Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 21(1): 582–601.
DOI: 10.3934/mbe.2024025
Received: 23 October 2023
Revised: 02 December 2023
Accepted: 10 December 2023
Published: 15 December 2023

Research article

Diagnosis of musculoskeletal abnormalities based on improved lightweight network for multiple model fusion

Zhigao Zeng[1,2], Changjie Song[1,2], Qiang Liu[1,2,*], Shengqiu Yi[1,2] and Yanhui Zhu[1,2]

[1] School of Computer Science, Hunan University of Technology, Zhuzhou 412007, China

[2] Hunan Key Laboratory of Intelligent Information Perception and Processing Technology, Zhuzhou 412007, China

* **Correspondence:** Email: liuqiang@hut.edu.cn.

Abstract: This paper introduces a solution to address the intricacy of the model employed in the deep learning-based diagnosis of musculoskeletal abnormalities and the limitations observed in the performance of a single deep learning network model. The proposed approach involves the integration of an improved EfficientNet-B2 model with MobileNetV2, resulting in the creation of FusionNet. First, EfficientNet-B2 is combined with coordinate attention (CA) to obtain CA-EfficientNet-B2. Furthermore, aiming to minimize the model parameter count, we further enhanced the mobile inverted residual bottleneck convolution module (MBConv) employed for feature extraction in EfficientNet-B2, resulting in the development of CA-MBC-EfficientNet-B2. Next, the features extracted from CA-MBC-EfficientNet-B2 and MobileNetV2 are fused. Finally, the final diagnosis of musculoskeletal abnormalities was performed by using fully connected layers. The experimental results demonstrate that, first, compared to EfficientNet-B2, CA-MBC-EfficientNet-B2 not only significantly improves the diagnostic performance of musculoskeletal abnormalities, it also reduces the parameter count and storage space by 17%. Moreover, as compared to other models, FusionNet demonstrates remarkable performance in the area of anomaly diagnosis, particularly on the elbow dataset, achieving a precision of 92.93%, an AUC of 93.89% and an accuracy of 87.10%.

Keywords: diagnosis of musculoskeletal abnormalities; EfficientNet-B2; MobileNetV2; coordinate attention; mobile inverted residual bottleneck convolution

1. Introduction

Orthopedic imaging is one of the key tools for diagnosing and treating orthopedic diseases. Utilizing orthopedic imaging techniques, physicians can accurately assess abnormalities in the musculoskeletal system and develop optimal treatment plans. However, traditional methods for diagnosing muscu-

loskeletal abnormalities are influenced by subjective experience and professional expertise, leading to issues such as misdiagnosis and missed diagnosis. Moreover, accurate diagnosis of abnormalities is crucial for the subsequent treatment of musculoskeletal disorders. Therefore, it is necessary to study an accurate and efficient automated method for diagnosing musculoskeletal abnormalities.

Traditional diagnostic methods for musculoskeletal disorders are usually employed by manually extracting specific features. For instance, Zhang et al. [1] suggested a computer-aided image classification method for diagnosing finger joint osteoarthritis. The method can accurately identify the signs of osteoarthritis through optical image analysis and feature extraction, which provides a powerful auxiliary diagnostic tool for doctors. Al-Ayyoub et al. [2] introduced a system that utilizes machine learning to automatically detect fracture types in long bones based on X-ray images. Mahendran and Baboo [3] put forward a technique for the detection of long bone fractures, which involves the fusion of classification methods. Chai et al. [4] proposed a gray-level co-occurrence matrix method for evaluating the efficacy of femur long bone fractures. The extraction of features in the aforementioned algorithm primarily relies on the researchers' expertise, potentially resulting in a certain amount of information loss.

In comparison, deep learning, as an automated learning approach, possesses the ability to partially overcome the aforementioned limitations. One of its strengths lies in its ability to autonomously acquire features from data without manual intervention. Additionally, deep learning models are adaptable to diverse tasks and data distributions.

In the past couple of years, there has been a significant surge in interest in the application of deep learning models, specifically convolutional neural networks (CNNs), in the field of medical image analysis. This surge can be attributed to the rapid advancements observed in deep learning technology.

To date, there have been some studies on musculoskeletal X-ray images that use deep learning methods. For instance, Cohen et al. [5] used an artificial intelligence algorithm with deep learning to analyze and diagnose wrist X-rays; they compared the performance of radiologists in the detection of wrist fracture X-rays. Nam et al. [6] presented a novel approach that utilizes EfficientNet-B7 to diagnose nasal bone fractures automatically. The study can quickly and accurately classify and judge new X-ray images by learning a large number of X-ray images of nasal bone fractures and non-fractures. Oka et al. [7] suggested a novel network architecture utilizing the VGG-16 [8] to diagnose distal radius fractures. This method can automatically analyze and diagnose biplane X-ray images by learning fracture characteristics and patterns. In the work of He et al. [9], a methodology was presented that combines calibrated deep learning to detect abnormalities in radiographs of skeletal muscles. Singh et al. [10] suggested a novel CNN-based hybrid architecture, ComDNet-512, to effectively detect skeletal abnormalities in patient musculoskeletal X-rays. Yi et al. [11] used a pre-trained ResNet [12] for transfer learning to quickly and accurately classify children's skeletal X-ray images. Yao et al. [13] proposed a deep learning framework to enhance the diagnostic speed of orthopedic diseases based on X-rays. This framework incorporated a two-stage approach for bone classification and anomaly detection. Cheng et al. [14] applied the DenseNet-121 [15] model for the identification and localization of hip fractures in pelvic X-rays. Choi et al. [16] suggested a dual-input network model utilizing the ResNet architecture, which automatically detected supracondylar fractures in children on conventional X-rays showing both anterior and lateral elbow X-rays. Cheng et al. [17] put forward PelviXNet, a multi-scale deep learning algorithm designed specifically for detecting pelvic and hip fractures in plain X-ray radiographs. The study showed that PelviXNet showed comparable performance to radiologists

and orthopedic surgeons. Thian et al. [18] leveraged the Inception-ResNet [19] and Faster R-CNN [20] models to identify and localize fractures on X-rays of the wrist. The study proved that object detection CNNs had high sensitivity and specificity when detecting and locating fractures of the radius and ulna in wrist X-rays. Ye et al. [21] built a deep learning network model based on DenseNet-169 architecture to differentiate between acetabular fractures on anterior-posterior pelvic X-rays, and according to the experimental outcomes, the model demonstrated diagnostic performance that is on par with, or even superior to, that of the clinician. Proposed by Wang and Wang [22], a method utilizing U-net [23] was introduced for the detection of rib fractures. This approach leverages pixel-level rib fracture features to achieve the rapid and precise detection of rib fractures. Jin et al. [24] proposed a CNN model called FracNet for detecting and segmenting rib fractures. Ghosh et al. [25] put forward a CNN model that utilizes a patch-based image analysis method and transfer learning via ResNet to detect fractures of the ribs in frontal X-rays of children under the age of 2 years.

Although the CNN-based methods mentioned above possess the ability to automatically extract features and tackle complex challenges in musculoskeletal abnormality diagnosis, the high accuracy of CNNs often relies on complex network architectures [26], which means that a larger amount of computation and parameters are required [27]. Expanding the depth and width of a CNN is generally known to improve its performance; however, it also results in a proliferation of network parameters [28, 29]. Consequently, the network becomes excessively complex to deploy on edge devices. This is the main disadvantage of the aforementioned method based on a CNN. In addition, these anomaly diagnosis methods typically use a single network. Due to the limited performance of a single network, it may be difficult to capture feature information at different scales, which could potentially impact the diagnostic results.

To date, many scholars have conducted research on lightweight model architectures. For example, Chen et al. [30] proposed a lightweight garbage classification model, GCNet, which achieved an average accuracy of 97.9% on a self-built dataset with only 1.3M parameters. Versaci et al. [31] proposed an innovative fuzzy classification procedure based on fuzzy similarity calculation. This method can group similar images together in a fuzzy manner and extract representative images from each individual group, demonstrating the ability to reduce computational load. Chen et al. [32] proposed a progressive lightweight network, BrightsightNet, for enhancing low-light images. The model has only 2.6K parameters and was shown to achieve a single inference time of 0.052 seconds. Angelov and Gu [33] put forward an image classifier based on fuzzy rules of deep learning, which combines deep learning with fuzzy rules and greatly improves classification performance. Chen et al. [34] proposed an efficient railway track area segmentation network, ERTNet, based on an encoder-decoder architecture. The model achieves a balance between segmentation accuracy and computational efficiency. Feng et al. [35] proposed a lightweight and efficient railway area extraction model, LRseg, which provides technical support for foreign object detection on railways. The model has size and memory requirements of only 2.98 MB and 37.5 MB, respectively. Tan and Le [36] proposed a new network architecture called EfficientNet, which provides an effective and simple method for scaling CNNs, and they achieved better performance and higher efficiency. Sandler et al. [37] introduced a lightweight network architecture called MobileNetV2, which maintains high image classification accuracy and computational efficiency even on resource-constrained mobile devices. Chen et al. [38] proposed an improved single shot multibox detector (SSD) algorithm. The algorithm utilizes MobileNetV2 as the backbone feature extraction network for the SSD, enhancing the real-time performance of the algorithm. The above

literature provides us with very good new ideas for designing lightweight models.

Considering the lightweight structure and high performance of EfficientNet-B2 and MobileNetV2, we propose an improved EfficientNet-B2 network model. This model has become more lightweight and higher-performing than the baseline model. To achieve more accurate results in the detection of musculoskeletal abnormalities, a classification algorithm based on the improved EfficientNet-B2 and MobileNetV2 multi-model fusion is proposed. We have used the public dataset for musculoskeletal radiography (MURA) [39] to verify the above method.

To provide a summary of the contributions offered by this paper, they can be described as follows:

1). Our proposed CA-MBC-EfficientNet-B2 involves integrating the coordinate attention (CA) module, which effectively mitigates the loss of positional information during feature extraction and enhances the network's expressive capacity.

2). We propose an improved mobile inverted residual bottleneck convolution (MBConv) module that reduces the amount of EfficientNet-B2 parameters and improves the performance of EfficientNet-B2.

3). Our proposed CA-MBC-EfficientNet-B2 model has shown higher accuracy and better lightweight performance than the traditional EfficientNet-B2 model when applied for the diagnosis of musculoskeletal abnormalities, providing empirical evidence for the performance improvement of musculoskeletal disease diagnosis in the medical field.

4). Our proposed FusionNet combines two different network architectures to provide more comprehensive and accurate abnormal diagnosis results, opening up new possibilities for musculoskeletal disease diagnosis in the medical field.

The organization of the subsequent sections of this paper can be described as follows. Section 2 focuses on the dataset utilized in this study, along with a comprehensive explanation of the enhanced EfficientNet-B2 classification model. Additionally, it delves into the multi-model fusion classification algorithm, which is based on both the improved EfficientNet-B2 and MobileNetV2 architectures. Section 3 provides the experimental design in this study and the analysis of the experimental results. Section 4 encompasses the conclusions derived from the findings of this study.

2. Datasets and proposed methods

2.1. Datasets

MURA is a large skeletal X-ray dataset that is used to evaluate whether the corresponding parts of the X-ray images fed into the network model are abnormal or normal. It includes seven upper limb radiological study sites: shoulder, elbow, fingers, forearm, hand, humerus, and wrist. Using these seven datasets, we can train deep learning models to automatically identify abnormalities in musculoskeletal images, helping doctors to make more accurate and rapid diagnoses. The anomaly detection task on the MURA dataset involves a binary classification problem. An upper limb X-ray film serves as the input to the model, which produces a binary label $y \in \{0, 1\}$ as its output. This label indicates whether the corresponding body part is classified as normal or abnormal. Table 1 visualizes the distribution of the MURA dataset, presenting the breakdown of different categories within the dataset. In Figure 1, a collection of X-ray images from the MURA dataset is presented, revealing a range of normal and abnormal cases.

Table 1. The distribution of normal and abnormal data in the dataset.

Datasets	Training set		Testing set		Total
	Normal	Abnormal	Normal	Abnormal	
Wrist	5769	3987	364	295	10415
Elbow	2925	2006	235	230	5391
Shoulder	4211	4168	285	278	8942
Forearm	1164	661	150	151	2126
Finger	3138	1968	214	247	5567
Hand	4059	1484	271	189	6003
Humerus	673	599	148	140	1560

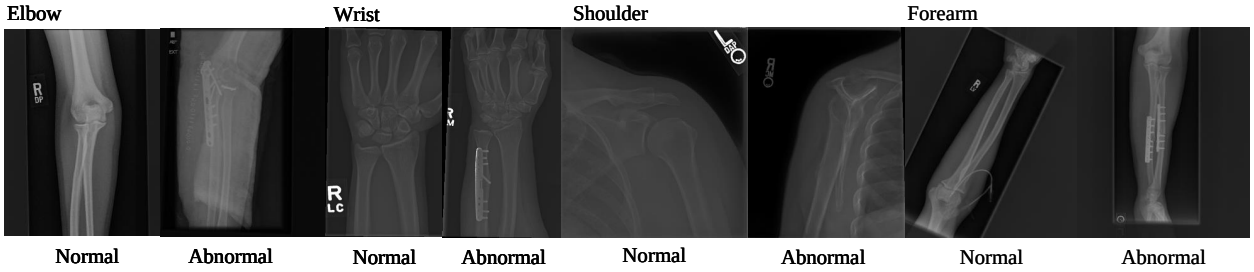


Figure 1. Examples of X-rays in the MURA dataset.

2.2. Improved EfficientNet-B2 classification model

Introduced in 2019 by the Google Brain team, EfficientNet is a CNN architecture. It leverages composite scaling coefficients to optimize the network's depth, width, and resolution simultaneously. By employing automatic model scaling, EfficientNet can adaptively adjust to datasets of varying sizes, leading to impressive performance across different computer vision tasks and datasets. As a result, EfficientNet has gained immense popularity and is widely deployed in diverse computer vision applications. The EfficientNet architecture comprises a series of eight models, denoted as B0-B7. As the depth and width of the models increase, their complexity also escalates, demanding more advanced experimental equipment. In this study, given the constraints of the laboratory equipment, EfficientNet-B2 was chosen as the baseline model. EfficientNet-B2 employs the MBConv module for feature selection. Figure 2 displays the MBConv module.

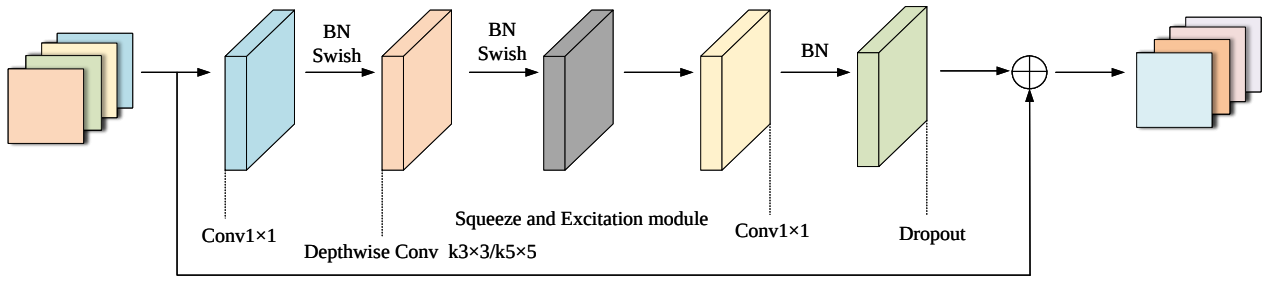


Figure 2. MBConv module.

The inclusion of an attention mechanism [40] in the network model typically results in a focus on essential feature points while disregarding less significant ones. The channel attention mechanism within the MBConv module, known as the squeeze and excitation (SE) module [41], contributes to enhancing network performance by dynamically adjusting the channel weights of the feature map. Nonetheless, the SE module operates on a global scale within each channel, overlooking spatial information interaction and failing to fully exploit its underlying information potential. Figure 3 shows the structure of the SE module.

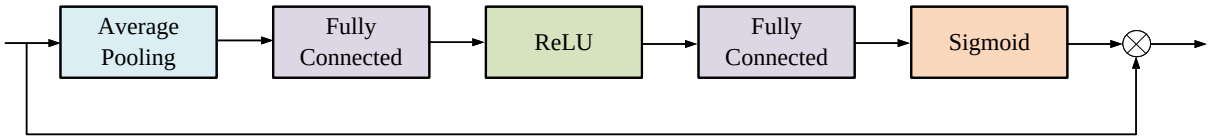


Figure 3. SE module.

The integration of channel and spatial attention mechanisms in the CA module [42] is achieved by incorporating positional information into the channel attention. This strategy effectively prevents the loss of positioning information during two-dimensional global pooling. Hence, it is integrated with EfficientNet-B2 to better allocate feature weights and enhance model performance. The structure of the CA module is depicted in Figure 4.

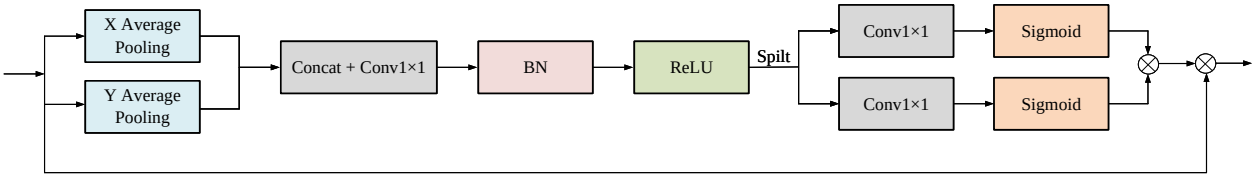


Figure 4. CA module.

The inclusion of two fully connected layers in the SE module for the purpose of computing the channel attention weights results in an expansion of the parameter count for the MBConv module. EfficientNet-B2 is a network built by stacking MBConv modules, which increases the complexity of

EfficientNet-B2. The efficient channel attention (ECA) module [43] does not contain the fully connected layers, and it calculates the channel attention weight through a learnable one-dimensional convolution (1D convolution) operation to achieve efficient calculation. By applying a 1D convolution operation on the channel dimension, the ECA module can not only capture local channel interdependencies, it can also involve only a few parameters. The modeling of this local relationship can result in the extraction of important channel features more effectively. The ECA module is shown in Figure 5. As compared to the SE module, the ECA module has a better lightweight network design. As a means of improving model performance and simultaneously reducing model parameters, the SE module within the MBConv module is substituted with the ECA module. Figure 6 shows the improved MBConv module. The structure of the CA-MBC-EfficientNet-B2 model is shown in Figure 7.

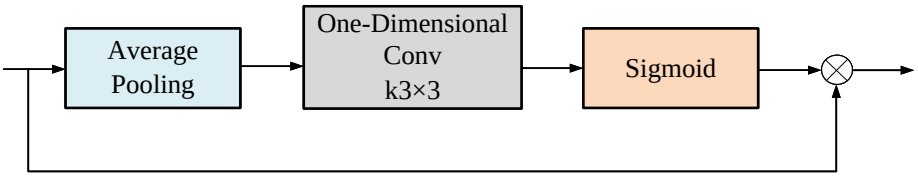


Figure 5. ECA module.

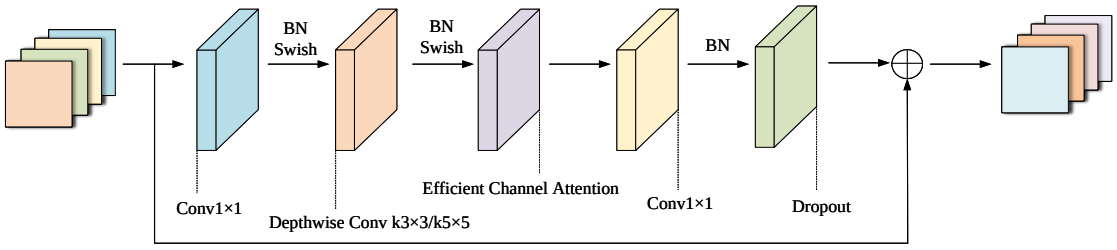


Figure 6. Improved MBConv module.

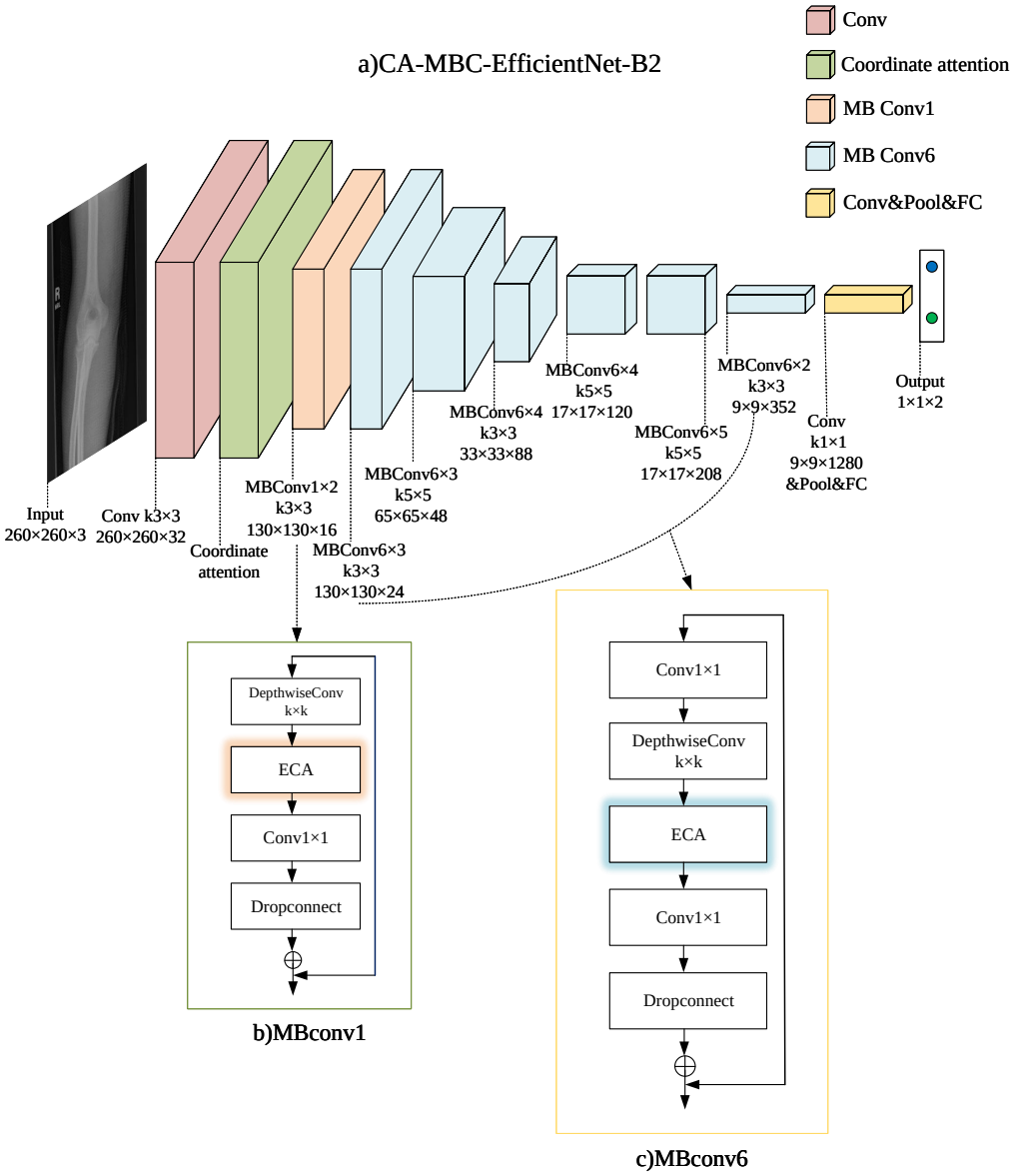


Figure 7. CA-MBC-EfficientNet-B2 model structure.

2.3. Classification algorithm based on multi-model fusion of CA-MBC-EfficientNet-B2 and MobileNetV2

The diagnosis of musculoskeletal diseases in current deep learning techniques primarily relies on the utilization of a single network. Compared with traditional disease diagnosis techniques, this method offers progress in areas of diagnostic accuracy and diagnostic speed. However, due to the limitation of the performance of a single network, the diagnosis result may be affected to some extent. To address these issues, we suggest the implementation of a classification algorithm that utilizes the multi-model fusion of CA-MBC-EfficientNet-B2 and MobileNetV2. MobileNetV2 and EfficientNet-B2 have similar features, and both of them are lightweight networks. MobileNetV2 is a lightweight CNN proposed

by Google in 2018. One of the main ideas of MobileNetV2 is to replace traditional convolutional layers with deep separable convolution [44] to decrease computational effort and model size. Meanwhile, MobileNetV2 introduces the linear bottleneck and inverse residual structure as part of its architecture enhancements, with the primary objective of enhancing network performance. The linear bottleneck is employed in MobileNetV2 for feature selection, and Figure 8 illustrates its linear bottleneck structure.

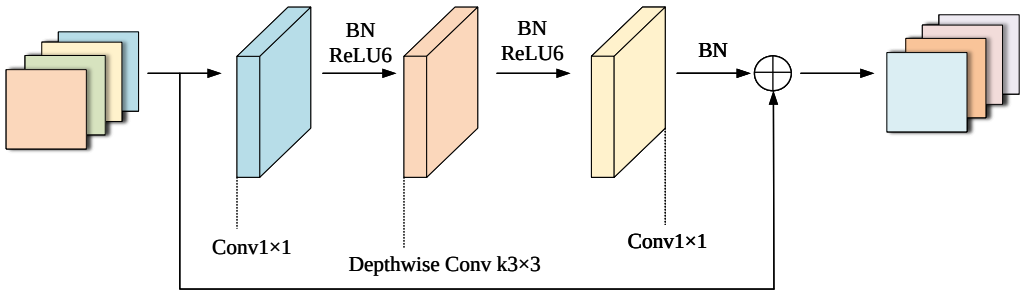


Figure 8. Bottleneck module.

Algorithm 1 describes the process of image classification via a multi-model fusion approach based on CA-MBC-EfficientNet-B2 and MobileNetV2. First, the training dataset is fed into FusionNet for 150 epochs, continuously optimizing it based on the cross-entropy loss function. Then, the instance images are classified by using the optimized model. The structure of FusionNet is shown in Figure 9. The specific form of the cross-entropy loss function is as follows:

$$L(y, \hat{y}) = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{2.1}$$

where $L(y, \hat{y})$ represents the loss function, y denotes the true label, $\hat{y}$ represents the predicted value by the model, N represents the number of classes, y_i refers to the i-th element of the true label, and $\hat{y}_i$ indicates the i-th element of the predicted value.

The computational complexity of Algorithm 1 primarily depends on the model structure and the size of the input images, which is typically measured by using the metric of floating-point operations. The detailed parameters of the CA-MBC-EfficientNet-B2 structure are shown in Table 2, and the detailed parameters of the MobileNetV2 structure are shown in Table 3. Here, "Input" represents the size of the input feature map, "Operator" represents the operation performed on the input feature map, "Channels" represents the number of output feature map channels and "Layers" represents the number of times each operation is performed. When the input image size is 260×260, based on the model structure parameters, the computed floating-point operations for CA-MBC-EfficientNet-B2 are 2.04 GFLOPs (floating point of operations), for MobileNetV2 are 0.94 GFLOPs, and for FusionNet are 2.98 GFLOPs.

Algorithm 1: Model fusion algorithm based on CA-MBC-EfficientNet-B2 and MobileNetV2

Input: Training dataset, X Predicting samples, Y

Output: Classification results, R

begin

Initialize CA-MBC-EfficientNet-B2 network and MobileNetV2 network.

/* $f(x)_{\texttt{CA\_MBC}} \overset{Improved\ MBConv}{\longleftarrow} f(x)_{\texttt{CA}} \overset{CA}{\longleftarrow} f(x)_1 \leftarrow$ `EfficientNet-B2` */

/* $f(x)_2 \leftarrow$ `MobileNetV2` */

Fuse $f(x)_{\text{CA\_MBC}}$ and $f(x)_2$ and initialize.

/* $f(x)_{\texttt{fusion}} \leftarrow f(x)_{\texttt{CA\_MBC}} + f(x)_2$ */

while *epoch* < 150 **do**

Substitute X into $f(x)_{\text{fusion}}$.

Optimize the model $f(x)_{\text{fusion}}$ through the cross-entropy loss function:

/* `Cross-Entropy loss function:` $H(\texttt{Label}, f(x)) = -\sum_x \texttt{Label} \times \log(f(x))$ */

$f(x)_{\text{fusion}} \leftarrow f(x)_{\text{fusion}}$

end while

Save model $f(x)_{\text{fusion}}$.

Substitute Y into model $f(x)_{\text{fusion}}$ to get the model classification results R.

return R.

end

Table 2. CA-MBC-EfficientNet-B2 model structure. Conv: convolution, FC: fully connected.

Input	Operator	Channels	Layers
$260 \times 260 \times 3$	Conv 3×3	32	1
$130 \times 130 \times 32$	CA module	32	1
$130 \times 130 \times 32$	MBConv1, k3 $\times$3	16	2
$130 \times 130 \times 16$	MBConv6, k3 $\times$3	24	3
$65 \times 65 \times 24$	MBConv6, k5 $\times$5	48	3
$33 \times 33 \times 48$	MBConv6, k3 $\times$3	88	4
$17 \times 17 \times 88$	MBConv6, k5 $\times$5	120	4
$17 \times 17 \times 120$	MBConv6, k5 $\times$5	208	5
$9 \times 9 \times 208$	MBConv6, k3 $\times$3	352	2
$9 \times 9 \times 352$	Conv 1 $\times$1 & Pooling & FC	1280	1

Table 3. MobileNetV2 model structure.

Input	Operator	Channels	Layers
$260 \times 260 \times 3$	Conv 3×3	32	1
$130 \times 130 \times 32$	Bottleneck	16	1
$130 \times 130 \times 16$	Bottleneck	24	2
$65 \times 65 \times 24$	Bottleneck	32	3
$33 \times 33 \times 32$	Bottleneck	64	4
$17 \times 17 \times 64$	Bottleneck	96	3
$17 \times 17 \times 96$	Bottleneck	160	3
$9 \times 9 \times 160$	Bottleneck	320	1
$9 \times 9 \times 320$	Conv 1×1	1280	1
$9 \times 9 \times 1280$	Avgpool 7×7	-	1
$1 \times 1 \times 1280$	Conv 1×1	32	-

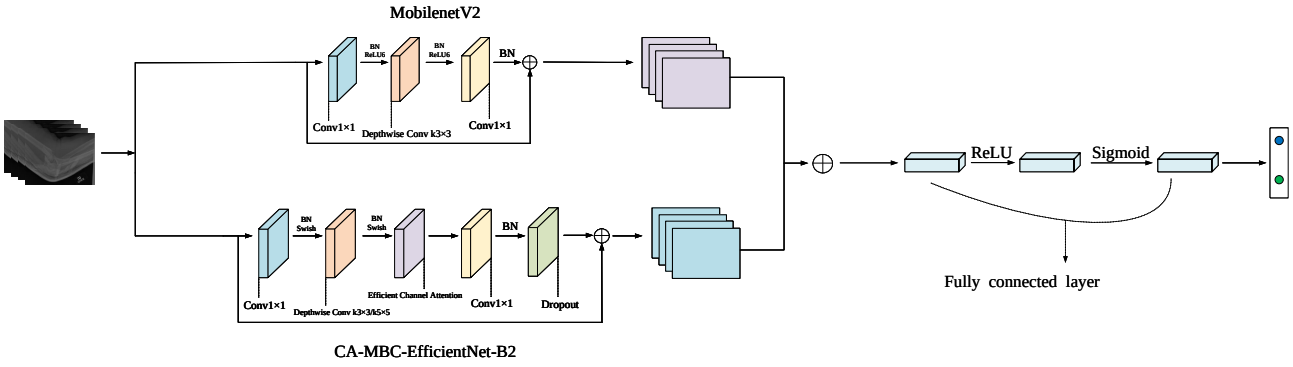


Figure 9. FusionNet based on the multi-model fusion of CA-MBC-EfficientNet-B2 and MobileNetV2.

2.4. Model training and testing procedures

The study involved two processes, i.e., 1) training FusionNet and CA-MBC-EfficientNet-B2 separately on seven datasets in MURA to generate an optimized model, and 2) testing FusionNet and CA-MBC-EfficientNet-B2 separately on seven datasets from MURA. The training process is as follows:

1.) The samples intended for input into the training model are divided into training and validation sets, maintaining a ratio of 8:2.

2.) For the training set, first of all, randomly crop the size of the picture to 260×260. A horizontal flip is then performed to enhance the diversity of the data set. A data type conversion is then performed to convert the image to a tensor. Finally, to ensure uniformity and comparability of pixel values across all channels of the image, a standardization process is performed. For the validation set, first adjust the image size to 260×260. Then perform a center cropping operation to crop the image from the center

to 260 × 260. Then, data type conversion is performed to convert the image into a tensor. Finally, the pixel values of each channel in the image are standardized through a process that aims to achieve uniform range and distribution.

3.) The processed images are taken as input and the corresponding labels are output; additionally, FusionNet and CA-MBC-EfficientNet-B2 are used for training. The precision, AUC and accuracy indexes on the verification set are monitored and saved, and the weights with the best performance on these indexes are selected as the final weights to obtain the optimized model.

The following outlines the testing process:

1.) Select the test sample and adjust the picture size to 260 × 260. The image is then cropped at the center, reducing it to 260 × 260 in size, thus preserving the central portion. A data type conversion is then performed to convert the image to a tensor. Finally, to achieve comparable range and distribution, the pixel values in each channel of the image undergo a standardization process.

2.) Using the processed image as input, the optimization model is utilized to determine the image label based on the model's output.

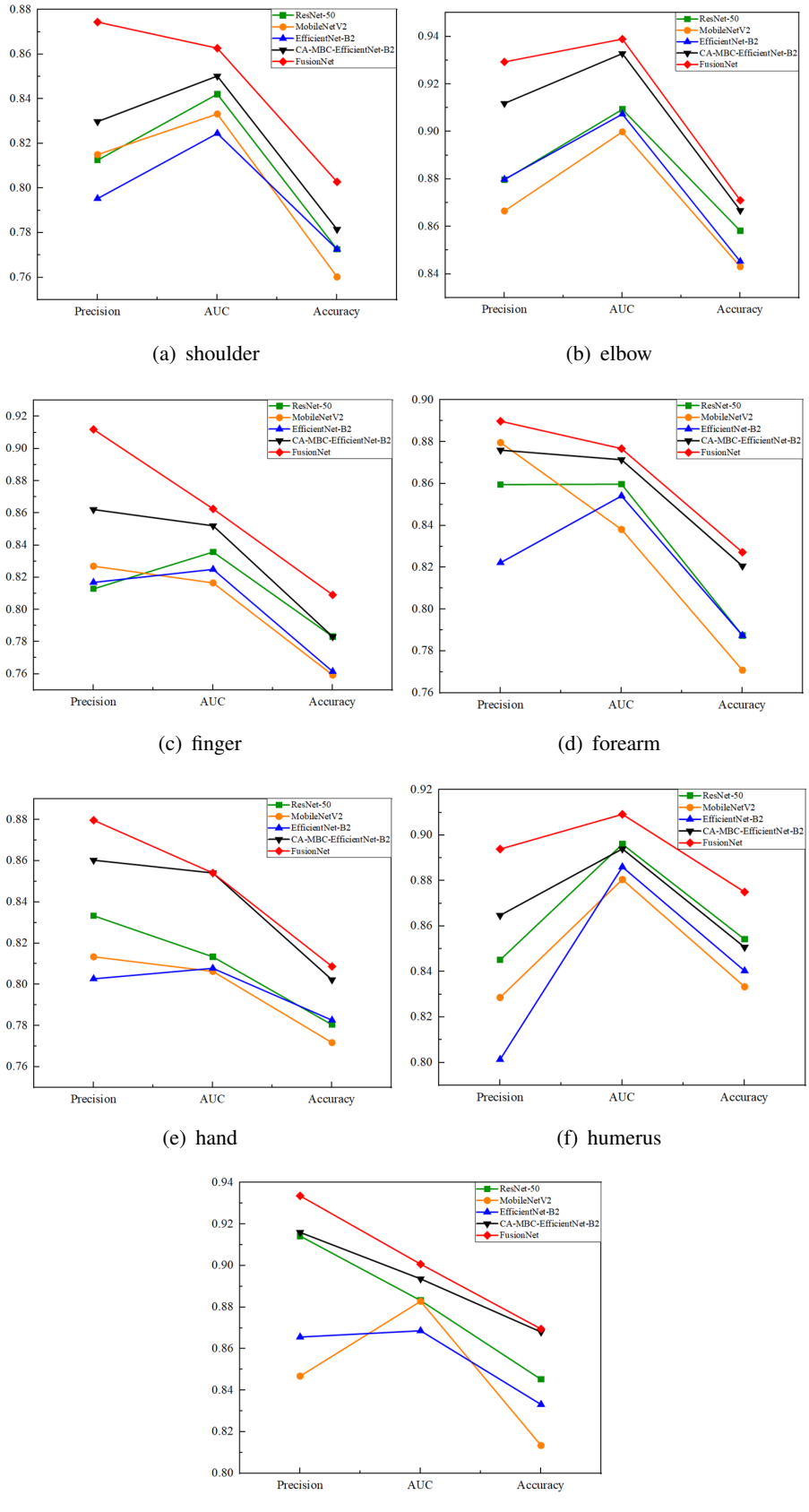
3. Experiments

3.1. Experimental environment and hyperparameter selection

The network model training in this study was conducted on hardware consisting of an Intel Xeon Gold 5320 processor, an Nvidia RTX A4000 graphics card, 32 GB of memory and the Linux operating system; additionally, Pytorch is the deep learning framework utilized in this study. To maintain objectivity in a comparison of the performance of each network model and to prevent biases from being introduced by the training process, the experiment was conducted with uniform parameter settings. All network models were trained for 150 epochs on the same dataset, with a consistent sample batch size of 16.

3.2. Experimental results and analysis

To accurately verify the performance of FusionNet and CA-MBC-EfficientNet-B2, this experiment compares the abnormal diagnostic performance of FusionNet with that of CA-MBC-EfficientNet-B2, EfficientNet-B2, MobileNetV2 and ResNet-50 on seven datasets from MURA, respectively. Precision, AUC and accuracy were selected as performance evaluation indexes. Precision is a metric used to assess the accuracy of positive case predictions in a classification model, AUC is the measure of the classification ability and differentiation of the model and accuracy is a metric employed to evaluate the overall accuracy of a classification model. Figure 10 illustrates the experimental results, with Figure 10(a)–(g) representing the outcomes for each model on the datasets corresponding to the shoulder, elbow, finger, forearm, hand, humerus, and wrist, respectively.



(a) shoulder (b) elbow

(c) finger (d) forearm

(e) hand (f) humerus

(g) wrist

Figure 10. Performance comparison for different models on MURA datasets.

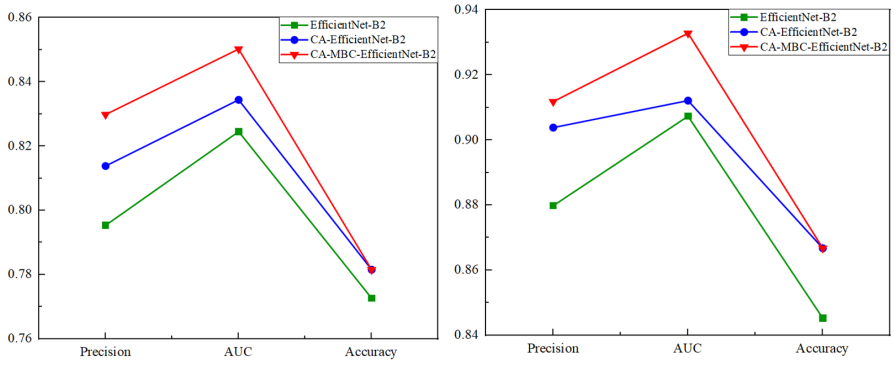
The results of the experiments show that FusionNet achieves significantly better performance on the seven datasets (a) to (g), particularly on the elbow dataset, where the precision, AUC and accuracy reach 92.93%, 93.89% and 87.10%, respectively, while EfficientNet-B2 and MobileNetV2 do not perform well in terms of the evaluation metrics. This means that compared to EfficientNet-B2 and MobileNetV2, FusionNet overcomes the limitations of individual network performance while incorporating features from multiple networks and extracting more effective features, ultimately improving overall performance. CA-MBC-EfficientNet-B2 performs worse than FusionNet but generally better than other models on the seven datasets shown in Figure 10(a)–(g). This means that by combining the CA module and using the improved MBConv module, the performance of EfficientNet-B2 is significantly improved. All in all, FusionNet and CA-MBC-EfficientNet-B2 demonstrated excellent performance in the diagnosis of musculoskeletal abnormalities.

To accurately validate the lightweight design of FusionNet and CA-MBC-EfficientNet-B2, Table 4 presents the parameter count, model size and FLOPs of FusionNet and CA-MBC-EfficientNet-B2 as compared to other models. The parameter count is usually used to measure the intricacy of the model, model size is usually used to measure the storage requirements and computing resource consumption and FLOPs is commonly employed as a metric to assess the computational complexity and speed of a model.

Table 4. Comparison of parameter count, model size and FLOPs of different models.

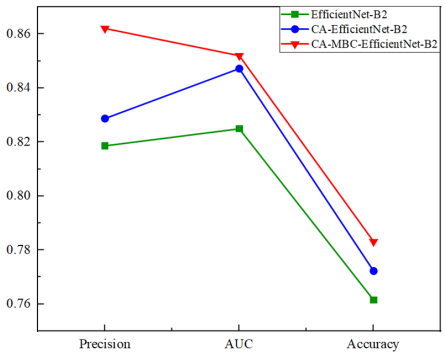
Model	No. of parameters	Model size	FLOPs
ResNet-50	23.51M	89.68 MB	12.00G
MobileNetV2	2.23M	8.49 MB	0.94G
EfficientNet-B2	7.70M	29.38 MB	2.04G
CA-MBC-EfficientNet-B2	6.37M	24.34 MB	2.04G
FusionNet	9.51M	36.28 MB	2.98G

As can be seen from the table above, CA-MBC-EfficientNet-B2 reduced the parameter count and model size by 1.33M and 5.04 MB, respectively, relative to EfficientNet-B2. Compared with ResNet-50, the parameter count, model size, and FLOPs are only 27%, 27% and 17% of its size, respectively. This means that CA-MBC-EfficientNet-B2 has extremely low complexity. Although FusionNet exhibited larger values for the number of parameters, model size and FLOPs compared to EfficientNet-B2 and MobileNetV2, its performance has been greatly improved. This means that FusionNet strikes a favorable balance between performance and having a lightweight structure, steering clear of the extremes characterized by low model complexity and low performance.

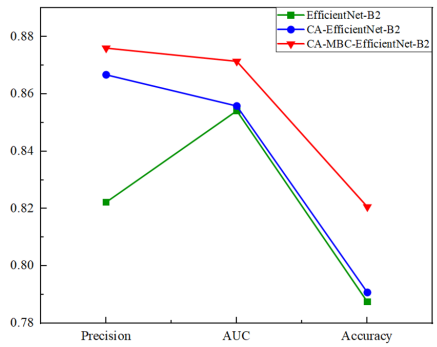


(a) shoulder

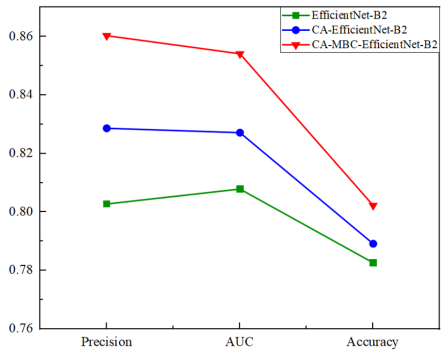
(b) elbow



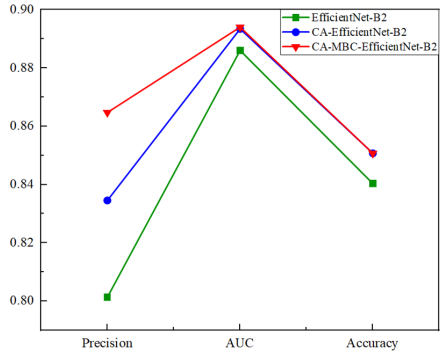
(c) finger



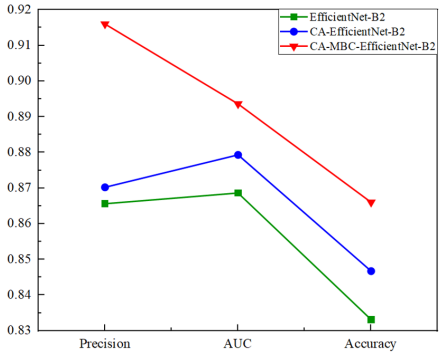
(d) forearm



(e) hand



(f) humerus



(g) wrist

Figure 11. Performance comparison for different models on MURA datasets.

3.3. Ablation studies

We conducted ablation studies on seven datasets within MURA to investigate the components of CA-MBC-EfficientNet-B2.

To investigate the contributions of the CA module and improved MBConv module in terms of performance, we conducted an experiment to compare the abnormal diagnosis performance of EfficientNet-B2, CA-EfficientNet-B2 and CA-MBC-EfficientNet-B2 on seven datasets from MURA. The selected performance evaluation indices were precision, AUC and accuracy. Figure 11 displays the experimental results. The experimental results demonstrate a substantial improvement in the performance of CA-EfficientNet-B2 relative to EfficientNet-B2. This suggests that EfficientNet-B2 fails to preserve positional feature information during the process of feature extraction, thereby impacting the model's performance. By incorporating location information within the channel attention mechanism, the CA modules effectively address the issue of feature information loss, leading to notable improvements in the model's performance. The superior performance of CA-MBC-EfficientNet-B2 compared to CA-EfficientNet-B2 indicates that the enhanced MBConv module is more proficient in extracting crucial features.

To investigate the contributions of the CA module and improved MBConv module in terms of lightweight design, Table 5 presents the parameter count, model size and FLOPs of EfficientNet-B2, CA-EfficientNet-B2 and CA-MBC-EfficientNet-B2, respectively.

Table 5. Comparison of parameter count, model size and FLOPs of different models.

Model	No. of parameters	Model size	FLOPs
EfficientNet-B2	7.70M	29.38 MB	2.04G
CA-EfficientNet-B2	7.70M	29.41 MB	2.04G
CA-MBC-EfficientNet-B2	6.37M	24.34 MB	2.04G

As can be seen from the table above, as compared to EfficientNet-B2, the number of parameters and FLOPs of CA-EfficientNet-B2 remain the same, and the model size is only increased by 0.03 MB. This implies that the CA module can almost enhance the complexity of EfficientNet-B2. CA-MBC-EfficientNet-B2 reduces the parameter count and model size by 17% relative to CA-EfficientNet-B2 and EfficientNet-B2, which shows that the improved MBConv involves significantly fewer parameters. It also means that CA-MBC-EfficientNet-B2 can be deployed on mobile, embedded and edge devices.

4. Conclusions

We have proposed an improved EfficientNet-B2 model, CA-MBC-EfficientNet-B2, which is fused with MobileNetV2 to obtain FusionNet. CA-MBC EfficientNet-B2 and FusionNet were used for the diagnosis of musculoskeletal abnormalities. CA-MBC-EfficientNet-B2 can simultaneously consider the relationships between different channels and different spatial positions. The utilization of the enhanced MBConv module not only allows for parameter reduction, but it also allows further performance optimizations to be made to the model. By overcoming the performance limitations of a single network and leveraging the features of multiple networks simultaneously, FusionNet efficiently extracts features that in turn enhance the overall performance of the model. The experimental results showcased

the efficacy of our proposed model, CA-MBC-EfficientNet-B2 and FusionNet, as a tool to realize the accurate diagnoses of musculoskeletal abnormalities across all seven datasets of MURA. These findings highlight the applicability of our model in the field of musculoskeletal abnormality diagnosis. The lightweight characteristic of CA-MBC-EfficientNet-B2 makes it easy to deploy the model on mobile devices, embedded devices and edge devices. FusionNet strikes a balance between performance and complexity, avoiding the extreme case of low complexity but low performance. Although this study has achieved certain research results in the area of musculoskeletal abnormality diagnosis, it cannot avoid the inherent limitations of deep learning, namely low model interpretability. So our next research direction will entail the use of some methods to improve the interpretability of the model to make up for this shortcoming.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence tools in the creation of this article.

Acknowledgments

This study was supported by the Major Project for New Generation of AI (Grant no. 2018AAA0100400), the Scientific Research Fund of the Hunan Provincial Education Department, China (Grant no. 21A0350, 21C0439, 22A0408, 22A0414, 23C0194) and the National Natural Science Foundation of Hunan Province, China (Grant no. 2022JJ50051).

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. J. Zhang, J. Z. Wang, Z. Yuan, E. S. Sobel, H. Jiang, Computer-aided classification of optical images for diagnosis of osteoarthritis in the finger joints, *J. Xray. Sci. Technol.*, **19** (2011), 531–544. https://doi.org/10.3233/XST-2011-0312
2. M. Al-Ayyoub, D. Al-Zghool, Determining the type of long bone fractures in X-ray images, *WSEAS Trans. Inf. Sci. Appl.*, **10** (2013), 261–270.
3. S. K. Mahendran, S. S. Baboo, An enhanced tibia fracture detection tool using image processing and classification fusion techniques in X-ray images, *Glob. J. Comput. Sci. Technol.*, **11** (2011), 22–28.
4. H. Y. Chai, L. K. Wee, T. T. Swee, S. Hussain, Gray-level co-occurrence matrix bone fracture detection, *WSEAS Trans. Syst.*, **10** (2011), 7–16.
5. M. Cohen, J. Puntonet, J. Sanchez, E. Kierszbaum, M. Crema, P. Soyer, et al., Artificial intelligence vs. radiologist: Accuracy of wrist fracture detection on radiographs, *Eur. Radiol.*, **33** (2023), 3974–3983. https://doi.org/10.1007/s00330-022-09349-3

6. Y. Nam, Y. Choi, J. Kang, M. Seo, S. J. Heo, M. K. Lee, Diagnosis of nasal bone fractures on plain radiographs via convolutional neural networks, *Sci. Rep.*, **12** (2022), 21510. https://doi.org/10.1038/s41598-022-26161-7
7. K. Oka, R. Shiode, Y. Yoshii, H. Tanaka, T. Iwahashi, T. Murase, Artificial intelligence to diagnosis distal radius fracture using biplane plain X-rays, *J. Orthop. Res.*, **16** (2021), 694. https://doi.org/10.1186/s13018-021-02845-0
8. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, preprint, arXiv: 1409.1556. https://arxiv.org/abs/1409.1556
9. M. He, X. Wang, Y. Zhao, A calibrated deep learning ensemble for abnormality detection in musculoskeletal radiographs, *Sci. Rep.*, **11** (2021), 9097. https://doi.org/10.1038/s41598-021-88578-w
10. G. Singh, D. Anand, W. Cho, G. P. Joshi, K. C. Son, Hybrid deep learning approach for automatic detection in musculoskeletal aadiographs, *Biology*, **11** (2022), 665. https://doi.org/10.3390/biology11050665
11. P. H. Yi, T. K. Kim, J. Wei, J. Shin, F. K. Hui, H. I. Sair, et al., Automated semantic labeling of pediatric musculoskeletal radiographs using deep learning, *Pediatr. Radiol.*, **49** (2019), 1066–1070. https://doi.org/10.1007/s00247-019-04408-2
12. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 770–778.
13. J. Yao, Z. Guo, W. Yu, Enhanced deep residual network for bone classification and abnormality detection, *Med. Phys.*, **49** (2022), 6914–6929. https://doi.org/10.1002/mp.15966
14. C. T. Cheng, T. Y. Ho, T. Y. Lee, C. C. Chang, C. C. Chou, C. C. Chen, et al., Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs, *Eur. Radiol.*, **29** (2019), 5469–5477. https://doi.org/10.1007/s00330-019-06167-y
15. G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 2261–2269.
16. J. W. Choi, Y. J. Cho, S. Lee, J. Lee, S. Lee, Y. H. Choi, et al., Using a dual-input convolutional neural network for automated detection of pediatric supracondylar fracture on conventional radiography, *Invest. Radiol.*, **55** (2020), 101–110. https://doi.org/10.1097/RLI.0000000000000615
17. C. T. Cheng, Y. Wang, H. W. Chen, P. M. Hsiao, C. N. Yeh, C. H. Hsieh, et al., A scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs, *Nat. Commun.*, **12** (2021), 1066. https://doi.org/10.1038/s41467-021-21311-3
18. Y. L. Thian, Y. Li, P. Jagmoha, D. Sia, V. E. Y. Chan, R. T. Tan, Convolutionalneural networks for automated fracture detection and localization on wrist radiographs, *Radiol. Artif. Intell.*, **1** (2019), e180001. https://doi.org/10.1148/ryai.2019180001
19. C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in *Proceedings of the AAAI Conference on Artificial Intelligence*, (2017), 4278–4284. https://doi.org/10.1609/aaai.v31i1.11231

20. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *Adv. Neural Inf. Process. Syst.*, **39** (2015), 1137–1149.

21. P. Ye, S. Li, Z. Wang, S. Tian, Y. Luo, Z. Wu, et al., Development and validation of a deep learning-based model to distinguish acetabular fractures on pelvic anteroposterior radiographs, *Front. Physiol.*, **14** (2023), 1146910. https://doi.org/10.3389/fphys.2023.1146910

22. X. Wang, Y. Wang, Composite attention residual U-Net for Rib fracture detection, *Entropy*, **25** (2023), 466. https://doi.org/10.3390/e25030466

23. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

24. L. Jin, J. Yang, K. Kuang, B. Ni, Y. Gao, Y. Sun, et al., Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet, *EBioMedicine*, **62** (2020), 103106. https://doi.org/10.1016/j.ebiom.2020.103106

25. A. Ghosh, D. Patton, S. Bose, M. K. Henry, M. Ouyang, H. Huang, et al., A patch-based deep learning approach for detecting rib fractures on frontal radiographs in young children, *J. Digit Imaging*, **36** (2023), 1302–1313. https://doi.org/10.1007/s10278-023-00793-1

26. T. Guo, C. Xu, S. He, B. Shi, C. Xu, D. Tao, Robust Student Network Learning, *IEEE Trans. Neural Networks Learn. Syst.*, **31** (2020), 2455–2468. https://doi.org/10.1109/TNNLS.2019.2929114

27. L. Yang, G. Yuan, H. Wu, W. Qian, An ultra-lightweight detector with high accuracy and speed for aerial images, *Math. Biosci. Eng.*, **20** (2023), 13947–13973. doi: 10.3934/mbe.2023621

28. E. Parcham, M. Fateh, HybridBranchNet: A novel structure for branch hybrid convolutional neural networks architecture, *Neural Networks*, **165** (2023), 77–93. https://doi.org/10.1016/j.neunet.2023.05.025

29. O. Eminaga, M. Abbas, J. Shen, M. Laurie, J. D. Brooks, J. C. Liao, et al., PlexusNet: A neural network architectural concept for medical image classification, *Comput. Biol. Med.*, **154** (2023), 106594. https://doi.org/10.1016/j.compbiomed.2023.106594

30. Z. Chen, J. Yang, L. Chen, H. Jiao, Garbage classification system based on improved ShuffleNet v2, *Resour. Conserv. Recycl.*, **178** (2022), 106090. https://doi.org/10.1016/j.resconrec.2021.106090

31. M. Versaci, G. Angiulli, P. Crucitti, D. De Carlo, F. Laganà, D. Pellicanò, et al., A fuzzy similarity-based approach to classify numerically simulated and experimentally detected carbon fiber-reinforced polymer plate defects, *Sensors*, **22** (2022), 4232. https://doi.org/10.3390/s22114232

32. Z. Chen, J. Yang, C. Yang, BrightsightNet: A lightweight progressive low-light image enhancement network and its application in "Rainbow" maglev train, *J. King Saud Univ-Com*, **35** (2023), 101814. https://doi.org/10.1016/j.jksuci.2023.101814

33. P. Angelov, X. Gu, A cascade of deep learning fuzzy rule-based image classifier and SVM, in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, (2017). https://doi.org/10.1109/SMC.2017.8122697

34. Z. Chen, J. Yang, L. Chen, Z. Feng, L. Jia, Efficient railway track region segmentation algorithm based on lightweight neural network and cross-fusion decoder, *Autom. Constr.*, **155** (2023), 105069. https://doi.org/10.1016/j.autcon.2023.105069

35. Z. Feng, J. Yang, Z. Chen, Z. Kang, LRseg: An efficient railway region extraction method based on lightweight encoder and self-correcting decoder, *Expert Syst. Appl.*, **238** (2024), 122386. https://doi.org/10.1016/j.eswa.2023.122386

36. M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in *Proceedings of the 36th International Conference on Machine Learning(ICML)*, (2019), 6105–6114. https://arxiv.org/abs/1905.11946

37. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. C. Chen, MobileNetV2: Inverted residuals and linear bottlenecks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2018), 4510–4520.

38. Z. Chen, H. Guo, J. Yang, H. Jiao, Z. Feng, L. Chen, et al., Fast vehicle detection algorithm in traffic scene based on improved SSD, *Measurement*, **201** (2022), 111655. https://doi.org/10.1016/j.measurement.2022.111655

39. P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, et al., Mura: Large dataset for abnormality detection in musculoskeletal radiographs, preprint, arXiv: 1712.06957.

40. K. Xu, J. Ba, R. Kiros, K. Cho, A. C. Courville, R. Salakhutdinov, et al., Show, Attend and Tell: Neural omage caption generation with visual attention, in *International Conference on Machine Learning(ICML)*, (2015), 2048–2057.

41. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2018), 7132–7141.

42. Q. Hou, D. Zhou, J. Feng, Coordinate attention for efficient mobile network design, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, (2021), 13708–13717.

43. Q. Wang, B. Wu, P. Zhu, P. Li, Q. Hu, ECA-Net: Efficient channel attention for deep convolutional neural networks, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020).

44. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, et al., MobileNets: Efficient convolutional neural networks for mobile vision applications, preprint, arXiv: 1704.04861.

AIMS Press