IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 4, 2021, accepted March 1, 2021, date of publication March 4, 2021, date of current version March 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3063912

Artificial Intelligence-Based Assessment System for Evaluating Suitable Range of Heel Height

SI-HUEI LEE[1,2], BOR-SHING LIN[3], (Senior Member, IEEE), HSIANG-CHEN LEE[3], XIAO-WEI HUANG[3], YA-CHU CHI[3], BOR-SHYH LIN[4], (Senior Member, IEEE), AND KAORU ABE[5]

[1]Department of Physical Medicine and Rehabilitation, Taipei Veterans General Hospital, Taipei 112, Taiwan
[2]Faculty of Medicine, National Yang Ming Chiao Tung University, Taipei 112, Taiwan
[3]Department of Computer Science and Information Engineering, National Taipei University, New Taipei 23741, Taiwan
[4]Institute of Imaging and Biomedical Photonics, National Yang Ming Chiao Tung University, Tainan 71150, Taiwan
[5]Department of Prosthetics, Orthotics and Assistive Technologies, Niigata University of Health and Welfare, Niigata 950-3198, Japan

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

This work was supported in part by the Ministry of Science and Technology in Taiwan, under Grant MOST 108-2221-E-009-054-MY2, Grant MOST 109-2314-B-305-001, and Grant MOST 109-2221-E-305-001-MY2; in part by the Higher Education Sprout Project of the National Chiao Tung University and Ministry of Education (MOE), Taiwan, through the University System of Taipei Joint Research Program under Grant USTP-NTPU-TMU-109-03, and in part by the Faculty Group Research Funding Sponsorship by National Taipei University under Grant 2020-NTPU-ORDA-02.

ABSTRACT High-heeled shoes of excessive height can severely injure shoe users. Such shoes may cause various injuries, including musculoskeletal pain, osteoarthritis, and hallux valgus. A physician can estimate an appropriate heel height limitation for an individual wearer by touching calcaneus to estimate deformation of the calcaneal varus. It would typically be impractical for a woman to seek the professional assistance of her physician when buying high-heeled shoes. A novel system was developed in this study for evaluating the maximum safe height of high-heeled shoes for female wearers. In this study, images of Achilles tendons, medial longitudinal arches, lateral longitudinal arches, and plantar pressure distributions served as the inputs in the proposed system. After the system had been trained with those images, the system could output the maximum height of high-heeled shoes for each individual wearer. In this study, two crucial methods were used for performing the evaluating system. First, Basic CNN, VGG16, and MobileNetV2 were used to evaluate images of feet. Through the experiments, the proposed artificial intelligence (AI) model achieved an accuracy of 0.88. Next, a statistics algorithm was used to modify the results obtained from the AI model. Subsequently, the error of the system declined. The mean absolute error of the proposed system which was used for evaluating the maximum height of high-heeled shoes was 1.21 cm, which is less than the typical increment for commercially available high-heeled shoes.

INDEX TERMS High-heeled shoes, calcaneal varus, plantar pressure, convolutional neural network (CNN), artificial intelligence (AI).

I. INTRODUCTION

Women generally wear high-heeled shoes, which position the heels higher than the toes, to appear taller and increase their perceived attractiveness [1]. Moreover, many women feel pressure to wear high-heeled shoes for long periods for social or job requirements. According to related research, wearing inappropriate high-heeled shoes causes severe burdens and effects on the body. Of all injuries related to high-heeled shoes, >72% are related to either the ankle or foot [2], including hallux valgus, musculoskeletal pain, and osteoarthritis [3]. In addition, wearing unsuitable high-heeled shoes for a long time may change the distribution of plantar pressure, which affects height of the longitudinal arch and tends to flatten the arch of the foot [4]. Women who wear high-heeled shoes must know their bodily limits and wear shoes that will not injure them. Because wearing high-heeled shoes will affect change of gaits [5], studies have mainly focused on the effects on health when people wear high-heeled shoes by analyzing gaits, and changes of the center of pressure (COP) and plantar pressure while walking in high-heeled shoes. In 2002, [6] evaluated the level of

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

muscle fatigue caused by walking in high-heeled shoes through plantar pressure and surface electromyography and noted that the lateral and medial gastrocnemius muscle activities of the females who frequently wear high-heeled shoes are unbalanced. These findings might be related to the abnormal lateral shift of sole COPs. In 2005, [7] found that walking while wearing high-heeled shoes increased displacement of COP by 200% compared with bare feet and that varying speed of COP was twice that of bare feet. Analyses of change of plantar pressure can obtain states of walking and comfortable levels when subjects are wearing high-heeled shoes. When walking in high-heeled shoes, plantar pressures of the heel and mid-foot will shift to the inner side of the forefoot. The higher the heels are, the more plantar pressure is transferred to the inner side of the forefoot and the greater the pain is in the feet [8]–[10]. Reference [11] performed comprehensive analysis of pressure distribution and change of COP and noted increased pressure on the right foot when the subjects walked in high-heeled shoes, which was due to an unstable gait.

When a high-heeled shoe user is evaluated at a clinic, the physician observes heel deformation and medial and lateral longitudinal arch changes and determines whether the heel height of the shoes is suitable for the user. The deformation of heels is caused by the tibialis anterior muscle and tibialis posterior contracting excessively. If the subject's heel bones are deformed, it means the heel heights of her shoes are inappropriate for her. In the early 21st century, deep learning has been widely used to classify photographs; starting from a dataset of foot measurements, the execution process of a deep learning classification program may be the same as the process of a physician's diagnosis. Therefore, this study used deep learning to make judgments about the suitability of wearing high-heeled shoes of various heights.

For classifying photographs (including photographs of humans), convolutional neural networks (CNNs) have demonstrated excellent performance [12]. In 2018, [13] applied CNNs to sign language recognition and determined letters with different gestures. In addition to the Basic CNN model, numerous simple and effective image classification models have also been developed. Transfer learning is a common method for improving models' accuracy by using a pretrained model, such as VGG16 and MobileNetV2 [14], [15]. In 2017, [16] adopted VGG16 for transfer learning to classify different gestures and obtained markedly high accuracy (0.93). In 2019, [17] used VGG16 for transfer learning and reached notably high accuracy (0.96) in American sign language recognition. In 2019, [18] used MobileNetV2 developed by Google for transfer learning. The accuracy reached 0.99 in sign language recognition. MobileNetV2 can also deliver rapid results on mobile devices and embedded systems because of its small size.

To the best of our knowledge, no relevant system that automatically and objectively evaluates the height limit of high-heeled shoes for female shoe users has been reported thus far. Thus, this study proposed an evaluating platform to automatically provide the highest heel height that a physician would advise for female users. Artificial intelligence (AI) models were trained; foot image and plantar pressure data were collected at each heel height as training data. Three models, including the Basic CNN, VGG16, and MobileNetV2, were selected as the cores of AI models. By using the proposed system, all female shoe users can obtain their suitable heel height range without physicians' assistance.

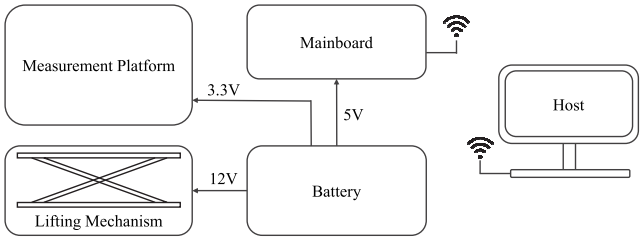


FIGURE 1. System overview.

II. SYSTEM ARCHITECTURE AND DESIGN

A. SYSTEM OVERVIEW

This study proposed an AI-based system for evaluating a suitable height range of high-heeled shoes. As depicted in Fig. 1, the system comprises five parts: measurement platform, mainboard, lifting mechanism, battery, and host. The measurement platform is used to collect images of plantar pressure, medial longitudinal arch, lateral longitudinal arch, and deformation of Achilles tendon under different heel heights. The mainboard is mainly used to collect force data from force-sensing resistor (FSR) sensors, to control the electric lifting jack in the lifting mechanism, and to read the height value of the digital caliper. The host provides a graphical user interface (GUI) used to control the whole system, collect data, display the suitable height range of high-heeled shoes, and train AI models. The system must be used indoors with sufficient lighting and have little shadows on the measurement platform.

B. MEASUREMENT PLATFORM

In this study, the size of the measurement platform was approximately $40 \times 40 \times 18.6$ cm^3 excluding camera frame parts. It was mainly used to collect foot data. Its architecture is illustrated in Fig. 2. Six universal serial bus (USB) webcams (C270, Logitech, Lausanne, Switzerland) were set up around the measurement platform, with 270p resolution, a 30-fps sampling rate, and a fixed focal length. The image data were collected through a USB hub and then transmitted to the host. 42 FSR sensors (FlexiForce A301, Tekscan, Boston, MA, USA; with thickness 0.203 mm, length 25.4 mm, width 14 mm, radius of sensing area 9.53 mm) were installed on the plane of the measurement platform to detect plantar pressure. The locations of 42 FSR sensors is based on related research [8]–[11]. The force range of each FSR sensor was 0 to 111 N (0–25 lb.). For this type of FSR, the greater the force is, the lower the resistance is; the force is proportional

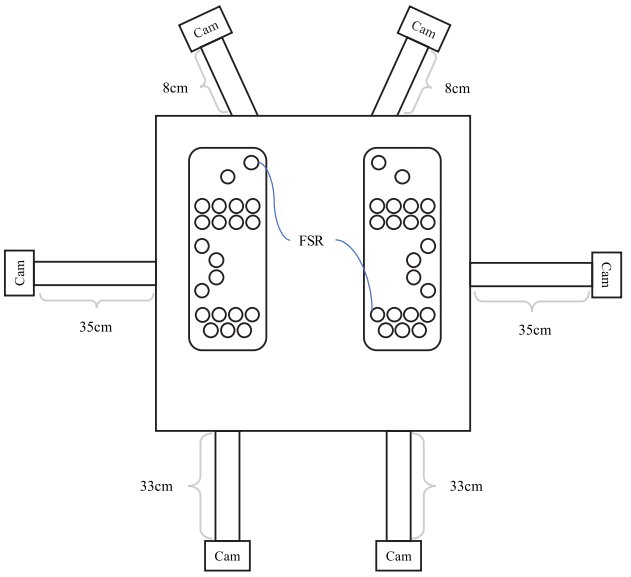


FIGURE 2. Measurement platform architecture.

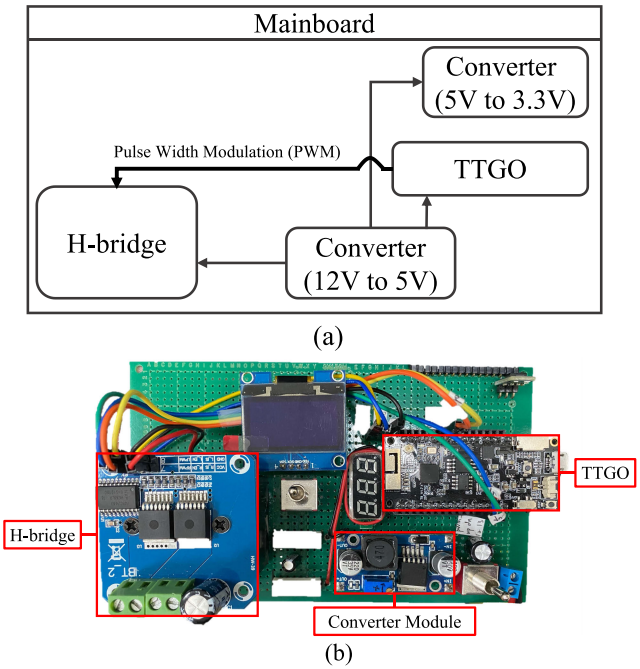


FIGURE 3. (a) Mainboard and lifting mechanism architecture. (b) Photograph of mainboard.

to the conductance. Two multiplexers (CD74HC4067, Texas Instruments, Dallas, TX, USA) integrated 42 FSR sensors and sent the analog values of FSRs to the analog-to-digital converter in the TTGO. When the measurement platform was lifted by the electric lifting jack, photographs of Achilles tendon, medial longitudinal arch, and lateral longitudinal arch were taken at every 0.5 cm of motion. The USB webcams taking pictures of Achilles tendons were 33 cm away from the platform plane, 25 cm high, and parallel to the ground. The USB webcams on both laterals of the platform were 35 cm away from the platform plane, 20.5 cm high, and tilted 10° upward. The USB webcams in front of the platform were 8 cm away from the platform plane, 39.8 cm high, and tilted 15° downward.

C. MAINBOARD AND LIFTING MECHANISM

The architecture of the mainboard is presented in Fig. 3(a), including a TTGO module (TTGO ESP32 16MB Flash 8MB PSRAM module, Shenzhen Xin Yuan Electronic Technology Co., Shenzhen, China), a dc–dc step-down power supply module (LM2596S), and an H-bridge (BTS7960). The overall size is 15 × 9 cm2, and a photograph is presented in Fig. 3(b). The first function of the TTGO is to collect data from 42 FSRs and the digital caliper and transmit it to the host through the ESP32 Wi-Fi module in TTGO; the second function is to control the electric lifting jack. The lifting mechanism includes an electric lifting jack and a digital caliper. When adjusting the height of the measurement platform, the host transmits a target height to TTGO through Wi-Fi, and then the TTGO instructs the H-bridge through pulse width modulation to control the speed and up/down direction of the electric lifting jack. To exactly lift or lower the measurement platform to the specified height, the TTGO first reads the value of the digital caliper through an integrated circuit and raises or lowers the lifting jack to the specified height. The power of the entire system was supplied by a 12-V 7800-mAh lithium-ion battery module. The battery module consisted of 12 lithium-ion batteries (3.65 V and 2600 mAh; 18650, Shenzhen Baiguan Battery Co., Ltd., Shenzhen, Guangdong, China) in series and parallel, which supplied 5 V to the mainboard, 12 V to the electric lifting jack, and 3.3 V to FSRs. When the battery module was supplying power to the mainboard and FSRs, 12 V power was converted to 5 V and 3.3 V through two independent dc–dc step-down power supply modules.

D. HOST

The host was a personal computer equipped with an Intel Core i7@3 GHz CPU, 64GB DRAM, and a graphics card (GeForce RTX 2080 Ti, MSI Computer Corp., New Taipei City, Taiwan); the operating system was Windows 10. The host provided a GUI to control the whole system, collected data, displayed the suitable height ranges, and trained AI models. When collecting data, the host received values of plantar pressure and foot photographs transmitted from the TTGO and webcams through Wi-Fi and USB respectively to form a dataset for training AI models. This system entered different combinations of datasets as inputs for three AI models; the optimal combination of AI models and input datasets was selected. The three AI models were executed in Tensorflow (Version 1.14.0, Google Brain, Mountain View, CA, USA) and Keras (Version 2.3.1) using the Python programming language [19], [20].

III. SOFTWARE DESIGN

The software architecture is presented in Fig. 4. It consists of four parts, including the GUI, image preprocessing, proposed CNNs, and statistical algorithm. The GUI is applied to

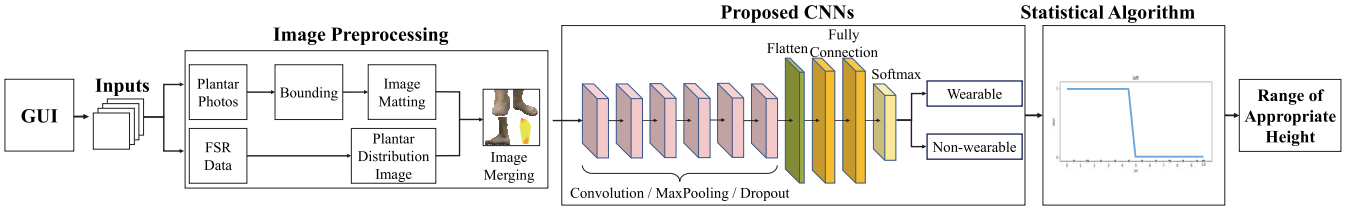


FIGURE 4. Software architecture.

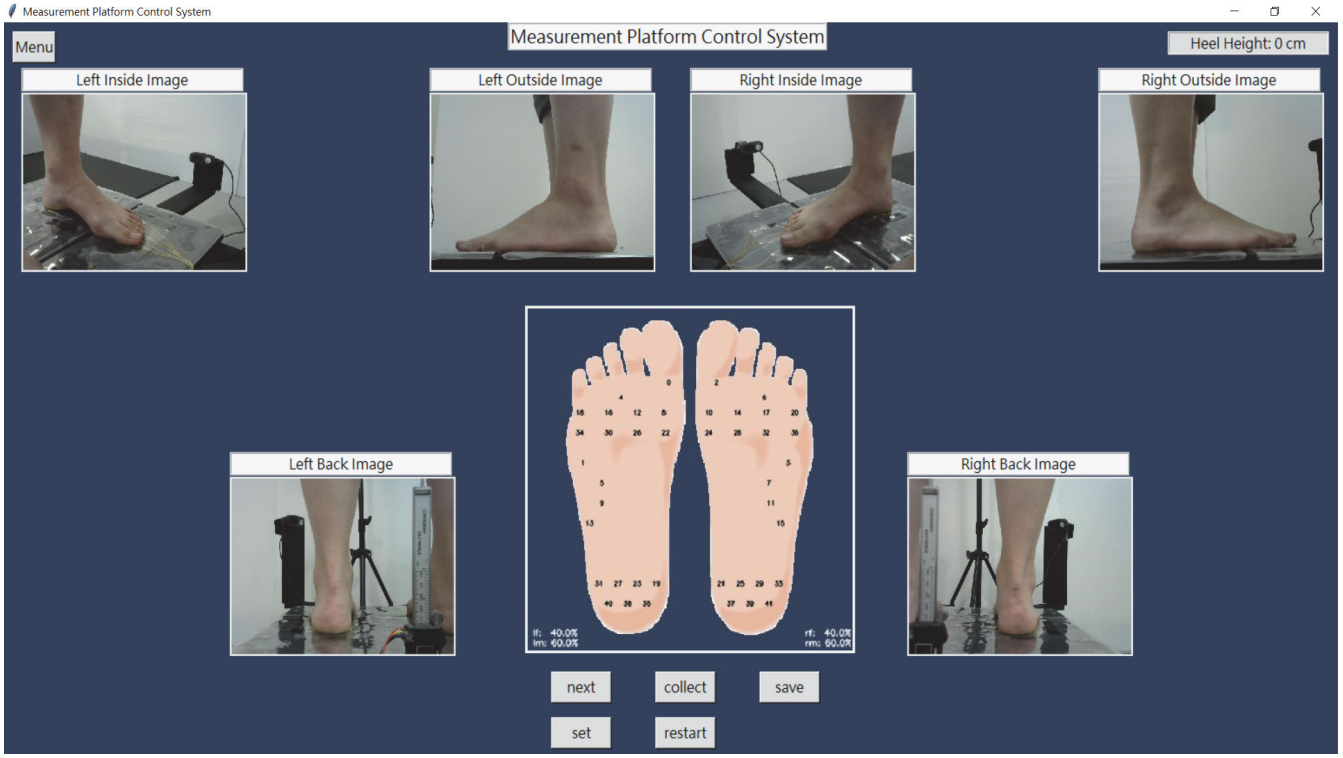


FIGURE 5. GUI in collecting data mode.

collect training data for AI models and displays a maximum suitable height of high-heeled shoes. Image preprocessing involves generating images of plantar pressure distribution, foot image preprocessing, and image merging. In generating images of plantar pressure distribution, collected pressure values from FSR sensors would convert into a continuous distribution map. The foot image preprocessing is used to bound the precise area of the foot in the image and remove the background. The last part is to merge images, and the dataset inclusive of merged images is applied to training CNNs. After model training, a CNN is conducted to classify images into two categories: wearable and nonwearable. After acquiring the result of CNN, the statistical algorithm modifies errors in 21 increments of heights which are represented wearable or nonwearable status to obtain maximum wearable height.

A. GRAPHICAL USER INTERFACE

A GUI was implemented in Python. That GUI consists of three modes: collecting data mode, physician evaluation mode, and AI evaluation mode. The collecting data mode has five buttons for operating and an interface to display foot images, plantar pressure values, and current height of the platform, as illustrated in Fig. 5. The ''Collect'' button, ''Next'' button, ''Save'' button, ''Set'' button, and ''Restart'' button are applied to collect photographs, raise the platform by 0.5 cm, save foot images and plantar pressure values to the host, set the height of the platform, and reset the platform to the horizontal state respectively. The physician evaluation mode provides an explicit interface through which doctors can set the height of the platform and stores the height value after evaluating. The AI evaluation mode determines the maximum of suitable heel heights for users after measurement by the platform.

B. IMAGE PREPROCESSING

Image Preprocessing comprises three parts: generation of plantar pressure distribution images, foot image bounding and matting, and image merging. For generating of plantar

pressure distribution images, the collected data from FSR sensors cannot directly be used as training data for CNNs because the data are discrete values rather than images. Therefore, an interpolation method is used to make up the areas that are not covered by the FSR sensors to fill in the regions with no values. When making a plantar pressure distribution image, we first made a base map with the foot shape, and the size of this base map is larger than the size of any subject's foot. The distribution of plantar pressure shows on the base map, which can ensure that there isn't any pressure value ignored. Because of no value being ignored, the consistency of images can be ensured. Eventually, discrete values are represented as corresponding to different colors on a color ramp; they are used to plot the image of plantar pressure distribution with Python's Matplotlib (Version: 2.2.2). The transformed image of plantar pressure distribution is presented in Fig. 6, which yellow means no value of plantar pressure and darker color means the larger value of plantar pressure. Images of plantar pressure distribution are used to observe the trend of change in pressure; values generated from the interpolation method are not used as training data. Therefore, it is helpful to use images of plantar pressure distributions as training data.

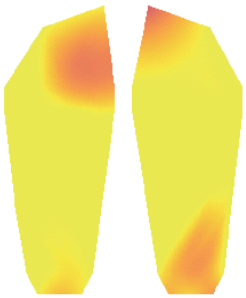


FIGURE 6. Plantar pressure distribution.

For foot image bounding and matting, Python's OpenCV package (Version 4.2.0.34) was used to process foot images taken by six cameras. Original images contain redundant information around the foot, which reduces the accuracy of classification. To eliminate redundant information and retain the region inclusive of the foot, cropping can be used to make foot in the center of the images. Moreover, setting color threshold was used to remove the non–skin-colored region; reserving the foot area in the whole picture minimized the possibility of erroneous judgment due to background interference.

After generating images of plantar pressure distribution and preprocessing of foot images, the process must merge images of the heel, the medial longitudinal arch, the lateral longitudinal arch, and the image of plantar pressure distribution into a single image. The image size must be 224 × 224 pixels, as illustrated in Fig. 7. In this study, we experimented on images with and without matting to confirm whether image matting would increase the accuracy in proposed models.

FIGURE 7. Merged image with matting.

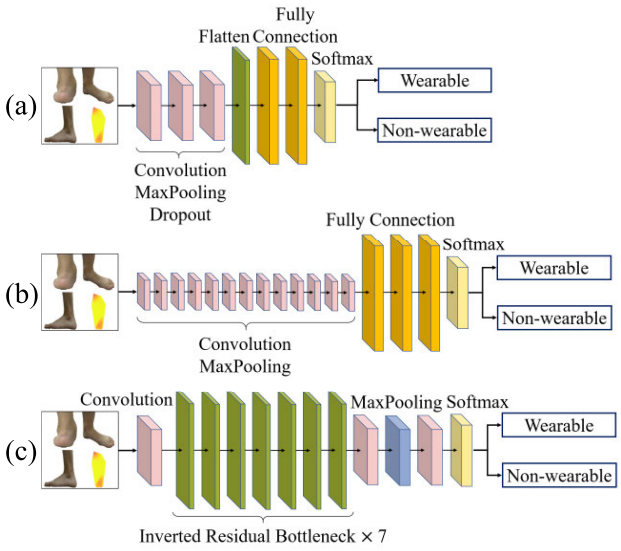


FIGURE 8. CNN architectures of the system: (a) Basic CNN. (b) VGG16. (c) MobileNetV2.

C. PROPOSED CNNS

A typical CNN architecture includes an input layer, several convolutional and pooling layers, one flattening layer, one or more fully connected layers, and one output layer. The size of input image is set in the input layer. After the size of the input image has been set, the features of the input image can be extracted by convolutional layers, and noise can be reduced through pooling layers. After a two-dimensional feature vector has been acquired, it can be converted to a one-dimension feature vector through a flattening layer. The fully connected layer is used to calculate the weights of neurons and to select significant features to map final output. In this study, the three CNN architectures presented in Fig. 8 (Basic CNN, VGG16, and MobileNetV2) were used to experiment. The Basic CNN model was based on Keras and modified for this study. The VGG16 and MobileNetV2 models were based on high-performance image classification system architectures presented in other studies [14]–[18]. To select the CNN model with the best performance, numerous experiments were conducted with these CNN architectures. The Basic CNN model was equipped with an input layer, three convolutional, max

TABLE 1. Different network configurations.

CNN Architecture	Parameter Setting of Layers
Basic CNN	C([32, (3×3)))]D(0.25)P(2×2) - C([64, (3×3))))]D(0.25)P(2×2) - C([64, (3×3))))]D(0.25)P(2×2) - F(4000) - F(100) - S(2)
VGG16	C([64, (3×3))))]C([64, (3×3))))]P(2×2) - C([128, (3×3))))]C([128, (3×3))))]P(2×2) - C([256, (3×3))))]C([256, (3×3))))]C([256, (3×3))))]P(2×2) - C([512, (3×3))))]C([512, (3×3))))]C([512, (3×3))))]P(2×2) - C([512, (3×3))))]C([512, (3×3))))]C([512, (3×3))))]P(2×2) - F(4096) - F(4096) - F(1000) - S(2)
MobileNetV2	C([32, (3×3)))] - IR([16, (3×3)))] - IR([24, (3×3)))] - IR([32, (3×3)))] - IR([64, (3×3)))] - IR([96, (3×3)))] - IR([160, (3×3)))] - IR([320, (3×3)))] - C([1280, (1×1)))] - P(7×7) - C([100, (1×1)))] -S(2)

pooling, and dropout layers, one flattening layer, two fully connected layers, and an output layer using Softmax. The size of input image was set as 224 × 224 pixels. Thus, the system is not affected by the camera resolution. For parameter setting, the optimizer was set according to Adam [21], and the activation function was set as ReLu. Moreover, the number of channels was three due to red–green–blue (RGB) channels in the input images. In this study, the batch size and the number of epochs were set as 16 and 200 respectively. The kernel size of convolutional layer was 3 × 3, as listed in Table 1. Table 1 offers additional details regarding the parameter settings for three CNN architectures; C denotes the convolutional layer, and the values in the parentheses following C are sequentially the number of filters and kernel size; C followed by D denotes the dropout layer, and the value in the parentheses is the proportion of dropping data; P denotes the pooling layer that follows the convolution layer, and the value in the parentheses is the size of max pooling layer; the hyphen denotes the connection between two adjacent layers. For example, the first layer of Basic CNN model is presented as C[32, (3 × 3)] D(0.25) P(2 × 2), indicating that 32 filters with the kernel size 3 × 3 were used in the convolutional layer which was followed by a dropout function and a max pooling. In the VGG16 architecture, original layers comprised 13 convolutional layers with pooling layers among convolutional layers and three fully connected layers; an output layer using Softmax was added finally. In the MobileNetV2 architecture, the original layers comprised three convolutional layers, one pooling layer, and seven inverted residual bottleneck layers; an output layer using Softmax was added. The details regarding the parameter settings for VGG16 and MobileNetV2 architectures are also listed in Table 1, where IR denotes the inverted residual bottleneck layer, and the values in the parentheses are sequentially the number of filters and kernel size.

All proposed CNN models in this study were used to determine whether the heights of tested high-heeled shoes were appropriate according to the collected data. Any shoes that a physician would describe as suitable were noted as appropriate, whereas any shoes that a physician would describe as injurious were noted as inappropriate. The maximum suitable heights were acquired from clinical diagnoses by physicians, and these results were used as models' labels. To obtain maximum suitable height, first 21 judgments that represented wearable or nonwearable were predicted by trained CNN model. The maximum suitable height was obtained for each case because it occurs at a critical point between the sets of wearable and nonwearable heights. However, the predicted results from proposed CNN models may be less than perfectly precise because none of the models had a 100% accuracy regarding the critical point. Therefore, a statistical algorithm was used to modify errors to improve the accuracy of the proposed system.

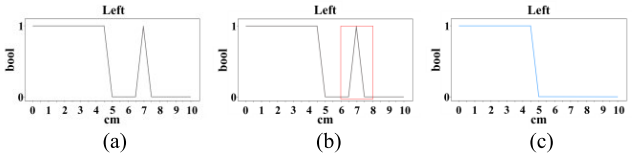


FIGURE 9. Statistical algorithm graphs. (a) Original model-predicted result. (b) Algorithmic modification process. (c) Modified result.

D. STATISTICAL ALGORITHM

The statistical algorithm considered 21 judgments evaluated by the model and determined each critical point more exactly. Fig. 9(a) depicts the original result predicted by a model. The vertical axis of Fig. 9 represents whether the heel height is suitable; the values one and zero denote that it is wearable and nonwearable respectively; the horizontal axis represents the unit of height. As illustrated in Fig. 9(a), the results contain scattered errors, not continuous. Therefore, each data point can be set to the mode of the five-member set containing its adjacent four data points and itself, as presented in Fig.9(b). Taking Fig. 9(b) as an example, the data at 7 cm should be modified because continuous data should not vary for a single subject. As illustrated in Fig. 9(b), The values from left to right for 6, 6.5, 7, 7.5, and 8 are 0, 0, 1, 0, and 0; the value at 7 cm is modified from one to zero because the mode is zero. Fig. 9(c) indicates the modified results. As revealed in Fig 9(a) and Fig. 9(c), the number of critical points decreases from two to one. The mode algorithm can more exactly determine the critical point without interference from model errors, and the system can evaluate the maximum height more precisely. The system evaluates maximum wearable height according to single critical point. As Fig. 9(c) for example, the maximum wearable height is 4.5 cm.

IV. EXPERIMENT DESIGN

The flow chart of the experiment design in this study is presented in Fig. 10. It consists of three parts, including selection and exclusion of subjects, evaluation of suitable range of heel height, and data collection process. Subjects are firstly screened whether they are suitable for all conditions

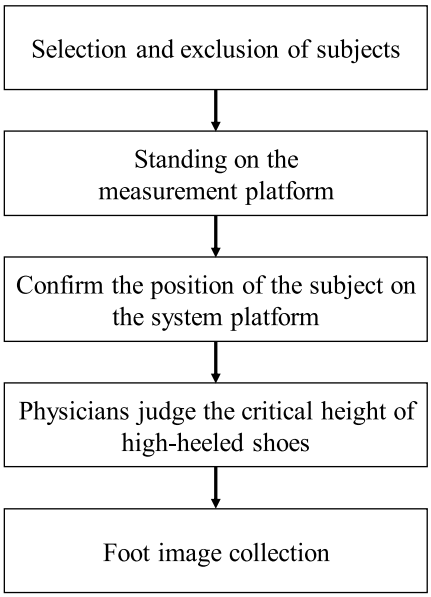


FIGURE 10. The flow chart of the experiment design.

mentioned in next section. The selected subjects then stand at the designated position on the measurement platform to allow physicians to judge the critical height of high-heeled shoes. After physicians finishing to estimate the maximum safe height of high-heeled shoes for female wearers, data of subjects are collected by using the system platform.

A. SELECTION AND EXCLUSION OF SUBJECTS

In this study, data of female subjects were collected after the subjects had been screened. During condition screening, the recruited subject was asked to stand naturally, with the inner sides of the feet parallel, with the backs of the heels on a single horizontal line. The screening and exclusion conditions based on [22]–[24] and included the following points:

- Arch rate = 11%–14% [calculated as (the height from the most convex point of the scaphoid bone to the ground/foot length) × 100]
- Skew angle of the back heel $\leq 4°$
- Valgus of the big toe $\leq 20°$
- Varus of the little toe $\leq 20°$
- No congenital scoliosis
- No calluses on the soles of feet
- No history of serious lower limb injuries

B. EVALUATION OF SUITABLE RANGE OF HEEL HEIGHT

When physicians want to determine the height limit of high-heeled shoes that the subject can wear, the subject must naturally stand on the system platform with her hands hanging beside her thighs, and the position of her transverse arch of the feet should be at the turning point of platform when the platform rises, as indicated by the arrow in Fig. 11. Because the contact point between the jack and the platform is fixed, the subject's heels may not likely be on this contact point during the ascent process, and the actual height of the subject's

FIGURE 11. Position of the transverse arch (red arrow), which is the turning point when the platform rises.

heel may not match the returned value from the digital caliper. To adjust the height of each rise that can actually raise the heel of the subject by 0.5 cm, this study sets four specifications of A, B, C, and D according to the common foot length range of Asian women, which are marked on the platform for researchers to determine what specification of the subject's foot length is. The four specifications are 13.2 cm, 13.8 cm, 14.4 cm, and 15 cm respectively from the axis of rotation on the platform, as illustrated in Fig. 12.

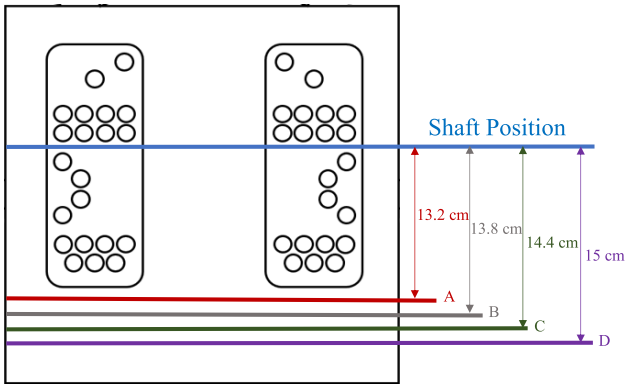


FIGURE 12. Schematic of the four specifications at the position of the platform. A, B, C, and D are at 13.2, 13.8, 14.4, and 15 cm from the shaft, respectively.

FIGURE 13. (a) Natural stance and (b) inversion of the calcaneus during elevation.

Physicians estimate the critical height of high-heeled shoes by diagnosing whether the subject's calcaneus is inverse during the continuous elevation of heels of the subject, as depicted in Fig. 13. The entire diagnosis process is repeated three times to avoid misjudgments and averaged to reduce

errors. The tibialis anterior muscle attaches from the front of the ankle to the medial foot, and the tibialis posterior attaches from the posterior surface of the ankle to the medial foot. When a female wears high-heeled shoes in a standing position, the ankle will be locked in plantar flexion. These two muscles contract simultaneously to act as antagonist muscles and thus maintain ankle plantar flexion. When the height of shoes just exceeds the threshold, two muscles contract so strongly that they cause calcaneus inversion. When the calcaneus is inverse, the program converts the height value of the digital caliper to actual height of the heels according to a conversion table and records the value. The aforementioned movements must be repeated three times for the left and right feet; the system calculates the average. In a later stage, this average serves as a label for training AI models.

C. DATA COLLECTION PROCESS

When data are collected from a subject, the subject naturally stands on the system platform with hands hanging beside the thighs, and the data starts to be collected after determining the specifications of the subject's foot length (A - D). When the experimenter operates the apparatus, the GUI must be switched to data collection mode. Each time the platform rises, it automatically transforms the height, which is converted to a 0.5-cm rise on the heel of the subject. The researcher uses the GUI to control the platform, from 0 cm to 10 cm; the platform rises in 0.5-cm steps. Each time the platform rises, the photographs from six cameras and 42 FSR sensor values are collected and stored in the host, as material for AI training; the overall collecting process requires approximately 12 minutes. The actual situation of collecting data are presented in Fig. 14.

(a) (b)

FIGURE 14. Actual data collection: (a) platform height = 0 cm and (b) the rising platform.

V. RESULTS

A. DATA DISTRIBUTION OF ALL PARTICIPANTS

In this study, we recruited 175 female subjects. The experiment procedure had been approved by the Institutional Review Board (IRB) of En Chu Kong Hospital (IRB No. ECKIRB1090404). All subjects provided written informed consent before participating in the trial. A total of 100 subjects passed the screening, with average age of 21 (standard deviation (SD) = 1.245), an average height of 159 cm (SD = 5.42 cm), an average weight of 51 kg (SD = 6.38 kg), and an average foot size of 24 cm (SD = 1.02 cm). With the system set to physician evaluation mode, 200 labeling data on maximum left and right heel heights were labeled by physicians; the distribution of labels is illustrated in Fig. 15. In data collection mode, we collected 42 items of left-foot and right-foot data from each subject, and collected 4200 pieces of data from 100 subjects in total. The data distribution was close to a normal distribution, and the ratio of wearable and nonwearable data was 1.1:1, which (statistically speaking) is close to 1:1.

TABLE 2. Accuracy obtained when different CNN architectures were used with and without image preprocessing.

	Basic CNN	VGG16	MobileNetV2
without Image Preprocessing	0.88	0.88	0.86
with Image Preprocessing	0.86	0.87	0.86

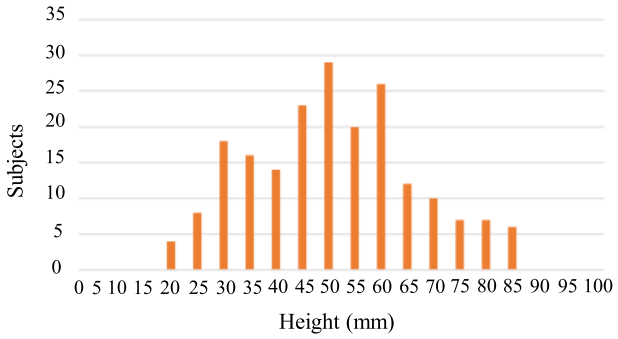


FIGURE 15. Distribution of 200 labeling data on maximum heel height of left and right feet labeled by physicians.

B. MODEL PERFORMANCE WITH AND WITHOUT IMAGE PREPROCESSING

This experiment explores whether the model can perform better with image preprocessing. Two datasets were used for categorizing the heights of the high-heeled shoes as wearable or unwearable. The foot images in one dataset were cropped but not image matted. The foot images in the other dataset were cropped and image matted. Three CNN architectures (Basic CNN, VGG16, and MobileNetV2) were trained on these two datasets; we verified the accuracy through 10-fold cross-validation. The performance of three models is listed in Table 2. The accuracies of Basic CNN, VGG16, and MobileNetV2 for preprocessing cropped but not image-matted datasets were 0.88, 0.88, and 0.86, respectively, all higher than or equal to cropped and image-matted datasets.

C. MODEL PERFORMANCE WITH DIFFERENT INPUT IMAGES

To compare performance of different models with different input combinations, this experiment used the same three CNN

architectures, namely Basic CNN, VGG16, MobileNetV2, and nine datasets (of which the details are summarized in Table 3). The foot and plantar pressure distribution images in nine datasets were used without image matting and data augmentation. Each dataset contains 4200 pictures, which 90% of dataset for training and 10% of dataset for testing through three CNN architectures. Each model was verified by 10-fold cross-validation.

Three CNN architectures and nine datasets were used to generate nine binary classifiers for training and testing to classify wearable and unwearable. Therefore, the study obtained 27 models and used them for 10-fold cross-validation. Then, this experiment compared the advantages and disadvantages of models by averaged accuracy, loss, precision, recall, and F1-score. The experiment results are displayed in Table 3. It can be observed that the performance of the model decreased when it used plantar pressure distribution as a training dataset, and the dataset including Achilles tendon, medial longitudinal arch, and lateral longitudinal arch had high performance. In general, Basic CNN had the best performance among all models when it used a dataset containing Achilles tendon, medial longitudinal arch, and lateral longitudinal arch data for training. The time for this trained model to analyze a set of 21 judgments was about 5 seconds.

D. RESULTS OF MODE ALGORITHM

According to the results in Table 3, the dataset containing cropped pictures of Achilles tendon, medial longitudinal arch, and lateral longitudinal arch without augmentation was selected to train Basic CNN as model in this study. Basic CNN was used to testing every 420 data for ten runs because of 10-fold cross-validation. The experiment had 20 sets of left and right feet for 10 subjects, and each set had 21 prediction results of wearable and unwearable judgments of different height. Next, a set of 21 prediction results of model were converted in a single batch job through a mode algorithm to reduce error. We compared the modified result with answers evaluated by physicians; the overall average mean absolute error (MAE) for each of the ten tests was 1.21 through 10-fold cross-validation.

VI. DISCUSSION

This study proposed an AI-based system for evaluating the appropriate heights of high-heeled shoes using foot images. Two experiments were conducted to evaluate whether the system requires image preprocessing and to determine which combination of input dataset and CNN model had the highest accuracy.

In the first experiment, the datasets processed image matting and nonmatting respectively were evaluated to choose which one could provide higher accuracy for AI models. Table 2 shows that image nonmatting yields a higher accuracy in judging whether the height is wearable. The image matting achieved a lower accuracy than image nonmatting in judging whether such height is wearable because the system used a threshold method of image matting so that some foot-color-like components around the foot were reserved. That is, the image had some irregular noise elements that could influence the judgment of a typical AI model. A typical AI model could perceive the changes of the feet between different images because the backgrounds of the nonmatted images were consistent.

Table 3 lists the results of training three different CNN architectures using nine different datasets. It can be observed from Table 3 that when images of plantar pressure distribution are used individually as the dataset, the accuracy of the model is lower than that of other datasets. This result is caused by two reasons. The first is that when a person is standing, the center of gravity is divided into left or right foot. That is, the subject is upheld primarily by the pivot foot, and the nonpivot foot bear less pressure. The tilted center of gravity would cause inaccurate measurement of the values of plantar pressure, which causes irregular changes in plantar pressure distribution images and affects AI model classification. The second reason is that each subject had feet of an idiosyncratic size. The subjects with small feet could not step on all the sensors or could only half-step on some sensors, but those with large feet could step on all the sensors. This situation leads to the fact that not all pressure values used in illustrating images of the plantar pressure distribution are derived from all sensors. When the specifications are inconsistent, the trend changes on images of the plantar pressure distribution of each subject are also different, which reduces the performance of AI models.

In Table 3, in addition to acquiring the result that the poor accuracy is achieved when using images of plantar pressure distribution individually as a training dataset, it could also be found that the datasets including parts or all of Achilles tendon, medial longitudinal arch, and lateral longitudinal arch achieve high accuracy. Although some datasets include images of plantar pressure distribution, the characteristics of other foot pictures are sufficient for the model to make correct judgments. As revealed in Table 3, datasets with only pictures of Achilles tendon, or medial longitudinal arch, or lateral longitudinal arch for training could achieve high performance; the accuracy was equivalent to use of a dataset containing more than three pictures. Therefore, it is only necessary to select one of the pictures from Achilles tendon, the medial longitudinal arch, and the lateral longitudinal arch as the training dataset. By reducing the types of images in the dataset, we can also reduce the complexity of the measurement platform. Finally, we use Achilles tendon images individually as the training dataset, because the physician determines the height of subject's suitable high-heeled shoes based on the changes in Achilles tendon during clinical evaluation. Therefore, our AI model uses Achilles tendon pictures as the training dataset that it is the most identical evaluation method close to the physician.

As listed in Table 3, by using images of Achilles tendon as the training dataset, the accuracy of Basic CNN and VGG16 is higher than that of MobileNetV2. This is because MobileNetV2 slightly sacrifices accuracy in pursuit of a

TABLE 3. Overall results of different CNN architectures with different datasets.

	Dataset	Accuracy	Loss	Precision	Recall	F1-score	AUC
Basic CNN		0.88	0.27	0.88	0.90	0.89	0.97
VGG16	AMLP	0.88	0.63	0.89	0.89	0.89	0.93
MobileNetV2		0.86	0.63	0.87	0.87	0.87	0.93
Basic CNN		0.88	0.26	0.88	0.90	0.89	0.96
VGG16	AML	0.87	0.63	0.88	0.89	0.88	0.93
MobileNetV2		0.87	0.63	0.87	0.90	0.88	0.94
Basic CNN		0.88	0.27	0.88	0.90	0.89	0.97
VGG16	AMP	0.87	0.63	0.88	0.89	0.88	0.93
MobileNetV2		0.86	0.63	0.85	0.91	0.87	0.92
Basic CNN		0.88	0.27	0.89	0.87	0.88	0.96
VGG16	ALP	0.88	0.63	0.88	0.90	0.88	0.93
MobileNetV2		0.86	0.63	0.87	0.88	0.87	0.93
Basic CNN		0.88	0.27	0.90	0.87	0.88	0.97
VGG16	MLP	0.87	0.63	0.88	0.88	0.88	0.93
MobileNetV2		0.87	0.63	0.87	0.89	0.88	0.93
Basic CNN		0.88	0.28	0.86	0.92	0.89	0.96
VGG16	A	0.88	0.63	0.89	0.88	0.88	0.94
MobileNetV2		0.87	0.62	0.87	0.89	0.88	0.94
Basic CNN		0.88	0.28	0.85	0.93	0.89	0.96
VGG16	M	0.88	0.63	0.88	0.89	0.88	0.93
MobileNetV2		0.85	0.63	0.86	0.87	0.86	0.92
Basic CNN		0.87	0.28	0.88	0.89	0.88	0.96
VGG16	L	0.87	0.63	0.87	0.89	0.88	0.92
MobileNetV2		0.87	0.62	0.87	0.89	0.88	0.94
Basic CNN		0.81	0.48	0.84	0.80	0.81	0.90
VGG16	P	0.80	0.64	0.81	0.85	0.82	0.87
MobileNetV2		0.79	0.64	0.81	0.81	0.80	0.84

AMLP = Achilles tendon, medial longitudinal arch, lateral longitudinal arch, and plantar pressure distribution; AML = Achilles tendon, medial longitudinal arch, and lateral longitudinal arch; AMP = Achilles tendon, medial longitudinal arch, and plantar pressure distribution; ALP = Achilles tendon, lateral longitudinal arch, and plantar pressure distribution; MLP = medial longitudinal arch, lateral longitudinal arch, and plantar pressure distribution; A = Achilles tendon; M = medial longitudinal arch; L = lateral longitudinal arch; P = plantar pressure distribution.

smaller architecture and faster execution efficiency, whereas Basic CNN and VGG16 have higher accuracy but have larger model architectures. VGG16 and MobileNetV2 have overfitting because the models are more complex, resulting in higher loss. Basic CNN uses dropout function [25] to solve the problem of overfitting, so the loss is lower. Overall, Basic CNN has the best performance. In Table 3, we observed the values of indicators when Basic CNN only uses Achilles tendon images as the training dataset. Precision is 0.86 and recall is 0.92, which means this model has the best image recognition; F1-score is 0.89, which means this model has the most stable performance; area under the curve (AUC) is 0.96, which means this model has the best performance for binary classification. Thus, this study used Basic CNN as its favored classification model.

The MAE of the system used to evaluate the height of high-heeled shoes suitable for subjects in this study is 1.21 cm. High-heeled shoes can be classified into three classes according to height of heels: shoes with heel height of $<$1 inch ($<$2.54 cm), 1–2.5 inches (2.54–6.35 cm), and $>$2.5 inches ($>$6.35 cm) are termed low-, mid-, and high-heeled shoes, respectively [26]. The MAE (1.21 cm) of this system is $<$0.5 inches, which indicates that the system has ability to differentiate these three classes. Moreover, this system can even be applied to choose high-heeled shoes whose increment is only 1 inch high; the system does not affect user's choice of suitable high-heeled shoes.

VII. CONCLUSION

The study proposed an innovative system to evaluate the heel height limitations of high-heeled shoes for female shoe users. Nine datasets and three CNN architectures were used to evaluate the comprehensive performance of the system. The experimental results indicated that the proposed system attained the highest performance and faster computing speed as using Basic CNN model with the dataset including only Achilles tendon. As Basic CNN with the dataset including only Achilles tendon was used as the training model, multiple experimental results were obtained through 10-fold cross-validation, inclusive of accuracy (0.88), loss (0.28), precision (0.86), recall (0.92), F1-score (0.89), and AUC (0.96). The experimental results also prove that comprehensive performance of the system declines when the system uses undesirable images of distribution of plantar pressure. Next, the MAE of the system could reach only 1.21 cm through mode algorithm. The most significant contribution of the study is that we are the first team to propose an AI system to evaluate the suitable range of heel height of high-heeled shoes for females. Moreover, the MAE of the system is only 1.21 cm (i.e., $<$0.5 inches). Therefore, this system is practical and

effective for the prediction of heel height. This system makes females more convenient to realize their heel height limitation of high-heeled shoes without assistance from physicians. However, the system is only suitable for particular females; the dataset of the system was limited because data were only collected from females who were eligible for the screening conditions and whose plantar lengths were between 22 and 26 cm. Furthermore, this system must be used in a bright environment to avoid erroneous judgment from AI model. In the future, the measurement platform may be improved to reduce measurement time by using Wi-Fi cameras or other types of cameras, which can improve user comfort and make it easier for subjects to maintain the same position. The current approach is to provide a video for subjects to watch, which is used to distract their attention to mitigate their discomfort. In addition, more foot images from females whose plantar lengths are not between 22 and 26 cm or who are unsuitable for the screening conditions could be collected by increasing the number of FSR sensors and arranging the positions of FSR sensors more densely such that our proposed system can be suitable for more extensive groups of female shoe users.

ACKNOWLEDGMENT

Prof. Bor-Shing Lin thanks Kuan-Hsun Ho for his technique support. And this manuscript was edited by Wallace Academic Editing.

REFERENCES

[1] P. Prokop and J. Švancárová, "Wearing high heels as female mating strategy," *Pers. Individual Differences*, vol. 152, Jan. 2020, Art. no. 109558.
[2] J. X. Moore, B. Lambert, G. Jenkins, G. McGwin, and Jr., "Epidemiology of high-heel shoe injuries in U.S. women: 2002 to 2012," *J. Foot Ankle Surg.*, vol. 54, no. 4, pp. 615–619, 2015.
[3] M. S. Barnish and J. Barnish, "High-heeled shoes and musculoskeletal injuries: A narrative systematic review," *BMJ Open*, vol. 6, no. 1, pp. 1–8, Jan. 2016.
[4] C.-M. Yin, X.-H. Pan, Y.-X. Sun, and Z.-B. Chen, "Effects of duration of wearing high-heeled shoes on plantar pressure," *Hum. Movement Sci.*, vol. 49, pp. 196–205, Oct. 2016.
[5] E. Di Sipio, G. Piccinini, C. Pecchioli, M. Germanotta, C. Iacovelli, C. Simbolotti, A. Cruciani, and L. Padua, "Walking variations in healthy women wearing high-heeled shoes: Shoe size and heel height effects," *Gait Posture*, vol. 63, pp. 195–201, Jun. 2018.
[6] A. Gefen, M. Megido-Ravid, Y. Itzchak, and M. Arcan, "Analysis of muscular fatigue and foot stability during high-heeled gait," *Gait Posture*, vol. 15, no. 1, pp. 56–63, Feb. 2002.
[7] W. H. Cho and H. Choi, "Center of pressure (COP) during the postural balance control of high-heeled woman," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, Shanghai, China, Jan. 2005, pp. 1–7.
[8] W. H. Hong, Y. H. Lee, Y. C. Pei, and C. Y. Wu, "Influence of heel height and shoe insert on comfort perception and biomechanical performance of young female adults during walking," *Amer. Orthopaedic Foot Ankle Soc.*, vol. 26, no. 12, pp. 1042–1048, 2005.
[9] 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.
[10] T. Albon, "Plantar force distribution for increasing heel height within women's shoes," Wooster Physics Junior Theses, Dept. Phys., College Wooster, Wooster, OH, USA, 2011, pp. 1–5.
[11] 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. Automat. (ROMA)*, Ipoh, Malaysia, Sep. 2016, pp. 1–5.
[12] 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.
[13] M. R. Islam, U. K. Mitu, R. A. Bhuiyan, and J. Shin, "Hand gesture feature extraction using deep convolutional neural network for recognizing American sign language," in *Proc. 4th Int. Conf. Frontiers Signal Process. (ICFSP)*, Poitiers, France, Sep. 2018, pp. 115–119.
[14] M. U. Hassan. (2018). *VGG16-Convolutional Network for Classification and Detection*. Accessed: Aug. 18, 2020. [Online]. Available: https://neurohive.io/en/popularnetworks/vgg16
[15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 4510–4520.
[16] S. Hussain, R. Saxena, X. Han, J. A. Khan, and H. Shin, "Hand gesture recognition using deep learning," in *Proc. Int. SoC Design Conf. (ISOCC)*, Seoul, South Korea, Nov. 2017, pp. 48–49.
[17] F. Nugraha and E. C. Djamal, "Video recognition of American sign language using two-stream convolution neural networks," in *Proc. Int. Conf. Electr. Eng. Informat. (ICEEI)*, Bandung, Indonesia, Jul. 2019, pp. 400–405.
[18] N. F. P. Setyono and E. Rakun, "Recognizing word gesture in sign system for Indonesian language (SIBI) sentences using DeepCNN and BiLSTM," in *Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS)*, Bali, Indonesia, Oct. 2019, pp. 199–204.
[19] (2016). *Google Brain, Build a Convolutional Neural Network Using Estimators*. Accessed: Aug. 18, 2020. [Online]. Available: https://www.tensorflow.org/tutorials/estimators/cnn
[20] Keras. (2018). *Keras: The Python Deep Learning Library*. Accessed: Aug. 18, 2020. [Online]. Available: https://keras.io
[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2014, pp. 1–15. Accessed: Aug. 18, 2020. [Online]. Available: https://arxiv.org/abs/1412.6980
[22] H. Okuda, S. Juman, A. Ueda, T. Miki, and M. Shima, "Factors related to prevalence of hallux valgus in female university students: A cross-sectional study," *J. Epidemiol.*, vol. 24, no. 3, pp. 200–208, 2014.
[23] E. Sobel, S. Levitz, M. Caselli, Z. Brentnall, and M. Q. Tran, "Natural history of the rearfoot angle: Preliminary values in 150 children," *Foot Ankle Int.*, vol. 20, no. 2, pp. 119–125, Feb. 1999.
[24] G. S. Murley, H. B. Menz, and K. B. Landorf, "A protocol for classifying normal- and flat-arched foot posture for research studies using clinical and radiographic measurements," *J. Foot Ankle Res.*, vol. 2, no. 1, pp. 1–13, Dec. 2009.
[25] 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, no. 1, pp. 1929–1958, 2014.
[26] Wikipedia. (2020). *High-Heeled Shoes*. Accessed: Aug. 13, 2020. [Online]. Available: https://zh.wikipedia.org/wiki/%E9%AB%98%E8%B7%9F%E9%9E%8B

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. From May 2007 to June 2007, she was an Observer of geriatric medicine with Guy's and St Thomas' Hospitals, U.K. Since 2008, she has been an Attending Physiatrist of physical medicine and rehabilitation with the Taipei Veterans General Hospital, Taiwan. Her research interests include the areas of physical medicine and rehabilitation, especially in smart medical rehabilitation.

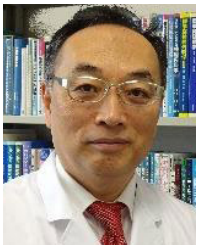
BOR-SHING LIN (Senior Member, IEEE) received the B.S. degree from National Cheng Kung University, Taiwan, in 1997, and the M.S. and Ph.D. degrees from National Taiwan University, Taiwan, in 1999 and 2006, respectively, all in electrical engineering. From December 2007 to June 2009, he served as a Principal Engineer with MStar Semiconductor. Since 2009, he has been an Assistant Professor with the Department of Computer Science and Information Engineering, National Taipei University, Taiwan. He is currently a Professor with the Department of Computer Science and Information Engineering, National Taipei University. He is also the Director of the Computer and Information Center, National Taipei University. His research interests include the areas of smart medicine, embedded systems, wearable systems, biomedical signal processing, biomedical image processing, and portable biomedical electronic system design.

HSIANG-CHEN LEE is currently pursuing the bachelor's 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.

XIAO-WEI HUANG is currently pursuing the bachelor's 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.

YA-CHU CHI is currently pursuing the bachelor's 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-SHYH LIN (Senior Member, IEEE) received the B.S. degree from National Chiao Tung University, Hsinchu, Taiwan, in 1997, and the M.S. and Ph.D. degrees from the Institute of Electrical Engineering, National Taiwan University, Taipei, Taiwan, in 1999 and 2006, respectively. He is currently a Distinguished Professor and the Director of the Institute of Imaging and Biomedical Photonics, National Yang Ming Chiao Tung University. His current research interests include biomedical circuits and systems, biomedical signal processing, and biosensor. He is also a Fellow of the Institution of Engineering and Technology (IET), U.K.



KAORU ABE is currently a Professor with the Department of Prosthetics, Orthotics and Assistive Technologies, Niigata University of Health and Welfare, Niigata, Japan. His research interests include the areas of practice of shoes human science, shoes adjustment science, development and evaluation of functional arch support (shoe insole) and footwear, lower extremities orthosis for hemiplegia, and foot pressure (F-scan) analysis.