ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 03, March 2025

Efficient Footwear Classification: Memory-Optimized Transfer Learning for E-Commerce Applications

[1] T. Hithin Vaibhav, [2] T. Rohith Kumar, [3] Dr. K. Venkatraman

[1] [2] [3] Department Of CSE, Amrita Vishwa Vidyapeetham

Corresponding Author Email: [1] hithin2003@gmail.com, [2] rohithkumar2003@gmail.com, [3] kvrforu@gmail.com

***Abstract*— In the era of digital transformation, the footwear industry has experienced a significant shift toward e-commerce, creating a demand for automated image classification systems. This research presents a memory efficient, generalization-focused transfer learning model for footwear classification designed to handle unseen data while ensuring scalability. The approach leverages pre-trained Convolutional Neural Networks (CNNs) such as ResNet152V2, DenseNet201, NASNetMobile, and InceptionResNetV2, along with data augmentation and dropout techniques to address classification challenges. Transfer learning fine-tunes the features of these models, improving accuracy with limited labeled data. Among the models tested, InceptionResNetV2 achieved the highest accuracy of 98.78%. Applications include automated category labeling, improved recommendation systems, and enhanced search results on e-commerce platforms, ultimately making the online shopping experience more efficient and accurate.**

***Index Terms*— CNN, Data Augmentation, E-commerce, Footwear Classification, Memory Efficiency, Transfer Learning.**

I. INTRODUCTION

The quick expansion of e-commerce and digital retail has transformed the way customers buy, especially in the fashion industry. The classification of footwear presents distinct problems due to the wide range of styles, materials, and complicated designs. Manual product categorization, which is still frequently employed in the C2C and B2C platforms, is both time consuming and prone to human error, resulting in incorrect classification and inefficiencies in product retrieval. As a result, the demand for automatic, accurate, and scalable image classification systems has increased dramatically. The suggested system has numerous real-world applications, such as automatic product tagging, recommendation systems, and visual search engines for online stores. This can significantly improve the overall customer experience.

Deep learning, specifically CNN, has demonstrated extraordinary performance in picture recognition and classification. However, training CNNs from scratch requires large-scale labeled datasets and significant computational resources, making it difficult for many enterprises and researchers. Transfer learning, which uses pre-trained models, is a more efficient option since it allows for feature extraction and fine-tuning with less data and computational overhead. This strategy not only reduces training time but also increases generalization, enabling models to perform well on unseen data.

This research proposes a memory-efficient and generalization-focused transfer learning framework for footwear image classification. The study explores state-of-the-art CNN architectures, including ResNet152V2, DenseNet201, and others, evaluating their effectiveness in terms of precision, scalability, and computational power. The framework employs strategic data augmentation, batch normalization, and dropout techniques to enhance robustness and prevent overfitting. Among the models, InceptionResNetV2 achieves the highest classification accuracy (98.78%), demonstrating its superiority in extracting fine-grained features from footwear images.

II. LITERATURE SURVEY

This section reviews existing works and models that have made notable advances in the area of image classification, especially within the domain of footwear categorization, and some papers related to data augmentation.

Onalaja, J.Shahra, E. Q.Basurra, & S.Jabbar (2024) [1] The research compares SVM and CNN for sneaker authentication in the resale market. Using the classification of real and counterfeit sneakers, the study demonstrates that CNNs outperform SVMs, achieving a precision of more than 95%. The findings highlight CNNs' ability to capture complex image patterns, making them more effective than traditional methods. This research provides practical insights into enhancing operational efficiency, reducing human verification errors, and speeding up authentication processes, offering potential for broader applications beyond the fashion industry, including healthcare and security.

S. Kumar (2024) [2] The research paper "Adaptive Fusion-Enhanced CNN Architectures for Advanced Image Classification in E-Commerce" introduces a dynamic fusion model leveraging CNN architectures VGG16, ResNet50, and MobileNetV2. Addresses challenges in image classification by dynamically integrating the strengths of these models to adapt to diverse image characteristics. This fusion model employs a concatenation strategy, enhancing accuracy and computational efficiency. The experimental results on the Fashion-MNIST dataset reveal a training accuracy of 97.50% and test accuracy of 89.27%. model excels in distinguishing nuanced categories and surpasses traditional single-model approaches, validated through precision, recall, F1-score, and ROC analysis. Key contributions include the adaptive feature integration framework and improved computational efficiency, suitable for real-time applications. This work establishes a benchmark in adaptive CNN fusion and proposes solutions for resource-constrained environments. It emphasizes robust performance across diverse datasets, paving the way for advancements in efficient image classification systems.

B. Suvarna (2024) et al. [3] This study focuses on optimizing footwear image classification using the UTZappos dataset, leveraging CNN models like ResNet, MobileNet, EfficientNet and VGG19. EfficientNet achieved the highest accuracy of 90.02%, outperforming others due to its compound scaling approach. The research aims to enhance ecommerce platforms by improving image-based recommendation systems, offering personalized and intuitive shopping experiences. Despite the dataset's size of 50,000 images across 20 classes, noise from underrepresented classes was addressed by excluding them from analysis. The literature review highlights advancements in zero-shot learning, segmentation accuracy, and compositional tasks using deep learning models. Methodologies such as Mask-RCNN and ensemble classifiers demonstrate effective feature extraction and classification. The study's findings contribute to recommendation system improvements and underline EfficientNet's efficiency in computational and recognition tasks.

P. Shourie, V. Anand and S. Gupta (2024) [4] The classification of footwear types specifically, shoes, sandals, and boots using CNN is the subject of this research. 15,000 photos of these shoe categories make up the dataset, with 5,000 photos of each category. By learning discriminative features from raw picture data, CNNs efficiently complete the classification process. Many convolutional layers, max-pooling layers, dropout for regularization, and completely linked layers are all included in the suggested CNN model. With an accuracy of 97.19% on the training data and 97.08% on the validation set, the model demonstrated excellent performance. The model's efficacy was demonstrated through epoch-wise evaluation, which monitored gains in accuracy and loss throughout training. The study also identifies issues such as dataset size and class imbalance.

M. Hameed (2024) et al. [5] This study explores the use of AI techniques to enhance e-commerce clothing sector by developing a deep learning-based image classification system. The research focuses on classifying images of clothing products, such as shirts, dresses, pants, and shoes, using advanced deep learning models embedded into an e-commerce web application. The process includes requirement gathering, deep learning model training using Convolutional Neural Networks (CNNs), and testing the models' accuracy on various image datasets. Specifically, the study compares the performance of two CNN architectures, Xception and VGG-19, finding that the VGG-19 outperforms Xception in terms of accuracy. The trained VGG-19 model is integrated into a web application for clothing stores, where it efficiently classifies clothing products. The system's accuracy is further verified through testing with in-lab sample images.

K. S. Gill, A. Sharma, V. Anand and R. Gupta (2023) [6] This study focuses on classifying smart shoes using AI, specifically with the EfficientNetB3 model, aiming for high precision in categorizing shoe types. The research addresses both commercial and medical perspectives, noting the importance of shoe classification for personalized footwear design, health monitoring, and injury prevention. The machine learning model leverages data from multiple shoe angles (front, left, right, rear) for accurate predictions. The EfficientNetB3 model achieved an accuracy of 88.8% using the Adam optimizer. The study uses transfer learning for enhanced model performance, validating a large dataset with 10,000 images across five shoe classes: Ballet Flat, Boat, Brogue, Clog, and Sneaker. Findings aim to reduce shoe manufacturing errors, support personalized shoe designs, and contribute to health studies by identifying appropriate footwear for specific conditions.

Jayaseeli (2023) et al. [7] The study explores the Federated Averaging algorithm for image classification, highlighting its ability to train machine learning models across decentralized devices while preserving data privacy. Federated Learning eliminates the need to transfer raw data to central servers, addressing privacy concerns, especially in sensitive fields like healthcare and finance. Using the FashionMNIST dataset with 70,000 images across ten categories, the approach achieved an accuracy of 87.27%, comparable to traditional centralized methods. Challenges such as device heterogeneity and secure aggregation are discussed, along with strategies like FedProx and FedQuant to enhance efficiency and performance. The research underscores Federated Learning's potential for secure, inclusive, and efficient machine learning, with applications in healthcare, IoT, and finance.

Oliveira (2023) et al. [8] propose two deep learning solutions for online footwear retail. One tool uses Mask-RCNN for automatic shoe background removal, improving catalog updates. The other, a multi-label classifier with ResNet101 and Xception, enables visual search for personalized recommendations. The system achieved 90% accuracy for shape, 84.03% for color, and 85.97% for texture. Both modules are accessible via REST endpoints. A synthetic dataset can enhance robustness. These tools improve communication, catalog management, and user experience, with promising preliminary results for digital transformation in footwear marketplaces.

M. Sari and W. F. Al Maki (2023) [9] This paper discusses improving the performance of K-Nearest Neighbor for footwear classification, which can aid in criminal investigations and e-commerce. The authors apply KNN with Leave One Out Cross Validation to enhance classification accuracy. Footwear is crucial in identifying suspects, and the proposed method helps categorize footwear types, such as boots, shoes, and sandals, from crime scenes or e-commerce images. The pre-processing includes image labeling, converting images to grayscale, and extracting features using Histogram of Oriented Gradients (HOG). The use of LOOCV increases KNN's accuracy from 94% to 98%, making the method more effective in footwear classification. The dataset used contains 1500 footwear images from Kaggle, with 500 images for each class. The research aims to improve the performance of the model by addressing the challenges related to optimal k values and low accuracy in previous studies. Through these methods, the paper enhances classification accuracy, making it more reliable for both crime detection and e-commerce applications.

Z. Zhang, Q. Gao, L. Liu and Y. He (2023) [10] In this work, a dual GAN is used to generate images of rice leaf disease using a high-quality image augmentation HQIA technique. The technique generates initial pseudo-data using a Wasserstein GAN with Gradient Penalty and then improves it with and Opt-RealESRGAN for super resolution ,improving image clarity. This dual GAN approach addresses the challenge of insufficient training samples in deep learning for the detection of rice disease. The augmented data set was tested using the ResNet18 and VGG11 models, achieving an improvement in recognition accuracy of 4.57% and 4. 1%, respectively, compared to the original data set. Furthermore, it showed an improvement of 3. 08% and 3. 55% on the increase only with WGAN-GP. The strategy efficiently expands training datasets for disease recognition by combining image production and enhancement techniques. Experimental results demonstrate the effectiveness of this approach in addressing sample scarcity in image recognition of agricultural diseases.

J. You, G. Huang, T. Han, H. Yang and L. Shen, (2023) [11] propose a unified approach for facial image restoration and data augmentation using pretrained GANs. Image restoration enhances clarity, while augmentation expands datasets. Unlike traditional separate methods, this system uses a modified U-Net to predict biases in latent codes, improving fidelity. Linear interpolation aids augmentation, especially for imbalanced data. A Difference Extractor refines feature and latent code adjustments. Experiments show superior restoration quality and high-quality synthetic image generation, effectively addressing both data quality and quantity.

III. DATASET DESCRIPTION

The Shoe vs Sandal vs Boot Image Dataset is employed in this study for the classification task involving three categories of footwear: shoes, sandals, and boots. Dataset comprises a total of 15,000 images, with 5,000 images per class, providing a balanced distribution of data across the categories. The images are of uniform resolution, 136x102

pixels, and are represented in the RGB color model, ensuring consistency in data input for deep learning models. This dataset is publicly available and can be accessed on the Hugging Face website. It is designed for multiclass image classification tasks and serves as an ideal resource for deep neural network training, particularly CNNs. Additionally, the dataset is compatible with several machine learning frameworks, including TensorFlow, Keras, PyTorch, and Scikit-learn

A. Dataset Composition

- Shoe: Includes various types of shoes, such as formal shoes, sneakers, and flats.
- Sandal: Contains images of different sandal styles, including open-toe and slip-on variants.
- Boot: Features images of various boots, such as ankle boots, knee-high boots, and work boots.

The dataset provides an excellent platform for experiments with different neural network architectures and is particularly useful for applications in footwear classification, product recommendation systems, and e-commerce platforms. It enables researchers to explore advanced techniques in multiclass classification and transfer learning, using both traditional and modern deep learning approaches.

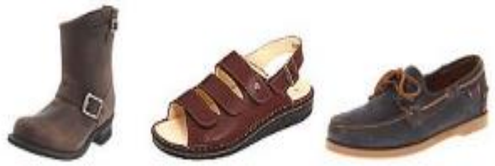


Figure 1. Sample images from the dataset.

B. Data Preprocessing Methods

Data Rescaling: Images are rescaled by dividing pixel values by 255 using rescale, normalizing them to a [0, 1] range for improved stability during training and give better performance.

Data Augmentation: Augmentation techniques are applied to the training data to improve robustness. Zooming is used by applying random zoom, which creates size variations for better scale learning. Shifting involves both horizontal and vertical shifts to simulate variations in object placement, and fill mode ensures that these shifts fill empty areas by replicating nearby pixels. These augmentations increase variability, reduce overfitting, and help the model generalize better.

C. Data Splitting

To facilitate effective model evaluation and avoid overfitting, the dataset is divided into two subgroups:

Training Subset (80%): This subset is utilized to train the model by updating its weights during backpropagation.

Validation Subset (20%): This is a hold-out set used to evaluate the model's efficacy on unseen data during training, providing an early indication of overfitting or underfitting.

D. Batch Size and Target Size

Image Resizing: Images are resized to 100x100 for consistency and compatibility with deep learning models. This reduces preprocessing time and ensures uniform input shapes for efficient processing.

Batch Processing: A batch size of 64 is used, allowing 64 images to be processed simultaneously, improving computational efficiency, optimizing memory, and stabilizing weight updates.

Figure 2. Images after data augmentation

IV. METHODOLOGY

A. Introduction

This research employs a deep learning to classify footwear images into different categories. The methodology leverages transfer learning to utilize pre-trained models that have been trained ImageNet, to increase the accuracy of the footwear classification task. Transfer learning allows the use of pre-trained feature extractors and the adaptation of these models for our specific dataset, thus minimizing the requirement for massive volumes of labeled data. It involves fine-tuning the pre-trained models to extract relevant features from footwear images, followed by training the models on the labeled dataset to make accurate predictions. The main target of the methodology is to build a robust classification model capable of accurately categorizing footwear images into three classes, which include sneakers, sandals, and boots. The methodology involves preprocessing the images, applying transfer learning to pre-trained models, and optimizing the model using a series of fine-tuning steps. The final model will be assessed using several criteria like accuracy, loss and others.

B. Transfer Learning Process

Base Model: The pre-trained models, DenseNet201, NASNetMobile, and InceptionResNetV2, are used as feature extractors in this research. The top layers of these models, which are responsible for high-level classification, are removed to ensure that the models can be fine-tuned for the specific task of footwear classification. The base models extract relevant features from the input images, which are then passed to the newly added layers for classification.

Fine-tuning: The weights are frozen, meaning the feature extractor layers do not get updated. Only the new layers added to the model, such as fully connected layers, are trained. This approach allows the model to learn from the new dataset without distorting the features learned from the original dataset. After training the new layers, some layers of the pre-trained models may be unfrozen and fine-tuned to improve performance further.

Model Customization: After using the pre-trained base models, several modifications are made Adjust the model to the footwear categorization problem:

- Batch Normalization: Batch normalization is applied after the base model to normalize the activations and improve the stability of the learning process. This helps to accelerate training and allows the model to converge faster.
- Fully Connected Layers: Two fully connected layers with 256 units each and ReLU activations are added after the base model, enables the model to learn complicated patterns and features from the pre-trained layers, specifically tailored to the footwear classification task.
- SoftMax Output Layer: This is the final layer, which produces a probability distribution across the classes. This output layer ensures that the model can classify input images into one of three footwear classes: sneakers, sandals, or boots.

C. Model Compilation and Training

Compilation: The model is compiled using the Adamax optimizer (learning rate: 0.001) and categorical cross-entropy loss, ensuring stable weight updates and improved accuracy in multi-class classification.

Training: The model is trained on preprocessed and augmented data, with validation on a separate dataset. Key parameters:

- Epochs and Batch Size: 10 epochs, batch size of 64 for balanced convergence.
- Callbacks: Early stopping halts training if loss stagnates; learning rate reduction refines convergence.
- Model Saving and History: The trained model is saved in keras format, with accuracy and loss metrics recorded for analysis.

V. PERFORMANCE EVALUATION AND RESULTS

A. Final Training & Validation Results

Training Results: Training metrics such as accuracy, loss and others were tracked.

Table I: Final Training Metrics (Epoch 10)

Model	Train Loss	Train Accuracy
Resnetet152V2	0.02775	99.01%
DenseNet201	0.01850	99.36%
NASNetMobile	0.01491	99.47%
InceptionResNetV2	0.01486	99.48%

Testing Results: The trained models were evaluated on an unseen test dataset. Key metrics were recorded.

Table II: Final Validation Metrics (Epoch 10)

Model	Validation Loss	Validation Accuracy
Resnetet152V2	0.17155	94.97%
DenseNet201	0.08420	97.45%
NASNetMobile	0.37029	91.51%
InceptionResNetV2	0.03862	98.78%

Table III: Other Performance Metrics

Model	Precision	Recall	F1-Score
Resnetet152V2	95.22%	95.03%	95.04%
DenseNet201	97.75%	97.66%	97.67%
NASNetMobile	92.90%	91.46%	91.43%
InceptionResNetV2	98.74%	98.73%	98.73%

B. Epoch-wise Trends in Accuracy & Loss

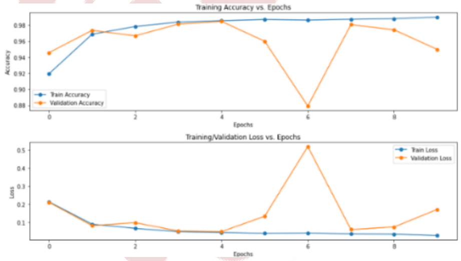


Figure 3. Accuracy and Loss vs. Epochs for Resnetet152V2

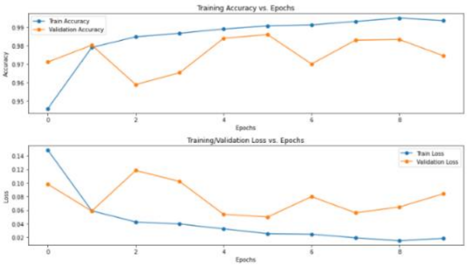


Figure 4. Accuracy and Loss vs. Epochs for DenseNet201

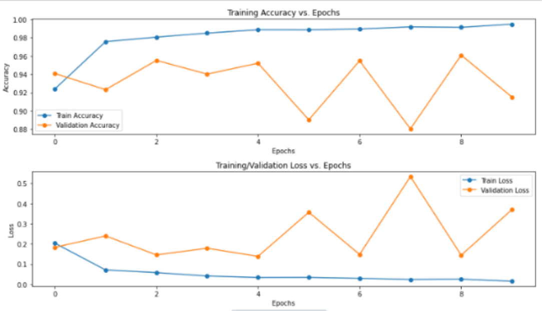


Figure 5. Accuracy and Loss vs Epochs for NASNetMobile

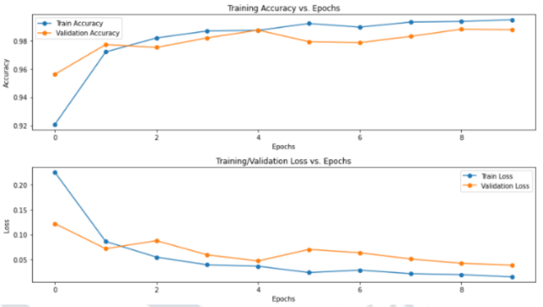
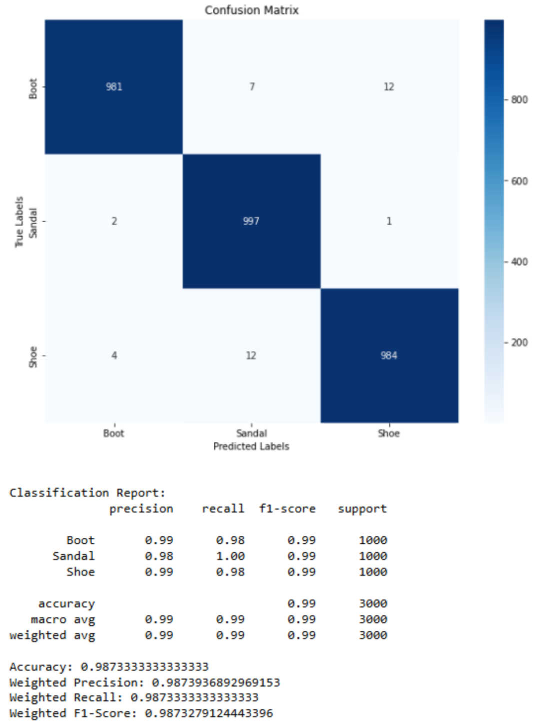


Figure 6. Accuracy and Loss vs Epoch for InceptionResNetV2

C. Confusion matrix of most accurate model



```
Classification Report:
              precision    recall  f1-score   support

        Boot       0.99      0.98      0.99      1000
      Sandal       0.98      1.00      0.99      1000
        Shoe       0.99      0.98      0.99      1000

    accuracy                           0.99      3000
   macro avg       0.99      0.99      0.99      3000
weighted avg       0.99      0.99      0.99      3000

Accuracy: 0.9873333333333333
Weighted Precision: 0.9873936892969153
Weighted Recall: 0.9873333333333333
Weighted F1-Score: 0.9873279124443396
```

Figure 7. Confusion Matrix for InceptionResNetV2 Model

VI. CONCLUSION AND FUTURE WORK

This study utilized transfer learning with pre-trained models (resnetet152V2, DenseNet201, NASNetMobile, and InceptionResNetV2) to classify footwear images. The models showed strong performance in both training and testing, with InceptionResNetV2 providing the best results. Fine-tuning the models and adding fully connected layers allowed them to effectively classify footwear images. Overall, transfer learning proved to be an effective approach for this task.

Future improvements include:

- Model Optimization: Experimenting with optimizers like ADAM, SGD and fine-tuning hyperparameters to enhance accuracy and convergence speed.
- Ensemble Methods: Using techniques like boosting to combine predictions, reducing bias and variance for more reliable results.
- Real-Time Deployment: Optimizing model size via quantization and using TensorFlow Lite or TensorFlow.js for efficient mobile/web inference.
- Dataset Expansion: Adding more footwear categories to improve generalization and classification accuracy.

Exploring Advanced Models:Testing architectures like EfficientNet and Xception to further enhance performance.

REFERENCES

[1] Onalaja, J., Shahra, E. Q., Basurra, S., & Jabbar, W. A. (2024). Image Classifier for an Online Footwear Marketplace to Distinguish between Counterfeit and Real Sneakers for Resale. Sensors, 24(10), 3030. https://doi.org/10.3390/s24103030.

[2] S. Kumar, "Adaptive Fusion-Enhanced CNN Architectures for Advanced Image Classification in E-Commerce," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-7, doi: 10.1109/ICCCNT61001.2024.10724724.

[3] B. Suvarna, A. M. S. S. Chandra, Y. J. S. Ganesh, S. M. Kaif and B. S. Teja, "Optimizing Footwear Image Classification with Hyperparameter Tuning: Insights from the UTZappos Dataset," 2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC), Gwalior, India, 2024, pp. 1034-1039, doi: 10.1109/AIC61668.2024.10731099.

[4] P. Shourie, V. Anand and S. Gupta, "Footwear Fusion: Convolutional Neural Network for Shoe, Sandal, and Boot Classification," 2024 4th International Conference on Sustainable Expert Systems (ICSES), Kaski, Nepal, 2024, pp. 1508-1511, doi: 10.1109/ICSES63445.2024.10763197.

[5] M. Hameed, A. Al- Wajih, M. Shaiea, M. Rageh and F. A. Alqasemi, "Clothing Image Classification Using VGG-19 Deep Learning Model for E-commerce Web Application," 2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA), Sana'a, Yemen, 2024, pp. 1-7, doi: 10.1109/eSmarTA62850.2024.10639001.

[6] K. S. Gill, A. Sharma, V. Anand and R. Gupta, "Smart Shoe Classification Using Artificial Intelligence on EfficientnetB3 Model," 2023 International Conference on Advancement in Computation \& Computer Technologies (InCACCT), Gharuan, India, 2023, pp. 254-258, doi: 10.1109/InCACCT57535.2023.10141787

[7] J. D. D. Jayaseeli, D. Malathi, B. Aljaddouh, F. Alaswad, A. Shah and D. Choudhary, "Image Classification Using Federated Averaging Algorithm," 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2023, pp. 675-680, doi: 10.1109/ICCCIS60361.2023.10425013.

[8] João Oliveira, Rui Gomes, Dibet Gonzalez, Nuno Sousa, Somayeh Shahrabadi, Miguel Guevara, Maria José Ferreira, Pedro Alves, Emanuel Peres, Luís Magalhães, Telmo Adão, Footwear segmentation and recommendation supported by deep learning: an exploratory proposal, Procedia Computer Science, Volume219, 2023, Pages 724-735, ISSN1877-0509, https://doi.org/10.1016/j.procs.2023.01.345.

[9] M. Sari and W. F. Al Maki, "Improving K-Nearest Neighbor Performance in Footwear Classification Using Leave One Out Cross Validation," 2023 3rd International Conference on Intelligent Cybernetics Technology \& Applications (ICICyTA), Denpasar, Bali, Indonesia, 2023, pp. 55-60, doi: 10.1109/ICICyTA60173.2023.10428849.

[10] Z. Zhang, Q. Gao, L. Liu and Y. He, "A High-Quality Rice Leaf Disease Image Data Augmentation Method Based on a Dual GAN," in IEEE Access, vol. 11, pp. 21176-21191, 2023, doi: 10.1109/ACCESS.2023.3251098.

[11] J. You, G. Huang, T. Han, H. Yang and L. Shen, "A Unified Framework from Face Image Restoration to Data augmentation Using Generative Prior," in IEEE Access, vol. 11, pp. 2907-2919, 2023, doi:10.1109/ACCESS.2022.3233868

[12] Andyrasika, "Shoe vs Sandal vs Boot Image Dataset," Hugging Face. Available: https://huggingface.co/datasets/Andyrasika/ShoeSandalBootimages. [Accessed:15-Oct-2024].