



OPEN ACCESS

EDITED BY

Naomichi Ogihara,
The University of Tokyo, Japan

REVIEWED BY

Romany Mansour,
The New Valley University, Egypt
Fu-Lien Wu,
University of Nevada, Las Vegas,
United States

*CORRESPONDENCE

Jian Chen,
✉ chenjiansci@163.com
Jianquan He,
✉ hejianquan08@163.com
Yiqun Hu,
✉ hyq0826@xmu.edu.cn

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 13 June 2023

ACCEPTED 30 August 2023

PUBLISHED 11 September 2023

CITATION

Zhao Y, Zhou J, Qiu F, Liao X, Jiang J, Chen H, Lin X, Hu Y, He J and Chen J (2023), A deep learning method for foot-type classification using plantar pressure images.

Front. Bioeng. Biotechnol. 11:1239246.
doi: 10.3389/fbioe.2023.1239246

COPYRIGHT

© 2023 Zhao, Zhou, Qiu, Liao, Jiang, Chen, Lin, Hu, He and Chen. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A deep learning method for foot-type classification using plantar pressure images

Yangyang Zhao^{1†}, Jiali Zhou^{1†}, Fei Qiu^{2†}, Xuying Liao³, Jianhua Jiang², Heqing Chen², Xiaomei Lin², Yiqun Hu^{4*}, Jianquan He^{2*} and Jian Chen^{2*}

¹School of Medicine, Xiamen University, Xiamen, Fujian, China, ²Department of Rehabilitation, Zhongshan Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, Fujian, China, ³The School of Clinical Medicine, Fujian Medical University, Fuzhou, Fujian, China, ⁴Department of Gastroenterology, Zhongshan Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, Fujian, China

Background: Flat foot deformity is a prevalent and challenging condition often leading to various clinical complications. Accurate identification of abnormal foot types is essential for appropriate interventions.

Method: A dataset consisting of 1573 plantar pressure images from 125 individuals was collected. The performance of the You Only Look Once v5 (YOLO-v5) model, improved YOLO-v5 model, and multi-label classification model was evaluated for foot type identification using the collected images. A new dataset was also collected to verify and compare the models.

Results: The multi-label classification algorithm based on ResNet-50 outperformed other algorithms. The improved YOLO-v5 model with Squeeze-and-Excitation (SE), the improved YOLO-v5 model with Convolutional Block Attention Module (CBAM), and the multilabel classification model based on ResNet-50 achieved an accuracy of 0.652, 0.717, and 0.826, respectively, which is significantly higher than those obtained using the ordinary plantar-pressure system and the standard YOLO-v5 model.

Conclusion: These results indicate that the proposed DL-based multilabel classification model based on ResNet-50 is superior in flat foot type detection and can be used to evaluate the clinical rehabilitation status of patients with abnormal foot types and various foot pathologies when more data on patients with various diseases are available for training.

KEYWORDS

flat feet, deep learning, ResNet-50, YOLO-v5, multilabel classification, foot type recognition

Introduction

The feet comprise bones, muscles, and ligaments and enable seamless movement and direct contact with the ground during walking. However, congenital foot shape problems or walking with an uncomfortable gait, such as in-toeing and out-toeing gait, can result in concentrated ground pressure on specific parts of the foot, resulting in permanent deformation of the foot and may cause knee joint or back pain (Arunakul et al., 2013; Izraelski, 2013). The feet may get easily deformed due to wrong walking postures; moreover, foot deformations not only pose a threat to foot health but also cause fatigue and pain while

walking and can even result in spine deformation. Therefore, accurate diagnosis of foot deformations is crucial.

Flatfoot is a common orthopedic condition characterized by the collapse of the medial longitudinal arch (MLA) and is often accompanied by calcaneal valgus and talonavicular joint abduction. It may cause plantar pain or fatigue after exercise (Arunakul et al., 2013; Carr et al., 2016). In children, developmental flat foot can be caused by many factors and may be symptomatic or asymptomatic and flexible or rigid. For example, the cause may be abnormal bone and joint development, such as with a tarsal coalition, a congenital vertical talus, or an accessory navicular bone. The soft tissue of generalized ligamentous laxity from Marfan's or Ehlers–Danlos can also result in flat foot deformity (McCormack et al., 2001; Flores et al., 2019). Adult flat foot can be categorized as either residual flat foot deformity from a developmental cause or as an acquired flat foot. Acquired flat foot is associated with a tight triceps surae or isolated gastrocnemius tightness, posterior tibial tendon dysfunction, midfoot laxity, abduction of the forefoot, external rotation of the hindfoot, subluxation of the talus, traumatic deformities, ruptured plantar fascia, Charcot's foot, and neuromuscular imbalance (polio, cerebral palsy, closed head injury, or following a cerebrovascular accident) (McCormack et al., 2001; O'Leary et al., 2013).

Flat foot deformity is often overlooked due to difficulties in accurate diagnosis, often leading to severe consequences. Attempts have been made to diagnose flat foot deformity by using the footprint index, MLA, and arch height index (Adoración Villarroja et al., 2008; Chang et al., 2014; Drefus et al., 2017). Nikolaidou and Boudolos, (2006) proposed a classification method based on static footprints to distinguish different foot types; however, these classifications are subjective, imprecise, and cannot quantify the changes before and after treatment. The 3D measurement system based on the footprint index has drawbacks, including variations caused by observer bias, equipment costs, and instrument calibration disparities. Although two-dimensional plantar image detection is user-friendly, it has low sensitivity (Su et al., 2017).

Foot deformity detection involves a classification process; therefore, improving the accuracy of classification is crucial. Deep learning (DL) methods are machine learning (ML) techniques that enable the computer to learn from data by extracting features from the data without human intervention, which is beneficial for professional data analysis applications (LeCun et al., 2015). Deep neural networks can extract features and identify and locate targets through backpropagation and parameter tuning (Deng et al., 2020; Lin et al., 2020). Region-based convolutional neural network (R-CNN) is a typical DL target recognition model and employs a multiscale feature pyramid and sliding window method for region proposal and bounding box detection. This two-stage pipeline greatly improves recognition accuracy (Redmon et al., 2016). You Only Look Once (YOLO) is a novel DL target detection model that uses anchors to concurrently localize and classify targets during convolutional feature extraction. Anchors leverage prior knowledge to design multiscale fixed reference boxes covering all positions and scales in an image. Each anchor predicts targets by using intersection over union (IoU) as a measure of detection accuracy; if the IoU exceeds a threshold, the number of learnable parameters reduces, and the efficiency improves (Albahli et al., 2020). Region-based

convolutional neural network (R-CNN) and YOLO have been used in various fields; for example, in real-time vehicle recognition, R-CNN identifies and locates vehicles separately in an image, whereas YOLO performs one-stage target detection, concurrently localizing, classifying, and detecting vehicles. Although static footprint images can be used to detect foot morphology and pressure patterns (Yu et al., 2019), existing image classification methods rely on individual features and lack consistency, resulting in reduced precision and sensitivity. Therefore, we proposed a DL model for predicting foot type based on pressure distribution of the foot to more accurately and efficiently identify flat foot deformity and provide guidance for clinical diagnosis, treatment, and prognosis.

Material and methods

Data preparation

We retrospectively collected plantar pressure images and clinical data from flat foot patients and healthy volunteers during June 2021–June 2022. In case two physiotherapists disagreed on the diagnosis, a third researcher (Chen J) was consulted. The FeetMapping[®] Plantar Pressure Plate system (NeuCognic, Jiangsu, China) was used to obtain the plantar pressure images (Yu et al., 2019). This system comprises a pressure board with an effective area of 0.11 m², a power interface, a network interface, a power switch, and an indicator light. During the test, subjects stood naturally for 30 s, and the tester collected raw data based on the plantar pressure distribution and stored it on a computer (Figure 1). An arch score of 20%–26% was considered normal. This research was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Zhongshan Hospital, which is affiliated to Xiamen University.

Diagnostic criteria

MLA collapse, physical examination, and X-ray examination were used to diagnose flat foot (Meehan and Brage, 2003). The midpoint of the calcaneus, the midpoint of the inner and outer malleolus, and the midpoint of the lower third of the calf were traced and connected. The normal range of the Angle obtained was -5° to $+5^{\circ}$, and the angle less than -5° was the deformity of foot varus (Flores et al., 2019). Moreover, A drop of 10 mm or more of the navicular bone will be interpreted as a flat foot (Persiane et al., 2021). Therefore, further X-rays are taken. The X-ray film is made from the lowest point of the calcaneus to the lowest point of the talar bone in a straight line, and then from the lowest point of the talar bone to the lowest point of the first metatarsal head in a straight line, the normal range of the Angle between the two lines is 113° – 130° , more than 130° will be diagnosed as flat foot (Ueki et al., 2019; Caravaggi et al., 2021).

Inclusion and exclusion criteria

Inclusion criteria: 1) age range: 5–60 years; 2) patients with a normal gait and no motor system diseases; 3) no history of foot

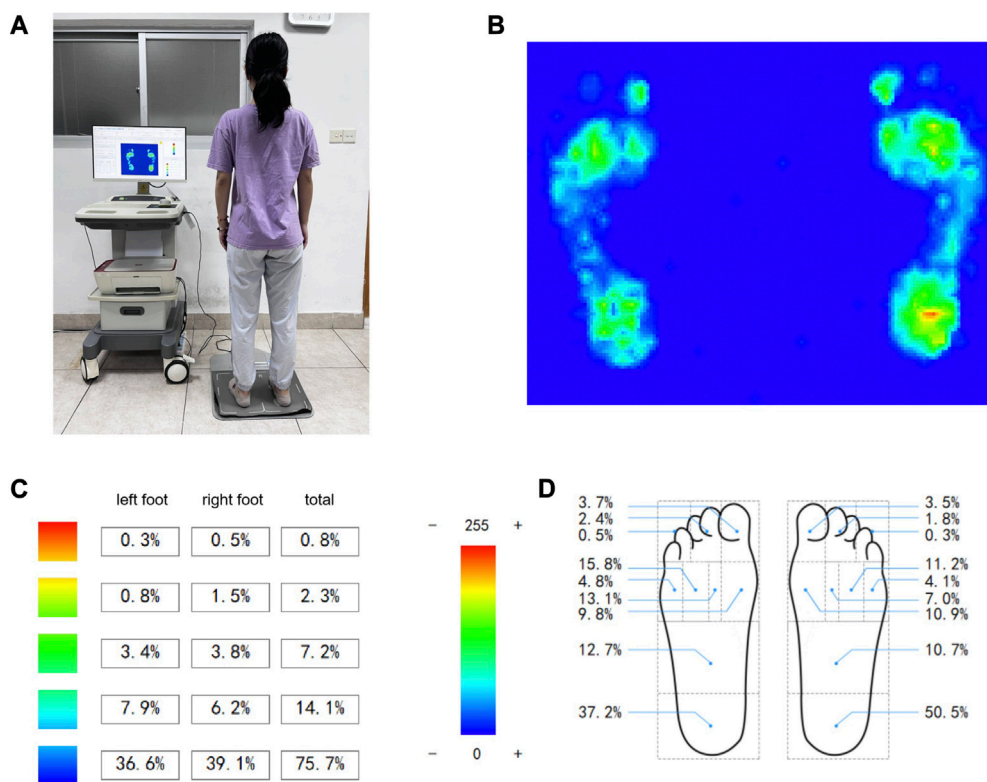


FIGURE 1 Collection of subjects' plantar pressure and proportional distribution. (A) FeetMapping® Plantar Pressure Plate system. (B) Plantar pressure image. (C) Image pressure distribution ratio. (D) Proportional distribution of pressure in each area of plantar.

trauma or operation; 4) all arches were determined by X-ray examination.

Exclusion criteria: 1) obvious foot deformity; 2) presence of high arches; 3) a history of lower limb and foot fractures or ankle sprain resulting in ligament damage or articular cartilage damage; 4) leg length discrepancy >2 cm (Vogt et al., 2020); 5) patients with cognitive dysfunction who are unable to cooperate with the test; 6) missing data.

For healthy subjects, the inclusion and exclusion criteria were consistent, except for not meeting the diagnostic criteria for flat feet.

Image dataset processing and classification

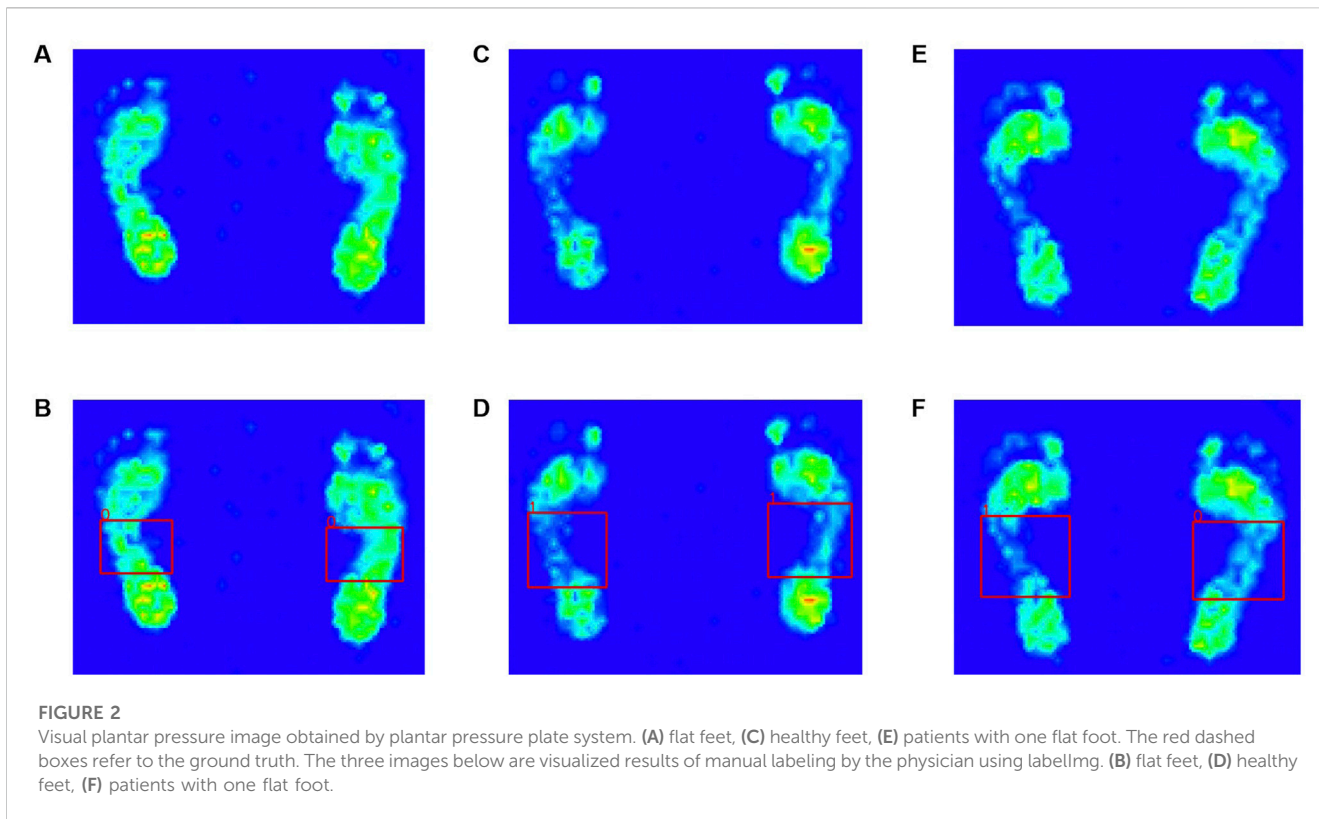
After anonymizing the collected plantar pressure images, two professionally trained rehabilitation therapists labeled them; labelImg(1.8.6) was used to select the midfoot of the images for labeling classification (Figure 2). A text file containing the category, target center coordinates, target boundary frame, and aspect ratio information of the overall image corresponding to the original image was generated. Images were labeled as flat foot (class 0) or healthy foot (class 1), denoted by values of 0 and 1, respectively. The four corners of the middle foot region were selected as the center points in each image and recorded in the txt file by using (x,y) coordinates. A random seed 0) was used to divide the total dataset into a training set and a test set in the ratio of 8:2.

Target detection algorithms can be classified as two-stage detectors with region proposal networks and one-stage detectors. Although two-stage detectors yield high accuracy, their size makes deployment on embedded devices difficult. To improve abnormal foot-type screening using plantar pressure systems, we explored the use of YOLO, a one-stage detector, for its performance and efficiency. The YOLOv5 model has four variants, namely, YOLOv5s, YOLOv5 m, YOLOv5 l, and YOLOv5x; among these, YOLOv5s has a relatively small model size and is thus suitable for embedded applications. The model architecture can be divided into three parts: CSPDarkNet53 for feature extraction (Bochkovskiy et al., 2020), FPN module and PAN module for feature fusion and transmission, and the category and position prediction module. By utilizing YOLOv5s, we aimed to achieve real-time foot type detection on low-power devices. During inference, when there is a target in the specified cell, the IoU between the bounding box and the true target can be calculated using Eq. 1:

$$IoU_{pred}^{truth} = \frac{\text{ground truth box} \cap \text{predicted bounding box}}{\text{ground truth box} \cup \text{predicted bounding box}} \quad (1)$$

Predicted category information and target box confidence are multiplied to obtain the category confidence score for each target box (Eq. 2):

$$\text{Class specific confidence score} = \Pr_{(class_i|object)} \times \Pr_{(object)} \times IOU_{pred}^{truth} \quad (2)$$



However, YOLOv5s has limited learning capacity. Attention mechanisms enhance object detection models by improving their capability to learn representations. Attention modules require very few additional parameters, thereby enhancing capabilities without substantially increasing model complexity.

Improved YOLOv5s algorithm

Squeeze-and-Excitation Networks (SENet) (Hu et al., 2018) and Convolutional Block Attention Module (CBAM) (Woo et al., 2018) are commonly used attention mechanism modules. SENet improves feature learning by modeling channel relationships, whereas CBAM focuses on informative regions and suppresses irrelevant regions by sequentially inferring spatial and channel attention. The simple and efficient designs of SE, CBAM, and other attention modules enable easy integration into different layers of YOLOv5s, thereby yielding considerable performance gains. Incorporating attention mechanisms into the C3 module (the third residual block of the CSPDarknet53 backbone network) enhances feature representation. Therefore, in this study, we incorporated SE, CBAM, and other attention modules into the C3 module. The SE module involves squeeze and excitation operations. The squeeze operation performs global average pooling to reduce the output feature map of the C3 block with size $H \times W \times C$ to a feature map Z of size $1 \times 1 \times C$ (Eq. 3):

$$Z = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_C(i, j) \tag{3}$$

The excitation operation learns a channel weight vector S . σ denotes the sigmoid activation function, W_1 and W_2 are learnable parameter matrices, and δ denotes the ReLU activation function for channel recalibration. Channel recalibration is performed using the excitation operation (Eq. 4):

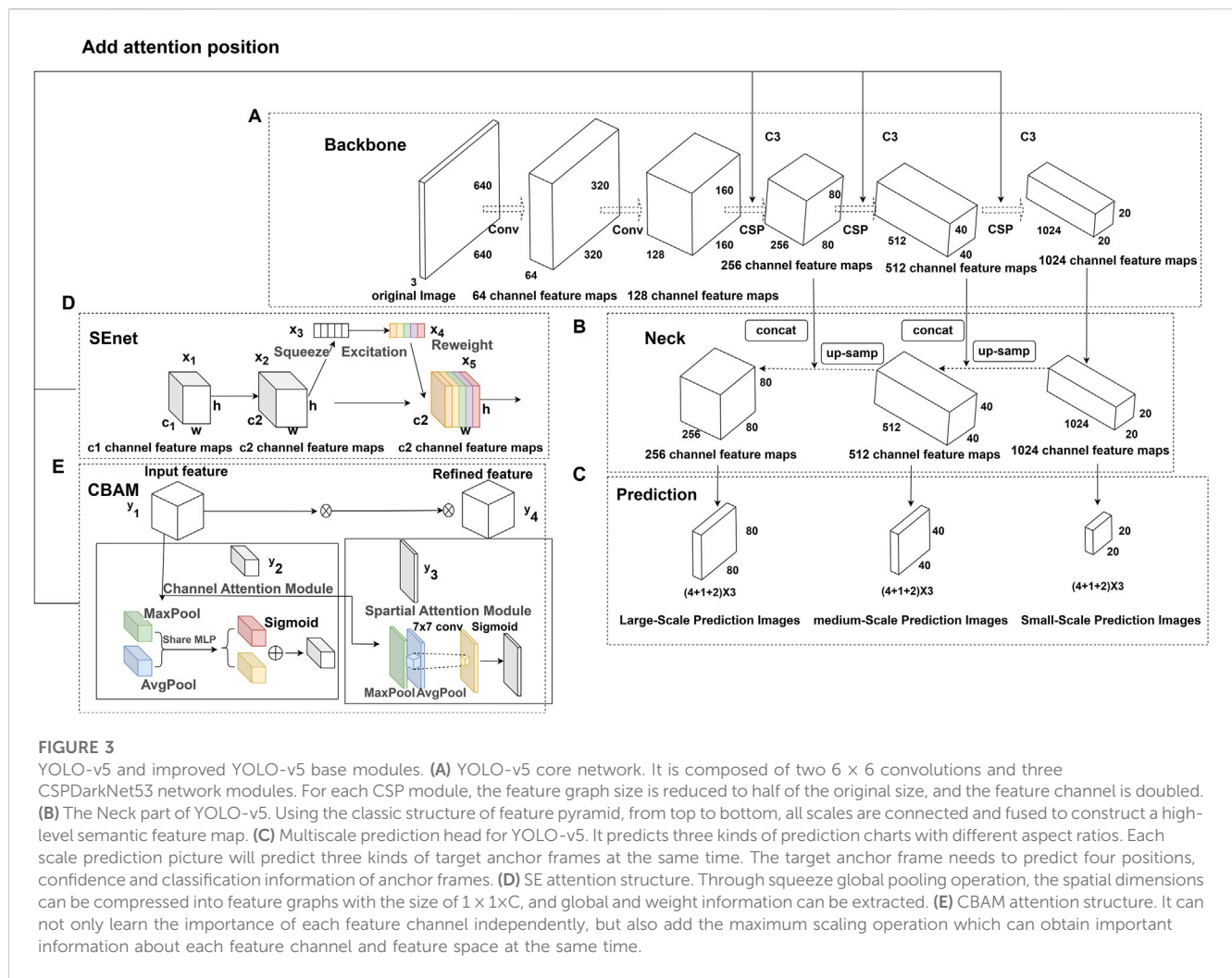
$$S = \sigma(W_1 \delta W_2(Z)) \tag{4}$$

After processing using the SE module, channel attention improves important features while suppressing irrelevant ones.

The CBAM module contains channel attention and spatial attention branches. We denoted the output feature map of the C3 block as U with size $H \times W \times C$. The channel attention branch comprises global max pooling and average pooling operations, which transform U into two one-dimensional descriptor vectors. These vectors are processed by respective multilayer perceptrons and combined. Next, a sigmoid activation generates channel weights from 0 to 1. Finally, the learned channel weight coefficients are multiplied by the input feature map u_X to obtain the channel-wise attended output feature map Z_C (Eq. 5):

$$Z_C = \sigma(MLP(F_{max}(u_X)) + MLP(F_{avg}(u_X))) \times u_X \tag{5}$$

First, the spatial attention branch compresses the input u_X along the channel dimension by applying global max pooling and average pooling across the channels to generate two 2D feature maps. Next, these feature maps are concatenated channel-wise to form a tensor with two channels, which is further convolved using a 7×7 kernel to reduce to a single channel. A sigmoid activation then generates spatial attention coefficients. Finally, the input u_X is multiplied by



these coefficients to obtain the spatially attended output feature map Z_X (Eq. 6):

$$Z_X = \sigma(\text{Conv}_{7 \times 7}(F_{\max}(u_X)), \text{Conv}_{7 \times 7}(F_{\text{avg}}(u_X))) \times u_X \quad (6)$$

Two attention mechanisms are added to the CSPDarkNet53 module, as shown in Figure 3.

Other strategies

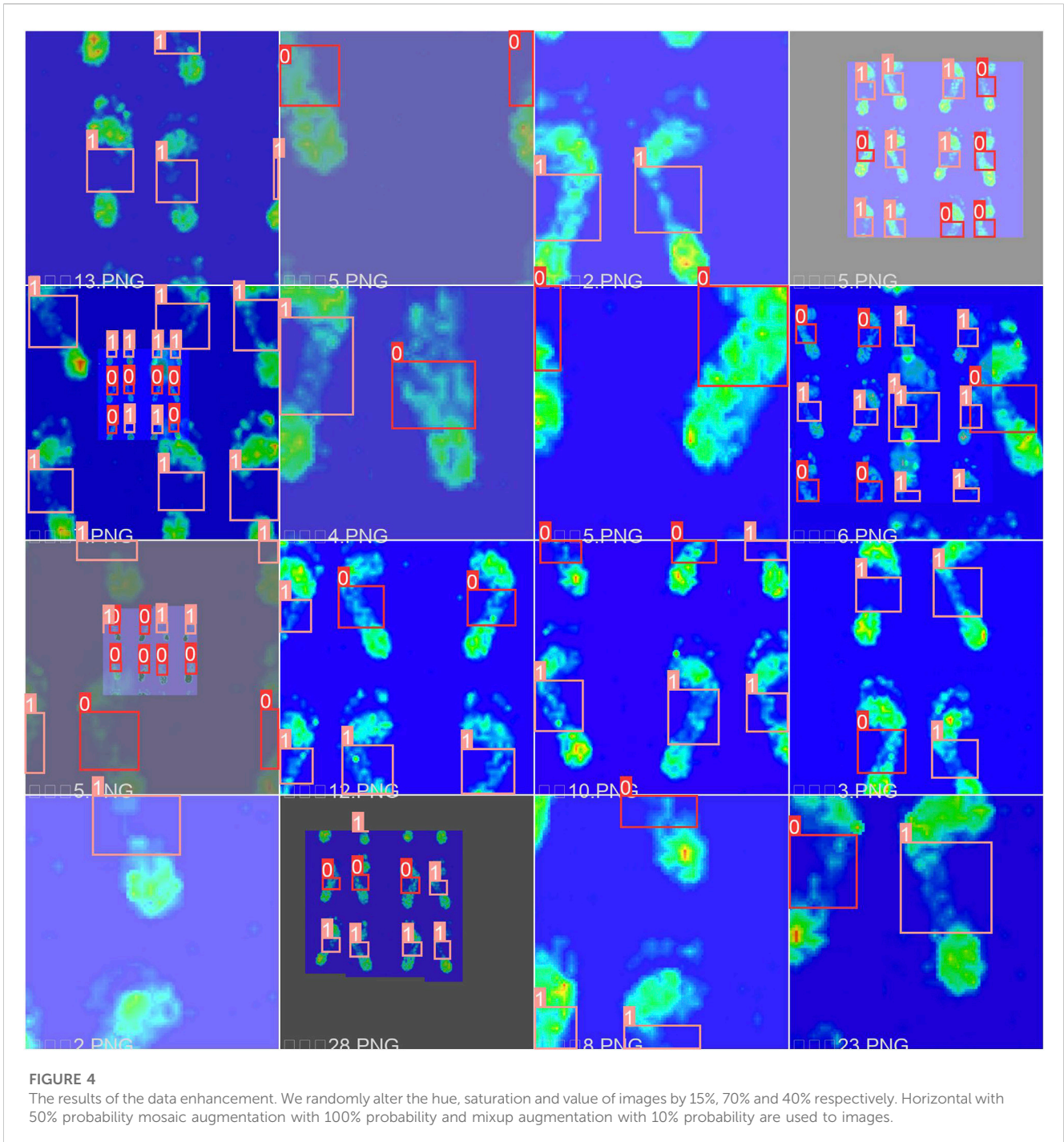
For data augmentation, we randomly cropped the training set to images of size 640 × 640 × 3. The sizes of nine prior boxes were obtained through K-means clustering: 10 × 13, 16 × 30, 33 × 23, 30 × 61, 62 × 45, 59 × 119, 116 × 90, 156 × 198, and 373 × 326. For training, we divided the feature map into grids of the same size as the prior box. When the target center was within a grid, that specific grid unit was responsible for target detection. It outputted a prediction box based on the initial anchor box and then compared it with the real box, calculated the discrepancy between the two, and then iteratively updated the network parameters.

In this study, we improved the generalizability by using enhancement techniques. Previous studies have mainly utilized random rotation, scaling, panning, flipping, and illumination to augment data. However, such techniques have limitations, such as

poor generalization, information loss, and noise interference. Therefore, to maximize algorithm performance, we adopted mosaic data augmentation following YOLOv4 and YOLOX (Zheng et al., 2022), random affine transformation, mixup augmentation (Zhang et al., 2018), and HSV enhancement (Redmon et al., 2016). Experimental results demonstrated that these enhancement methods improved detection accuracy and enhanced model generalization in the YOLOV5 framework. The augmentation results are shown in Figure 4.

Multilabel classification

Studies have demonstrated the importance of the midfoot region in flatfoot recognition (Welte et al., 2023). However, Buldt et al. (2018) studied the difference in the ratio of plantar pressure distribution between normal and flat feet and found that the ratio of plantar pressure distribution between the two foot types differs in different regions. Therefore, to determine whether these additional regions aid flatfoot identification, we employed a multilabel classification model to extract features and categorize plantar pressure images. We used ResNet-50 in the backbone network because it is a dynamic convolution neural network architecture with good performance. We resized the image



size of the dataset to 256×256 pixels, assigned two labels to the single original image, namely flat foot and healthy foot, and then performed one-hot encoding. Healthy feet were assigned the code (0,0), whereas flat feet were assigned (1,1). In cases where only one arch collapsed, the coding label was set as 0, 1 (or 1, 0), and training was conducted for a total of 100 epochs, as depicted in Figure 5.

Statistical analysis

We used SPSS 26.0 for the statistical analysis of participants' general information and expressed the measurement data as Mean \pm Standard

Deviation. In addition, we used independent sample's *t*-test to verify the differences between groups, with statistically significance set at $p < 0.05$.

Results

The inclusion criteria yielded 125 subjects: 61 with normal feet, 61 with flat feet, and 3 with single flat foot. Each person underwent the tests 13 times. A total of 1,573 plantar pressure images were obtained: 748 flat feet, 51 single flat foot, and 774 normal feet. Basic information is presented in Table 1, revealing no statistical difference between the two groups ($p > 0.05$).

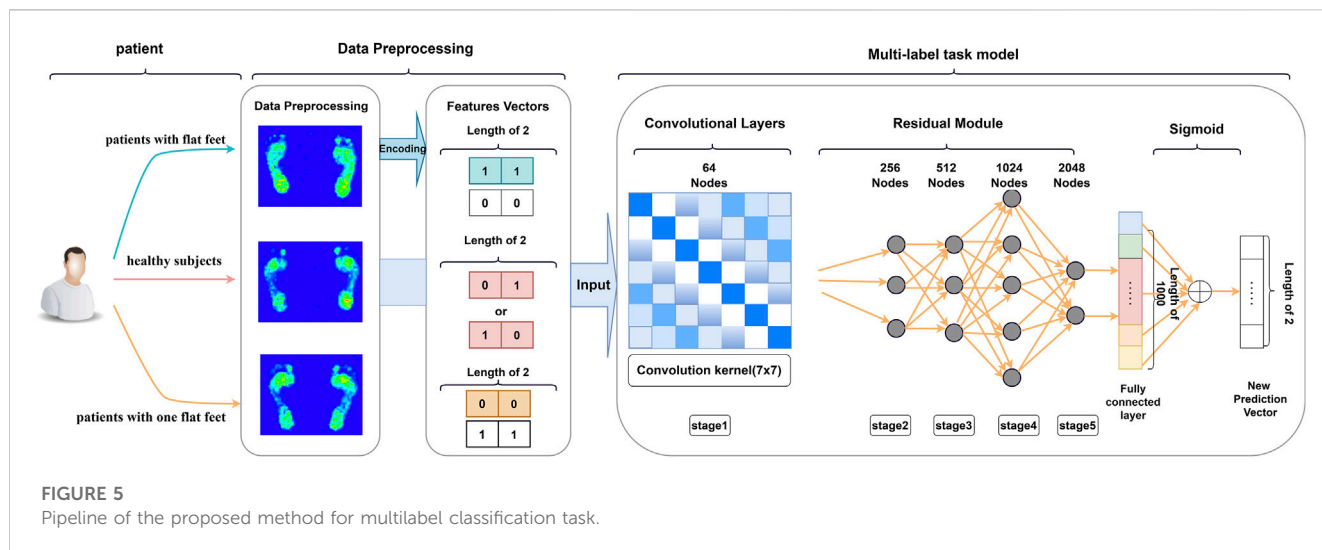


FIGURE 5 Pipeline of the proposed method for multilabel classification task.

TABLE 1 General information of subjects.

	Normal foot (<i>n</i> = 61)	Flat foot (<i>n</i> = 64)	<i>P</i>
Female, <i>n</i>	37	37	
Age, <i>y</i>	32.64 ± 14.85	29.56 ± 17.09	0.286
Body mass (kg)	55.83 ± 12.94	54.00 ± 17.44	0.508
Body mass index (kg/m ²)	20.92 ± 2.86	21.72 ± 4.40	0.231

Experimental environment

The laboratory setup comprised Ubuntu 20.04.1 LTS, an Intel Xeon E5–2,620 v4 2.10 GHz processor, and four NVIDIA TITAN Xp-12 GB GPUs. Python 3.7.5 was the development environment, and PyTorch 1.12.1 was the DL framework, and CUDA 11.3/ CUDNN 8.2 was used for image processing.

Evaluation indicators

We used accuracy, precision, recall, F1-score, average precision (AP), and mean average precision (mAP) to evaluate the merits of models (Eq. 7):

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + FN + FP + TN} \\
 \text{Precision} &= \frac{TP}{TP + FP} \\
 \text{Recall} &= \frac{TP}{TP + FN} \\
 \text{F1 - score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\
 \text{AP} &= \int_0^1 \text{Precision}(\text{Recall})d\text{Recall} \\
 \text{mAP} &= \frac{1}{C} \sum c \int \text{AP}(c)
 \end{aligned}
 \tag{7}$$

True positives (TPs) are correctly classified positive samples, false positives (FPs) are incorrectly classified positives, and false

negatives (FNs) are incorrectly classified negatives. Accuracy is the proportion of correctly classified samples. Precision is the ratio of TPs to total predicted positives. Recall is the ratio of TPs to total actual positives. F1-score is the harmonic mean of precision and recall. AP and mAP are commonly used in target detection to assess a model’s detection effectiveness and performance (Zhao et al., 2022). AP integrates the precision *P* at each discrete recall *R* point from 0 to 1 by calculating the area under the precision-recall curve. A higher AP value indicates better detection performance for a certain class. Let *C* denote the total number of classes in the detection model, *c* denote each class, and AP(*c*) denote the AP for each class *c*. mAP is computed by first calculating the AP for each class and then averaging over the AP across all classes, resulting in a comprehensive evaluation metric that reflects the model’s detection effectiveness for all categories.

Model training

Training of attention-enhanced YOLOv5s

The following steps are performed to train the model. To obtain better training results, we initialized the YOLOv5s model by using weights pretrained on the COCO train 2017 dataset and optimally trained the model by using adaptive weight (AdaW) (Loshchilov and Hutter, 2019). Next, we resized the input images to 640 × 640 pixels. We employed AdamW for parameter optimization and adopted a maximum learning rate of 0.1, with a batch size of 16 images per

TABLE 2 Ratio of pressure distribution in the plantar area between the two groups (%).

	Normal foot (<i>n</i> = 61)	Flat foot (<i>n</i> = 64)	<i>P</i>
Left toe	6.49 ± 3.06	4.53 ± 1.73	<0.001
Left metatarsal	42.64 ± 6.41	38.13 ± 9.74	0.003
Left mediopodium	12.97 ± 6.69	21.56 ± 7.68	<0.001
Left heel	37.91 ± 7.60	35.79 ± 8.59	0.147
Right toe	6.84 ± 3.00	4.95 ± 2.10	<0.001
Right metatarsal	41.82 ± 7.30	36.97 ± 9.66	0.002
Right mediopodium	13.15 ± 6.02	21.99 ± 7.12	<0.001
Right heel	38.17 ± 9.46	36.11 ± 8.71	0.205

iteration. The first three epochs employed frozen training, where the weights of the feature extraction layers were maintained constant. This was followed by 97 epochs of unfrozen training for fine-tuning all layers in the model. The IoU threshold and momentum were set as 0.2 and 0.937, respectively. We used the same training method for the other versions of YOLOv5s in this study.

Training of multilabel classification

To fit the dataset, we adopted a fine-tuning approach by modifying the last fully-connected layer of ResNet-50. We loaded the model on the GPU for training by using binary cross-entropy loss. Furthermore, we utilized the stochastic gradient descent optimizer with differentiated learning rates, where the newly added classifier layers assume a 10× larger rate of 0.001 for rapid adaptation. The learning rates were reduced by 10× every 5 epochs by using step decay scheduling. The model was trained for 100 epochs to obtain an adapted network suited for the specific dataset.

Plantar pressure

Compared with normal subjects, subjects with flat feet exhibited lower pressure in the forefoot (toe and metatarsal region) and higher pressure in the middle foot ($p < 0.05$); in addition, no significant difference was observed in the heel pressure ratio between the two groups ($p > 0.05$) (Table 2).

Model performance evaluation

Comparison between the deep learning model and traditional machine learning model

To verify the test performance of the model and the FeetMapping[®] plantar pressure measuring system (NeuCognic, Jiangsu, China), we inputted the test dataset into the improved trained YOLO-v5 model and the FeetMapping[®] plantar pressure measuring system (NeuCognic, Jiangsu, China) for statistical

verification. The prediction accuracy of YOLOv5 model based on the attention mechanism was higher than that of traditional ML in both healthy and flat foot patients (Table 3).

Performance comparison between improved YOLO-v5 and original YOLO-v5

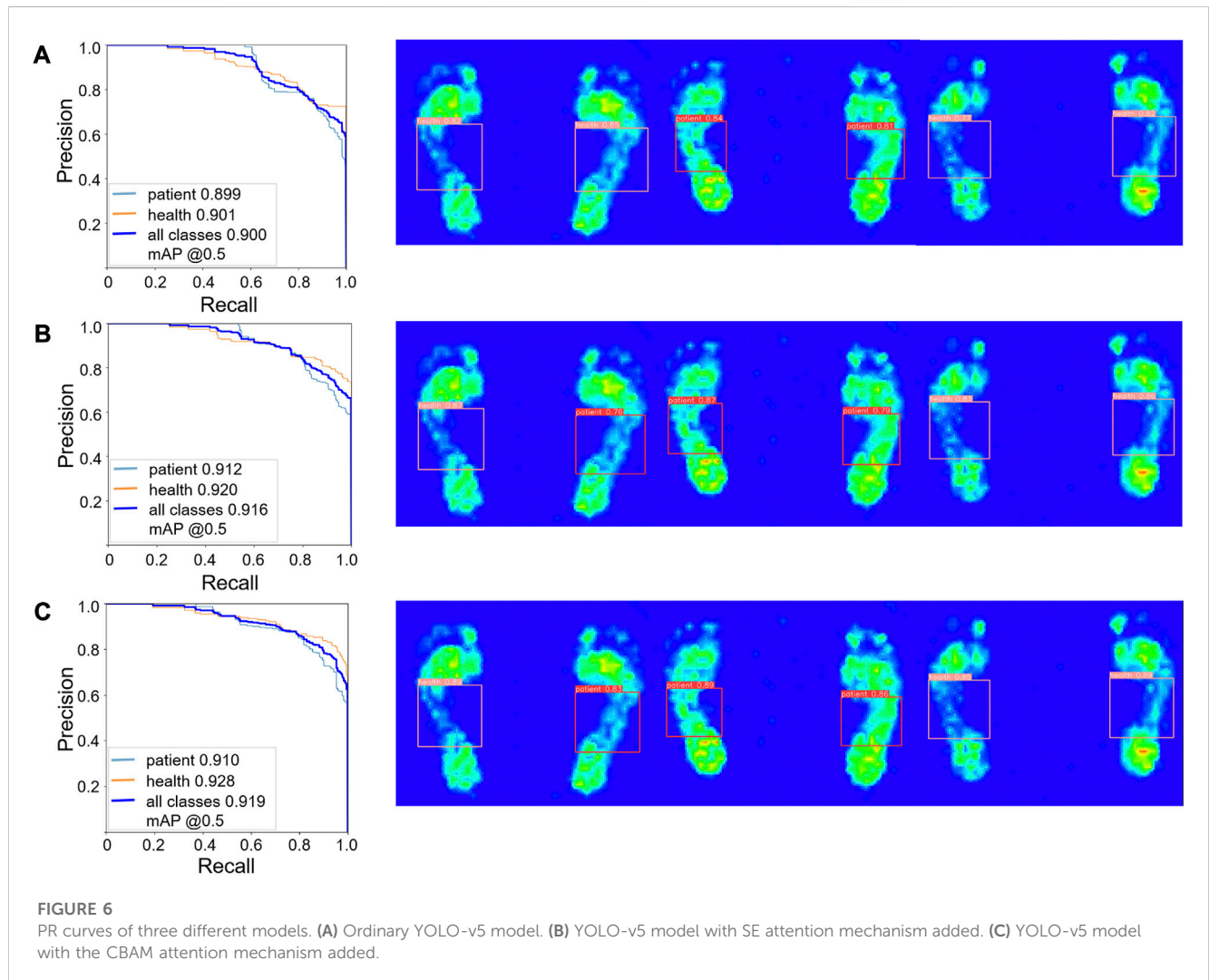
The improved YOLO-v5 model included SE and CBAM attention mechanisms in the C3 module. We compared the performance of the original YOLO-v5 model with the improved YOLO-v5 model by using precision-recall curve (PR curve). PR curve is commonly used when the distribution of data categories is uneven. The horizontal axis represents recall, and the vertical axis represents accuracy. Similar to the ROC curve, when the PR curve is closer to the upper right, it indicates that the model performance is better. The PR curve of YOLO-v5 model based on C3CBAM attention mechanism exhibited the best performance (Figure 6). The average accuracy and sensitivity of YOLO-v5 based on the CBAM attention mechanism were 84.7%, 86.4%, respectively, and the mAP for different IoU thresholds (0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, and 0.95) was 91.9%. The average accuracy was 8.5% higher than that of YOLO-v5 with SE and 4.7% higher than that of YOLO-v5 without attention mechanism. Furthermore, the sensitivity was 5.4% lower than that of YOLO-v5 with SE and 2% lower than that of YOLO-v5. The average mAP for different IoU thresholds was 0.3% higher than that of YOLO-v5 with SE and 1.9% higher than that of YOLO-v5.

Performance evaluation of multilabel classification tasks

The distribution ratio of the participants' plantar pressure regions revealed that the target detection task focused only on extracting features from the middle foot and overlooked the features of the toe, metatarsal bone, and hindfoot due to the specific labels assigned. To investigate the effect of plantar pressure on foot type detection capability, we generated thermal maps by using Grad-CAM in a multilabel classification task based on ResNet-50 (Figure 7). Thermal maps enable backpropagation through category assignment, obtaining gradient information from

TABLE 3 Accuracy of deep learning models and traditional machine learning models (%).

	YOLO-v5 based on attention mechanism	FeetMapping [®] machine
All	84.7	76.5
Patient	86.5	76.0
Health	82.9	75.8



extracted features, and weighting the elements that significantly contribute to algorithmic recognition. When the recognition object was set as a normal foot, the highlighted regions were concentrated in the metatarsal and middle foot regions. Similarly, when the recognition object was set as a flat foot, the highlighted regions were primarily concentrated in the metatarsal and middle foot areas.

To further assess the accuracy of the multilabel sorting task, we randomly selected 20% of the dataset as the test set and used the remaining 80% for training. We employed different DL models for foot-type recognition of plantar pressure images based on the label features extracted in this study. We compared the performance of four models, namely, YOLO-v5, improved YOLO-v5 incorporating

different attention mechanisms (SE and CBAM), and multilabel classification task based on ResNet-50. The results showed that although the accuracy, recall rate, and F1-score of the algorithm improved after adding attention mechanisms in YOLO-v5, the performance of the multilabel classification algorithm based on ResNet-50 was significantly superior to other algorithms (Table 4).

Model verification

To validate the feasibility and accuracy of the proposed model in clinical practice, we freshly re-collected plantar pressure images of 46 participants from June 2022 to August 2023 as a new dataset,

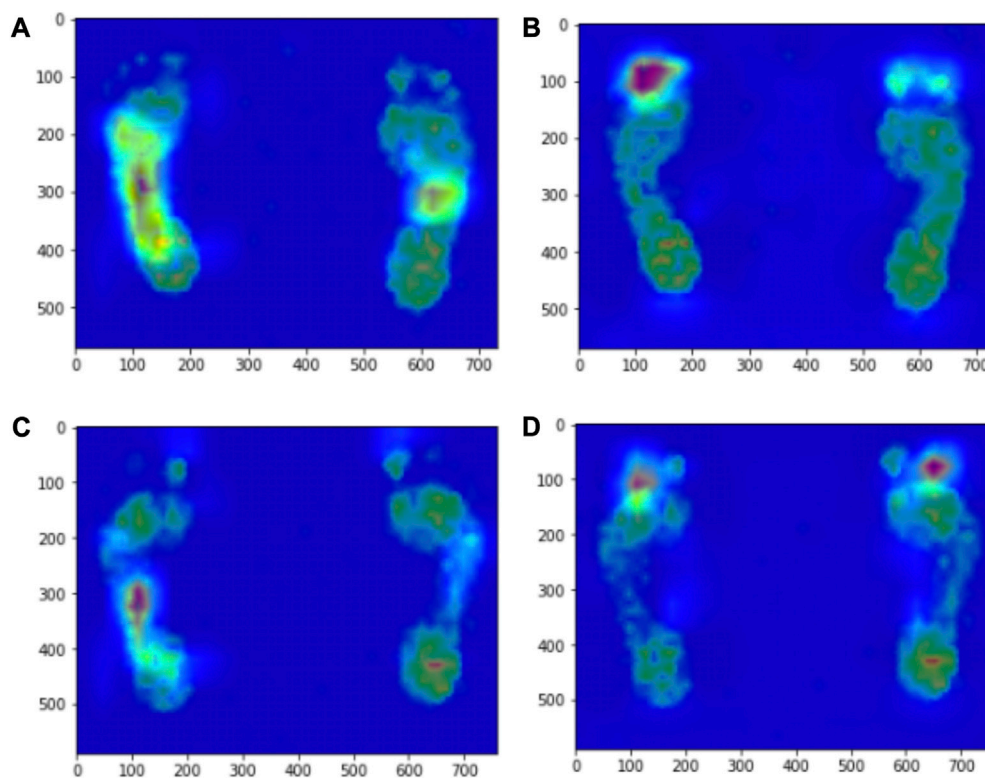


FIGURE 7 Grad-CAM heat map based on ResNet-50 model. ResNet-50 model sets the recognition object to normal foot, (A) the high activation region of pixel sensitivity of visual prediction images is located in the middle foot region, (B) the high activation area of pixel sensitivity of visual prediction images is located in the metatarsal region. ResNet-50 model sets the recognition object as flat feet, (C) the high activation region of pixel sensitivity of visual prediction images is located in the middle foot region, (D) the high activation area of pixel sensitivity of visual prediction images is located in the metatarsal region.

TABLE 4 Precision comparison of deep learning models.

Model	Precision (%)	Recall (%)	F1-score (%)
YOLO-v5	77.6	83.8	80.6
YOLO-v5_C3SE	76.2	91.8	83.3
YOLO-v5_C3CBAM	84.7	86.4	85.6
Multilabel_ResNet-50	91.8	92.9	92.3

including 16 children and 30 adults, with the same inclusion and exclusion criteria as before (Supplementary Table S1). We inputted the datasets into different DL models and compared the obtained results with the original ML and clinical diagnosis results. The accuracy of FeetMapping® plantar pressure measurement system (NeuCognic, Jiangsu, China), ResNet-50-based multilabel classification algorithm, improved YOLO-v5 with SE or CBAM attention mechanism, and original YOLO-v5 were 0.630, 0.826, 0.652, 0.717, and 0.652, respectively (Figure 8).

Discussion

In this study, we aimed to improve the YOLO-v5 DL model for foot classification based on plantar pressure images. We

extracted middle foot features and incorporated SE and CBAM attention mechanisms. A comparison of the two modeling methods revealed that the modified YOLO-v5 model and the multilabel classification task greatly improved the performance of the flat foot classification system. Clinical verification showed that both models performed well, with the multilabel classification task yielding higher accuracy. Therefore, the ResNet-50-based multilabel classification algorithm is suitable for foot diagnosis and treatment prognosis.

This study highlighted the limitations of existing models in identifying soft tissue health conditions. Although the YOLO model excels in bone X-ray detection (Jiang et al., 2022; Li et al., 2023), it cannot identify soft tissue health issues. In this study, we used plantar pressure image data to conduct early screening for flat feet because they provide valuable information regarding lower limb dynamics and foot health.

The YOLO series is a one-stage target detection algorithm. YOLO-v1 divides the input image into a grid with uniform size; considers the prediction of the target and the target boundary box in the grid as a regression problem; and obtains the location, confidence, and category of the object (Redmon et al., 2016). However, YOLO-v1 poses challenges in accurately positioning objects and yields a low recall rate. Therefore, k-means clustering is incorporated in YOLO-v2, and the prior anchor frame is used for migration constraint

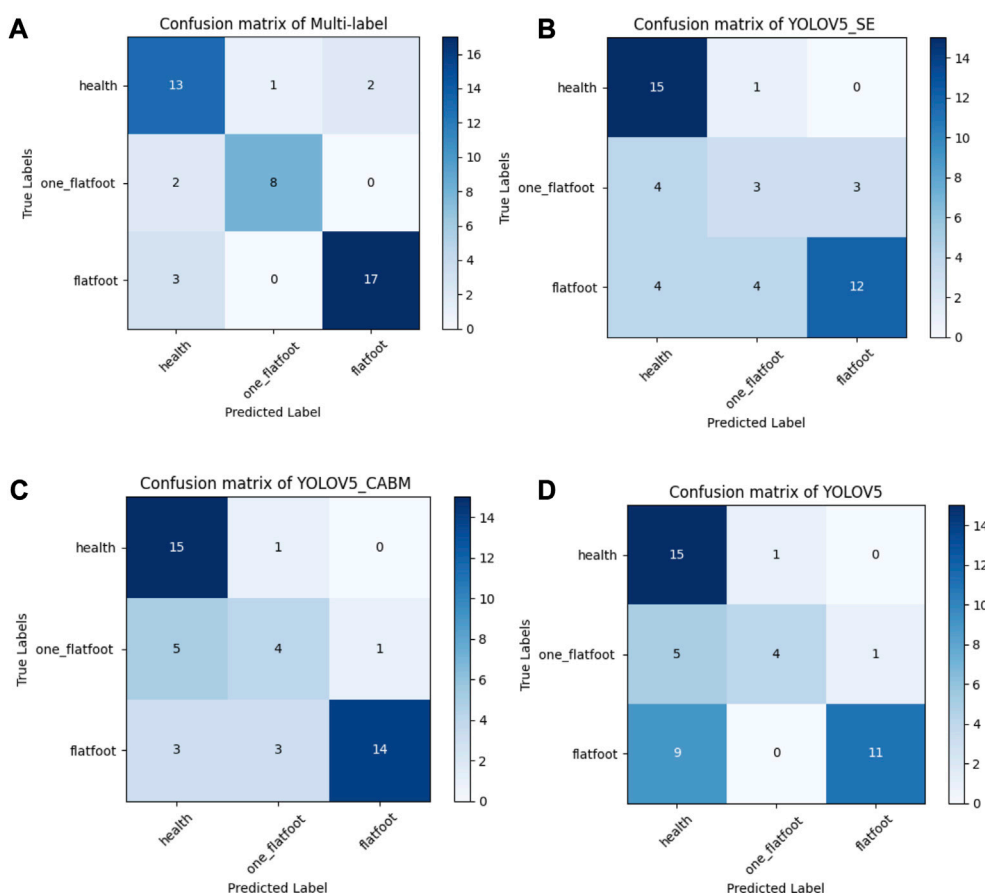


FIGURE 8

The confusion matrix of four different models. (A) For the confusion matrix of multilabel task model, the sensitivity and specificity of flatfoot recognition were 89.5% and 83.3%, respectively, and the specificity of one-foot flat recognition was 97.2%. (B) For the confusion matrix of YOLO-v5_SE, the sensitivity and specificity of flatfoot recognition were 100% and 68%, respectively, and the specificity of one-foot flat recognition was 96.9%. (C) For the confusion matrix of YOLO-v5_CBAM, the sensitivity and specificity of flatfoot recognition were 100% and 73.9%, respectively, and the specificity of one-foot flat recognition was 97%. (D) For the confusion matrix of YOLO-v5, the sensitivity and specificity of flatfoot recognition were 100% and 65.4%, respectively, and the specificity of one-foot flat recognition was 97.2%.

and confidence prediction of target position; this further improves the performance of YOLO-v1 (Redmon and Farhadi, 2017). Nevertheless, during target detection, the regression parameters of the center position of the prior anchor frame in YOLO-v2 are not effectively constrained, which may cause the prediction anchor frame to appear in any position of the original image, resulting in low prediction accuracy. Furthermore, SoftMax activation classification is applicable only to a single target and not for multiple classifications. As such, logistic regression is incorporated in YOLO-v3 for parameters in the central location, limiting them to the range of 0–1, thereby greatly improving the detection accuracy. Furthermore, multiple independent logistic regression classifiers are used to replace SoftMax to improve the accuracy of multiobjective classification (Redmon and Farhadi, 2018). Based on YOLO-v3, YOLO-v4 incorporates advanced techniques such as Bag-of-Freebies (data enhancement, regularization, and loss function improvement) and Bag-of-Specials (enhancement of model sensitivity field, introduction of attention mechanism, feature integration, and post-processing method) methods to further

improve detection efficiency (Bochkovskiy et al., 2020). The prior anchor frame of YOLO-v5 is similar to that of YOLO-v3 and YOLO-v4; however, YOLO-v5 incorporates the training prediction anchor frame into the network (Wang et al., 2022). During training, the optimal anchor frame value of different training sets is calculated adaptively, making the model applicable to various datasets and greatly improving the positioning accuracy for different tasks and datasets.

In this study, we also addressed the phenomenon of unilateral arch collapse observed in some flat feet patients as it can lead to asymmetrical force distribution and potential complications (Zhu et al., 2021; Kim and Lee, 2022). To address this issue, we adjusted the task strategy by focusing on the recognition of the foot arch and employed multilabel classification using ResNet-50, which yielded good results. Thermal maps generated using Grad-CAM revealed the importance of force distribution in the metatarsal region in addition to the midfoot. This is in agreement with previous studies (Szczepanowska-Wołowiec et al., 2021).

Common abnormal foot types include flat foot and high-arched foot. However, Buldt et al. (2018) observed no significant difference in plantar pressure between high arch feet

and normal feet. Therefore, to avoid introducing bias to the model, we focused on the plantar pressure distribution of flat and normal feet in this study. Flat feet can lead to various health problems and affect daily functioning and quality of life (Chou et al., 2021; 2021; Moon and Jung, 2021). However, the latent early clinical manifestations of flat feet are often overlooked, and traditional diagnostic methods such as X-rays are less motivating for patients due to radiation and cost concerns. The use of plantar pressure measurement systems provides a simple screening method for flat feet but has limitations in terms of accuracy and sensitivity. In this study, we proposed a DL model and a multilabel classification algorithm as a novel and more accurate approach for foot type discrimination.

We observed differences in plantar pressure distribution between flat and healthy feet but no statistically significant difference in pressure distribution on the heel. This can be attributed to long-term compensatory or equilibrium mechanisms in the lower limbs of flat feet patients. The small number of participants included in this study posed challenges in model prediction, necessitating the need for further investigations to explore the relationship between these conditions and the subjects' mental or physical health and to improve data collection methods.

Limitations of the study

Limited datasets: In this study, data collected from a single healthcare facility by using specific equipment and evaluators were used. This limits the generalizability of the training model to datasets from different centers, devices, and evaluators. In addition, high arched feet were not included in this study, and a model suitable for high arched feet needs to be developed. To prevent overfitting and ensure robustness, future studies should include larger and more diverse datasets from multiple centers and devices.

Conclusion

In this study, we clinically validated that the multilabel classification task and the improved YOLO-v5 model improve the performance of the plantar flat press system in foot classification. The multilabel classification task achieved higher accuracy. The pressure ratios of all plantar regions contribute to foot-type recognition, not limited to the arch of the foot. The multilabel classification algorithm based on ResNet-50 is suitable for foot diagnosis and treatment prognosis.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Member of Ethics Committee of Zhongshan Hospital Affiliated to Xiamen University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

Conception and design of the study: YZ, JZ, and FQ. Collection of data: YZ, JZ, FQ, JJ, and XiL. Analysis and interpretation of data: YZ, JZ, FQ, XuL, and HC. Drafting of the manuscript: YZ, JZ, FQ, XuL, JJ, HC, XiL, and JH. Critical revision of the manuscript for important intellectual content: YH, JH, and JC. Administrative support and study supervision: YH, JH, and JC. All authors revised the final draft. All authors read and approved the final manuscript.

Funding

This study was funded by National Natural Science Foundation of China (82004433), Science Foundation of Fujian Province in China (2021J011327), and Key Clinical Specialty Discipline Construction Program of Fujian, P.R.C.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2023.1239246/full#supplementary-material>

References

- Adoración Villarroya, M., Manuel Esquivel, J., Tomás, C., Buenafé, A., and Moreno, L. (2008). Foot structure in overweight and obese children. *Int. J. Pediatr. Obes.* 3, 39–45. doi:10.1080/17477160701520298
- Albahli, S., Nida, N., Irtaza, A., Yousaf, M. H., and Mahmood, M. T. (2020). Melanoma lesion detection and segmentation using YOLOv4-DarkNet and active contour. *IEEE Access* 8, 198403–198414. doi:10.1109/ACCESS.2020.3035345
- Arunakul, M., Amendola, A., Gao, Y., Goetz, J. E., Femino, J. E., and Phisitkul, P. (2013). Tripod index: A new radiographic parameter assessing foot alignment. *Foot Ankle Int.* 34, 1411–1420. doi:10.1177/1071100713488761
- Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). *YOLOv4: Optimal speed and accuracy of object detection*. doi:10.48550/arXiv.2004.10934
- Buldt, A. K., Forghany, S., Landorf, K. B., Levinger, P., Murley, G. S., and Menz, H. B. (2018). Foot posture is associated with plantar pressure during gait: A comparison of normal, planus and cavus feet. *Gait Posture* 62, 235–240. doi:10.1016/j.gaitpost.2018.03.005
- Caravaggi, P., Rogati, G., Leardini, A., Ortolani, M., Barbieri, M., Spasiano, C., et al. (2021). Accuracy and correlation between skin-marker based and radiographic measurements of medial longitudinal arch deformation. *J. Biomech.* 128, 110711. doi:10.1016/j.jbiomech.2021.110711
- Carr, J. B., Yang, S., and Lather, L. A. (2016). Pediatric pes planus: A state-of-the-art review. *Pediatrics* 137, e20151230. doi:10.1542/peds.2015-1230
- Chang, C.-H., Chen, Y.-C., Yang, W.-T., Ho, P.-C., Hwang, A.-W., Chen, C.-H., et al. (2014). Flatfoot diagnosis by a unique bimodal distribution of footprint index in children. *PLoS One* 9, e115808. doi:10.1371/journal.pone.0115808
- Chou, M.-C., Huang, J.-Y., Hung, Y.-M., Perng, W.-T., Chang, R., and Wei, J. C.-C. (2021). Flat foot and spinal degeneration: evidence from nationwide population-based cohort study. *J. Formos. Med. Assoc.* 120, 1897–1906. doi:10.1016/j.jfma.2020.12.019
- Deng, Y., Wu, H., and Zhu, H. (2020). Recognition and counting of citrus flowers based on instance segmentation. *Trans. Chin. Soc. Agric. Eng.* 36, 200–207.
- Drefus, L. C., Kedem, P., Mangan, S. M., Scher, D. M., and Hillstrom, H. J. (2017). Reliability of the arch height index as a measure of foot structure in children. *Pediatr. Phys. Ther.* 29, 83–88. doi:10.1097/PEP.0000000000000337
- Flores, D. V., Mejía Gómez, C., Fernández Hernando, M., Davis, M. A., and Pathria, M. N. (2019). Adult acquired flatfoot deformity: anatomy, Biomechanics, staging, and imaging findings. *Radiographics* 39, 1437–1460. doi:10.1148/rg.2019190046
- Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-Excitation networks. Available at: https://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html (Accessed May 23, 2023).
- Izraelski, J. (2013). Human locomotion: the conservative management of gait-related disorders. *J. Can. Chiropr. Assoc.* 57, 366.
- Jiang, L., Chen, D., Cao, Z., Wu, F., Zhu, H., and Zhu, F. (2022). A two-stage deep learning architecture for radiographic staging of periodontal bone loss. *BMC Oral Health* 22, 106. doi:10.1186/s12903-022-02119-z
- Kim, J., and Lee, W.-C. (2022). Joint preservation surgery for varus and posterior ankle arthritis associated with flatfoot deformity. *Foot Ankle Clin.* 27, 115–127. doi:10.1016/j.fcl.2021.11.010
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi:10.1038/nature14539
- Li, J., Li, S., Li, X., Miao, S., Dong, C., Gao, C., et al. (2023). Primary bone tumor detection and classification in full-field bone radiographs via YOLO deep learning model. *Eur. Radiol.* 33, 4237–4248. doi:10.1007/s00330-022-09289-y
- Lin, P., Lee, W. S., Chen, Y. M., Peres, N., and Fraise, C. (2020). A deep-level region-based visual representation architecture for detecting strawberry flowers in an outdoor field. *Precis. Agric.* 21, 387–402. doi:10.1007/s11119-019-09673-7
- Loshchilov, I., and Hutter, F. (2019). *Decoupled weight decay regularization*. doi:10.48550/arXiv.1711.05101
- McCormack, A. P., Ching, R. P., and Sangeorzan, B. J. (2001). Biomechanics of procedures used in adult flatfoot deformity. *Foot Ankle Clin.* 6, 15–23. doi:10.1016/s1083-7515(03)00072-x
- Meehan, R. E., and Brage, M. (2003). Adult acquired flat foot deformity: clinical and radiographic examination. *Foot Ankle Clin.* 8, 431–452. doi:10.1016/S1083-7515(03)00019-6
- Moore, D., and Jung, J. (2021). Effect of incorporating short-foot exercises in the balance rehabilitation of flat foot: A randomized controlled trial. *Healthc. (Basel)* 9, 1358. doi:10.3390/healthcare9101358
- Nikolaïdou, M. E., and Boudolos, K. D. (2006). A footprint-based approach for the rational classification of foot types in young schoolchildren. *Foot* 16, 82–90. doi:10.1016/j.foot.2006.02.001
- O’Leary, C. B., Cahill, C. R., Robinson, A. W., Barnes, M. J., and Hong, J. (2013). A systematic review: the effects of podiatric deviations on nonspecific chronic low back pain. *J. Back Musculoskelet. Rehabil.* 26, 117–123. doi:10.3233/BMR-130367
- Persiane, A. S., Negrão, D. M. G., Alves, R. D. P., Freitas, D. G. D., Cazarini, C., and Alves, V. L. D. S. (2021). Subtalar joint in neutral and relaxed positions for evaluation of medial longitudinal arch. *Acta Ortop. Bras.* 29, 177–180. doi:10.1590/1413-785220212904240893
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only Look once: unified, real-time object detection. Available at: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html (Accessed August 5, 2023).
- Redmon, J., and Farhadi, A. (2017). YOLO9000: better, faster, stronger. Available at: https://openaccess.thecvf.com/content_cvpr_2017/html/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.html (Accessed August 5, 2023).
- Redmon, J., and Farhadi, A. (2018). *YOLOv3: An incremental improvement*. doi:10.48550/arXiv.1804.02767
- Su, S., Mo, Z., Guo, J., and Fan, Y. (2017). The effect of arch height and material hardness of personalized insole on correction and tissues of flatfoot. *J. Healthc. Eng.* 2017, 8614341–8614349. doi:10.1155/2017/8614341
- Szczepanowska-Wołowiec, B., Sztandera, P., Kotela, I., and Zak, M. (2021). Assessment of the foot’s longitudinal arch by different indicators and their correlation with the foot loading paradigm in school-aged children: A cross sectional study. *Int. J. Environ. Res. Public Health* 18, 5196. doi:10.3390/ijerph18105196
- Ueki, Y., Sakuma, E., and Wada, I. (2019). Pathology and management of flexible flat foot in children. *J. Orthop. Sci.* 24, 9–13. doi:10.1016/j.jos.2018.09.018
- Vogt, B., Gosheger, G., Wirth, T., Horn, J., and Rödl, R. (2020). Leg length discrepancy- treatment indications and strategies. *Dtsch. Arztebl Int.* 117, 405–411. doi:10.3238/arztebl.2020.0405
- Wang, D., Zhang, D., Yang, G., Xu, B., Luo, Y., and Yang, X. (2022). SSRNet: in-field counting wheat ears using multi-stage convolutional neural network. *IEEE Trans. Geoscience Remote Sens.* 60, 1–11. doi:10.1109/TGRS.2021.3093041
- Welte, L., Holowka, N. B., Kelly, L. A., Arndt, A., and Rainbow, M. J. (2023). Mobility of the human foot’s medial arch helps enable upright bipedal locomotion. *Front. Bioeng. Biotechnol.* 11, 1155439. doi:10.3389/fbioe.2023.1155439
- Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). Cbam: convolutional block attention module. Available at: https://openaccess.thecvf.com/content_ECCV_2018/html/Sanghyun_Woo_Convolutional_Block_Attention_ECCV_2018_paper.html (Accessed May 23, 2023).
- Yu, B., Yu, H., Meng, Q., Meng, Q., and Cao, W. (2019). Study on gait symmetry based on simulation and evaluation system of prosthesis gait. *J. Biomed. Eng.* 36, 924–929. doi:10.7507/1001-5515.201811038
- Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. (2018). *mixup: Beyond empirical risk minimization*. doi:10.48550/arXiv.1710.09412
- Zhao, J. L., Zhang, X. Z., and Dong, H. Y. (2022). Defect detection in transmission line based on scale-invariant feature pyramid networks. *Comput.* 58, 289–296.
- Zheng, H., Wang, G., and Li, X. (2022). YOLOX-Dense-CT: A detection algorithm for cherry tomatoes based on YOLOX and DenseNet. *Food Meas.* 16, 4788–4799. doi:10.1007/s11694-022-01553-5
- Zhu, F., Hong, Q., Guo, X., Wang, D., Chen, J., Zhu, Q., et al. (2021). A comparison of foot posture and walking performance in patients with mild, moderate, and severe adolescent idiopathic scoliosis. *PLOS ONE* 16, e0251592. doi:10.1371/journal.pone.0251592