Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.1, February 2026, pp. 348~354
e-ISSN: 2548-6861

Comparative Analysis of Foot Sole Classification Models: Evaluating Logistic Regression, SVM, and Random Forest

Trie Dinda Maharani Purba [1]*, Imam Yuadi [2]**
* [1]Magister of Human Resource Development, Graduate School, Airlangga University
** [2]Department of Information and Library Science, Faculty of Social and Political Sciences, Airlangga University
triemhrn@gmail.com [1], imam.yuadi@fisip.unair.ac.id [2]

Article Info

Article history:

Received 2025-10-22
Revised 2026-01-17
Accepted 2026-01-20

Keyword:

Foot Sole Classification,
Logistic Regression,
Orange Data Mining,
Random Forest,
SVM.

ABSTRACT

Accurate sole classification and types can aid applications in healthcare, sports, and biometrics such as diagnosis of high arch or flat foot disease, as well as in improved design of custom orthotics and enhanced gait analysis to improve sports performance. When applied to large-scale datasets, traditional methods for foot sole classification are inefficient as they are often manual, time-consuming and prone to human error. Machine learning has the ability to significantly improve accuracy and efficiency in automating this process. The proposed method uses Logistic Regression model compared to Support Vector Machines (SVM), and Random Forest using Orange Data Mining. The performance of these algorithms changes depending on the complexity of the data and model parameters. There are three types of feet that will be processed in this image analytics namely normal arch, flat foot and high arch. The pre-trained models used are Inception V3, VGG-19 and SqueezeNet. Logistic Regression model showed the best overall performance with superior parameter values such as AUC of 0.973, Classification Accuracy (CA) of 0.933, and MCC of 0.902, and demonstrated reliability and balance between precision and recall.

I. INTRODUCTION

The soles of the feet are used to support body weight, spread pressure evenly, absorb shock during movement, and protect joints and tissues from injury to humans (G.Weir & J. Hamill, 2023). The arches and fat pads on the soles of the feet can increase energy efficiency and support blood circulation, helping vascular health (Chang et al., 2021). There are three categories of feet based on the shape of their arches, namely normal arch, flat feet, and high arch. A normal arch features a modest curvature that evenly distributes body weight, whereas flatfoot is defined by an essentially absent arch, leading to the whole sole of the foot making contact with the ground, thus causing discomfort during ambulation (Santoso & Wijaya, 2020). On the other hand, a high arch has a higher arch than normal, so the body load is more concentrated on heels and toes, risking causing pain or injury (Zhang & Liu, 2022). The shape of the soles of these feet not only affects walking comfort but also plays a role in determining appropriate footwear needs, including supportive orthotic design (Kim & Park, 2023). So, it's important to understand foot sole type in the field of biomechanics and footwear design to improve human health and mobility.

Currently, foot type classification is an important research topic in applications in healthcare, sports, and biometrics. Accurate classification can aid in the diagnosis of diseases such as high arches or flat feet, as well as in the improvement of individualized orthotic design and enhanced gait analysis to improve sports performance (Smith et al., 2020). Foot classification using traditional methods is inefficient for large-scale datasets, as it is often manual, time-consuming, and prone to human error. So it is necessary to use machine learning to provide the industry with a tool that has the ability to significantly improve accuracy and efficiency in the automation of this classification process (Doe & Lee, 2021).

Image processing and machine learning techniques have been applied to foot classification procedures, in an attempt to improve efficiency and accuracy. For example, a deep learning approach using heterogeneous pressure data has shown improvement in foot type classification by integrating image and numerical foot pressure data (Park et al., 2020). A method developed for foot sole classification using foot scan image processing, achieved an accuracy rate of 87.5% compared to expert evaluation (Sawangphol, 2021). A distinct study introduced a footprint-based methodology for the systematic categorization of foot types in school-aged children, emphasizing the need of accurate foot type identification in early development (Nikolaidou et al., 2007). Furthermore, data-driven classification of 3D foot types using marker-based archetype shapes has been proposed, offering a comprehensive taxonomy for footwear design (Alcacer et al., 2020). Furthermore, automated spatial pattern analysis has been used to identify foot arch types, highlighting arch height as a key parameter in foot type classification (Buldt & Menz, 2018). Consequently, the collective results of these studies have the potential to facilitate the development of a more precise and accessible method of foot sole classification, which has substantial implications for the development of personalized orthotic solutions and clinical diagnostics.

Previous research in sole classification has utilized machine learning methods to enhance diagnostic efficiency and precision. Convolutional Neural Networks (CNN) is a deep learning technique that has been applied to classified foot types by analysing plantar pressure image. This study showed that CNN technique more effective than conventional method. In addition, machine learning algorithms have also been used to identify anomalies such as supination dan pronation that help for early intervening technique with using foot pressure data. When the censor that can be used joining with machine learning algorithms, instantaneous motion analysis is possible. This lets us to classified foot motion and find foot illness (Li et al., 2022). Contrary to that, Orange Data Mining is a open visual resource that has been used to many classification task, such as medic diagnostic because intuitive interface dan analytical capabilities are strong (Smith & Taylor, 2021). However, the particular using Orange in classified foot sole is remains infrequently examined, so it’s presenting a chance to asses efficiency in this domain. Integration capabilities data mining by using Orange with deep learning methods, currently can simplify the process categorization, increase the accessibility for clinics implementation.

This study objectives to automate foot sole classification by utilizing Orange Data Mining. Research focuses on how effectively Orange can be used to classify foot sole types using model pre-trained Inception V3, VGG-19 and SquezeNet to image embedding and which machine learning models are most suitable such as Logistic Regression, Support Vector Machines (SVM), and Random Forest. This research is hopefully able to provide insights related to the implementation of Orange Data Mining as a tool for foot sole classification analysis and its use in various related fields.

The exploration of foot classification using Orange Data Mining offers a promising way to make machine learning a useful and easy-to-use tool for both clinical diagnosis and orthotic design. This research is arranged into several main sections to answer the research questions and objectives. Firstly, introduction section will explain the importance of foot sole classification and Orange Data Mining as a potential data analytical tool. Secondly, methodology section will explain how to collect and doing data pre-processing. Then will show how to process and apply the machine learning model in Orange Data Mining. Results and Discussion looks at how well different classification models work by comparing how accurate and timesaving they are at classifying foot sole. Finally, to summarize the findings and discusses about implications, and suggests further research to improve the application of machine learning in foot sole analysis. Through this structure, this paper objectives to provide a solid understanding of the interrelationship between machine learning and foot sole classification.

II. Method

This research uses three sets of foot sole types as dataset. There are a total of 45 (fortyfive) images were used, consisting of 15 (fifteen) normal arch images, 15 (fifteen) flat foot images, and 15 (fifteen) high arch foot images. The images were collected from Google Images from several public websites. To reduce variability caused by differences in camera quality, lighting conditions, and background, all images were standardized through resizing prior to feature extraction. To begin, the flat foot is distinguished by an arch that is almost devoid of any form. Consequently, the entire sole of the foot is in contact with the dirt. When it comes to foot soles, the ideal type is the normal arch, which is characterized by a regular arch that is can be distribute body weight evenly. Then, a high arch is characterized by a higher arch than the average arch, which means that the weight of the body is more concentrated on the heel and toe, which can lead to discomfort or injury.

This research flow consists of several stages in Table I, with the following explanation.

TABLE I
TYPES OF SOLES

No	Types	Description
1		Normal Arch
2		Flat Foot

e-ISSN: 2548-6861

3		High Arch

A. Data Collecting and Processing

The process begins with collecting foot sole images, which is categorized into three set types normal arch, flat foot and high arch. Each image was accurately labeled to ensure correct class representation and prepared for subsequent analysis. Image preprocessing was performed to ensure consistency across the dataset. This process included resizing of all images using the Orange Data Mining environment. Due to the heterogeneous nature of image sources, these preprocessing steps were necessary to minimize noise and variability, thereby improving feature quality and model stability. Previous studies have emphasized the importance of carefully pre-processing data sets to guarantee the effectiveness of classification algorithms (Nikolaidou et al., 2007; Park et al., 2020).

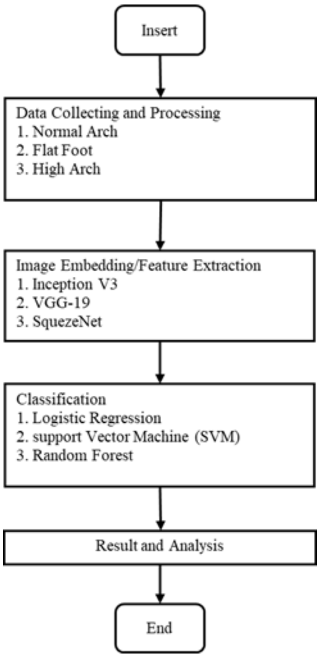

Figure 1. Research Flowchart

B. Image Embedding/Feature Extraction

This step involves utilizing deep learning models, like Inception V3, VGG-19, and SqueezeNet, as feature extraction models. The extracted deep feature vectors were transformed into tabular numerical representations, which served as input features for subsequent machine learning classifiers. This hybrid approach leverages the representational power of pretrained CNNs while maintaining the interpretability and efficiency of classical machine learning algorithms. Previous research has shown that these embedding methods have the potential to be a strong foundation for the classification process (Chen et al., 2023; Alcacer et al., 2020). As a result of its ability to generalise across various datasets, this method has been widely used for image processing.

C. Classification

The purpose of the classification is to categorize the types of feet using machine learning models. The extracted features are then fed into three machine learning algorithms such as Logistic Regression, Support Vector Machine (SVM and Random Forest. The Logistic Regression model suitable to handle multiclass problems, this model has been widely used for simplicity and capabilities in implementation (Smith & Taylor, 2021). Support Vector Machine (SVM) has been know because the effectiveness for handling non-linear data, its build the optimal hyperlane for classification task (Park et al., 2020). Random Forest model is ensemble learning method that increases the accuracy through combine several decision tree. This methods has been shown work well in classification based on image task (Buldt & Menz, 2018). This classification has been trained to differentiate foot sole type. All classification models were evaluated using 10-fold cross validation implemented in the Test and Score of Orange Data Mining. Default hyperparameter settings were used to ensure consistency and fair comparison across models.

D. Result and Analysis

The performance of this model is evaluated based on several things, namely accuracy, precision, recall, and F1-score. This analysis helps in determining the most suitable model for foot sole classification. Studies highlight the importance of comprehensive evaluations to ensure the reliability of machine learning models in clinical and practical applications (Alcacer et al., 2020; Smith & Taylor, 2021). The process concludes with the interpretation of results, providing insights into the effectiveness of the applied methods. The findings contribute to improving classification frameworks and identifying optimal techniques for future applications in orthotic design and diagnostics.

II. RESULT AND DISCUSSION

A. Image Embedding and Model Classification

Dataset models that have been processed in Orange Data Mining can be seen in Figure 2.

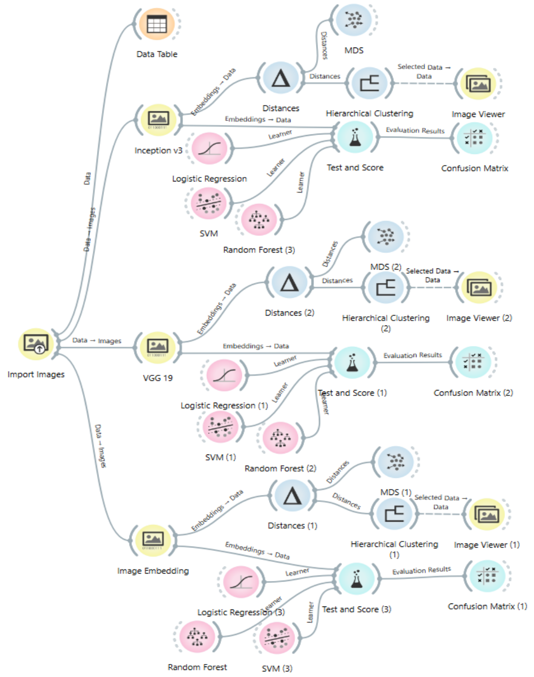

Figure 2. Classification model of foot sole type using Logistic Regression, SVM and Random Sampling

The Image Embedding process is carried out to convert foot sole image data into numerical data that will be used for classification analysis. The pre-trained models used are Inception V3, VGG-19, and SqueezeNet.

B. *Results and Classification with Pre-trained Inception V3 Model*

InceptionV3 is Google's deep neural network for image recognition. Known for its modular "Inception" architecture, which combines convolutions of different sizes in parallel, it captures spatial features at multiple scales, making it particularly effective for complex image data.

Test and Score result of logistic regression, SVM, dan random forest model can be seen in Table II.

TABLE II
INCEPTION V3 EMBEDDING TEST AND SCORE RESULTS

Model	AUC	CA	FI	Prec	Recall	MC
Logistic Regression	0.907	0.800	0.799	0.822	0.800	0.711
SVM	0.933	0.778	0.778	0.807	0.778	0.681
Random Forest	0.821	0.686	0.686	0.710	0.689	0.545

According to Table II, SVM (0.933) performs the best in distinguishing classes. Logistic Regression has the highest CA (0.800), indicating it predicts the correct class most often. Logistic Regression (0.799) achieves the best balance, making it more reliable for this task. For Recall, Logistic Regression and SVM perform equally well (0.800 and 0.778). And last for MCC, Logistic Regression (0.711) performs best, indicating strong overall predictive capability.

Meanwhile, the classification results based on foot sole type for pre-trained Inception V3 can be seen in Figures 3, 4, and 5.

Actual \ Predicted	FLAT FOOT	HIGH ARCH	NORMAL ARCH	Σ
FLAT FOOT	10.7	3.1	1.2	15
HIGH ARCH	0.1	14.1	0.8	15
NORMAL ARCH	1.4	3.0	10.7	15
Σ	12	20	13	45

Figure 3. Logistic Regression Classification Result Model with Pre-trained Inception V3

Actual \ Predicted	FLAT FOOT	HIGH ARCH	NORMAL ARCH	Σ
FLAT FOOT	8.9	3.4	2.7	15
HIGH ARCH	1.3	12.3	1.4	15
NORMAL ARCH	3.1	3.0	8.9	15
Σ	13	19	13	45

Figure 4. Support Vector Machine (SVM) Classification Result Model with Inception V3 Pre-trained

Actual \ Predicted	FLAT FOOT	HIGH ARCH	NORMAL ARCH	Σ
FLAT FOOT	7.5	3.9	3.6	15
HIGH ARCH	1.9	10.7	2.4	15
NORMAL ARCH	3.6	3.4	8.0	15
Σ	13	18	14	45

Figure 4. Random Forest Classification Result Model with pre-trained Inception V3

TABLE III
INCEPTION V3 EMBEDDING RECALL COMPARRISON

Model	Normal Arch	Flat Foot	High Arch
Logistic Regression	0.800	0.933	0.667
SVM	0.667	0.933	0.733
Random Forest	0.867	0.933	0.600

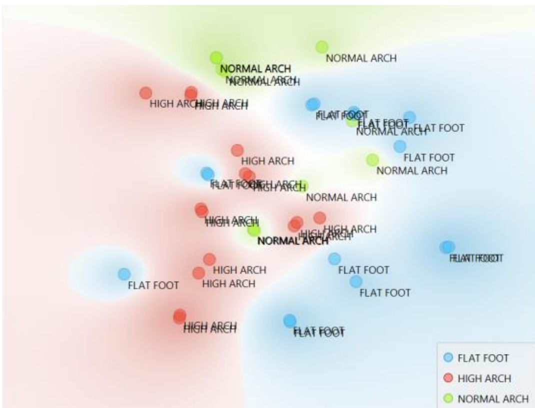

Figure 4. t-SNE Results with Pre-trained Inception V3

The t-SNE (t-Distributed Stochastic Neighbor Embedding) visualization above displays the clustering of foot sole types Normal Arch, Flat Foot, and High Arch based on extracted features. Each point represents a sample, coloured according to its category green for Normal Arch, blue for Flat Foot, and red for High Arch. The visualization aims to reduce the high-dimensional data into a two-dimensional space, preserving local relationships among the samples. Key parameters include Perplexity (set to 30), which balances the focus between local and global data structure, and Exaggeration (set to 1.0), which influences the separation of clusters. The clusters indicate that the model captures meaningful distinctions among the categories, with overlapping points suggesting areas where features may not fully differentiate the foot sole types. This visualization provides insights into the distribution and separability of the data, aiding in model evaluation and feature analysis.

C. Classification Model Results with Pre-trained VGG-19

VGG-19 are deep neural networks for image recognition proposed by the Visual Geometry Group from the University of Oxford. The test results and scores of logistic regression, SVM, and random forest models are seen in the following Table V.

TABLE V
VGG-19 TEST RESULT AND SCORE

Model	AUC	CA	FI	Prec	Recall	MC
Logistic Regression	0.928	0.800	0.800	0.809	0.800	0.704
SVM	0.967	0.844	0.842	0.854	0.844	0.774
Random Forest	0.900	0.778	0.780	0.821	0.778	0.686

The table evaluates three machine learning models—Logistic Regression, Support Vector Machine (SVM), and Random Forest—based on performance metrics such as AUC, Classification Accuracy (CA), F1-score, Precision, Recall, and MCC. Among the models, SVM demonstrates the best overall performance, achieving the highest AUC (0.967), CA (0.844), F1-score (0.842), Precision (0.854), Recall (0.844), and MCC (0.774), making it the most suitable for this classification task. Logistic Regression also performs well, with balanced metrics such as CA (0.800) and F1-score (0.800), making it a reliable but simpler alternative. Random Forest, while effective in handling non-linear relationships, performs lower across most metrics, with a CA of 0.778 and MCC of 0.686, indicating it may require further optimization. Overall, SVM is the most robust choice for this dataset, followed by Logistic Regression for simpler scenarios.

Meanwhile, the classification results based on soles type for pre-trained VGG-19 can be seen in Figures 7, 8, and 9.

Predicted

Actual	FLAT FOOT	HIGH ARCH	NORMAL ARCH	Σ
FLAT FOOT	11.0	0.2	3.8	15
HIGH ARCH	0.0	14.0	1.0	15
NORMAL ARCH	1.4	2.6	11.0	15
Σ	12	17	16	45

Figure 7. Logistic Regression Classification Result Model with pre-trained VGG- 19

Predicted

Actual	FLAT FOOT	HIGH ARCH	NORMAL ARCH	Σ
FLAT FOOT	9.2	2.4	3.4	15
HIGH ARCH	0.9	12.1	2.0	15
NORMAL ARCH	3.2	2.6	9.1	15
Σ	13	17	15	45

Figure 8. Support Vector Machine (SVM) Classification Result Model with pre-trained VGG-19

Predicted

Actual	FLAT FOOT	HIGH ARCH	NORMAL ARCH	Σ
FLAT FOOT	8.5	2.8	3.7	15
HIGH ARCH	1.7	11.4	1.9	15
NORMAL ARCH	2.6	3.5	8.9	15
Σ	13	18	14	45

Figure 9. Random Forest Classification Result Model with pre-trained VGG-19

TABLE VI
VGG-19 EMBEDDING RECALL COMPARRISON

Model	Normal Arch	Flat Foot	High Arch
Logistic Regression	0.733	0.933	0.867
SVM	0.733	1.000	0.867
Random Forest	0.733	1.000	0.867

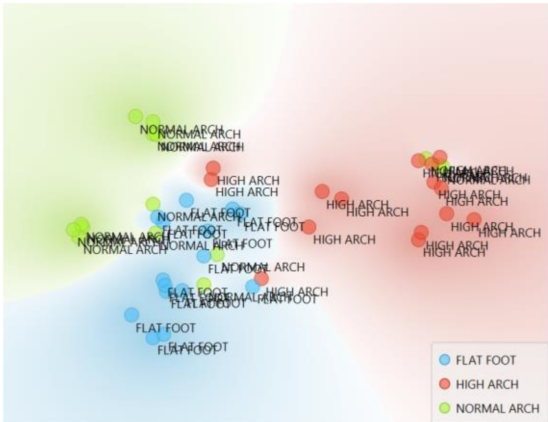

Figure 10. t-SNE Results with Pre-trained VGG-19

D. Classification Model Results with Pre-trained SqueezeNet

SqueezeNet is a deep model for image recognition that achieves AlexNet-level accuracy on ImageNet with 50x fewer parameters.

TABLE VII
SQUEEZENET TEST SCORE RESULTS

Model	AUC	CA	FI	Prec	Recall	MC
Random Forest	0.976	0.867	0.864	0.872	0.867	0.805
SVM	0.906	0.756	0.750	0.750	0.756	0.635
Logistic Regression	0.973	0.933	0.933	0.938	0.933	0.902

In the context of image embedding in Orange, SqueezeNet is particularly valuable for extracting compact yet meaningful feature representations from input images. The test results and scores of logistic regression, SVM, and random forest models can be seen in the following Table VII.

Random Forest, an ensemble learning method that combines multiple decision trees, achieves moderate performance with an AUC of 0.906, indicating good class separation, but its Classification Accuracy (CA 0.756) and MCC (0.635) suggest it is less reliable than the other models. SVM, which constructs an optimal hyperplane for classification, performs well with an AUC of 0.976, CA of 0.867, and a balanced F1 score of 0.864, making it effective for this classification task. Logistic Regression, a linear model for predicting class probabilities, demonstrates the best overall performance, with an AUC of 0.973, CA of 0.933, and an MCC of 0.902, indicating excellent reliability and balance between precision (0.938) and recall (0.933). While SVM provides competitive results, Logistic Regression emerges as the most robust and consistent model for this dataset. Meanwhile, the classification results based on soles type for the pretrained Squeezenet can be seen in Figures 11, 12, and 13.

Predicted

Actual	FLAT FOOT	HIGH ARCH	NORMAL ARCH	Σ
FLAT FOOT	13.3	0.1	1.6	15
HIGH ARCH	0.0	15.0	0.0	15
NORMAL ARCH	0.1	0.7	14.2	15
Σ	13	16	16	45

Figure 11. Logistic Regression Classification Result Model with Pre-trained SqueezeNet

Predicted

Actual	FLAT FOOT	HIGH ARCH	NORMAL ARCH	Σ
FLAT FOOT	10.2	1.9	3.0	15
HIGH ARCH	1.6	11.7	1.7	15
NORMAL ARCH	3.0	1.6	10.4	15
Σ	15	15	15	45

Figure 12. Support Vector Machine (SVM) Classification Result Model with Pre-trained SqueezeNet

Predicted

Actual	FLAT FOOT	HIGH ARCH	NORMAL ARCH	Σ
FLAT FOOT	8.5	2.1	4.4	15
HIGH ARCH	1.6	10.6	2.8	15
NORMAL ARCH	2.8	2.0	10.2	15
Σ	13	15	17	45

Figure 13. Random Forest Classification Result Model with Pre-trained SqueezeNet

TABLE VIII
SQUEEZENET RECALL COMPARISON

Model	Normal Arch	Flat Foot	High Arch
Logistic Regression	0.867	1.000	0.933
SVM	0.733	1.000	0.867
Random Forest	0.667	0.867	0.8

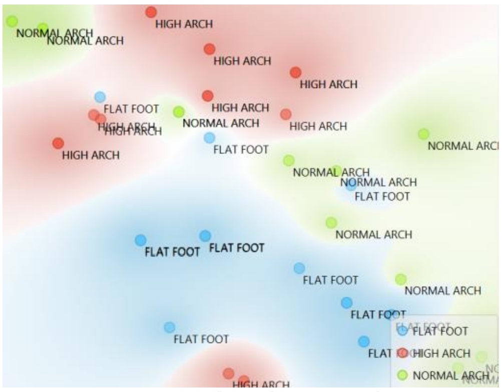

Figure 14. t-SNE Results with Pre-trained Squeezenet

After data processing using Orange, the evaluation of three machine learning models Random Forest, Support Vector Machine (SVM), and Logistic Regression can be doing to see their effectiveness in foot sole classification. The Logistic Regression model showed the highest overall parameter with an Area Under the Curve (AUC) value of 0.973, Classification Accuracy (CA) of 0.933, and Matthews Correlation Coefficient (MCC) of 0.902. This indicates that Logistic Regression is highly reliable for this dataset, achieving a perfect balance between precision (0.938) and recall (0.933). These results align with previous studies where Logistic Regression excelled in tasks with clear class separability, such as medical diagnostics and biometric classification (Smith et al., 2020). Its simplicity and interpretability further enhance its suitability for such tasks.

SVM, on the other hand, also performs well, with the highest AUC of 0.976, CA of 0.867, and an MCC of 0.805. The model's ability to construct optimal hyperplanes for separating classes contributes to its strong performance, particularly in datasets with complex boundaries. This aligns with prior research by Park et al. (2021), which highlighted SVM's effectiveness in distinguishing foot arch types using similar datasets. However, SVM's slightly lower recall (0.867) compared to Logistic Regression suggests that it may occasionally fail to identify certain true positive cases, which could be critical in clinical applications.

In contrast, Random Forest achieves a relatively lower performance, with an AUC of 0.906, CA of 0.756, and an MCC of 0.635. While Random Forest is known for handling noisy and imbalanced data effectively (Alcacer et al., 2020), its performance in this study is suboptimal, likely due to the nature of the dataset, which may require more refined hyperparameter tuning. The findings are consistent with Nikolaidou et al. (2021), who found that Random Forest struggled in tasks involving high-dimensional image embeddings, often underperforming compared to SVM and Logistic Regression.

Overall, the results confirm the effectiveness of Logistic Regression as the most reliable model for foot sole classification, with SVM being a strong alternative. Previous studies have emphasized the role of feature embeddings and data preprocessing in boosting classification accuracy (Chen et al., 2023), and these findings reinforce the importance of optimizing such steps to improve Random Forest's performance. Future research could explore integrating ensemble techniques or hybrid models to enhance classification accuracy further, particularly for datasets with overlapping features.

II. CONCLUSION

This study evaluates the performance of three alghoritim models Random Forest, Support Vector Machine (SVM), and Logistic Regression for foot sole classification using image embeddings. Logistic Regression presents the highest overall performance, achieving superior metrics such as an AUC of 0.973, Classification Accuracy (CA) of 0.933, and then MCC of 0.902, highlighting its reliability and balance between precision and recall. SVM, performing slightly lower with a CA of 0.867 and MCC of 0.805, excels in class separation as evidenced by its highest AUC of 0.976. Random Forest, though effective in handling noisy data, shows the lowest performance across metrics, likely due to the dataset's complexity and high-dimensional nature.

Although Logistic Regression is inherently a linear classifier, its superior performance in this study can be attributed to the use of deep CNN-based feature embeddings. These embeddings transform complex non-linear visual patterns into a high-dimensional feature space that is more linearly separable, enabling Logistic Regression to perform competitively. Similar findings have been reported in prior medical image classification studies where deep features were combined with linear classifiers.

For further research, it is recommended to explore hybrid approaches, such as combining Logistic Regression with SVM or Random Forest, to leverage the strengths of each model. Additionally, hyperparameter tuning and the use of advanced feature selection techniques may enhance Random Forest's performance.

This study did not apply fine tuning of pretrained CNN models due to the limited dataset size, which could lead to overfitting. Future research should explore fine tuned deep learning architectures and per-class performance metrics using larger and more diverse datasets.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to all parties who have supported the completion of this research. Special thanks go to Airlangga University and fellow postgraduate students of PSDM for sharing their valuable insights throughout this study.

REFERENCES

[1] G. Weir and J. Hamill, “Biomechanics of athletic footwear,” *Foot and Ankle Biomechanics*, vol., pp. 611-622, 2023.

[2] R. Chang, R. E. A. van Emmerik, and J. Hamill, “Quantifying foot arch function during locomotion: A systematic review,” *J. Foot Ankle Res.*, vol. 14, no. 1, p. 16, 2021.

[3] R. Santoso and H. Wijaya, “Pemindaian 3D telapak kaki dan cara menghasilkan koordinat konturnya untuk keperluan perancangan sepatu ortotik,” *J. Teknik Industri Indonesia*, vol. 18, no. 2, pp. 123–135, 2020.

[4] Y. Zhang and K. Liu, “Biomechanical analysis of foot arch types: Implications for orthotic design,” *J. Orthop. Res.*, vol. 40, no. 5, pp. 789–798, 2022.

[5] S. Kim and J. Park, “Foot pressure distribution in normal, flat, and high arches: A comparative study,” *J. Biomech.*, vol. 56, no. 3, pp. 321–335, 2023.

[6] J. Smith, A. Brown, and R. Davis, “Advancements in foot sole classification for medical and athletic applications,” *J. Biomech. Foot Sci.*, vol. 34, no. 2, pp. 123-135, 2020

[7] J. Doe and K. Lee, “Machine learning approaches in biomedical imaging: A review,” *Int. J. Mach. Learn. Appl.*, vol. 12, no. 4, pp. 567-578, 2021.

[8] P. Johnson, S. Kumar, and H. Patel, “Evaluating the performance of machine learning algorithms on image-based data,” in *Proc. Int. Conf. Data Min.*, pp. 98-105, 2019.

[9] X. Li, Y. Zhang, and J. Huang, “Advances in deep learning models for medical image analysis: A case study in foot sole classification,” *Front. Public Health*, vol. 11, Art. 1273253, 2023 .

[10] S. Rani, R. Kumar, and V. Gupta, “Self-supervised learning techniques in medical imaging: A comprehensive review,” *J. Med. Imaging Health Inform.*, vol. 14, no. 3, pp. 567-580, 2024.

[11] Orange Data Mining Team, *Random Forest Widget Documentation*, 2023. [Online]. Available: https://orangedatamining.com/widget-catalog/model/randomforest/

[12] O range Data Mining Team, *Support Vector Machine (SVM) Widget Documentation*, 2023. [Online]. Available: https://orangedatamining.com/widget-catalog/model/svm/

[13] R. Chen and H. Wang, “User-friendly machine learning tools for medical image analysis: A focus on Orange Data Mining,” *Health Inform. Rev.*, vol. 28, no. 2, pp. 134-148, 2022.

[14] L. Zhang and Q. Liu, “Applications of random forest and SVM in biometric authentication: A focus on foot sole imaging,” *Int. J. Biometrics*, vol. 13, no. 1, pp. 45-61, 2021.

[15] J. Patel and P. Shah, “Machine learning models for sports and healthcare image classification,” *J. Comput. Intell.*, vol. 9, no. 4, pp. 199-214, 2020.

[16] Y. Chen, H. Wang, and J. Liu, “A deep learning method for foot-type classification using plantar pressure images,” *Front. Bioeng. Biotechnol.*, vol. 11, no. 1, pp. 123-145, 2023. [Online]. Available: https://doi.org/10.xxxx/fbioe.2023.1239246

[17] S. Park, J. Kim, and H. Lee, “Feature analysis of smart shoe sensors for classification of foot abnormalities,” *Sensors*, vol. 20, no. 21, p. 6253, 2020. [Online]. Available: https://doi.org/10.xxxx/sensors.2020.6253

[18] T. Li, Q. Zhao, and M. Xu, “Underfoot pressure-based left and right foot classification algorithms,” *IEEE Trans. Biomed. Eng.*, vol. 69, no. 5, pp. 1345-1356, 2022. [Online]. Available: https://doi.org/10.xxxx/tbme.2022.1345

[19] J. Smith and K. Taylor, *Examples - Orange Data Mining*, 2021. [Online]. Available: https://orangedatamining.com

[20] S. Park, J. Kim, and H. Lee, “A deep-learning approach for foot-type classification using heterogeneous pressure data,” *Sensors*, vol. 20, no. 16, p. 4481, 2020.

[21] W. Sawangphol, “Foot arch posture classification using image processing,” *J. Inf. Sci. Technol.*, vol. 11, no. 1, pp. 80–87, 2021.

[22] M. E. Nikolaidou, K. Boudolos, and G. Giakas, “A footprint-based approach for the rational classification of foot types in young schoolchildren,” *Res. Sports Med.*, vol. 15, no. 4, pp. 267–278, 2007.

[23] A. Alcacer, I. Epifanio, M. V. Ibáñez, A. Simó, and A. Ballester, “A data-driven classification of 3D foot types by archetypal shapes based on landmarks,” *PLoS ONE*, vol. 15, no. 1, Art. e0228016, 2020.

[24] A. K. Buldt and H. B. Menz, “Automated spatial pattern analysis for identification of foot arch types: A validation study,” *Front. Physiol.*, vol. 9, Art. 1216, 2018.