IEEE Access*, vol. 11, pp. 35097–35110, 2023, doi: 10.1109/ACCESS.2023.3264582.

[10] R. Kijowski, J. Fritz, and C. M. Deniz, "Deep learning applications in osteoarthritis imaging," *Skeletal Radiol.*, vol. 52, no. 11, pp. 2225–2238, Nov. 2023, doi: 10.1007/s00256-023-04296-6.

[11] C. W. Yong, K. Teo, B. P. Murphy, Y. C. Hum, Y. K. Tee, K. Xia, and K. W. Lai, "Knee osteoarthritis severity classification with ordinal regression module," *Multimedia Tools Appl.*, vol. 81, no. 29, pp. 41497–41509, Dec. 2022, doi: 10.1007/s11042-021-10557-0.

[12] Y. X. Teoh, K. W. Lai, J. Usman, S. L. Goh, H. Mohafez, K. Hasikin, P. Qian, Y. Jiang, Y. Zhang, and S. Dhanalakshmi, "Discovering knee osteoarthritis imaging features for diagnosis and prognosis: Review of manual imaging grading and machine learning approaches," *J. Healthcare Eng.*, vol. 2022, pp. 1–19, Feb. 2022, doi: 10.1155/2022/4138666.

[13] M. H. A. Latif and I. Faye, "Automated tibiofemoral joint segmentation based on deeply supervised 2D-3D ensemble U-net: Data from the osteoarthritis initiative," *Artif. Intell. Med.*, vol. 122, Dec. 2021, Art. no. 102213, doi: 10.1016/j.artmed.2021.102213.

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

[15] F. Ambellan, A. Tack, M. Ehlke, and S. Zachow, "Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative," *Med. Image Anal.*, vol. 52, pp. 109–118, Feb. 2019, doi: 10.1016/j.media.2018.11.009.

[16] P. S. Q. Yeoh, K. W. Lai, S. L. Goh, K. Hasikin, X. Wu, and P. Li, "Transfer learning-assisted 3D deep learning models for knee osteoarthritis detection: Data from the osteoarthritis initiative," *Frontiers Bioengineering Biotechnol.*, vol. 11, Apr. 2023, Art. no. 1164655, doi: 10.3389/fbioe.2023.1164655.

[17] B. Norman, V. Pedoia, A. Noworolski, T. M. Link, and S. Majumdar, "Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs," *J. Digit. Imag.*, vol. 32, no. 3, pp. 471–477, Jun. 2019.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[19] A. Paszke, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Vancouver, BC, Canada, vol. 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alche-Buc, E. Fox, R. Garnett, Eds. San Francisco, CA, USA: NIPS, 2019, pp. 8024–8035. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

[20] 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)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[21] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New York, NY, USA, Jul. 2017, pp. 5987–5995, doi: 10.1109/CVPR.2017.634.

[22] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2018, pp. 6848–6856, doi: 10.1109/CVPR.2018.00716.

[23] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016*, vol. 9901, S. Ourselin, L. Joskowicz, M. Sabuncu, G. Unal, and W. Wells, Eds. Cham, Switzerland: Springer, 2016, pp. 424–432.

[24] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, New York, NY, USA, Oct. 2016, pp. 565–571, doi: 10.1109/3DV.2016.79.

[25] A.-M. Rickmann, A. Guha Roy, I. Sarasua, and C. Wachinger, "Recalibrating 3D ConvNets with project & excite," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2461–2471, Jul. 2020, doi: 10.1109/TMI.2020.2972059.

[26] Y. S. Jeon, K. Yoshino, S. Hagiwara, A. Watanabe, S. T. Quek, H. Yoshioka, and M. Feng, "Interpretable and lightweight 3-D deep learning model for automated ACL diagnosis," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2388–2397, Jul. 2021, doi: 10.1109/JBHI.2021.3081355.

[27] Y. Qiblawey, A. Tahir, M. E. H. Chowdhury, A. Khandakar, S. Kiranyaz, T. Rahman, N. Ibtehaz, S. Mahmud, S. A. Maadeed, F. Musharavati, and M. A. Ayari, "Detection and severity classification of COVID-19 in CT images using deep learning," *Diagnostics*, vol. 11, no. 5, p. 893, May 2021.

[28] V. Pedoia, B. Norman, S. N. Mehany, M. D. Bucknor, T. M. Link, and S. Majumdar, "3D convolutional neural networks for detection and severity staging of meniscus and PFJ cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects," *J. Magn. Reson. Imag.*, vol. 49, no. 2, pp. 400–410, Feb. 2019, doi: 10.1002/jmri.26246.

[29] A. Morales Martinez, F. Caliva, I. Flament, F. Liu, J. Lee, P. Cao, R. Shah, S. Majumdar, and V. Pedoia, "Learning osteoarthritis imaging biomarkers from bone surface spherical encoding," *Magn. Reson. Med.*, vol. 84, no. 4, pp. 2190–2203, Oct. 2020, doi: 10.1002/mrm.28251.

PAULINE SHAN QING YEOH received the B.Eng. degree (Hons.) from Universiti Malaya, Malaysia, in 2020, where she is currently pursuing the Ph.D. degree. She is also a visiting Ph.D. student with Xuzhou Medical University, China. Her research interests include artificial intelligence, image processing, medical imaging, and healthcare analytics.

SIEW LI GOH received the Ph.D. degree from the University of Nottingham for her work in investigating the efficacy of exercise in OA using network meta-analysis. She is currently a Clinician and a Medical Lecturer of sports medicine with the University of Malaya (UM). Her research interests include overuse injuries of the lower limb, exploring multidisciplinary work and evidence-based medicine.

KHAIRUNNISA HASIKIN received the B.Eng. and M.Eng.Sc. degrees in biomedical engineering from Universiti Malaya, Malaysia, and the Ph.D. degree from Universiti Sains Malaysia. She is currently a Senior Lecturer with the Department of Biomedical Engineering, Faculty of Engineering, Universiti Malaya. She is expanding her expertise in AI for sustainability management projects in intelligent environmental decision support system. Her research interests include artificial intelligence (AI), machine learning, and image processing in various fields, especially in medical and healthcare applications. She is also a Professional Engineer (P.Eng) of electrical engineering discipline registered with the Board of Engineer Malaysia (BEM).

XIANG WU (Member, IEEE) received the B.Eng. degree in information engineering and the M.S. and Ph.D. degrees in communication and information system from the China University of Mining and Technology, Xuzhou, China, in 2007, 2010, and 2014, respectively. He is currently the Deputy Dean of the School of Medical Information and Engineering and the Director of the Institute of Medical Information Security, Xuzhou Medical University, China. He is also a Visiting Professor and Doctoral Supervisor with Universiti Malaya. His research interests include privacy protection and information security.

KHIN WEE LAI (Senior Member, IEEE) received the B.Eng. degree (Hons.) from Universiti Teknologi Malaysia, Malaysia, and the Ph.D. degree in biomedical engineering from Technische Universitat Ilmenau, Germany, and Universiti Teknologi Malaysia, through the DAAD Ph.D. Sandwich Programme. He is the Programme Head of the M.E. degree (biomedical) with the Faculty of Engineering, Universiti Malaya. His research interests include computer vision, machine learning, medical image processing, and healthcare analytics. He is a registered Professional Engineer with Practicing Certificate (PEPC) at the Board of Engineers Malaysia (BEM), a fellow of the Engineers Australia (FIEAust), an APEC Engineer IntPE, Australia, and a Chartered Professional Engineer (CPEng.) at NER, Australia. He is a fellow of the Institute of Engineers Malaysia (IEM), a member of the Institution of Engineering and Technology (IET), and a U.K. Chartered Engineer (C.Eng.). He is currently an Associate Editor of IEEE Access.