Development of deep clustering model to stratify occurrence risk of diabetic foot ulcers based on foot pressure patterns and clinical indices

Xuanchen Ji[a], Yasuhiro Akiyama[a], Hisae Hayashi[b], Yoji Yamada[a], and Shogo Okamoto[a]

[a]Graduate School of Engineering, Nagoya University,Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan

ji.xuanchen@d.mbox.nagoya-u.ac.jp, akiyama-yasuhiro@mech.nagoya-u.ac.jp,

yoji.yamada@mae.nagoya-u.ac.jp, shogo.okamoto@mae.nagoya-u.ac.jp

[b]Faculty of Rehabilitation and Care, Seijoh University, 2-172 Fukinodai, Tokai City, 476-8588, Japan

hisaehayashi@seijoh-u.ac.jp

Abstract

In recent years, the number of patients suffering from diabetes mellitus has continued to increase. When diabetes becomes severe, ulcers may form on the feet of the patient. In the past few years, several researchers have focused on the risk factors and avoidance of ulceration. One effective method to predict the occurrence of diabetic foot ulcers is developing a machine learning model. However, few studies combine both clinical indices and mechanical data as the attributes of the training datasets. In this study, we developed a composite model of a convolutional neural network and K-means clustering to extract features from diabetic patients with or without ulceration as well as healthy individuals. Using a deep clustering model, the center of pressure (CoP) trajectory images were divided into three clusters. Furthermore, we evaluated the performance of the clustering by extracting the features from the CoP trajectory images in each cluster and combining them with the clinical indices of the patients. The results showed that patients with ulcers when walking tend to contact the ground with a narrow area of the plantar and apply a small force. Furthermore, it was found that patients undergoing diabetic neuropathy or with a toe amputation have a high potential of suffering from ulcers.

1. Introduction

The number of diabetic patients continues increasing especially in recent years. According to the International Diabetes Federation (IDF), the number of global diabetes patients reached 41.5 million in 2015 and is expected to reach 64.2 million in 2040 [1]. Diabetes can cause many other diseases, such as cardiovascular diseases [2], blindness [3] and stroke [4]. Among these, diabetic foot ulcer (DFU) is the most disabling complication arising from diabetes. Approximately 40% of patients with DFU experience a recurrence within a year, and this proportion increases to approximately 60% in 3 years and 65% in 5 years [5]. Foot deformities, duration of diabetes, and clinical treatment are responsible for the development of ulceration [6]. Preventive measures can be adopted by both patients and doctors. In therapeutic treatment, providing patients with preventive medicine and well-informed nursing care can reduce the incidence of DFUs by approximately 66% [7]. However, gait variations and restrictions in subtalar and first metatarsophalangeal joint are found in persons with diabetic neuropathy even before the onset of foot deformity, which can be regarded as the early risk marker of DFUs [8]. Therefore, regular evaluation of gait patterns is believed to be a more effective measure as patients can take the necessary precautions to prevent the development of DFUs [9]. For this purpose, it is important to observe the clinical and dynamical feature among DFUs patients, normal patients and healthy people, to distinguish them from each other by creating models each of which has high correlation with the corresponding feature.

Numerical studies on the discussion of diabetes and DFUs feature have focused on various types of methods and datasets. Available clinical information was commonly combined to predict DFUs occurrence using backward stepwise elimination [10]. Borzouei and Soltanian identified the most important demographic risk factors for Type 2 diabetes mellitus using a neural network model [11]. Botros *et al.* focused on the temporal dynamic plantar pressure and estimated the occurrence risk of DFUs using a support vec-

978-1-7281-9186-7/20/$31.00

tor machine (SVM) [12]. Bennetts *et al.* made an effort on using K-means cluster analysis to determine characteristic regional peak plantar pressure distributions for diabetic patient classification that may be useful for footwear prescription [13]. Sawacha *et al.* aimed to verify if three-dimensional gait data obtained from a high-resolution motion capture would be able to distinguish the diabetic neuropathy(DPN) patients and healthy people [14]. However, high-resolution measurement devices always require high cost and are hard to be diffused and few researchers have focused on the composite influence of patients' clinical indices and gait patterns on their occurrence risk of ulcer. The objective of the present study is to develop a composite model of convolutional neural network (CNN) and K-means clustering to classify and distinguish the feature among diabetic patients with ulceration, normal diabetic patients and healthy people using different types of datasets, including their body indices, anamnesis and gait pressure distributions obtained from four-sensor measurement.

The remainder of this paper is organized as follows. In Section 2, we describe the materials and methods used in this study. In Section 3, we analyze the processing and output of the center of pressure (CoP) trajectory images [15] using a deep embedded clustering model [16]. The results and evaluation of the model are displayed in Section 4. Discussions and limitations are presented in Section 5, and we conclude with a summary of our findings in Section 6.

2. Materials and methods

The gait data and clinical indices of patients and healthy individuals were collected from several hospitals and pre-processed for further clustering. An unsupervised method can distinguish between areas that may not be obvious to the human eye. To investigate the features of the gait patterns of the subjects, a composite model of a CNN and K-means clustering was developed as an unsupervised clustering model. After classifying the dataset, we extracted several parameters from the gait data and clinical indices and conducted the Kruskal–Wallis test and chi-squared test to investigate the features and differences among the clusters.

2.1. Data acquisition

Since March 2019, we have been collecting the clinical indices and gait data of diabetic patients from the datasheets of a prospective trial conducted at the Nagoya Kyoritsu Hospital, Masuko Memorial Hospital, Matsunami General Hospital, and Seijoh University. The data is in an anonymized manner and will not be made public. 73 patients have enrolled in this experiment: 26 females (age, 71.0 $\pm$ 10.3 years; height, 150.6$\pm$ 6.7 cm; weight, 51.1 $\pm$ 10.3 kg), and 47 males (age, 71.5 $\pm$ 9.7 years; height, 165.9 $\pm$ 6.9 cm; weight, 62.9 $\pm$ 11.9 kg). Among these, 23 individuals were found to have ulcerations at the timing of measurement while the others did not. Among them, 2 individuals developed ulceration on both two feet. Therefore, 25 feet were found to get the ulceration. Moreover, we also conducted the same measurement of gait data and body indices on 25 healthy persons since August 2018. Among them, 20 females (age, 74.0$\pm$6.4 years; height, 151.3$\pm$7.1 cm; weight, 50.7$\pm$5.4 kg), and 5 males (age, 77.8$\pm$4.7 years; height, 164.6$\pm$4.1 cm; weight, 61.4$\pm$7.8 kg). We processed gait data of healthy persons and patients at the same epoch to investigate the feature of foot patterns

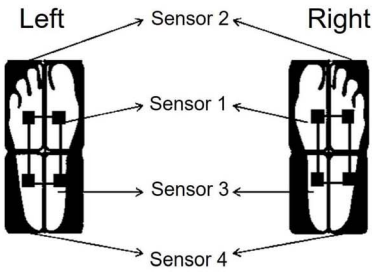

Figure 1: Sensor distribution.

The device used to measure the gait data was a pair of force-measuring shoes (Balance Aid, Leimac Ltd., Japan). The sole of each shoe was divided into four parts. A force sensor was installed on each part to obtain the ground reaction force (GRF) data (Figure 1). The subjects were required to wear these force-measuring shoes and walk for 10 m in a straight line, with and without a walking aid. The sensors on the soles recorded the GRF data of the subjects as they walked. However, the data were obtained only in the vertical direction to ensure realistic data collection.

2.2. Architecture of deep clustering

After collecting the gait and clinical information of the subjects, we transformed the gait patterns to images. For further investigation of the gait features, deep clustering, which was constructed using K-means clustering and a CNN, was used to divide the image dataset into several clusters by an unsupervised method. This is because it can fit small datasets well and provide spectrally distinct areas, which may not be notable to a human eye. Because three types of subjects were included in the dataset, we defined the number of clusters as three.

2.2.1 Convolutional autoencoders

The CNN is an improvised variant of a multilayer perceptron. The most extensively used neural network in deep clustering algorithms is a convolutional autoencoder (CAE); it can be trained in an end-to-end manner and was designed for learning features from unlabeled images [17]. The designed CAE is superior to stacked autoencoders ow-

ing to the incorporation of the spatial relationships between the pixels in the images.

A CAE is generally composed of two types of layers, corresponding to encoder $f_{\mathbf{W}}(\cdot)$ and decoder $g_{\mathbf{U}}(\cdot)$ respectively. It aims to identify a code for each input sample by minimizing the mean squared error (MSE) between the input and the output over all the samples. For a fully connected autoencoder,

$$\begin{aligned} f_{\mathbf{W}}(x_i) &= \sigma(\mathbf{W}x) \equiv h \\ g_{\mathbf{U}}(h) &= \sigma(\mathbf{U}h) \end{aligned} \tag{1}$$

where an m-dimensional vector h, which is regarded as embedded code of layers, serves as the new representation of input sample x, and σ is activation function. The activation functions used in our study are ReLU and sigmoid. After training, the embedded code h is regarded as the new representation of input image samples.

2.2.2 K-means clustering

K-means clustering is an unsupervised model and a method of vector quantization. Given a set of data samples $x_i (i = 1, \cdots, N)$ where $x_i \in \mathbb{R}^M$, the aim of this method is to partition N observations into K clusters, with each observation belonging to the cluster with the nearest mean. K-means clustering achieves this by optimizing the following cost function:

$$\begin{aligned} &\min_{\mathbf{Q}\in\mathbb{R}^{M\times K}, s_i\in\mathbb{R}^K} \sum_{i=1}^{N} \|x_i - \mathbf{Q}s_i\|_2^2 \\ &s.t.\ s_{i,j} \in \{0,1\}, \mathbf{1}^T s_i = 1\ \forall i, j, \end{aligned} \tag{2}$$

where $\mathbf{1}$ is the all-one column vector, s_i is the assignment vector of data point i which has only one non-zero element, $s_{i,j}$ denotes the jth element of s_i ($1 \le j \le k$), and the kth column of $\mathbf{Q}$, i.e., q_k, denotes the centroid of the kth cluster.

2.2.3 Overview of deep clustering model

The proposed deep convolutional embedded clustering model iteratively groups visual features using a standard clustering algorithm. Moreover, it uses subsequent assignments as supervisions to update the weights of the network. Its structure is composed of a CAE and a clustering layer, as shown in Figure 2. The clustering model in this study contains three convolutional layers in the encoder layers. We input 400 × 400 original images to the model, and finally, obtain an output from the encoder layers of size 100 × 100, which are regarded as the features of the images and input of the K-means clustering part. Subsequently, the K-means clustering model is used to classify the images into several clusters according to their features.

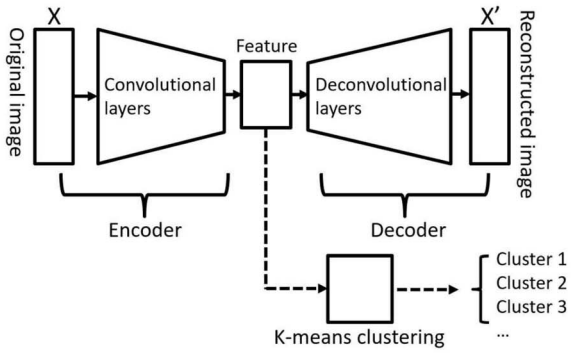

Figure 2: Structure of deep embedded clustering model. CAE contain encoder and decoder layers. The original image X is embedded using encoder layers to obtain the feature image, which can be reconstructed and recovered as X'.

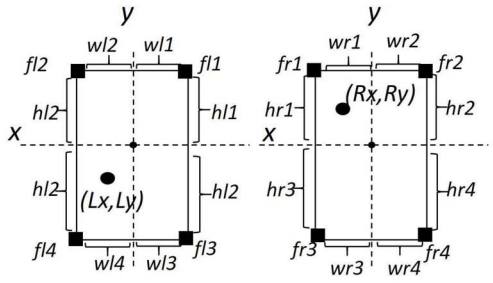

Figure 3: Position and pressure parameters of sensors. wr_p and hr_p are distances of right sensor p from the vertical and horizontal central axes, respectively, while fr_p represents pressure of sensor p. Parameters of the left foot are similarly labeled.

3. Analysis of GRF data

The GRF data of the patients obtained from the experiment constitute a significant factor reflecting their ulcer condition. To confirm the ulcer position in the sole of a patient, the CoP trajectory was used to explain the plantar pressure distribution. In this study, we obtained the CoP trajectory of each patient from the GRF data and clustered the trajectory images into three groups using the deep clustering model. The CoP trajectory was determined based on the position and pressure value of each sensor. Figure 3 illustrates the position and pressure parameters of the sensors. The coordinates of CoP are defined according to the following formulae:

$$\begin{aligned} R_x &= \frac{fr2\cdot wr2 + fr4\cdot wr4 - fr1\cdot wr1 - fr3\cdot wr3}{\sum_{p=1}^{4} fr_p} \\ R_y &= \frac{fr1\cdot hr1 + fr2\cdot hr2 - fr3\cdot hr3 - fr4\cdot hr4}{\sum_{p=1}^{4} fr_p} \end{aligned} \tag{3}$$

Where, R_x and R_y represent the CoP coordinates of the right foot, wr_p and hr_p represent the distances of sensor

Category	Cluster 1 n =111	Cluster 2 n =56	Cluster 3 n =29
Patients' feet with ulcer	4 feet	7 feet	14 feet
Patients' feet without ulcer	65 feet	42 feet	14 feet
Healthy persons' feet	42 feet	7 feet	1 feet

Table 1: Clustering result

p on the right foot from the vertical and horizontal central axes, respectively. fr_p which is normalized by patient's weight, represents the pressure value of sensor p. The parameters of the left foot are labelled similarly by replacing R and r with L and l respectively in formula (4) and (5).

The CoP trajectory images were drawn with reference to the CoP coordinates obtained from the GRF data. They were prepared as follows: (1) Four steps each from the start and end of the walking duration were selected. The CoP coordinates of each step were then calculated, (2) The CoP trajectory figure of each phase was drawn and placed into a 3 × 3 panel picture using MATLAB, (3) The plot size was changed to express the magnitude of pressure.

Figure 4 shows samples of the CoP trajectories of two different types of subjects, which reflect their different gait conditions. The subjects started to walk at t_0 and stopped at t_e. The number of plots shows the time period of the current step while the size of each plot reflects the magnitude of pressure. Large plots represent high pressure. The pressure values were standardized with reference to the weight of each subject. A comparison between the left foot trajectories of a healthy person and a patient is shown in Figure 4. The gait of the healthy person presents a more regular shape, from which the parts of the foot from heel to toe can be discerned, while that of the patient is too irregular to present any recognizable part of the foot. Before clustering, all the CoP trajectory images were pre-processed to obtain the following second moment of area about the geometric center, C_{xy}, of CoP trajectory in each image.

$$V_{rx} = \sum_{t=t_0}^{t_e} (R_x(t) - C_{xy})^2$$
$$V_{ry} = \sum_{t=t_0}^{t_e} (R_y(t) - C_{xy})^2 \quad (4)$$

Where V_{rx} and V_{ry} are the second moment of area about C_{xy}, $R_x(t)$ and $R_y(t)$ represent R_x and R_y in formula (4) and (5) at time t, respectively, similar manner of labeling holds for V_{lx} and V_{ly}.

4. Results and feature analysis

The CoP trajectory images obtained from the gait data were divided into three clusters using deep clustering

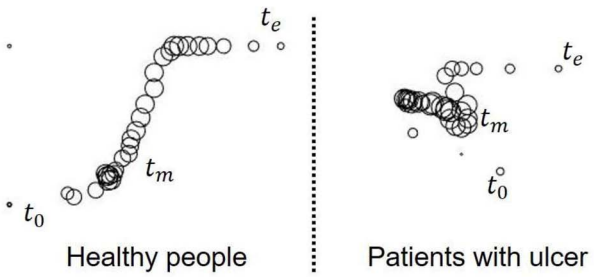

Figure 4: Comparison between CoP trajectories of healthy subjects and patients with ulcer. t_0 is the time at which the subjects start walking. t_e is the time when the subjects stop walking. $t_m = 1/2 * (t_0 + t_e)$

model. To determine the trajectory features of each cluster, trajectory images of 25 healthy persons were processed in the same period. Thus, the dataset used for clustering contains information on 73 patients and 25 healthy persons. Right and left feet of all the participants subjects are processed as dataset.

4.1. Clustering results

We input 196 CoP trajectory images into the deep clustering model, and the clustering results are summarized in Table 1. It is noteworthy that there exists a possibility that the right and left feet were divided into two different clusters, even though they belong to one subject category. As summarized in Table 2, Cluster 1 contains 4 ulcer feet of patients, 65 no-ulcer feet of patients, and 42 feet of healthy persons. Cluster 2 contains 7 ulcer feet, 42 no-ulcer feet, and 7 healthy feet. Compared with Clusters 1 and 2, Cluster 3 has 14 ulcer cases, which is the most among the three clusters, 14 no-ulcer feet, and no healthy foot.

4.2. Feature extraction

Each of the CoP trajectory images contains eight steps in total, and the gait patterns may differ from each other. To investigate the concrete differences among the three clusters more accurately, several parameters of the patients representing the cluster features were extracted from the gait data for each step.

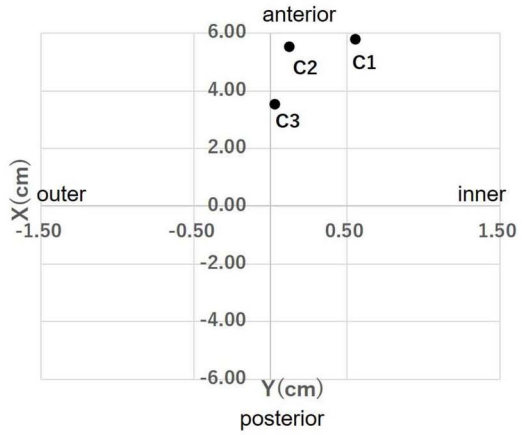

Figure 5: X-Ycoordinates of P_c: P_c is the CoP plot at the timing when opposite foot contacts the ground. C1 represents Cluster 1 (0.55***, 5.81***), C2 represents Cluster 2 (0.12, 5.55*) while C3 represents Cluster 3 (0.03, 3.55). *=p-value <0.05, **=p-value <0.01, ***=p-value <0.001 vs Cluster 3 Kruskal-Wallis test

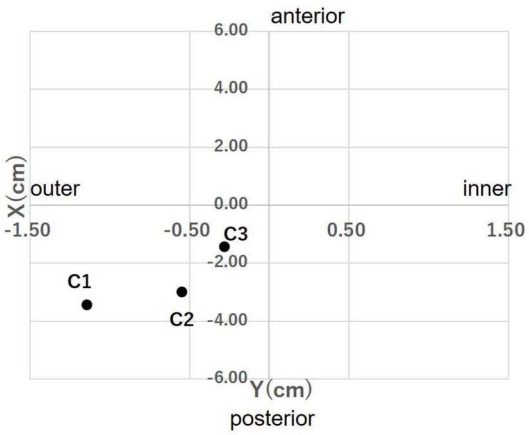

Figure 6: X-Ycoordinates of P_f: P_f is the CoP plot at the timing when the opposite foot starts swinging. C1 represents Cluster 1 (-1.14***, -3.42**), C2 represents Cluster 2 (-0.55, -2.98) while C3 represents Cluster 3 (-0.28, -1.43). *=p-value <0.05, **=p-value <0.01, ***=p-value <0.001 vs Cluster 3 Kruskal-Wallis test

4.2.1 X-Y coordinates of CoP plot

The x-y coordinates of the CoP plot at a specific timing can be a typical feature of gait patterns. This parameter shows the position of the CoP that reflects the pattern of applying the body mass on foot. In this study, we extracted two types of coordinates at different times to investigate the features among clusters. The first is defined as the x-y coordinates of the CoP plot when opposite foot contacts the ground in one gait cycle. We denote this plot as P_c. The second is defined as the x-y coordinates of the CoP plot at the time the opposite foot starts swinging, and this point is denoted as P_f. We computed the average value of the coordinates in each cluster shown in Figures 5 and 6, and p-value was computed using Kruskal-Wallis test.

4.2.2 Gait parameters

To investigate the differences among the clusters, we extracted several gait parameters that reflect the value of the GRF and walking speed, including peak GRF in one gait cycle, and the period of one gait cycle. The peak GRF is normalized by each subject's weight and is defined by the following equation:

$$Peak = \frac{Ground\ Reaction\ Force_{max}}{Weight} \quad (5)$$

Moreover, we tried to investigate the variance of GRF distribution by computing the second central moment of it, which is defined as 'Moment_2' according to the equation below:

$$Moment\_2 = \sum_{p=1}^{n}((X_p - X_c)^2 + (Y_p - Y_c)^2) \cdot F_p \quad (6)$$

Where, X_c and Y_c are the geometric center of the whole CoP trajectory, X_p and Y_p are the X-Y coordinates of each CoP while F_p is the GRF of it. Noted that GRF is normalized by patient's weight. This parameter describes GRF distribution.

The gait parameters of each cluster are provided in Table 2. Moreover, Kruskal-Wallis test was utilized to check the significant differences of these parameters among three clusters

4.2.3 Clinical indices

In addition to gait parameters, clinical indices of subjects in each cluster also have a significant influence on gait patterns. After evaluating several clinical indices, we selected 5 parameters including age, Body Mass Index (BMI), whether use walking aids or not, condition of Diabetic Peripheral Neuropathy (DPN), toe amputation and whether feel a pain in feet or not. Table 3 displays the clinical indices of subjects in three clusters, we computed the mean and standard deviation of each index. Kruskal-Wallis test was utilized to check the significant differences of age and BMI among three clusters; As for walking aids, DPN, toe amputation and pain, Chi-Squared Test was conducted to examine the p-value two clusters by two clusters.

5. Discussion and limitation

Based on this clustering result in Table 1, we hypothesize that the samples in Cluster 1 has the lowest potential of suffering from DFUs, whereas those in Cluster 3 have the highest potential for this. In this section, we discuss the

Category	Cluster 1 n =111	Cluster 2 n=56	Cluster 3 n=29	p-value
Peak Ground Reaction Force	0.99±0.20*	0.95±0.18	0.90±0.18	<0.05
Period(s)	1.16±0.20**	1.22±0.22	1.30±0.25	<0.05
Moment_2(cm^2)	430.43±233.82***	385.47±132.78***	209.80±150.03	<0.001

Table 2: Gait Parameters: *=p-value <0.05, **=p-value <0.01, ***=p-value <0.001 vs Cluster 3 Kruskal-Wallis test

Category	Cluster 1 n =111	Cluster 2 n=56	Cluster 3 n=29	χ^2	p-value
Age(years)	72.39±8.74	71.66±9.01	72.52±11.40		0.795
BMI(kg/m^2)	22.83±3.46	22.80±2.57	20.99±3.84		<0.05
Walking Aids(Yes)	23 cases (20.7%)	11 cases (19.6%)	10 cases (34.5%)	2.856	0.240
Diabetic peripheral neuropathy(Yes)	31 cases (44.9%)*	27 cases (55.1%)	20 cases (83.3%) †	10.610	0<0.01
Toe Amputation(Yes)	4 cases (5.7%)*	4 cases (7.7%)*	15 cases (53.5%)†	38.182	<0.001
Pain(Yes)	4 cases (3.6%)*	3 cases (5.4%)	6 cases (20.6%)†	38.182	<0.01

Table 3: Clinical indices: Applied Kruskal-Wallis test on age and BMI; Applied Chi-Squared Test on walking aids, DPN, toe amputation and pain; Adjusted residuals *:≤ -1.96, †: 1.96≥

features of the three clusters based on this hypothesis from the perspective of the x–y coordinates of the CoP plot, gait parameters, and clinical indices, respectively.

5.1. X-Y coordinates of CoP plot

Under the condition of P_c, the position of Cluster 3 is quite different from others especially the y coordinates of Cluster 3, which is lower than other two clusters, as shown in Figure 5. This distribution indicates that when the opposite foot contacts the ground, the CoPs of healthy persons and low-risk patients are already reached the toe area, whereas the CoP of a patient with ulcers is still in the center area of the soles. Furthermore, p-value of x-y coordinates of P_c shows that x-y coordinates of Cluster 3 have a significant difference from others, which can support the hypothesis that subjects in Cluster 3 are quite different from other 2 clusters in terms of gait patterns

Under the second condition, when we extracted x-y coordinates of P_f, all the plots are distributed in the posterior of the sole. The position of Cluster 3 is distributed in the center part of sole, while in Cluster 1 and Cluster 2 are distributed on the outer side. This phenomenon indicates that patients with ulcers are more likely to apply their body weight to the center area of the sole but not the heel part like healthy persons and low-risk patients when the opposite foot starts swinging, similar to that in the previous condition. Similar results that stride length of diabetic foot ulcer patients is reduced are proposed by Drerup *et al.* [18]. The p-value of x-y coordinates of P_f indicates this parameter has a significant difference among the three clusters.

Based on the results shown in Figure 5 and Figure 6, it can be inferred that patients with ulcers apply body weights to a limited area mainly in the center part of soles while healthy persons and low-risk patients tend to use the whole sole when walking.

5.2. Gait parameters

From Tables 1 and 2, it is clear that the peak GRF of Cluster 1 is larger than those of the other two clusters, whereas its period is the shortest. This indicates that the feet of healthy persons or affected patients having a lower risk of suffering DFUs tend to apply a larger force on their foot and walk faster.

Cluster 3 has the lowest peak GRF and longest period, indicating that people with serious DFUs tend to walk slowly. This phenomenon agrees with Youdas's study [19]. One possible reason for this phenomenon is that healthy persons and patients without ulcers do not have difficulties in walking; therefore, they tend to apply more force on their feet.

As for the 'Moment_2', the values of Clusters 1 and 2 are far larger than those of Cluster 3, and there is a significant difference among the three clusters in this parameter. The results of this parameter show that the GRF distributions along the CoP trajectories of Clusters 1 and 2 are wider than that of Cluster 3. This indicates that the feet of patients having a high risk of developing ulcers tend to apply force on a narrow area around the geometric center of the plantar. In comparison, generally, healthy feet and low-risk feet will apply a force from the heel to the toe when walk-

ing [13]. This result also supports the findings regarding the X–Y coordinate distribution presented in Section 4.2.1.

5.3. Clinical indices

It can be inferred through Table 3 that Cluster 3 has the most subjects who have DPN and got toe amputation. Besides, 20.6% of them felt a pain in their feet. The result corresponds to the hypothesis that the subjects in Cluster 3 have the highest risk of suffering from DFUs. These indicates that the DPN, toe amputation, and pain significantly influence the risk of contracting DFUs [10] [20] [21]. However, as for age and walking aids, it seems no significant difference based on this dataset.

5.4. Feature conclusion

We can concluded the feature of each cluster according to the results in section 4.2.1-4.2.3 that:

1. Cluster 1 contains the most healthy persons and least feet with ulcer samples which is shown in Table 1. Thus, we can consider that subjects in Cluster 1 is under the lowest risk of contracting DFUs. Based on this hypothesis, Table 2 shows that most patients tend to apply larger force on their foot and walk faster. The GRF distributions of feet in this cluster are longer than those in other clusters in y direction, showing that healthy feet and low-risk feet apply force more widely than feet with ulcer. Moreover, according to the results in Table 3, the subjects in this cluster have the lowest percentage of having DPN, getting toe amputation or feeling a pain in feet. This indicates they are in a low risk of getting DFUs in clinical explanation [10].

2. According to Table 1, most samples in Cluster 2 are patients' feet without ulcer, with the percentage of 75%. As for the gait parameters and clinical indices in Table 2 and Table 3, the values of Cluster 2 are always placed between Cluster 1 and Cluster 3. Among the no-ulcer feet in Cluster 2, the opposite feet of 26.2% of them are classified to Cluster 3, which is regarded as a high-risk cluster. It indicates that if opposite foot is a high-risk foot, it will influence gait pattern of no-ulcer foot [22]. Based on this phenomenon, we consider that Cluster 2 is a preliminary group of Cluster 3. However, since right and left feet of 1 individual are processed separately, the laterality of them becomes meaningless. To deal with the problem, the laterality study of subjects' gait and utilization of supervised model will be conducted as the plan of next stage.

3. Based on the clustering results shown in Table 1. We proposed the hypothesis that Cluster 3 is considered as the group with the highest potential of having ulcers. Through Figure 5∼6, it is known that subjects in Cluster 3 tend to apply weight to the center part of the plantar at the timing when the opposite foot starts swinging and contact ground. Comparing gait parameters with the results for other clusters in Table 2, subjects are more likely to apply lower force on their foot and walk more slowly. 'Moment_2' indicates patients' feet with a high potential of having an ulcer tend to apply force to a narrow area around geometric center on plantar, as they cannot use ulcer area to contact ground *et al.* [13] [18]. The result of 'Moment_2' supports the distribution results shown in Figure 5-6. According to clinical indices displayed in Table 3, 53.5% of subjects got toe amputation , 83.3% of them have trouble with DPN and 20.6% of them felt a pain in their feet. These percentages are the highest compared with other clusters. This result supports the hypothesis that cluster 3 has the highest potential of having DFUs, proposed at the beginning of this paragraph.

However, the clustering results show that the proposed model has several limitations such as the feature and difference of age, BMI and whether use walking aids or not among the three clusters can't be reflected and three types of samples: ulcer cases, no-ulcer cases and healthy persons are not classified into three clusters completely. One possible reason is that the resolution of device used in the experiment is not high enough and there are only 25 feet with ulcers among the 73 participants at the time of measurement. The distribution of the dataset is irregular. Thus, to improve the dataset and test the stability of the model, we plan to enlarge it in the future by conducting more experiments and collecting more information in both clinical and dynamical field.

6. Conclusion

This study primarily aims to develop a machine learning model to investigate the differences among DFU patients, diabetic patients without DFUs, and healthy persons using information such as their clinical indices and gait patterns. To achieve this objective, we proposed a composite model of a CNN and K-means clustering to classify the CoP trajectory images of the subjects and subsequently extracted the gait and clinical parameters to discuss the clustering features. The clustering results show that patients with DFUs are more likely to apply smaller forces on their feet and walk slowly. The areas of the plantar used to contact the ground is narrow during walking. They also tend to apply weight on the inner sides of the plantar when the pressure attains the peak value. In clinical explanation, patients who underwent toe amputations, have difficulty with DPN, or feel a pain in their feet have a higher potential of suffering from DFUs. However, some clinical indices remain inconclusive as they are unexplainable from the clustering

results. Hence, we aim to conduct additional experiments to improve the dataset and collect more information to indentify more influencing factors.

7. Acknowledgments

We gratefully acknowledge the support of Nagoya Kyoritsu Hospital, Masuko Memorial Hospital and Matsunami General Hospital and Mr. Takayasu Koike with the donation of the patients' data used for this research and the support of Mr. Takaaki Namba for his technical contribution about machine learning. A part of research was supported by The Japan Health Foundation, Grant-in-Aid for Scientific Research (C)(19K12915)

References

[1] K. Ogurtsova, J. da Rocha Fernandes, Y. Huang, U. Linnenkamp, L. Guariguata, N. H. Cho, D. Cavan, J. Shaw, and L. Makaroff, "IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040," *Diabetes research and clinical practice*, vol. 128, pp. 40–50, 2017.

[2] N. J. Tripolt, S. H. Narath, M. Eder, T. R. Pieber, T. C. Wascher, and H. Sourij, "Multiple risk factor intervention reduces carotid atherosclerosis in patients with type 2 diabetes," *Cardiovascular diabetology*, vol. 13, no. 1, p. 95, 2014.

[3] World Health Organization, "Prevention of blindness from diabetes mellitus: Report of a who consultation in geneva, switzerland, 11 november, 2005," *Switzerland: World Health Organization*, 2006.

[4] A. Tuttolomondo, C. Maida, R. Maugeri, G. Iacopino, and A. Pinto, "Relationship between diabetes and ischemic stroke: analysis of diabetes-related risk factors for stroke and of specific patterns of stroke associated with diabetes mellitus," *Journal of Diabetes and Metabolism*, vol. 6, no. 5, pp. 544–551, 2015.

[5] T. del Pino-Sedeño, M. M. Trujillo-Martín, I. Andia, J. Aragón-Sánchez, E. Herrera-Ramos, F. J. Iruzubieta Barragán, and P. Serrano-Aguilar, "Platelet-rich plasma for the treatment of diabetic foot ulcers: A meta-analysis," *Wound Repair and Regeneration*, vol. 27, no. 2, pp. 170–182, 2019.

[6] Z.-H. Huang, S.-Q. Li, Y. Kou, L. Huang, T. Yu, and A. Hu, "Risk factors for the recurrence of diabetic foot ulcers among diabetic patients: a meta-analysis," *International Wound Journal*, vol. 16, no. 6, pp. 1373–1382, 2019.

[7] J. M. Malone, M. Snyder, G. Anderson, V. M. Bernhard, G. A. Holloway Jr, and T. J. Bunt, "Prevention of amputation by diabetic education," *The American journal of surgery*, vol. 158, no. 6, pp. 520–524, 1989.

[8] S. Gnanasundaram, P. Ramalingam, B. N. Das, and V. Viswanathan, "Gait changes in persons with diabetes: Early risk marker for diabetic foot ulcer," *Foot and Ankle Surgery*, vol. 26, no. 2, pp. 163–168, 2020.

[9] W. J. Jeffcoate and K. G. Harding, "Diabetic foot ulcers," *The lancet*, vol. 361, no. 9368, pp. 1545–1551, 2003.

[10] E. J. Boyko, J. H. Ahroni, V. Cohen, K. M. Nelson, and P. J. Heagerty, "Prediction of diabetic foot ulcer occurrence using commonly available clinical information: the seattle diabetic foot study," *Diabetes care*, vol. 29, no. 6, pp. 1202–1207, 2006.

[11] S. Borzouei and A. R. Soltanian, "Application of an artificial neural network model for diagnosing type 2 diabetes mellitus and determining the relative importance of risk factors," *Epidemiology and health*, vol. 40, 2018.

[12] F. S. Botros, M. F. Taher, N. M. ElSayed, and A. S. Fahmy, "Prediction of diabetic foot ulceration using spatial and temporal dynamic plantar pressure," in *2016 8th Cairo International Biomedical Engineering Conference (CIBEC)*, pp. 43–47, IEEE, 2016.

[13] C. J. Bennetts, T. M. Owings, A. Erdemir, G. Botek, and P. R. Cavanagh, "Clustering and classification of regional peak plantar pressures of diabetic feet," *Journal of biomechanics*, vol. 46, no. 1, pp. 19–25, 2013.

[14] Z. Sawacha, G. Guarneri, A. Avogaro, and C. Cobelli, "A new classification of diabetic gait pattern based on cluster analysis of biomechanical data," *Journal of Diabetes Science and Technology*, vol. 4, no. 5, pp. 1127–1138, 2010.

[15] X. Li, H. Huang, J. Wang, Y. Yu, and Y. Ao, "The analysis of plantar pressure data based on multimodel method in patients with anterior cruciate ligament deficiency during walking," *BioMed research international*, vol. 2016, p. 7891407, 2016.

[16] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 132–149, 2018.

[17] S. K. Kumaran, D. P. Dogra, P. P. Roy, and A. Mitra, "Video trajectory classification and anomaly detection using hybrid CNN-VAE," *CoRR*, vol. abs/1812.07203, 2018.

[18] B. Drerup, K. Ch, A. Koller, and H. Wetz, "Reduction of plantar peak pressure by limiting stride length in diabetic patients," *Der Orthopade*, vol. 33, no. 9, pp. 1013–1019, 2004.

[19] J. W. Youdas, B. J. Kotajarvi, D. J. Padgett, and K. R. Kaufman, "Partial weight-bearing gait using conventional assistive devices," *Archives of physical medicine and rehabilitation*, vol. 86, no. 3, pp. 394–398, 2005.

[20] A. Hartemann-Heurtier, J. Robert, S. Jacqueminet, G. Ha Van, J. Golmard, V. Jarlier, and A. Grimaldi, "Diabetic foot ulcer and multidrug-resistant organisms: risk factors and impact," *Diabetic Medicine*, vol. 21, no. 7, pp. 710–715, 2004.

[21] D. J. Margolis, J. Kantor, J. Santanna, B. L. Strom, and J. A. Berlin, "Risk factors for delayed healing of neuropathic diabetic foot ulcers: a pooled analysis," *Archives of dermatology*, vol. 136, no. 12, pp. 1531–1535, 2000.

[22] B. Iraj, F. Khorvash, A. Ebneshahidi, and G. Askari, "Prevention of diabetic foot ulcer," *International journal of preventive medicine*, vol. 4, no. 3, p. 373, 2013.