Current Medical Imaging, 2021, *17*, 981-987

RESEARCH ARTICLE

Comparative Study of Encoder-decoder-based Convolutional Neural Networks in Cartilage Delineation from Knee Magnetic Resonance Images

Ching Wai Yong[1], Khin Wee Lai[1,*], Belinda Pingguan Murphy[1] and Yan Chai Hum[1]

[1]*Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, 1900 Kampar, Perak, Kuala Lumpur, Malaysia*

ARTICLE HISTORY

Received: May 29, 2020
Revised: September 23, 2020
Accepted: October 14, 2020

DOI:
10.2174/1573405616666201214122409

This is an Open Access article published under CC BY 4.0
https://creativecommons.org/licenses/by /4.0/legalcode

Abstract: ***Background*:** Osteoarthritis (OA) is a common degenerative joint inflammation that may lead to disability. Although OA is not lethal, this disease will remarkably affect patient's mobility and their daily lives. Detecting OA at an early stage allows for early intervention and may slow down disease progression.

***Introduction*:** Magnetic resonance imaging is a useful technique to visualize soft tissues within the knee joint. Cartilage delineation in magnetic resonance (MR) images helps in understanding the disease progressions. Convolutional neural networks (CNNs) have shown promising results in computer vision tasks, and various encoder-decoder-based segmentation neural networks are introduced in the last few years. However, the performances of such networks are unknown in the context of cartilage delineation.

***Methods*:** This study trained and compared 10 encoder-decoder-based CNNs in performing cartilage delineation from knee MR images. The knee MR images are obtained from the Osteoarthritis Initiative (OAI). The benchmarking process is to compare various CNNs based on physical specifications and segmentation performances.

***Results*:** LadderNet has the least trainable parameters with the model size of 5 MB. UNetVanilla crowned the best performances by having 0.8369, 0.9108, and 0.9097 on JSC, DSC, and MCC.

***Conclusion*:** UNetVanilla can be served as a benchmark for cartilage delineation in knee MR images, while LadderNet served as an alternative if there are hardware limitations during production.

Keywords: Comparative study, convolutional neural network, encoder-decoder neural network, knee cartilage segmentation, magnetic resonance imaging, osteoarthritis.

1. INTRODUCTION

Osteoarthritis (OA) is a common degenerative joint inflammation that may lead to disability in severe cases [1]. Although OA is not lethal, it profoundly affects mobility [2] and the patient's quality of life. The incessant breaking down of cartilage and continuous bone deformation are the main causes that lead to joints failure. Patients of severe OA (end-stage) will experience excruciating pain as the joint cartilages degenerate and cause bone-to-bone frictions during movements. Arthroplasty or total knee replacement is the last option available for knee OA patients to regain their mobility. However, this clinical procedure is invasive and costly. Therefore, diagnosing OA at an early stage is crucial for clinical intervention in halting disease progression and mitigating disability in later stages.

Magnetic Resonance Imaging (MRI) is the safest and non-radioactive imaging technique to visualize the knee joint's internal derangement, especially in determining OA features of the asymptomatic uninjured knee [3]. The main advantage of MRI as compared with traditional radiography is its capability to evaluate the structural changes during disease progressions [4] and provide biomarkers for early OA diagnosis [5]. Degeneration of cartilage tissues is one of the main criteria for an early stage of OA as defined by Luyten *et al*. [6]. Thus, delineating cartilage in biomedical images is crucial because early detection of cartilage defects allows for early medical interventions and leads to better treatments [7-9].

In clinical practices, cartilage delineation is manually performed by a radiologist [2]. Manual delineation is not only a time-consuming [2, 10] task but is also prone to inter- and intra-observer variability [11, 12]. In recent years, deep convolutional neural networks (CNNs) demonstrated state-of-the-art performance in biomedical image analysis, such

*Address correspondence to this author at Department of Biomedical Imaging, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia; Tel: +603-7967 4580; Email: lai.khinwee@um.edu.my

 © 2021 Bentham Science Publishers

Current Medical Imaging

as breast cancer analysis [13], bone disease prediction [14], and age assessment [15]. Unlike most conventional machine learning techniques such as fuzzy logic [16], bi-histogram equalization [17], and image registrations [18], CNN requires no feature engineering but demands a handful of dataset annotation and computation power. Fortunately, the computation requirement is no longer a challenge to CNN training with the advancement of graphic cards and cloud computing.

Encoder-decoder pair is the main core component in most of the existing segmentation neural networks. The encoder harvests data into features, whereas the decoder decodes the features to perform pixel-based classification; the encoder is discriminative, whereas the decoder is generative. These encoder-decoder-based CNNs (EDCNNs) reported remarkable achievements in natural scene images. The current study aims to examine the performances of various EDCNNs in delineating cartilage tissues from MR images.

The contributions of this study are listed as follows. First, we propose to group the various EDCNNs into different variations or families. To the best of our knowledge, our study is pioneering as it provides a genealogical chart of EDCNNs. Second, we perform a benchmarking process and identify the best EDCNN in delineating knee cartilage tissue within MR images. This paper is organized as follows. Section 2 briefly explains about U-Net, the base version of ECDNN, and its variations. We grouped different ECDNNs into families according to their unique characteristics and natures. Section 3 illustrates the methodology, which includes the datasets, data pre-processing techniques applied, specifications of model training, and model assessment strategy in Section 3. Section 4 evaluates the performance of EDCNNs by reporting the comparative results. Section 5 summarizes the conclusion and future works.

2. BASE AND VARIATIONS OF EDCNNS

The general architecture of an EDCNN has two paths: a contracting path for context capturing and an expanding path to localize features precisely. U-Net [19] is the first neural network to employ the encoder-decoder pairing scheme into the network design for the segmentation task, making it the first EDCNN. This architecture was inspired by the Fully Convolutional Network [20], a CNN that can perform pixel-wise classification. The encoding path of U-Net is built with repeating blocks containing 3x3 convolutional layers, a rectified linear unit (ReLU) [21], and a 2x2 max-pooling layers with the stride of 2. For each successive block, the feature map resolutions are reduced by half, whereas the feature channels are doubled. By contrast, the decoding path of U-Net contains Up-convolution blocks to up-sample the feature maps while reducing the feature channels by half. Feature maps from the encoding path are concatenated to a decoding path after each respective down- and up-sampling process. These unique and symmetric paths yield a u-shaped architecture.

2.1. Variations of EDCNNs

In this study, we refer to U-Net as the "Base" for EDCNNs while grouping its expansions into four different variations: "Skip-Connections," "Weight-Initialized," "Auxiliary-Based," and "Cascaded" as shown in (Fig. **1**).

Base. The original U-Net has a huge drawback: the output resolution from the final layer is not the same as the input image. The feature maps were cropped from each level of the contracting path as the border pixels were lost during each convolution. To overcome this problem, we padded the feature maps in each of the convolution layers, ensuring that the output dimension is equivalent to the input size to produce a network known as UNetVanilla.

Skip-Connections. The degree of connectivity within a neural network determines the information flow from one layer to another. DenseNet [22] exploits the effects of shortcut connections by directly connecting all layers with one another and performing iterative concatenation of feature maps. The improvement in connectivity helps this network converge faster. Although DenseNet was created for the classification task, a segmentation version, namely, FC-DenseNet [23], was carefully extended. FC-DenseNet inherits the following advantages of DenseNet: parameter efficiency, implicit deep supervision, and feature reuse. FC-DenseNet mitigates a large number of parameters by only up-sampling the feature maps created at the previous dense blocks. FC-DenseNet has three variations: FC-DenseNet56, FC-DenseNet67, and FC-DenseNet103 with 56, 67, and 103 layers, respectively. Unlike FC-DenseNet, LinkNet [24] provides a different type of linkage between encoder and decoder, and the input of the encoder layer is bypassed to the corresponding decoder's output. This approach aims to recover the lost spatial information that can be utilized by the decoder and its up-sample operations. Moreover, the decoder uses fewer parameters as the decoders share knowledge learnt by the encoder at every layer.

Weight-Initialized. Neural networks are normally trained from scratch, and their weights are initialized randomly. A wrong initialization will lead to exploding or vanishing weights and gradients. Studies showed that deep neural networks could converge much earlier and prevent the aforementioned scenarios with a proper initialization strategy [25, 26]. These strategies initialize weights according to a specific distribution with a formulated pair of mean and standard deviation. Apart from the manual initialization, we can replace the encoder path with sequential convolution and ReLU layers from a pre-trained CNN. For example, TernausNet [27] and AlbuNet [28] are using pre-trained VGG [29] and ResNet-34 [30] as encoder in the contracting path.

Auxiliary-Based. Apart from introducing a new weight initialization strategy and skip-connections scheme, existing studies explore the potential of equipping EDCNNs with auxiliary elements such as Attention Gates (AGs) and recurrent residual modules. AG is commonly applied in image captioning [31], machine translation [32, 33], and classification tasks [34, 35]. With the help of self-attention

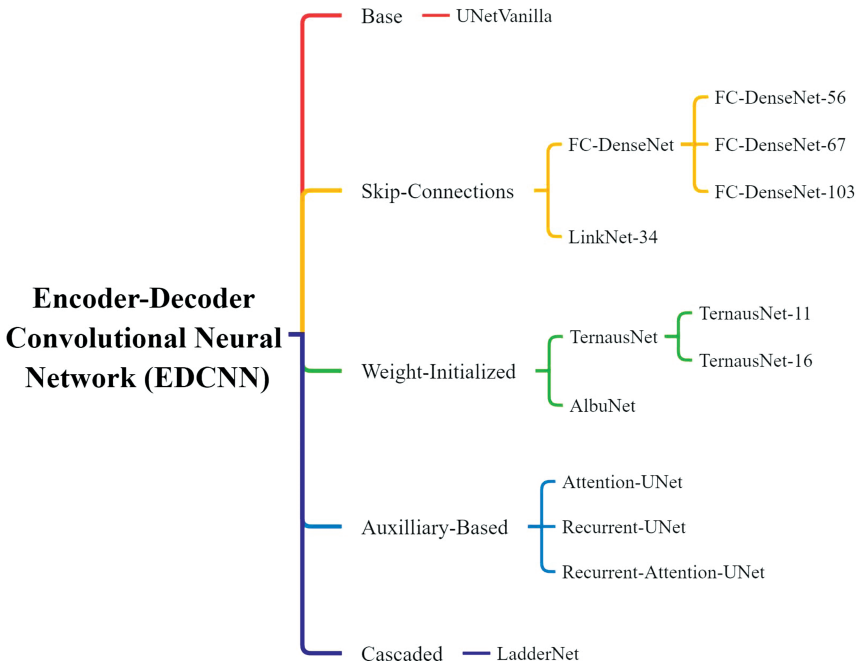

Fig. (1). Genealogical chart of EDCNNs. Various EDCNNs were grouped according to its distinctive functionality.

gating modules, AttentionUNet [36] shows that a network learns to focus on salient image regions and suppresses feature activation in irrelevant regions without introducing a substantial computational overhead. By contrast, RecurrentUNet [37] is using the recurrent residual module to accumulate features at different time-steps. This process allows the production of a relatively strong representation of features by extracting essential low-level features. RecurrentAttentionUNet [38] is introduced by combining both AGs and recurrent residual module into U-Net. This network takes advantages of three different cores: using U-Net to capture information at multiple scales while integrating low- and high--level features; stacking residual blocks to allow a network to go deeper; implementing attention modules to change the attention-aware features adaptively.

Cascaded. Conventional EDCNNs come with a single pair of encoder-decoder until the birth of LadderNet [39], an ensemble structure of multiple U-Nets. LadderNet concatenates encoder-decoder pairs, introducing additional paths for information flow and improving the capability of an EDCNN to capture complex features. A weight-sharing strategy was applied to the residual blocks to constrain the increase in trainable parameters due to the chaining of encoder-decoder pairs.

3. EXPERIMENTAL

3.1. Comparative Study

All the EDCNNs were trained using the Osteoarthritis Initiative (OAI) datasets, a longitudinal study of knee OA. This dataset contains 4,796 participants, with X-rays and Magnetic Resonance (MR) images of participants' knees. Although the size of this dataset is enormous, we arbitrarily chose 100 sets of Double Echo Steady State (DESS) MR images and subsequently annotated both femoral and tibial knee cartilages. Twenty sets of MR images were held out as a control set, while the remaining were partitioned into training and validation (ratio of 3:1). Isolating the control set will prevent the control set from exposure to the model during the training and validation process. The goal of a control set is to validate the models without any bias. The training and validation sets respectively contain 570 and 190 images, while the control set has 189 images.

Unlike natural scene images, medical images are usually stored as Digital Imaging and Communications in Medicine (DICOM). As OAI datasets are saved as DICOM files, extraction and format conversion are necessary. We performed MR slice extraction and format conversion through python scripting. The image dimensions of the MR slices were maintained at 384 height (pixels) and 384 width (pixels).

EDCNNs are prototyped using adaptive moment estimation, batch size 1 for 30 epochs, initial learning rate at 1e-3, and weight decay at 1e-4. The learning rate is controlled by a learning rate scheduler along the model training process. The scheduler reduces the learning rate by 0.1 if no improvement is seen on the validation loss for two consecutive epochs. We also utilized the early stopping algorithm to prevent a model from overfitting. We seized the model training process if the validation loss remained stagnant for the past two consecutive epochs, while the learning rate has been reduced to the lower bound at 1e-10. We conducted model training by using PyTorch.

In this study, the model's prediction output image was compared with manual annotations pixel by pixel. Through the pixel-wise comparison, a confusion matrix, as seen in

Table **1** was produced. With the four basic elements (*i.e.*, TP, FP, FN, and TN), different metrics can be used to analyze the model's performance. We evaluated the trained EDCNNs with the isolated control set with three different metrics.

Table 1. Confusion matrix and its four basic elements.

-	-	Manual Annotations	
-	Pixel's Class	Cartilage	Background
Model's Output	Cartilage	True Positive	False Positive
	Background	False Negative	True Negative

Jaccard Similarity Coefficient (JSC): JSC is used to gauge the similarity and diversity between two finite sample sets. It measures by dividing the size of the intersection with the size of the union of the sample sets. The formula is shown as equation 1:

$$JSC = \frac{TP}{TP + FP + FN} \quad (1)$$

Dice Similarity Coefficient (DSC): DSC is a harmonic mean between precision and recall. This value is in the range of [0, 1]. DSC is different from JSC, which only counts true positives once; however, both JSC and DSC do not take the true negatives into account. The formula is shown as equation 2:

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (2)$$

Matthew's Correlation Coefficient (MCC): MCC is commonly used in the field of machine learning as a measure to assess a binary classification task. Unlike JSC and DSC, MCC summarizes the confusion matrix elements into a value. MCC returns a value in the range [-1, 1], with perfect prediction labelled as 1; -1 indicates a completely incorrect prediction, while 0 represents that the prediction is no better than random. The formula is shown as equation 3:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (3)$$

Table **2** reports the JSC, DSC, and MCC values of each of the EDCNNs.

4. RESULTS AND DISCUSSION

The comparative study of EDCNNs can be split into two parts: based on the model's physical specifications and model's performances. The first part investigates the model size and the total trainable parameters for each EDCNN, while the second part focuses on the performance metrics.

The size of a model is generally proportional to its number of trainable parameters, *i.e.*, the more the trainable parameters, the bigger the model size. According to Table **2**, FCDenseNet-56 and LadderNet are the smaller models with approximately 5 megabytes (MB) and approximately 1.3 million trainable parameters. As mentioned in Section 2.1, the weight-sharing strategy in LadderNet reduces the total trainable parameters, although multiple encoder-decoder blocks are concatenated. By contrast, auxiliary-based EDCNNs (*i.e.*, AttentionUNet, RecurrentUNet, and RecurrentAttentionUNet) have the largest size with at least 34 million trainable parameters. However, a smaller model size is likely to improve the efficiency of model serving but does not necessarily generate better performances in terms of a model's accuracy and precision.

Table 2. Comparing the physical specifications of EDCNNs. The model sizes are represented in MB. The smallest model size is in bold numbers.

Variant & Architecture	Model Size (MB)	Trainable Parameters
UNetVanilla	118.0	31,045,441
FCDenseNet-56	5.39	1,374,865
FCDenseNet-67	13.40	3,460,353
FCDenseNet-103	36.00	9,319,521
LinkNet-34	83.20	21,794,721
TernausNet-11	87.40	22,927,393
TernausNet-16	111.00	29,306,465
AlbuNet	134.00	35,117,897
AttentionUNet	133.00	34,878,573
RecurrentUNet	149.00	39,091,393
RecurrentAttentionUNet	150.00	39,442,925
LadderNet	**5.28**	1,381,821

Following JSC, DSC, and MCC, UNetVanilla slightly outperformed FCDenseNet-56 and LadderNet. However, they come with a disadvantage because the former has 22 times more trainable parameters than the latter. With additional trainable parameters, the training process for UNetVanilla will be longer than FCDenseNet-56 and LadderNet. From Table **3**, UNetVanilla crowns all the performance metrics, although it is only a baseline model. The reasons are as follows.

First, we limited each EDCNN to 30 training epochs as stated in Section 3. In each of the epoch, each EDCNN iterates through all images within the training dataset and proceeds to validation at the end of the epoch. The model state with the lowest validation loss is then retrieved. However, the EDCNNs might not be at its optimum stage as we only limited the training to 30 epochs.

The second possible reason is the difference in the loss function. We implemented Binary Cross Entropy (BCE) with Logit loss as compared with Sorensen-Dice [36, 38] or custom-weighted loss function [19, 24, 27, 28]. BCE with Logit loss is numerically stable with log-sum-exp function. This feature might explain why EDCNNs could not surpass the performances of the baseline architecture.

The third reason is the inconsistency of the decoder block. Several methods can increase the size of feature map in the decoding path. Examples are interpolation, up-sam-

pling, and transpose convolution. Different approaches were chosen for the EDCNNs on the basis of their original works.

Overfitting is another potential cause for a model not to perform, especially models involving recurrent modules. The recurrent layer is well known for its high possibility of overfitting. Tables **2** and **3** do not report the results of RecurrentUNet and RecurrentAttentionUNet due to overfitting. Apart from the early stopping algorithm, we must implement strong mechanisms to reduce the chances of overfitting.

Moreover, the masking of all images is manually annotated, which is subject to a degree of errors due to intra- and inter-observer variability. Unlike natural scene images, each pixel from the MR images is not color-coded. Thus, segmenting the boundary of tissues is challenging, and the classification of a pixel near the tissue boundary is vague. Meanwhile, we considered the manual annotations as near "Ground Truth" level, accepting that minor mistakes may exist across the manual masking.

In general, all the EDCNNs reported high scoring in all performance metrics. The lowest and highest performance scores across EDCNNs range within 0.77-0.83 for JSC, 0.86-0.91 for DSC, and 0.87-0.90 for MCC. As seen in (Fig. **2**), all EDCNNs successfully predicted the cartilage regions. The slight imperfections are the FP and FN pixels at the tip of the cartilage as well as at the boundaries. By referring to the confusion matrix element images, FCDenseNets tend to have higher False Positive (red) pixels, while TernausNet-16 has the highest number of False Negative (green) pixels.

Table 3. Comparing the performances of EDCNNs in terms of JSC, DSC, and MCC onto the 20 sets of testing images. The scores are tabulated as mean and standard deviation. Results of RecurrentUNet and RecurrentAttentionUNet are excluded due to overfitting and did not result in any high confident results. The highest scores are in bold numbers.

Variant & Architecture	Jaccard-Similarity Coefficient	Dice Similarity Coefficient	Matthew's Correlation Coefficient
UNetVanilla	**0.8369 ± 0.0285**	**0.9108 ± 0.0172**	**0.9097 ± 0.0174**
FCDenseNet-56	0.8124 ± 0.0362	0.8956 ± 0.0225	0.8946 ± 0.0226
FCDenseNet-67	0.8017 ± 0.0323	0.8895 ± 0.0200	0.8898 ± 0.0193
FCDenseNet-103	0.7706 ± 0.0417	0.8696 ± 0.0269	0.8719 ± 0.0246
LinkNet-34	0.8305 ± 0.0389	0.9067 ± 0.0243	0.9057 ± 0.0243
TernausNet-11	0.8310 ± 0.0298	0.9072 ± 0.0181	0.9062 ± 0.0182
TernausNet-16	0.7873 ± 0.0430	0.8801 ± 0.0275	0.8796 ± 0.0272
AlbuNet	0.8357 ± 0.0308	0.9101 ± 0.0187	0.9090 ± 0.0188
AttentionUNet	0.8241 ± 0.0315	0.9028 ± 0.0195	0.9024 ± 0.0193
RecurrentUNet	-	-	-
RecurrentAttentionUNet	-	-	-
LadderNet	0.8253 ± 0.0373	0.9037 ± 0.0228	0.9025 ± 0.0232

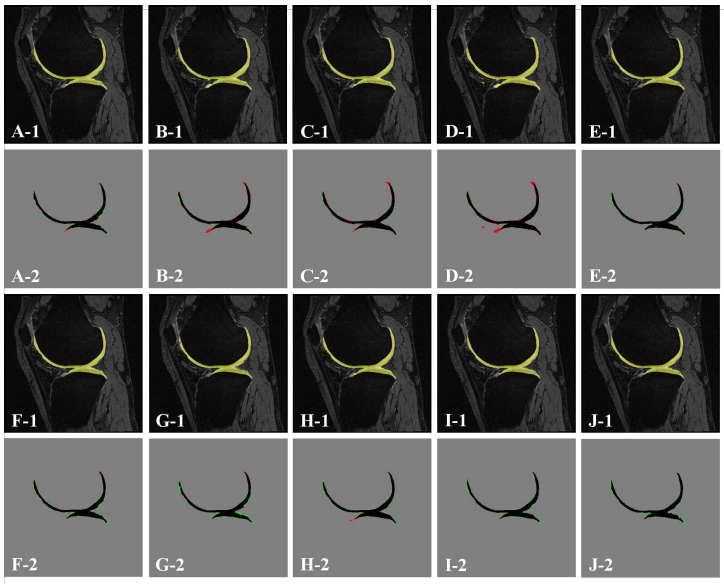

Fig. (2). Segmentation results from EDCNNs and comparison against manual annotations on a single MR image. Image with a label ending with "1" indicates the overlay results of EDCNNs onto MR images, while that label ending with "2" shows the results on the basis of the elements of confusion matrix: True Positive (black), False Positive (red), False Negative (green), and True Negative (gray). A to J are labels for UNetVanilla, FCDenseNet-56, FCDensetNet-67, FCDenseNet-103, LinkNet-34, TernausNet-11, TernausNet-16, AlbuNet, AttentionUNet, and LadderNet. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

CONCLUSION

This study provided a genealogical chart of EDCNNs by grouping architectures according to their characteristics. It then performed a benchmarking process onto 10 EDCNNs to identify the best architectures in segmenting cartilage tissue in MR images. In this comparison study, we compared EDCNNs from two perspectives: the model's physical specifications and its segmentation performances. On the one hand, LadderNet has the least trainable parameters, and the model size is only 5 MB. On the other hand, UNetVanilla crowned the best performances by having 0.8369, 0.9108, and 0.9097 on JSC, DSC, and MCC, respectively. Therefore, LadderNet is found to be the lightweight architecture, while UNetVanilla is the best performing architecture. The outcome of this study can serve as a guideline, reference, or even a comparison standard in the task of delineating knee cartilage tissue in MR images for OA analysis. We wish to expand this study in the future by including other variations and designs of EDCNNs and performing further in-depth comparative analysis.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study has been approved by the institutional review board of University Malaya, Malaysia (Approval Number: MECID.NO: 20165-2419).

HUMAN AND ANIMAL RIGHTS

No animals were used in this study. All the human procedures were is in accordance with the ethical standards of the committee responsible for human experimentation (institutional and national), and with the Helsinki Declaration of 1975, as revised in 2013 (http://ethics.iit.edu/ecodes/node/3931).

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Data and/or research tools used in the preparation of this manuscript were obtained and analyzed from the controlled access datasets distributed from the Osteoarthritis Initiative (OAI). OAI is a collaborative informatics system created by the National Institute of Mental Health and the National Institute of Arthritis, Musculoskeletal and Skin Diseases (NIAMS) to provide a worldwide resource to quicken the pace of biomarker identification, scientific investigation, and OA drug development.

FUNDING

This work was supported by the Fundamental Research Grant Scheme (FRGS), Ministry of Education, Malaysia. (Grant no. FRGS/1/2019/TK04/UM/01/2).

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Robin Poole A. Osteoarthritis as a whole joint disease. HSS J 2012; 8(1): 4-6.
[2] Pang J, Li P, Qiu M, Chen W. Automatic articular cartilage segmentation based on pattern recognition from knee MRI images. J Digit Imaging 2015; 28(6): 695-703.
[3] Culvenor AG, Øiestad BE, Hart HF, Stefanik JJ, Guermazi A. Prevalence of knee osteoarthritis features on magnetic resonance imaging in asymptomatic uninjured adults: a systematic review and meta-analysis. Br J Sports Med 2019; 53(20): 1268-78.
[4] Guermazi A, Roemer FW, Haugen IK, Crema MD, Hayashi D. MRI-based semiquantitative scoring of joint pathology in osteoarthritis. Nat Rev Rheumatol 2013; 9(4): 236-51. http://dx.doi.org/10.1038/nrrheum.2012.223
[5] Nagai K, Nakamura T, Fu FH. The diagnosis of early osteoarthritis of the knee using magnetic resonance imaging. Ann Joint 2018; 3: 110.
[6] Luyten FP, Denti M, Filardo G, Kon E, Engebretsen L. Definition and classification of early osteoarthritis of the knee. Knee Surg Sports Traumatol Arthrosc 2012; 20(3): 401-6. PMID: 22068268
[7] Xu J, Xie G, Di Y, Bai M, Zhao X. Value of T2-mapping and DWI in the diagnosis of early knee cartilage injury. J Radiol Case Rep 2011; 5(2): 13-8. PMID: 22470777
[8] Faisal A, Ng S-C, Goh S-L, Lai KWJM. Knee cartilage segmentation and thickness computation from ultrasound images. Med Biol Eng Comput 2018; 56(4): 657-9. PMID: 28849317
[9] Hossain MB, Pingguan-Murphy B, Chai HY, *et al.* Improved ultrasound imaging for knee osteoarthritis detection. medical imaging technology. Springer 2015; pp. 1-40.
[10] Lee S, Park SH, Shim H, Yun ID, Lee SUK. Optimization of local shape and appearance probabilities for segmentation of knee cartilage in 3-D MR images. Comput Vis Image Underst 2011; 115(12): 1710-20.
[11] Folkesson J, Dam EB, Olsen OF, Pettersen PC. Segmenting articular cartilage automatically using a voxel classification approach. IEEE Trans Med Imaging 2007; 26(1): 106-5.
[12] Li K, Millington S, Wu X, Chen DZ, Sonka M, Eds. Simultaneous segmentation of multiple closed surfaces using optimal graph searching. Inf Process Med Imaging. 19: 406-17. http://dx.doi.org/10.1007/11505730_34 PMID: 17354713
[13] Rakhlin A, Shvets A, Iglovikov V, Kalinin AA, Eds. Deep convolutional neural networks for breast cancer histology image analysis. Image Analysis and Recognition ICIAR 2018 Lecture Notes in Computer Science, vol 10882. Campilho A, Karray F, ter Haar Romeny B, Eds. http://dx.doi.org/10.1007/978-3-319-93000-8_83
[14] Tiulpin A, Thevenot J, Rahtu E, Lehenkari P. Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach. Sci Rep 2018; 8(1): 1727.
[15] Iglovikov VI, Rakhlin A, Kalinin AA, Shvets AA. Paediatric bone age assessment using deep convolutional neural networks.Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer 2018; pp. 300-8. http://dx.doi.org/10.1007/978-3-030-00889-5_34
[16] Salih AAM, Hasikin K, Isa ANAM. Adaptive fuzzy exposure local contrast enhancement. IEEE Access 2018; 6: 58794-806.
[17] Hum YC, Lai KW, Mohamad Salim MI. Multiobjectives bihistogram equalization for image contrast enhancement. Complexity 2014; 20(2): 22-36. http://dx.doi.org/10.1002/cplx.21499
[18] Ramli R, Idris MYI, Hasikin K, *et al.* Feature-based retinal image registration using D-saddle feature. J Healthc Eng 2017; 2017: 1489524. http://dx.doi.org/10.1155/2017/1489524 PMID: 29204257
[19] Ronneberger O, Fischer P, Brox T, Eds. U-net: Convolutional net-

works for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 Lecture Notes in Computer Science, vol 9351; Springer, Cham. http://dx.doi.org/10.1007/978-3-319-24574-4_28

[20] Long J, Shelhamer E, Darrell T, Eds. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. Boston, MA, USA.

[21] Nair V, Hinton GE, Eds. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel 2010.

[22] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, Eds. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu, HI, USA.

[23] Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y, Eds. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition workshops. Honolulu, HI, USA. http://dx.doi.org/10.1109/CVPRW.2017.156

[24] Chaurasia A, Culurciello E, Eds. Linknet: Exploiting encoder representations for efficient semantic segmentation. 2017 IEEE Visual Communications and Image Processing (VCIP) 2017. St. Petersburg, FL, USA.

[25] Glorot X, Bengio Y, Eds. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the thirteenth international conference on artificial intelligence and statistics PMLR. 9: 249-56.

[26] He K, Zhang X, Ren S, Sun J, Eds. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Proceedings of the IEEE international conference on computer vision 2015. Santiago, Chile. http://dx.doi.org/10.1109/ICCV.2015.123

[27] Iglovikov V. Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. arXiv:180105746 [csCV] 2018.

[28] Shvets AA, Iglovikov VI, Rakhlin A, Kalinin AA. Angiodysplasia detection and localization using deep convolutional neural networks. 2018 17th IEEE international conference on machine learning and applications (icmla). Orlando, FL, USA.

[29] Simonyan K. Very deep convolutional networks for large-scale image recognition. 2014. arXiv:1409.1556 [cs.CV].

[30] He K, Zhang X, Ren S, Sun J, Eds. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, NV, USA.

[31] Anderson P, He X, Buehler C, Teney D, Johnson M, Gould S, Eds. Bottom-up and top-down attention for image captioning and visual question answering. Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City, UT, USA. http://dx.doi.org/10.1109/CVPR.2018.00636

[32] Bahdanau D, Cho K. Neural machine translation by jointly learning to align and translate. 2014. arXiv:1409.0473 [cs.CL].

[33] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Eds. Attention is all you need. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA 2017.

[34] Jetley S, Lord NA, Lee N, Torr PH. Learn to pay attention. 2018. arXiv:1804.02391 [cs.CV].

[35] Veličković P, Cucurull G, Casanova A, Romero A, Lio P. Graph attention networks. 2017. arXiv:1710.10903 [stat.ML].

[36] Oktay O, Schlemper J, Folgoc LL, *et al.* Attention U-net: Learning where to look for the pancreas. 2018. arXiv:1804.03999 [cs.CV].

[37] Alom MZ, Hasan M, Yakopcic C, Taha TM. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. 2018. arXiv:1802.06955 [cs.CV].

[38] Jin Q, Meng Z, Sun C, Wei L. RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. 2018. arXiv:1811.01328 [cs.CV].

[39] Zhuang J. Laddernet: Multi-path networks based on u-net for medical image segmentation. 2018. arXiv:1810.07810 [cs.CV].