

RESEARCH ARTICLE

Artificial Intelligence–Based Mobile Application for Identifying Suitable Height Range of High Heels

SI-HUEI LEE^{1,2}, SHIN-WEI CHIU³, MAN-WEN SU³, SZU-YU WU³, YI-TING HUANG³, LI-YIN KANG³, AND BOR-SHING LIN³, (Senior Member, IEEE)

¹Department of Physical Medicine and Rehabilitation, Taipei Veterans General Hospital, Taipei 112, Taiwan

²Faculty of Medicine, National Yang Ming Chiao Tung University, Taipei 112, Taiwan

³Department of Computer Science and Information Engineering, National Taipei University, New Taipei City 237, Taiwan

Corresponding author: Bor-Shing Lin (bslin@mail.ntpu.edu.tw)

This work was supported in part by the National Science and Technology Council, Taiwan, under Grant NSTC 112-2221-E-305-001-MY3; in part by the University System of Taipei Joint Research Program under Grant USTP-NTPU-NTOU-111-01 and Grant USTP-NTPU-NTOU-112-01; in part by the Faculty Group Research Funding Sponsorship by the National Taipei University under Grant 2021-NTPU-ORDA-02 and Grant 2023-NTPU-ORD-01; and in part by the Taipei Veterans General Hospital, Taiwan, under Grant V111C-122.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of En Chu Kong Hospital under Approval No. ECKIRB1090404.


ABSTRACT The wearing of overly high high-heeled shoes can cause irreversible physiological damage. Doctors typically analyze the calcaneus when an individual is standing on tiptoes to observe changes in the Achilles tendon to determine the suitable high-heel height range. A mobile application system for evaluating the maximum heel height suitable for women was developed in this study. The system input is a smartphone-recorded video of an individual's while standing on tiptoes. The trained artificial intelligence (AI) evaluation system outputs the maximum heel height suitable for the wearer. This study used three core technologies to implement this AI evaluation system. First, EfficientDet was used to realize object detection, and area of the left heel, right heel, and a red cube were bounded. These bounding region pictures were then input into the convolutional neural network (CNN) model for binary classification; the accuracy rate of the integrated EfficientDet and CNN models was 0.86. Finally, the mode algorithm was applied to correct errors in model evaluation and improve overall accuracy. The system developed in this study can evaluate the maximum heel height of a high-heeled shoe with a mean absolute error of 0.8 cm. Only the operation of a smartphone is required without additional platform assistance. Thus, this system has high practicability and usability.

INDEX TERMS Height of high heel, artificial intelligence (AI), EfficientDet, convolutional neural network (CNN), mobile application.

I. INTRODUCTION

For modern women, high-heeled shoes are a popular fashion item that can enhance how long legs appear and improve the appearance of body proportions. However, wearing overly high heels can cause irreversible physiological damage. Many studies have reported that the wearing of overly high high-heeled shoes and the long-term wearing of high

heels adversely affect the human foot [1]. Studies have also demonstrated that long-term wearers of high-heeled shoes experience a considerable adverse effect when walking and are prone to muscle fatigue after a short period of time. When crossing over obstacles at heights up to 30% of the length of their leg, high-heel wearers experience a stronger impact than nonhigh-heel wearers, and their thrust to lift the foot is insufficient. Additionally, they are more likely to incur muscle injury or trip over obstacles [2]. In 2016, Karia et al. explored differences in plantar pressure when walking in flat

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno M. Garcia .

shoes and several types of high heels [3]. The sole of the foot was classified as the front, middle, or back, and the pressure changes were observed in each part. The results indicated that forefoot pressure increased when the heel height increased, resulting in a decrease in wearer comfort.

Most other studies have focused on the effect of high heels on human gait, joint mechanics, and balance control [4] or on statistically analyzing the distribution of plantar pressure when wearing high heels [5]. Some studies have developed high-heeled shoes for which the heel height that can automatically or manually be adjusted, enabling users to independently identify a suitable heel height. However, the definition of an appropriate height remains unclear, and these devices cannot offer objective and quantitative evaluations for users [6]. In 2021, Lee et al. proposed the first system that uses artificial intelligence (AI) to evaluate a suitable heel height for women's high-heeled shoes [7]. This system can quickly and accurately identify the most suitable high-heel height for a user and achieved an overall mean absolute error (MAE) of only 1.21 cm. However, the actual use of this system has some limitations. For example, the user's foot length must be between 22 and 26 cm, and the evaluation must be conducted using a platform specially designed for this research. However, this platform is large and costly and, in addition to the operational inconvenience, this system is not widely available on the market. To improve the system's applicability, a mobile app evaluation system that uses AI to determine the appropriate range of high heels was proposed. This system can be independently operated by the user. With a smartphone, users record videos of themselves gradually raising their rearfoot while standing on tiptoes and upload these videos to the server through the app for AI model-based prediction. Predictions are generated and provided to users on their smartphones to reveal the most suitable heel height for them, so that when choosing high heels, they can avoid choosing unsuitable high heels, so as not to affect their

personal health. The major contributions in this study are listed as follows: 1) This system can operate independently. It only requires a mobile app to predict the appropriate height of high heels for women without the need for additional equipment or assistance from other people. 2) This system can save medical resources. It usually takes an experienced physician 15 to 20 min to diagnose a woman's appropriate heel height, but our system's assessment process can be completed in 5 min. The remaining of this paper is organized as follows. Section II introduces the software and hardware architecture and AI model design. The process of collecting data is introduced in Section III. Section IV presents the results of the experiments. Discussions are conducted in Section V. Section VI concludes this paper.

II. METHODS

A. OVERALL SYSTEM ARCHITECTURE

The overall architecture of this system is depicted in Fig. 1. The software architecture consists of five modules, namely mobile application, image processing, automatic bounding, heel height judgment, and suitable heel height estimation modules. Except for the mobile app running on a smartphone, the other four modules run on a server with a high level of computing power. The app has a graphical user interface used to collect and upload the foot image data required for AI model prediction and return the user's maximum wearable heel height. The image processing module converts the collected videos into numerous photos at a sampling rate of 30 frames per second (FPS). In the training phase, the system feeds the photos into the automatic bounding and heel height judgment modules for separate training. Once trained, these two modules can be used in the prediction phase. In the prediction phase, the videos transmitted from the smartphone are first converted one by one into photos using the image processing module and then are fed into the trained automatic bounding and heel height judgment modules for separate

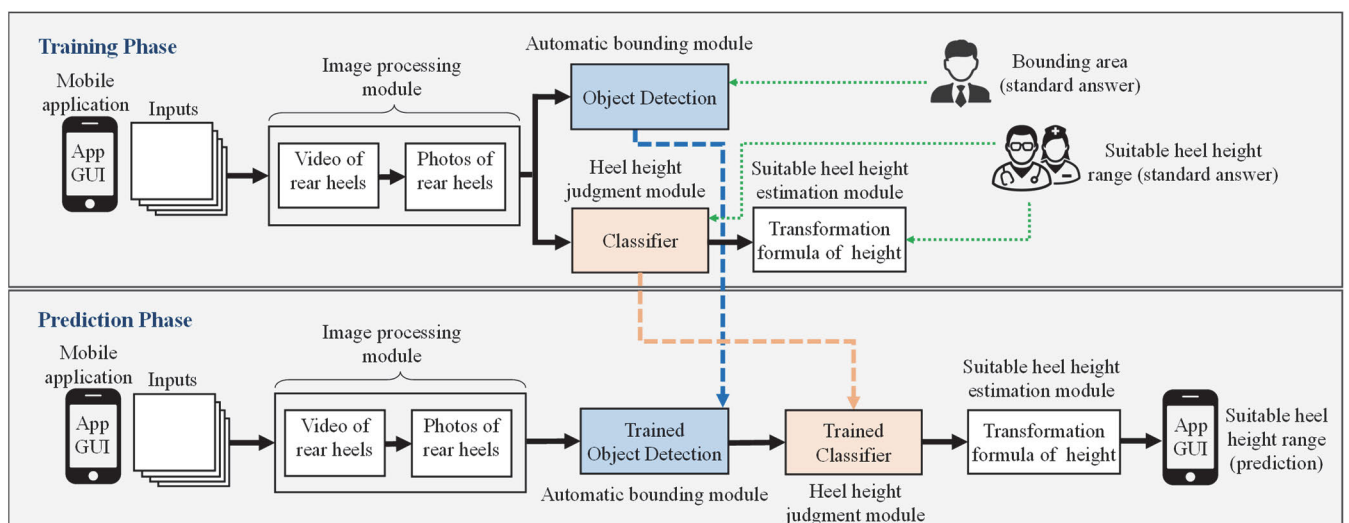


FIGURE 1. Overall architecture of the system.

processing. The automatic bounding and heel height judgment modules are both based on deep learning algorithms. The system then bounds the left- and right-foot areas using automatic bounding module and classifies the heel heights as wearable or nonwearable using heel height judgment module. Finally, the suitable heel height estimation module identifies the critical point of the wearable and nonwearable photos for the classified photos, and a built-in transformation formula of height is used to obtain the maximum wearable height.

B. MOBILE APP

The mobile app was developed on Android Studio, an integrated development environment for Android platform development programs released by Google; this software uses Java as the development language and operates with Android 4.1 Jelly Bean (application programming interface level 16). The app interface diagram is illustrated in Fig. 2. Presented on the interface is the video file in the app storage path and three operational buttons, namely the “Instructions,” camera, and upload buttons.

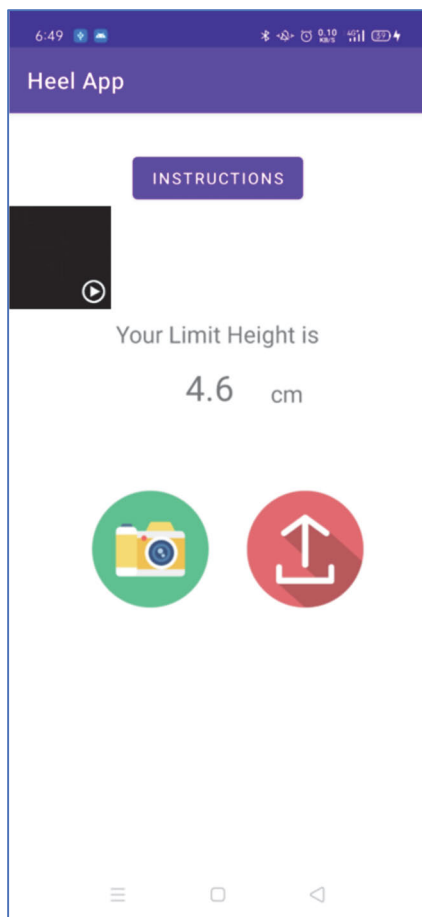


FIGURE 2. App user interface and returned results.

The application process of this app is as follows. First, users click the “Instructions” button to play a prerecorded demonstration video on YouTube to help them understand the operating process. Users then click the camera icon button to

turn on the built-in camera, allowing them to record videos. The recorded video is stored in the set specific path and is presented in the middle of the interface. After the user hears the prompt sound, she starts to raise her heel off the ground slowly, the app will simultaneously record the rear heels lifting as a video. Users select the video they wish to upload and click the upload button; the recorded video is then uploaded to the server. The server-side program predicts the maximum suitable heel height, and the prediction result is returned and displayed on the user’s app interface.

C. IMAGE PROCESSING MODULE

Videos uploaded from smartphones to the server must be processed into photos before they can be used as input data for the AI models. This system employs the Python OpenCV package (version 4.5.1.48) and cuts the uploaded video with a cutting frequency of 30 FPS. The server used in this study consisted of of a personal computer equipped with an Intel Core i7-4.8 GHz CPU, 64 GB DRAM, graphics processing unit card (GeForce RTX 3080, ASUSTeK Computer, Taipei City, Taiwan), and Windows 10 operating system.

D. AUTOMATIC BOUNDING MODULE

In this study, the EfficientDet object detection model [8] was used to realize automatic bounding. Fig. 3 presents the architectural diagram of EfficientDet. EfficientDet is an extension of EfficientNet and uses a new feature fusion method, weighted bidirectional feature pyramid network (BiFPN), and ImageNet pretrained EfficientNet-B0 as the base. After the BiFPN obtains features from the base, it performs two-way feature fusion from top to bottom and bottom to top, inputs the fused features into the class/box prediction network, and simultaneously predicts the bounding area and classification category. In the training phase of the model, we manually label each photo after image processing; the bounding boxes of the left heel, right heel, and a red cube box with a 5-cm side length are used by the model’s correct answer. The trained model has the function of automatically bounding out three different objects. Therefore, in the prediction stage of the model, we no longer input photos that have been manually tagged but use photos directly generated by the image processing module as inputs. The outputs are the bounding region pictures of the posterior left and right heels and the red cube area with a side length of 5 cm.

In addition, this module also contains a height conversion program, which can use the ratio of the red cube as a scale and bounding box of the foot to calculate the actual height of the heel from the ground. According to the preprocessing conversion of the video to pictures and EfficientDet model, we obtain k sets of the bounding coordinates of the left foot, right foot, and red cube. The bounding coordinates include the upper left corner and lower right corner coordinates of the bounding box. First, the Y coordinates of the lower right corners are subtracted from those of the upper left corners of the bounding coordinates of the k red cubes, and the absolute value is obtained. The heights of the k red boxes in pixels

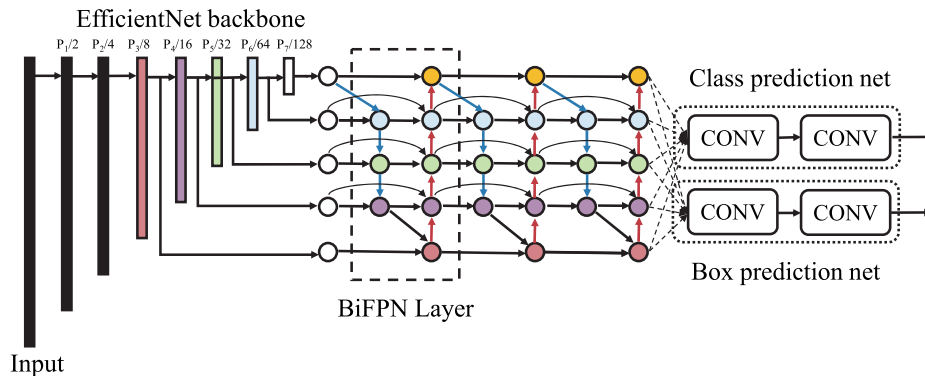


FIGURE 3. EfficientDet architecture diagram [8].

are calculated, and these heights are divided by k to obtain the average height of the red cube in pixels. The formula is written as follows:

$$\begin{aligned} & \text{Average height of the red cube in the photo (pixel)} \\ &= \frac{\sum_{t=1}^k |YRD_B(t) - YLU_B(t)|}{k}, \end{aligned} \quad (1)$$

where $YRD_B(t)$ and $YLU_B(t)$ represent the Y coordinates of the lower right and upper left corner, respectively, of the bounding coordinate of the t -th red cube, and k represents the k -th red cube.

The Y coordinate of the lower right corner of the right-foot bounding coordinates of the n -th picture is then subtracted from that of the lower right corner of the right-foot bounding coordinates of the first picture to obtain the raised height of the posterior heel in the n -th picture relative to the first picture (the units are pixels). After dividing this value by the average height (pixels) of the red cube, we can obtain the ratio of the height of the posterior heel off the ground of the n -th image relative to the first image to the average height of the red cube. Multiplying this ratio by the known true height of the red cube and by 5 cm, the real height of the posterior heel off the ground of the n -th picture relative to the first picture can be obtained. The same calculations are applied for the left foot, and are detailed in (2)–(5).

Height of the right heel off the ground of the n -th photo:

$$\begin{aligned} & \text{nth Right heel height}_{\text{pixel}} \\ &= |YRD_R(n) - YRD_R(1)| = |YLU_R(n) - YLU_R(1)| \end{aligned} \quad (2)$$

Actual height of the right heel off the ground:

$$\begin{aligned} & \text{nth Right heel height}_{\text{cm}} \\ &= \frac{\text{nth Right heel height}_{\text{pixel}}}{\text{Average cube height}_{\text{pixel}}} \times 5 \end{aligned} \quad (3)$$

Height of left heel off the ground of the n -th photo:

$$\begin{aligned} & \text{nth Left heel height}_{\text{pixel}} = \\ & |YRD_L(n) - YRD_L(1)| = |YLU_L(n) - YLU_L(1)| \end{aligned} \quad (4)$$

Actual height of the left heel off the ground:

$$\begin{aligned} & \text{nth Left heel height}_{\text{cm}} \\ &= \frac{\text{nth Left heel height}_{\text{pixel}}}{\text{Average cube height}_{\text{pixel}}} \times 5, \end{aligned} \quad (5)$$

where k is the number of photos of a video after cutting; n is the n -th picture in the k pictures, and the range is 1 to k ; (XLU_L, YLU_L) and (XRD_L, YRD_L) represent the XY coordinates of the upper left and lower right corner, respectively, of the bounding box of the left foot; and (XLU_R, YLU_R) and (XRD_R, YRD_R) represent the XY coordinates of the upper left and lower right corner, respectively, of the bounding box of the right foot. The bounding image is labeled as a left- or right-foot image and is fed into the corresponding heel height judgment module for prediction.

E. HEEL HEIGHT JUDGMENT MODULE

In this study, Keras 2.4.0 [9] was used a convolutional neural network (CNN) [10] to realize the heel height judgment module, which incorporates the maximum heel height provided by the physician as the correct answer for learning and classification to determine whether an image represents a wearable or nonwearable heel. The CNN architecture of the heel height judgment module is presented in Fig. 4. The CNN has a kernel size of 3×3 , pool size of 2×2 , and dropout ratio of 0.25. First, the bounding images of the left and right heels are fed into their respective CNN models for left- and right-foot predictions. After three layers of convolution, dropout, and maxpooling, the features are flattened and input into three fully connected layers for training and weight assignment. Finally, the model classifies the image as 1 or 0, which indicates a wearable or nonwearable height, respectively. This prediction module also uses the AI model MobileNetV2 for experiments. Compared with a traditional CNN, MobileNetV2 is a lightweight model more suitable for mobile terminals. The MobileNetV2 architecture used in this study has a kernel size of 3×3 and pool size of 2×2 ; the detailed architecture is illustrated in Fig. 5. Its operational mode is the same as that of the CNN model. First, the bounding left- and right-heel images are fed into their respective MobileNetV2 left- and right-foot-prediction models.

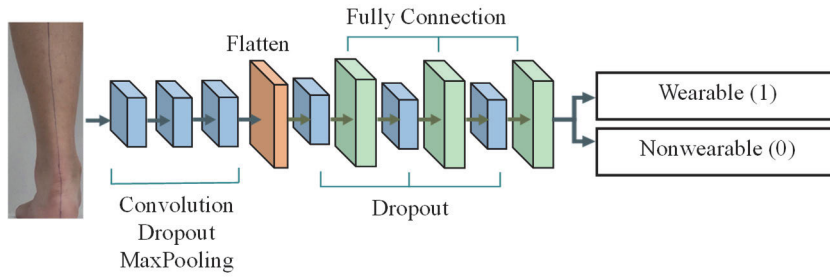


FIGURE 4. CNN architecture diagram.

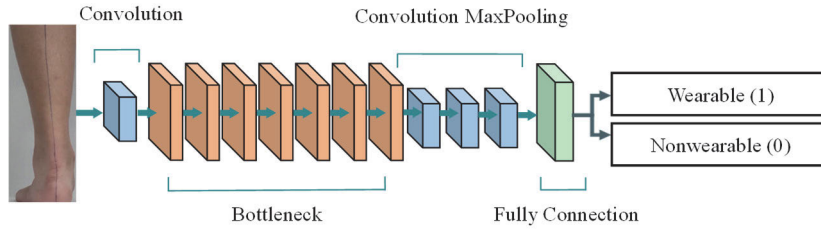


FIGURE 5. MobileNetV2 architecture diagram.

After one layer of convolution, seven layers of bottleneck, three layers of convolution and maxpooling, the images enter the fully connected layer for training and weight assignment. Finally, the model classifies the image as 1 or 0, indicating a wearable or nonwearable height, respectively. Because the system is designed to operate on smartphones, we compared the performance of a basic, common AI model with another more suitable for smartphones (i.e., MobileNetV2). In reference to the evaluation system developed by Lee et al. [7], the CNN was selected as the basic AI model, and the judgment results of the two models were evaluated.

In the training phase of the model, we manually labeled each photo after image processing and framed and cut out the left and right heel and the red cube box with a 5-cm side length to serve as the correct answer for this model.

The trained model has a classification function for judging the heel height obtained from an input image as either wearable or nonwearable. In the model prediction phase, we used the EfficientDet model trained in the automatic bounding module and used the prediction result output of the Efficient-Det model as the input for the heel height judgment module.

F. SUITABLE HEEL HEIGHT ESTIMATION MODULE

After being evaluated in the heel height judgment module, each picture can be assigned one of two values: 1 (wearable) or 0 (nonwearable). All pictures' values from the same video form a series that can be obtained through the suitable heel height estimation module to generate a numerical fault. For example, the 0/1 tomogram detailed in Fig. 6 can be used to determine that the batch of pictures showed a maximum heel height in the 50th photo. The critical images between the classification results of 1 and 0 can then be removed. Through the transformation formula of height incorporated into the

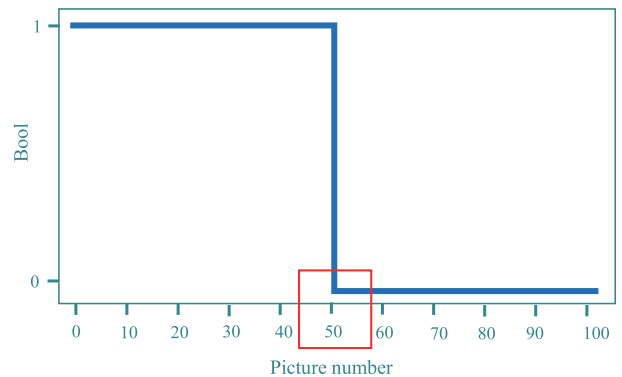


FIGURE 6. 0/1 tomogram and schematic of maximum heel height prediction.

EfficientDet model, the maximum wearable heel height can be calculated and returned to the user—displayed on their mobile app, as illustrated in Fig. 2.

Because the accuracy rate of the heel height judgment module is less than 100%, misjudgment may occur. Therefore, this module corrects the output results of the heel height judgment module to increase the accuracy of maximum heel height prediction. Fig. 7 illustrates the operation of the suitable heel height estimation module. Fig. 7(a) depicts a heel height misjudgment, with Fig. 7(b) presenting the corrected result. In Fig. 7(a) and (b), 0 and 1 on the vertical axis represent wearable and nonwearable, respectively, and the horizontal axis represents the picture number. As illustrated in Fig. 7(a), because misjudgment does not occur continually, the mode algorithm can thus applied to correct each data item and the four adjacent data items. For example, in the batch of photos depicted in Fig. 7(a), a single 0 appears in the 30th photo, and the adjacent photos are all 1; therefore, the 28th to 32nd photos in the red box from (1, 1, 0, 1, 1) are corrected to

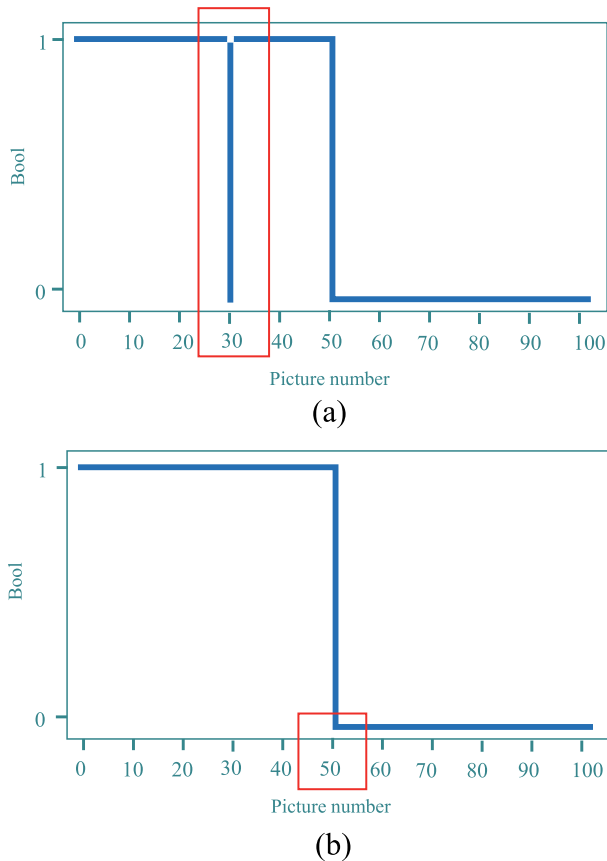


FIGURE 7. Schematic of the mode algorithm used in the suitable heel height estimation module: (a) AI model prediction results and revision process using the mode algorithm; (b) revised results.

(1, 1, 1, 1, 1) through the module, as presented in Fig. 7(b). The serial number of the photo that can be used to obtain the maximum heel height was corrected from 30 to 50. With the mode algorithm, the misjudgment of the AI model can be corrected and the overall accuracy of the system improved.

III. EXPERIMENTAL DESIGN

A. PARTICIPANT SCREENING CRITERIA

The experimental procedure was approved by the Institutional Review Board of En Chu Kong Hospital (Code: ECKIRB1090404), and each participant provided written informed consent before participating in the experiment. We recruited 100 women whose feet were in a healthy condition; participants' average age was 21.3 years. When screening for foot conditions, the participants stood in a natural position, with their hands hanging naturally by their thighs. The inside of their feet were parallel, with the heels of their feet aligned for the screening. The filtering criteria included the following points:

- The arch rate was between 11% and 14% (arch rate = the height from the most convex point of the scaphoid bone to the ground/foot length · 100%).
- Calcaneus varus or valgus angle of rearfoot was less than or equal to 4°.
- Hallux valgus angle was less than or equal to 20°.

- Quintus varus angle was less than or equal to 20°.
- No Unexplained scoliosis of the spinal cord.
- No calluses on the soles of the feet.
- No history of serious lower limb injuries.
- Aged from 20 to 35 years.

B. ASSESSMENT OF SUITABLE HEEL HEIGHT RANGE

When training the AI model, a rehabilitation physician must evaluate the height limit of each participant's high heels, which serves as the standard answer of the model. For diagnostic evaluation, the physician first drew a line marking the midpoint of the Achilles tendon, ankle bone, and calcaneus on the lower leg of the participant, and then the subject stood in a natural stance on the evaluation platform developed by Lee et al. [6], as illustrated in Fig. 8. On the platform, the participant aligned the transverse arch of their foot at the turning point while the platform was rising. This platform allows the physician to adjust the lifting height by using an electric jack inside the platform to simulate different heel heights and identify the height at which the participant's calcaneus becomes inverted. Physicians estimate the critical height of high-heeled shoes using hand placed on subject's calcaneus three times, and average them to diagnose whether the participant's calcaneus is inverse during the continuous elevation of heels of the participant.



FIGURE 8. Physician operating the evaluation platform.

C. DATA COLLECTION PROCESS

During the data collection process, a smartphone (OPPO R17, OPPO Guangdong Mobile Communications, Dongguan, Guangdong, China) equipped with Android 10 was used to record the heel videos. The rear camera lens is 16 million pixels, and the aperture is F1.7. The camera lens of the smartphone was fixed in a position 73 cm away from the heel of the participant and was aligned with the midpoint of the foot to maintain the shooting angle of view and avoid excessive distortion of the recorded image data. In addition, on the outer side of the right foot and 15.5 cm from the midpoint of the line between the feet, a red cube box with a side length of 5 cm served as a scale for future heel height estimation. The photographed surface of the cube box must be aligned with the line of the two posterior heels. Participants stood upright, and maintained a steady rate to gradually raise their rear heels off the ground to the highest point possible. The changes in

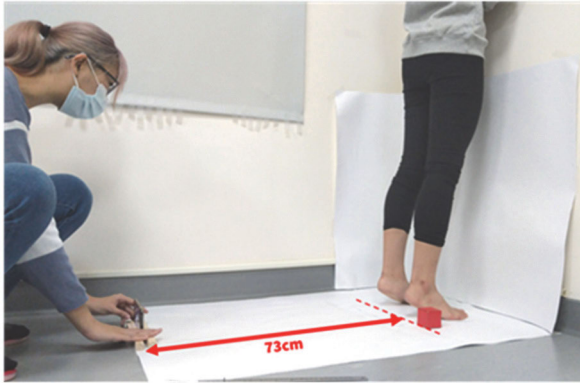


FIGURE 9. Videorecording of a participant raising their posterior heels off the ground.

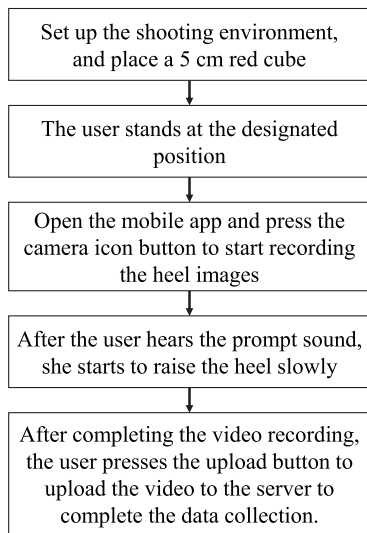


FIGURE 10. The flow chart of data acquisition process.

the bending angle of the heel bone and Achilles tendon were videorecorded. Fig. 9. depicts the actual filming of a video when the participant’s heels are raised off the ground.

When using a mobile app to predict the maximum height of high-heeled shoes in the future, the placement and location of the mobile phone and red cube box with a side length of 5 cm must be the same as the data collection environment presented in Fig. 9. The drawing of a line marking the posterior heel of the user during use is also necessary. The flow chart of data acquisition process is shown in Fig. 10.

IV. RESULTS

A. DATA DISTRIBUTION OF ALL PARTICIPANTS

We recruited 100 women whose feet were in a healthy condition. The average age, height, and weight of the participants were 21.29 ± 1.74 years, 160.61 ± 5.27 cm, and 51.86 ± 6.54 kg, respectively. Among participants, 78 women had a dominant right foot, and 22 women had a dominant left foot. After evaluation by the physician, the maximum heel heights that could be worn on the left and right feet of all participants approached a normal distribution, as illustrated in Fig. 11.

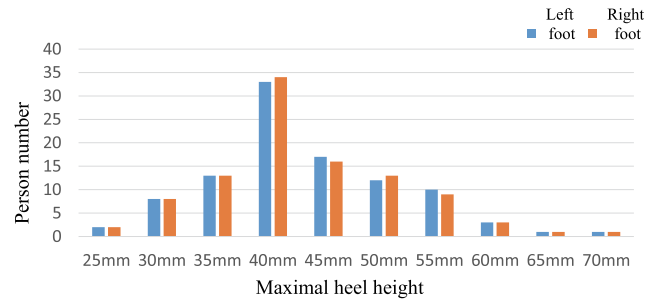


FIGURE 11. Distribution of the maximum heel height that can be worn on the left and right feet of all participants following the physician’s assessment.

B. PERFORMANCE EVALUATION RESULTS OF THE AUTOMATIC BOUNDING MODULE

In terms of automatic bounding, one EfficientDet model was used to simultaneously recognize the three objects, namely the two heels and cube box. Three EfficientDet models were also employed to detect the heels of the feet and the cube box separately. We compared the results of using one and three EfficientDet models, applying 10-fold cross-validation for evaluation and using average precision (AP) [11], [12] as the quantification standard. AP is a classic verification index in target detection. This evaluation index exhibits strong performance in evaluating the detection effect with satellite image datasets [13], and its formula is written as follows:

$$AP = \int_0^1 p(r) dr, \tag{6}$$

where $p(r)$ is the precision–recall curve [14] of each object, and r is the recall value.

Table 1 summarizes the performance comparison of the EfficientDet model for various detection objects. AP_{50} and AP_{70} in Table 1 represent the AP measurement values when the threshold of intersection over union (IoU) of the system prediction area and the real area are 0.5 and 0.7, respectively. AP_{AVE} is the overall average of ten values, AP_{50} , AP_{55} , ... AP_{90} , and AP_{95} . The calculation formula of the IoU is expressed as follows:

$$IoU = \frac{|A \cap B|}{|A \cup B|}, \tag{7}$$

where A and B represent the predicted and real area of the object, respectively.

TABLE 1. Comparison of the performance of the EfficientDet model in bounding different objects.

Object	AP_{AVE}	AP_{50}	AP_{70}	AP_S	AP_M	AP_L
B	94.4	100	100	-100	-100	94.4
R	81.3	100	97.8	-100	-100	81.3
L	77.5	98.0	93.7	-100	-100	77.5
BRL	75.4	99.6	88.5	-100	-100	75.4

Note: The unit of value is %, and -100 indicates an object could not be detected.

AP_S , AP_M , and AP_L refer to the AP measurement values of detected targets with a pixel area smaller than 32^2 ,

TABLE 2. Average effect after training with the CNN and MobileNetV2 models paired with 10-fold cross-validation.

Model	Dataset	Accuracy	Loss	Sensitivity	Specificity	Precision	Recall	F1-score	AUC
CNN	Left	0.9853	0.0359	0.983	0.9868	0.9786	0.983	0.9807	0.9849
CNN	Right	0.983	0.0507	0.98	0.985	0.977	0.98	0.9784	0.9825
MobileNetV2	Left	0.9755	0.0662	0.9674	0.9805	0.9678	0.9674	0.9675	0.9974
MobileNetV2	Right	0.9731	0.0664	0.9660	0.9777	0.9657	0.9660	0.9659	0.9973

TABLE 3. Overall results after the automatic bounding and heel height judgment modules were connected in series and tested with the two heel datasets.

Dataset	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-score
Left	0.8482	0.7740	0.9199	0.8872	0.7740	0.8267
Right	0.8686	0.8685	0.8653	0.8798	0.8685	0.8740
Average	0.8584	0.8213	0.8926	0.8835	0.8213	0.8504

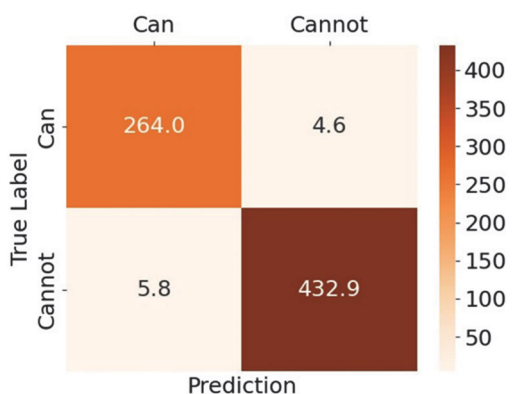


FIGURE 12. Confusion matrix of left feet dataset in CNN model.

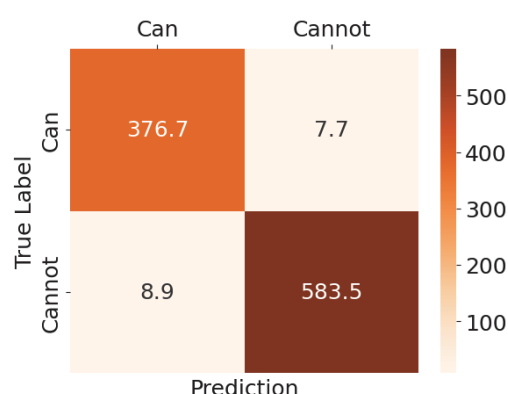


FIGURE 13. Confusion matrix of right feet dataset in CNN model.

ranging from 32^2 to 96^2 , and larger than 96^2 , respectively [15]. B , R , and L in Table 1 represent the cube box, right heel, and left heel, respectively. When one EfficientDet model recognized three objects at the same time, AP_{50} , AP_{70} , AP_S , AP_M , and AP_L were all lower than those for the other EfficientDet models that only detect one object. Therefore, we decided to use three EfficientDet models in the automatic bounding module to detect the right heel, left heel, and cube box separately.

C. PERFORMANCE EVALUATION RESULTS OF THE HEEL HEIGHT JUDGMENT MODULE

In the separate performance evaluation of the heel height judgment module, we used processed images and manually labeled the photos as input data. The CNN and MobileNetV2 models were used for training and testing separately, and 10-fold cross-validation was applied for verification. Because only the image of a single foot is input during the inputting of photos, the data of all participants can be divided into two datasets, namely the left- and right-foot datasets, to form four different models. Through measurement of the average accuracy, loss, sensitivity, specificity, precision, recall, F1 score, and area under the receiver operating characteristic curve, we compared the performance of the two models, with the experimental results detailed in Table 2. We also provide the confusion matrix of the average of the 10-fold cross-validation results for CNN model as the correctness

references. For the dataset used for the above analysis, we only used left feet data of 79 people, the other data of 21 people exists some problems so we deleted them, but the data of the right feet still uses the data of 100 people. The confusion matrixes of left and right feet dataset are shown in Figs. 12 and 13. A more effective performance was observed with the CNN, and we therefore used this model for the heel height judgment module.

We connected the EfficientDet model and CNN model in series and tested them with the left- and right-foot datasets, respectively. The overall experimental results are listed in Table 3.

D. PERFORMANCE EVALUATION RESULTS OF THE SUITABLE HEEL HEIGHT ESTIMATION MODULE

After the system passes through the automatic bounding and heel height judgment modules, the result is input into the suitable heel height estimation module. The mode algorithm is applied in the module to correct misjudgments, and the built-in height conversion program of the EfficientDet model is used to convert the corrected result to a heel height off the ground; this is then compared with the standard answer provided by the rehabilitation physician to obtain an MAE value. Based on a total of 16,841 data items from the feet of 100 participants, the correlations of high-heel heights from expert assessment and mobile app assessment can be shown

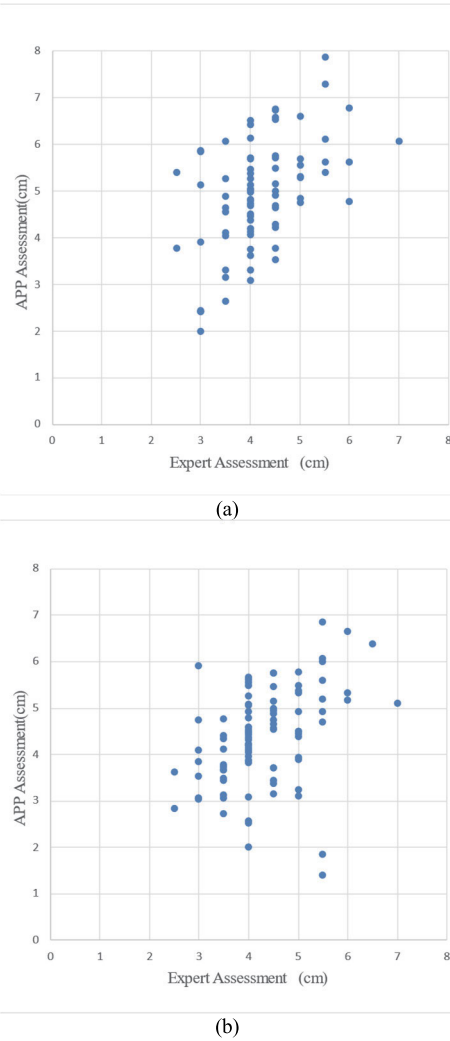


FIGURE 14. Correlation of high-heel heights from expert assessment and mobile app assessment for (a) left feet, and (b) right feet.

in Fig. 14. The average MAE of the left and right foot was computed to be 0.82 and 0.77 cm, respectively; the average MAE of the overall system was 0.80 cm.

Because the high heels of the left and right feet cannot be of different heights, our system will select the smaller height of the left and right feet as the output answer. This will further reduce the error generated, and the output value by the system will fall within the limit height. Fig. 15 is a plot of final output value versus physician assessment. Outputting the smallest answer of both feet can effectively reduce errors. Since the smallest unit of most high heels is inches (2.54 cm) as the cumulative unit, this system uses one inch as the allowable error. We draw two lines in Fig. 15 as a reference, the yellow part is the allowable error (± 2.54 cm), and the red part is the wearable height overestimated by the APP.

V. DISCUSSION

In this study, four experiments were implemented to evaluate which dataset had higher detection accuracy with the EfficientDet model; which classification model fed into the

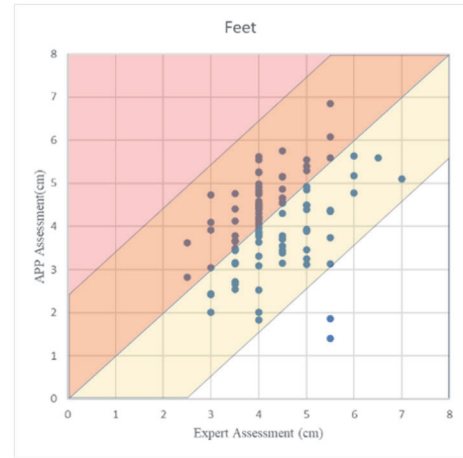


FIGURE 15. Data point distribution diagram of the smaller of the left and right heel heights falling within the allowable error (± 2.54 cm).

study dataset exhibited the highest accuracy; and whether the overall accuracy after cascading the EfficientDet module and CNN model could be maintained at a certain level. The MAE was used to present the overall performance evaluation of the whole system. In the first experiment, model analysis was performed with images containing only a single object or three objects as the dataset. As described in Table 1, when the pictures of the heels and cube box were separated as datasets for detecting, a high accuracy rate was noted. The datasets arranged from high to low according to the accuracy of the model were as follows: cube box, right heel, left heel, and cube box and the two heels. The cube box had the highest accuracy because this object did not change its position during the entire videorecording process, and thus, the image data were not deformed. For the three datasets *B*, *R*, and *L* that only contained one type of object, evaluation of the *B* dataset model was more effective than that of the other two datasets. Therefore, we speculated that although three types of objects were contained in the *BRL* dataset, the AI model conducted such highly accurate detection of the cube boxes that the overall judgment effect of the AI model increased for the *BRL* dataset. Therefore, the accuracy of the *L* and *BRL* datasets was close. Following a comparison of the effectiveness of the *L* and *R* datasets, the accuracy of the *L* dataset was revealed to be lower. However, the two ends of the marking line manually drawn on the feet of the participants for assisting the physician in diagnosis were also used as the basis for the EfficientDet model to manually mark the upper and lower bounds of the object range during the training phase. When filming, the left foot was farther away from the light source of the experiment site than the right foot, which may have caused the two ends of the marking line on the surface of the left foot to be relatively blurred in the image, reducing the effectiveness of model learning. In addition, AP_S and AP_M in Table 1 are both -100% , but AP_L is not equal to 100% . This phenomenon indicated that the four datasets in the experiment belong to the category with a pixel area greater than 96^2 , and in such a case, the display object must be sufficiently large for effective model evaluation.

In regards to the second experiment, Table 2 summarizes the results of the training of the CNN and MobileNetV2 models on the same two left- and right-foot datasets. Regardless of whether the dataset was for the left or right foot, the CNN model exhibited a higher level of accuracy than the MobileNetV2 model; the dropout layer in CNN model was adopted to eliminate overfitting, though it had a relatively low loss of the model [16]. The smaller size of the MobileNetV2 model compared with the CNN model facilitates installing the program and all AI models on a smartphone with small memory capacity. However, its accuracy was lower and its prediction speed approximately 2 s slower than that of the CNN model. To ensure accuracy and a quick prediction speed, we chose the CNN as the classification model for the heel height judgment module.

In the third experiment, we combined the trained automatic bounding module with the heel height judgment module to enable the left and right heels automatically bounded by the EfficientDet model to serve as the input data for the CNN model; the feet were separated into two datasets for testing. Through a comparison of Table 2 and 3, we observed that, when the CNN and EfficientDet models were connected in series, an average accuracy of 0.8584 was obtained, which is lower than that obtained when the CNN model used artificially labeled training data as input. Because the accuracy of the EfficientDet model itself is not 100%, misjudgment is still possible, which is also true for the CNN model. Therefore, when two models with an accuracy less than 100% are combined, a superposition of errors occur, resulting in reduced accuracy.

After completing the height calculation in the fourth experiment, the system used to evaluate the suitable height of high heels was employed to compare the maximum heel height calculated by the system with the standard answer provided by the physician. The MAE of the left and right foot was 0.82 and 0.77 cm, respectively, and the average MAE of both feet was 0.80 cm. The MAE of the left foot was 0.05 cm larger than that of the right foot, which may be attributable to the physician's use of Achilles tendon changes to determine the suitable high-heel height for participants during clinical evaluation. The left foot was the nondominant foot of most participants; when standing on tiptoes, the nondominant foot is prone to multistage calcaneal inversion, which can lead to misjudgment by the physician in the manual diagnosis of the left foot, resulting in incorrect standard answers being incorporated into the model. These small errors may have affected the learning effect of the AI model.

The MAE of the AI-based mobile app for evaluating the maximum height of high-heel shoes proposed in this study was 0.80 cm, which is 0.41 cm lower than the MAE (1.21 cm) of the evaluation system developed by Lee et al. [7]. The reason why the MAE is slightly higher than Lee et al. [7] may be that the number of photos our system took is much larger than that of [7]. If the CNN model is compared with [7] alone, the accuracy, precision, recall, and F1-score of our study are 0.98, 0.98, 0.98, and 0.98, respectively. The accuracy,

precision, recall, and F1-score of [7] were 0.88, 0.86, 0.92, and 0.89, respectively. But when the proposed system is running, our system will need to use both CNN and EfficientDet models. Table 3 shows that the accuracy, precision, recall, and F1-score are 0.86, 0.88, 0.82, and 0.85, respectively. The performance of our system is similar to that of Lee et al., but the final MAE is lower. The reason may be that Lee et al.'s study only took 21 pictures per operation, and only one picture was taken every 0.5 cm, but our system took 60 to 120 pictures per operation, that is, one picture was taken almost every 0.1 cm. This means that the fineness of the shooting of the heel area in this study is higher than [7], which is why the MAE is slightly lower than [7]. Apart from the excellent performance of the proposed system in terms of MAE, this system has many additional advantages such as low cost and quick evaluation speed. In addition, the assistance of a larger platform is not required, increasing its portability, and no restrictions apply to the foot length of participants. When the evaluation system developed by Lee et al. [7] collected the input data for model prediction, the filming process lasted approximately 12 min, and the basic CNN model employed required approximately 5 s to discriminate a set of 21 data items; by contrast, the recording process of our system was completed in 2 to 4 s, and the trained CNN model completed a set of judgments on an average of 97 data items in 1.98 s. Therefore, the data collection and execution speed of our AI model was 144 and 2.5 times, respectively, those of the system proposed by Lee et al., which markedly increases its usability.

VI. CONCLUSION

In this study, we developed a system that enables users to automatically and quickly predict the maximum height of high heels suitable for wear through a mobile app. We used a single independent object (B, R, L) and three objects (BRL), four datasets, an EfficientDet object recognition model, and CNN and MobileNetV2 AI classification models for overall performance evaluation. The test results demonstrated that the single object dataset of the right foot (R) collocated with the heel height judgment module constructed by the CNN model for object identification was most effective. Through use of the left- and right-heel image datasets with the CNN model, the average accuracy (0.86), precision (0.88), sensitivity (0.82), specificity (0.89), recall (0.82), and F1 score (0.85) were obtained through 10-fold cross-validation. In addition, when the subsequent correction of the misjudgment using the mode algorithm was completed, the MAE of the overall system was only 0.80 cm. Thus, when assessing their suitable heel height range, women no longer require professional assistance or a special measurement platform. Instead, users can easily complete the assessment with a smartphone. An additional advantage of the system is its ability to conserve medical resources. Typically, 15 to 20 min is required for diagnosis of a woman's suitable heel height by an experienced physician, but our system's evaluation process is completed within 5 min. At present, this system must be used in a place with no debris to avoid background objects

affecting the evaluation of the object detection model. In the future, we can collect more datasets that include videorecording heels in different backgrounds and collecting participants with minor foot problems to broaden the applicability of the system.

REFERENCES

- [1] C.-M. Yin, X.-H. Pan, Y.-X. Sun, and Z.-B. Chen, "Effects of duration of wearing high-heeled shoes on plantar pressure," *Human Movement Sci.*, vol. 49, pp. 196–205, Oct. 2016.
- [2] C.-Y. Hsu, "The effects on dynamic balance of wearing high heels with long term experience to stride over different heights of obstacle," M.S. thesis, Dept. Phys. Educ., Nat. Taipei Univ. Educ., Taipei, Taiwan, 2012.
- [3] S. Karia, S. Parasuraman, M. K. A. A. Khan, I. Elamvazuthi, N. Debnath, and S. S. A. Ali, "Plantar pressure distribution and gait stability: Normal VS high heel," in *Proc. 2nd IEEE Int. Symp. Robot. Manuf. Autom. (ROMA)*, Ipoh, Malaysia, Sep. 2016, pp. 1–5.
- [4] H.-L. Chien, "Lower limb joint mechanics and whole body balance control during high-heeled gait and obstacle-crossing," Ph.D. dissertation, Dept. Biomed. Eng., Nat. Taiwan Univ., Taipei, Taiwan, 2014.
- [5] X. Lu, "Changes in forefoot plantar pressure with shoe of different heel height," in *Proc. 3rd Int. Conf. Biomed. Eng. Informat.*, Yantai, China, Oct. 2010, pp. 791–794.
- [6] N. Kumar, S. N. Panda, and R. K. Kaushal, "High heel shoes with adjustable height of the heel," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 993, Apr. 2020, Art. no. 12120.
- [7] S.-H. Lee, B.-S. Lin, H.-C. Lee, X.-W. Huang, Y.-C. Chi, B.-S. Lin, and K. Abe, "Artificial intelligence-based assessment system for evaluating suitable range of heel height," *IEEE Access*, vol. 9, pp. 38374–38385, 2021.
- [8] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," 2019, *arXiv:1911.09070*.
- [9] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [10] Keras. (2018). *Keras: The Python Deep Learning Library*. Accessed: Aug. 18, 2020. [Online]. Available: <https://keras.io>
- [11] C. Zhu, Y. He, and M. Savvides, "Feature selective anchor-free module for single-shot object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 840–849.
- [12] S. Ren, K. He, R. Girshick, X. Zhang, and J. Sun, "Object detection networks on convolutional feature maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1476–1481, Jul. 2017.
- [13] A. Van Etten, "Satellite imagery multiscale rapid detection with windowed networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2019, pp. 735–743.
- [14] S. Cléménçon and N. Vayatis, "Nonparametric estimation of the precision-recall curve," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, Quebec, QC, Canada, Jun. 2009, pp. 185–192.
- [15] COCO. (2021). *Detection Evaluation*. Accessed: Jun. 29, 2021. [Online]. Available: <https://cocodataset.org/>
- [16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jun. 2014.



research interests include physical medicine and rehabilitation, especially in smart medical rehabilitation.

SI-HUEI LEE received the M.S. degree in Chinese medicine from China Medical University, Taichung, Taiwan, in 1999, and the Ph.D. degree from the Niigata University of Health and Welfare, Japan, in 2018. Since 2008, she has been an attending Physiatrist of physical medicine and rehabilitation with the Taipei Veterans General Hospital, Taiwan. She was an Observer of geriatric medicine with the Guy's and St Thomas' Hospitals, U.K., from May 2007 to June 2007. Her



SHIN-WEI CHIU is currently pursuing the degree with the Department of Computer Science and Information Engineering, National Taipei University, Taiwan. His research interests include smart medicine, wearable systems, and rehabilitation systems.



MAN-WEN SU is currently pursuing the degree with the Department of Computer Science and Information Engineering, National Taipei University, Taiwan. Her research interests include smart medicine, wearable systems, and rehabilitation systems.



SZU-YU WU is currently pursuing the degree with the Department of Computer Science and Information Engineering, National Taipei University, Taiwan. Her research interests include smart medicine, wearable systems, and rehabilitation systems.



YI-TING HUANG is currently pursuing the degree with the Department of Computer Science and Information Engineering, National Taipei University, Taiwan. Her research interests include smart medicine, wearable systems, and rehabilitation systems.



LI-YIN KANG is currently pursuing the degree with the Department of Computer Science and Information Engineering, National Taipei University, Taiwan. Her research interests include smart medicine, wearable systems, and rehabilitation systems.



BOR-SHING LIN (Senior Member, IEEE) received the B.S. degree in electrical engineering from the National Cheng Kung University, Taiwan, in 1997, and the M.S. and Ph.D. degrees in electrical engineering from the National Taiwan University, Taiwan, in 1999 and 2006, respectively. He is currently a Distinguished Professor with the Department of Computer Science and Information Engineering, and the Director of the Computer Center, National Taipei University, Taiwan. His research interests include smart medicine, embedded systems, wearable systems, biomedical signal processing, biomedical image processing, and portable biomedical electronic system design. He is a fellow of the Institution of Engineering and Technology (IET), U.K.; and the British Computer Society (BCS), U.K.

...