Modeling of women shoes sizing system based on 3D foot scanner result using machine learning approach

Jamila[a*], Eka Legya Frannita[a], Gilang Fatikhul Burhan[a], Windra Bangun Nuswantoro[a], Anwar Hidayat[a], Erlita Pramitaningrum[a], and Totok Yulaidin[b]

[a]*Department of Leather Product Processing Technology, Politeknik ATK Yogyakarta, Yogyakarta, Indonesia, 55188*

[b]*Department of Management, Universitas Bina Sehat PPNI Mojokerto, Jawa Timur, Indonesia, 61363*

*corresponding author: jamila@atk.ac.id

Modeling of women shoes sizing system based on 3D foot scanner result using machine learning approach

Abstract. Accurate shoe sizing plays a crucial role in ensuring comfort, performance, and consumer satisfaction, particularly for women whose foot shapes exhibit considerable anatomical variability. To address this challenge, this research proposes a data-driven modeling framework for developing a women's shoe sizing system based on three-dimensional foot scanner data. The study was carried out through a systematic process consisting of data preprocessing, clustering using the K-Means algorithm, and evaluation of the clustering performance. The clustering analysis identified four optimal clusters within the dataset, representing distinct patterns in foot dimension measurements. The evaluation result, with a Silhouette Score of 0.25, indicates a moderate yet acceptable level of cohesion and separation among the clusters. These findings demonstrate that the proposed model can effectively capture the underlying structure of women's foot morphology, providing a scientific foundation for establishing more accurate, customized, and ergonomically appropriate shoe sizing standards.

Keywords: anthropometric characteristics, shoe sizing system, 3D foot scanner, machine learning

1. Introduction

Designing new shoes begins by creating the design of the shoe last. Shoe last can be defined as human foot prototype made by wood or metal that plays an essential role in shoe production process particlarly in the fitness of the final product. In the shoe production, appropriateness of shoe last model extremelly impact the comfortability of the end product. Hence, a pair of shoes is required to cover the biomecanic funtion to create comfortability and adorability of the resulted shoes (Lin & Chen, 2015; Frannita et al., 2024).

In shoe production, the shoe last should also represent the anthropometric characteristics of the human foot and match the foot size. However, the footwear sizing systems commonly used worldwide, such as the Mondopoint, United Kingdom (UK), United States (US), European (EU), and other internationally recognized standards, were primarily developed based on the anthropometric characteristics of European populations. These systems define shoe sizes using measurements of foot length and, in some cases, width, assuming relatively consistent proportions across individuals. In practice, however, the shape and dimensional proportions of the human foot vary significantly across populations due to differences in genetics, lifestyle, habitual footwear use, and environmental factors. The anthropometric data underlying these sizing standards were largely collected from Western populations, particularly those of European descent, where individuals tend to have longer, narrower feet with higher arches (Shariff et al., 2019; Prananda et al., 2024; Frannita & Hidayahtullah, 2024).

In contrast, studies have demonstrated that Asian populations, including Indonesians, exhibit different foot morphology that deviates from these European-based standards. The characteristics of Asian feet include a broader forefoot, flatter arches, and shorter overall foot length relative to height. These differences produce significant implications for footwear design and comfort. When shoes designed according to European standards are used by individuals with different foot characteristics, the result is often a mismatch between the labeled size and the actual fit experienced by the user. Such differences may cause extreme anxiety in certain areas of the foot, discomfort during walking, and even long-term musculoskeletal problems if the footwear is used regularly (Shariff et al., 2019; Jandova & Mendricky, 2021).

For Indonesians, this mismatch can manifest in several practical problems. For instance, individuals may require to choose a larger shoe size to accommodate their foot width, which can lead to excessive length and heel slippage. Conversely, choosing a shoe size that fits the length may result in excessive tightness around the forefoot or instep. These issues not only impact comfort but can also affect posture, gait, and balance, which can potentially lead to foot deformities or injuries over time. From a manufacturing and design perspective, relying solely on international sizing standards without considering local anthropometric variation may limit market suitability and user satisfaction (Mueller et al., 2023; Limon et al., 2023; Dash et al., 2024).

To handle these hurdles, it is important to develop region-specific footwear sizing systems that are grounded in accurate and comprehensive anthropometric data of the local population. In recent years, advancements in three-dimensional (3D) scanning and modeling technologies have made it possible to capture detailed foot morphology across large sample sizes efficiently. The use of such technologies enables researchers to analyze critical parameters, including foot length, width, girth, instep height, and arch shape, with high precision. The resulting datasets can be used to design more accurate sizing conversions or to establish new local sizing standards that better reflect the population's unique characteristics (Dash et al., 2024; Kim et al., 2022).

Furthermore, combining local anthropometric data into footwear design processes would have significant ergonomic and economic benefits. From an ergonomic perspective, it is suitable for improving comfort, reducing the risk of injury, and promoting natural biomechanical movement. From an industrial standpoint, it allows manufacturers to gain specific markets and improve competitiveness and customer satisfaction. Accordingly, our research work aims to develop a model of women's shoe sizing system based on 3D foot scanner results using a machine learning approach.

2. Data and Methodology

2.1. Dataset

The dataset utilized in this study is a primary dataset collected using a foot scanner device with specifications as described in Table 1. The data gathering process involved sixteen female respondents within the age range of 18 to 25 years. The foot scanner generated three-dimensional (3D) foot models completed by corresponding metadata, as illustrated in Table 2.

Table 1. Foot scanner specification

Category	Information
Model	IFU-S-01
Type	3 Dimensional Foot Scanner
Scanner Dimensions	685 (L) × 400 (W) × 310 (H) mm
Power	100–240V ±10%, 50/60Hz, 150W max
Weight	24 kg
Scan Area	400 (L) × 200 (W) × 150 (H) mm
Scanning Speed	30 mm/sec
Data Processing Time	5 seconds per foot
Data Accuracy	1mm step: Y(0), Z(±0.1mm), X(±2mm); 0.5mm step: Y(0), Z(±0.1mm), X(±1mm)
Data Format	fbd (original data format)
Scan Target	Human foot & landmark points
Weight Limit	200 kg per person
Prohibition	Absorption and reflection of laser light
Operating Temperature	15°C – 40°C
Operating Humidity	35% – 60%
Placement	Indoor use only, flat floor
Supported OS	Windows 7 & 8 (32/64 bit)
CPU Requirement	Dual Core 1.66 GHz or higher
RAM Requirement	2 GB or more
USB Ports	2 × USB 2.0 ports (Scanner & Dongle)
Storage Requirement	80 GB HDD or more
Optical Drive	DVD±R / RW Drive
Monitor Resolution	1024 × 768 or higher

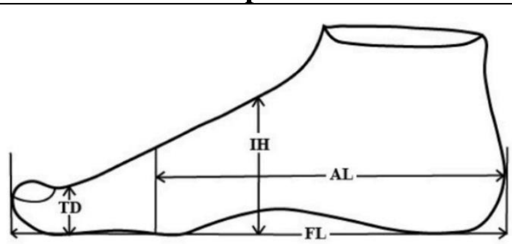
Category	Information
Main Software	Tester (scan conditioning), Measure (foot measuring), DB Controller (scan data)
Optional Accessories	Handrail, Calibration Tool, Carry Case (Pelican 1660), Marker, Foot Cover, Foot Stage
Optional Software	Dij+ (Homologous Foot Modeling), File Converter (fbd to DXF/CSV/STL/VRML)
PC Required	Yes, PC is required to operate the foot scanner
Manufacturer	I-Ware Laboratory Co., Ltd., Osaka, Japan

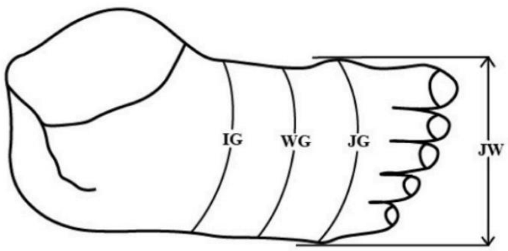
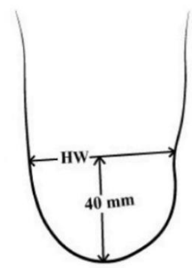
Table 2. Design of test problems

Component	Description
Type of data	3D image
Number of data	The dataset consists of 32 3D foot images, comprising 16 left foot and 16 right foot scans corresponding to each respondent
Metadata	foot length (FL), arch length (AL), instep height (IH), toe depth (TD), joint width (JW), joint girth (JG), instep girth (IG), waist girth (WG), heel width (HW).

Metadata in Table 2 illustrates nine key foot dimensions that represent critical anatomical points to be carefully considered in the analysis. These dimensions provide essential quantitative information for understanding foot morphology and serve as the fundamental parameters in evaluating fit, comfort, and ergonomic compatibility in footwear design. Table 3 illustrates the detailed position of each point.

Table 3. The detail description of each point (Limon et al., 2023).

Component		Description
	TD	Toe depth is determined by measuring the vertical distance from the ground surface to the uppermost point of the first metatarso-phalangeal joint.
	IH	Instep height is measured at the midpoint of the foot length, starting from the pternion as the reference point.

Component		Description
	AL	Arch length is calculated along the Brannock axis, extending from the pternion to the most medial prominence of the first metatarsal head
	FL	Foot length is calculated is from the pternion to the tip of the longest toe, representing the overall foot length along the longitudinal axis
	IG	Instep girth is calculated across the prominence of the middle cuneiform bone
	WG	Waist girth is calculated approximately at the midpoint of the metatarsal bones, taken in a vertical plane perpendicular to the Brannock axis, representing the midfoot girth
	JG	Joint girth is measured around the metatarso-phalangeal joints of both the big toe and the little toe, capturing the girth across the forefoot region
	JW	Joint width is measured between the first and fifth metatarso-phalangeal joints, representing the broadest region of the foot.
	HW	Heel witdh is determined 40 mm in front of the pternion

2.2. Research methodology

The proposed research comprises three major steps: data preprocessing, clustering, and evaluation. Each step plays a crucial role in forming accurate and meaningful cluster groups, where the preprocessing stage ensures data consistency and normalization, the clustering process groups similar foot measurements, and the evaluation stage assesses the quality and validity of the formed clusters using performance metrics such as the Silhouette Score.

2.2.1. Data preprocessing

Since the foot scanner machine only provides *STL format representing 3D images of the foot, data extraction into a numerical format is important to be conducted in order to enable analysis and application in machine learning models. The data extraction process is done by using the following pseudocode.

Algorithm 1. Pseudocode of data extraction process

```
Input: foot_model.stl
Output: foot_dimensions.csv
  load STL file as mesh
    sample 50,000 surface points → dataframe (x, y, z)
    calculate x_min, x_max, x_range
    FL ← x_range                                          // Foot Length
    AL ← distance in arch region (0.3–0.7 of FL)
    IH ← max height (0.4–0.6 of FL)
    TD ← toe height (0.9–1.0 of FL)
    JW ← width at joint region (0.25–0.35 of FL)
    HW ← width at heel region (0.0–0.1 of FL)
    JG ← π × JW                                           // Joint Girth
    IG ← π × width at instep region (0.55–0.65 of FL)
    WG ← π × width at waist region (0.45–0.55 of FL)
  store [FL, AL, IH, TD, JW, HW, JG, IG, WG] into CSV
return foot_dimensions.csv
```

2.2.2. Clustering

Clustering is one of the unsupervised learning techniques that is commonly applied to discover patterns in unlabeled data, segment data into meaningful groups, and identify outliers. In this study, K-Means clustering method is employed to systematically group and categorize new shoe sizes by identifying groups based on the available dataset. K-Means is chosen due to its efficiency, simplicity, and effectiveness in partitioning numerical data into well-defined clusters, making it suitable for analyzing foot measurement data.

2.2.3. Evaluation

To evaluate the clustering result, the Silhouette Score evaluation is carried out. The Silhouette Score can be defined as a metric evaluation used to justify the quality of a clustering result. It analyse how similar a data point is to its own cluster compared to other clusters. The score ranges from -1 to 1 with the following detail in Table 4.

Table 4. Silhouette Score

Score	Description
1	Result of clustering is close to 1 indicates that the data point is well-matched to its own cluster and poorly matched to neighboring clusters (good clustering)
0	The result of clustering is close to 0 indicates that the data point is on or very close to the boundary between two clusters
-1	The result of clustering is close to -1 indicates that the data point may have been assigned to the wrong cluster

3. Result and Discussion

3.1. Result data preprocessing

The experiment was started by converting the 3D model resulted from foot scanner machined to be numerical data. The converted result is described in Table 5.

Table 5. Converted result

File	FL	AL	IH	TD	JW	JG	IG	WG	HW
1	271	162	257	24	106	334	318	318	94
2	271	163	257	23	106	335	323	323	94
3	231	153	257	20	85	267	258	236	75
4	233	153	257	21	86	272	256	236	79
5	240	155	87	23	82	260	280	270	81
6	240	154	95	23	85	268	273	263	78
7	268	167	117	24	89	280	273	258	98
8	267	166	118	23	89	284	284	253	98
9	251	157	83	22	66	208	252	228	73
10	251	156	89	22	66	210.	255	231	73
...	...	...	...	...	...	...	...	...	...
32	240	140	116	27	76	241	272	244	74

3.2. Result of clustering process

The converted result from previous step was then used as input data for analysis process using K-Means clustering. The first importance step in K-Means utilization was determining the number of cluster (k). In this study, the number of k was determined using Elbow method by calculating sulhouette score. Figure 1 and Table 6 illustrates the result of Elbow method.

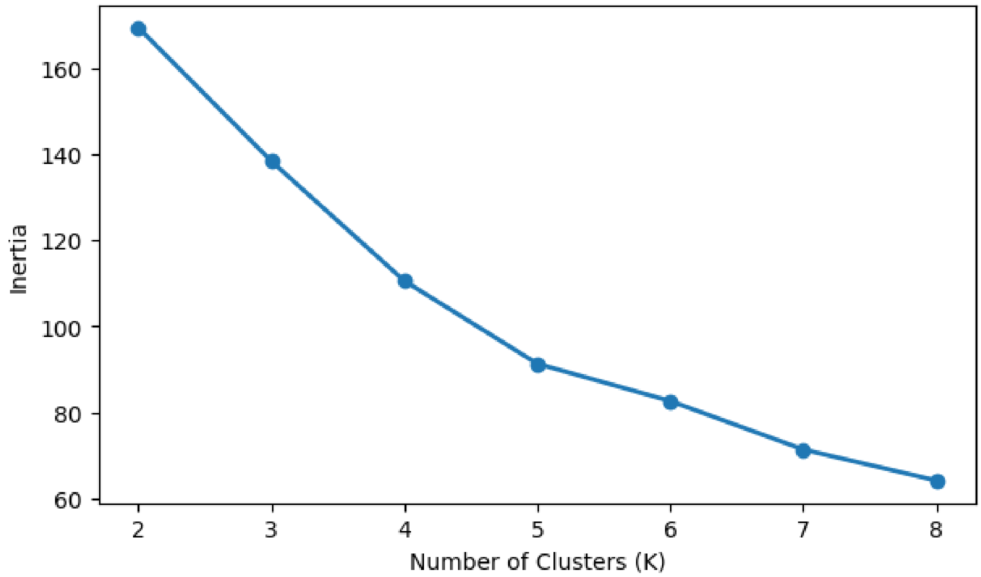


Figure 1. Result of Elbow method

According to Figure 1, the optimum number of cluster (k) is 4 with inertia (within-cluster sum of squares) of 110.58 and average Silhouette Score of 0.248. Based on the Elbow Method, k = 4 is the point indicating the optimal balance between cluster compactness and model simplicity. This result was then continued to clustering process.

In the research work, number of cluster (k) can be defined as the proposed shoe sizing system which can be defined as small (s), medium (m), large (l) and extra large (xl). The dataset was then utilized using K-Means with those number of cluster. The result of K-Means is depicted in Figure 2 and Figure 3.

Table 6. Average of each attibute per cluster

Cluster	Size	FL	AL	IH	TD	JW	JG	IG	WG	HW
0	Small (s)	237.2	81.15	142.4	21.78	86.29	271.1	270.7	250.9	80.38
1	Medium (m)	255.9	85.04	146.2	24.19	87.01	273.4	272.5	256.1	83.11
2	Large (l)	260.1	132.4	243.2	24.21	107.6	337.9	301.2	295.1	94.56
3	Extra-large (xl)	270.6	155.6	257.8	25.91	148.5	466.4	301.6	363.4	107.7

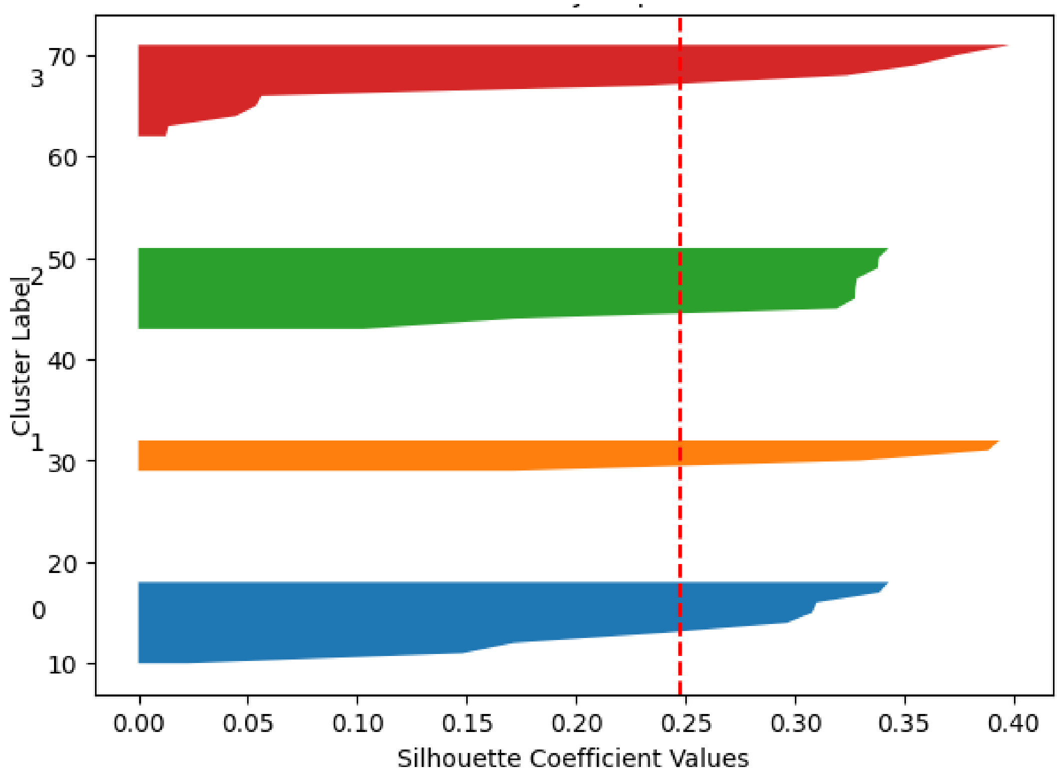


Figure 2. Silhouette analysis per cluster

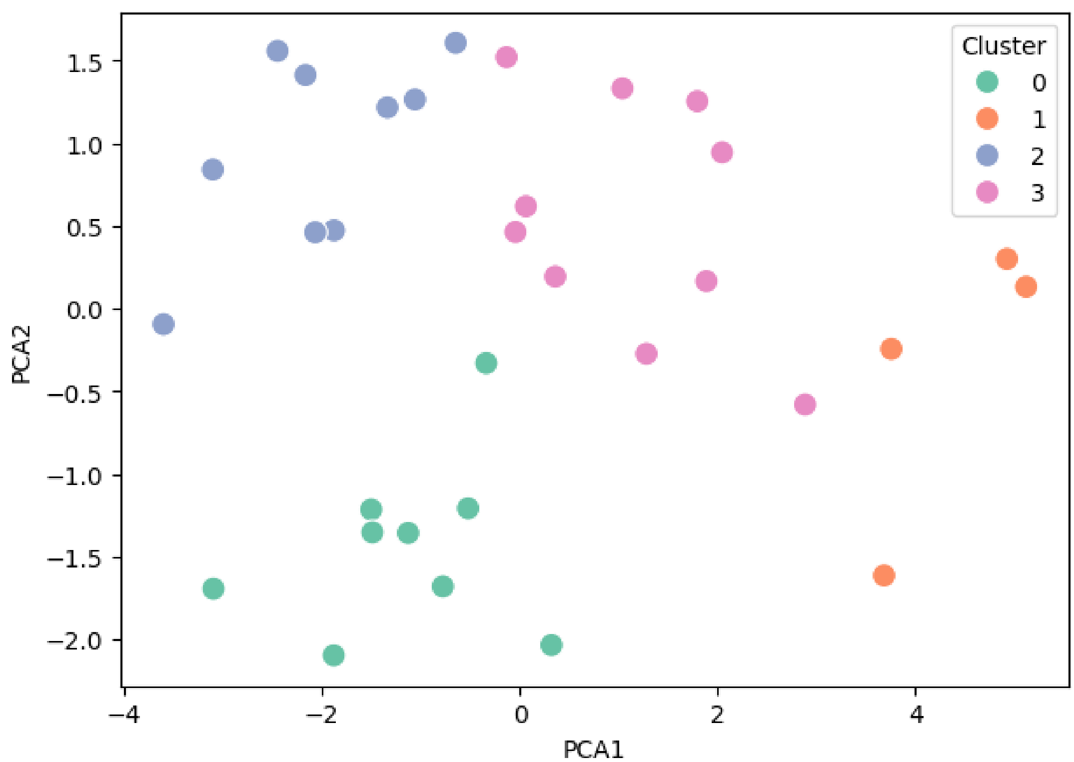


Figure 3. Foot size clustering visualization (PCA 2D)

Result depicted in Figure 2 illustrates the clustering performance for 4 clusters (labeled 0-3). The resulted graph indicates that almost all data points are reasonably well grouped. The relatively low average silhouette score implies that the cluster boundaries are not sharply defined. Figure 3 illustrates 2D visualization of the clustering results using principal component analysis (PCA), where PCA1 and PCA2 represent the first and second principal components, respectively. The plot shows that each cluster occupies distinct regions in the feature space, indicating a reasonable degree of separation among the groups. Beside of those results, in Figure 2, it can be found that the average of Silhouette Score is 0.25, which means that the clustering structure demonstrates only a moderate level of cohesion and separation. In other words, most data points are reasonably well assigned to their respective clusters.

3.3. Evaluation

The evaluation process was conducted by analysing the Silhouette Score. Since the resulted average Silhouette Score was 0.25, it can be concluded that the clustering results were reasonably well grouped which means that the modeling shoe sizing system was reasonably well designed.

4. Conclusions

This research work aimed to develop a modeling approach for a women's shoe sizing system based on anthropometric data resulted from foot scanner machine. The experiment involved several processes, including data preprocessing, clustering using the K-Means algorithm, and evaluation of the clustering performance. The experimental results achieved Silhouette Score of 0.25 (using k of 4) indicated a moderate but acceptable level of clustering quality, suggesting that the proposed approach successfully captures meaningful patterns in the data and can serve as a foundation for developing a more precise and data-driven women's shoe sizing system. Although the Silhouette Score of 0.25 is relatively low, it can still be considered moderate yet acceptable for several reasons. First, the dataset used in this study exhibits overlapping characteristics and lacks clearly separated cluster structures, which naturally leads to lower silhouette values. Second, scores in the range of 0.2-0.5 are commonly interpreted as indicating reasonable clustering performance for complex or noisy real-world data. In this context, a score of 0.25 demonstrates that the clusters are sufficiently distinct to provide meaningful separation, even though the margins between them are not perfectly sharp. Therefore, the clustering quality is considered moderate but still acceptable for exploratory analysis and for achieving the study's objectives.

References

Dash, B., Sarker, M. R., Alam, M. M., Asick, A. M., & Begum, A. (2024). Development of a footwear sizing system for diabetic feet. *Heliyon*, *10*(18), e37824. https://doi.org/10.1016/j.heliyon.2024.e37824

Frannita, E. L., & Hidayahtullah, M. C. (2024). Literature review in implementation of industry 4.0 for footwear industry. *Berkala Penelitian Teknologi Kulit, Sepatu, dan Produk Kulit*, *22*(2), 127–137. https://doi.org/10.58533/pqpnp243

Frannita, E. L., Pramitaningrum, E., & Hidayahtullah, M. C. (2024). *Study of organoleptic testing on sport shoe*. *Berkala Penelitian Teknologi Kulit, Sepatu, dan Produk Kulit*, *23*(2), 211–218.

Jandova, S., & Mendricky, R. (2021). Benefits of 3D printed and customized anatomical footwear insoles for plantar pressure distribution. *3D Printing and Additive Manufacturing*. https://doi.org/10.1089/3dp.2021.0002

Kim, M.-G., Kim, J., Chung, S. Y., Jin, M., & Hwang, M. J. (2022). Robot-based automation for upper and sole manufacturing in shoe production. In *Machines* (Vol. 10, Issue 4). https://doi.org/10.3390/machines10040255

Limon, M. I. K., Uddin, M. E., Hossin, M. M., & Rahman, M. S. (2023). Development of new shoe sizing system for women based on regression analysis of foot shapes. *International Journal of Industrial Ergonomics*, *94*(August 2022), 103408. https://doi.org/10.1016/j.ergon.2022.103408

Lin, L. Y., & Chen, C. H. (2015). Innovation and ergonomics consideration for female footwear design. *Procedia Manufacturing*, *3*, 5867–5873. https://doi.org/10.1016/j.promfg.2015.07.891

Mueller, J., Richter, M., Schaefer, K., Ganz, J., Lohscheller, J., & Mueller, S. (2023). How to measure children's feet: 3D foot scanning compared with established 2D manual or digital methods. *Journal of Foot and Ankle Research*, *16*(1), 21. https://doi.org/10.1186/s13047-023-00618-y

Prananda, A. R., Frannita, E. L., & Hidayat, A. (2024). AR-FootIN 4 . 0 : Aplikasi pengenalan teknologi industri 4.0 pada bidang alas kaki berbasis mobile augmented reality. *TIN : Terapan Informatika Nusantara*, *4*(10), 660–668. https://doi.org/10.47065/tin.v4i10.4956

Shariff, S. M., Merican, A. F., & Shariff, A. A. (2019). Development of new shoe-sizing system for Malaysian women using 3D foot scanning technology. *Measurement: Journal of the International Measurement Confederation*, *140*, 182–184. https://doi.org/10.1016/j.measurement.2019.03.057