# COMPARATIVE TRANSFER LEARNING FOR SIT AND STAND CLASSIFICATION: TOWARD SUSTAINABLE WORK-HEALTH SOLUTIONS

Kongfa Wamasing
Student of Master's Degree in Biomedical Engineering, Department of Biomedical Sciences and Biomedical Engineering, Faculty of Medicine, Prince of Songkla University
E-mail: 6810320011@psu.ac.th
Jermphiphut Jaruenpunyasak
Department of Biomedical Sciences and Biomedical Engineering, Faculty of Medicine, Prince of Songkla University
E-mail: jjermphi@medicine.psu.ac.th

**Abstract**

The development of sustainable health monitoring technologies is essential for improving patient care while reducing resource demands in healthcare systems. Furthermore, such technologies are crucial for addressing occupational health concerns among healthcare workers, who are at high risk of developing musculoskeletal disorders due to improper posture, prolonged standing, and incorrect sitting positions during work. Deep learning models have emerged as a promising tool for automatic posture classification and real-time monitoring of healthcare workers' activities. However, traditional deep learning approaches based on convolutional neural networks often require significant computational resources. To address this challenge, our research explores the potential of transfer learning techniques for efficient and accurate classification of sitting and standing postures among healthcare workers. In addition, our online dataset consisted of 4,000 annotated sitting and standing images with various hospital scenarios. We also evaluated pre-trained deep learning architectures on this dataset to assess their transferability for posture classification. Our results demonstrated that all transfer learning provided an efficient and sustainable solution, achieving high classification performance with all models exceeding 0.96 in accuracy, precision, recall, and F1-score. Notably, MobileNetV2 stands out as a highly efficient option, with only 3.5 million parameters and a weight size of 9 MB. In contrast, VGG16 and VGG19, despite their highest classification performance, were impractical for real-time applications due to their largest size exceeding 500 MB. Our findings contribute to the development of sustainable work-health solutions by enabling the integration of efficient deep learning models into monitoring systems for real-world hospital environments.

**Keywords:** Posture Classification, Transfer Learning, Deep Learning, Sustainable Work Health Solutions

## Introduction

Ergonomics is a scientific discipline focused on understanding the intricate relationship between human physical capabilities and limitations. Its primary goal is to optimize work efficiency, health, safety, comfort, and ease of use. However, poor ergonomic practices, such as prolonged sitting or standing without proper consideration, can overwhelm the body's coping mechanisms, leading to physical discomfort, emotional stress, reduced productivity, and an increased risk of musculoskeletal disorders (MSDs). According to Sirajudeen, et al. (2013), neglecting posture management in work environments can have particularly detrimental outcomes. Globally, MSDs affected approximately 494 million people in 2020, a staggering

increase of 123.4% from 221 million in 1990. By 2050, the prevalence of MSDs is projected to rise by 115%, reaching nearly 1.06 billion cases.

MSDs are particularly prevalent among females and those aged 65-69 years. In 2020, they were the sixth leading cause of Years Lived with Disability (YLDs), contributing to 42.7 million YLDs and 83,100 deaths. This trend highlights the growing public health concern (Gill, et al., 2023). Moreover, research on the link between MSDs and job performance in nursing staff has identified improper body posture and excessive strain as key contributors to the onset of MSDs, primarily affecting the shoulders, lower back, and neck. A study by Ou Y-K, et al. (2021) suggests that these physical ailments significantly impact nurses' work capacity and job performance, with a heavy workload exacerbating the problem. These findings highlight the critical need to promote occupational safety within healthcare environments. Ensuring the physical well-being of nursing staff is not only essential for reducing the risk of musculoskeletal disorders but also plays a vital role in maintaining the efficiency and quality of patient care delivery.

To mitigate the risk of MSDs in healthcare settings, it is crucial to develop effective solutions to optimize posture. Traditional methods for posture assessment, such as manual observation and basic sensor data, are limited by subjectivity and lack the necessary accuracy for reliable posture classification. In contrast, Convolutional Neural Networks (CNNs) have shown promising potential for automated, real-time posture classification. Nonetheless, training CNN-based models is time-consuming and requires substantial computational resources. Additionally, collecting large and diverse datasets to train these models in busy hospital settings can be particularly challenging due to the complexities of obtaining accurate data.

To address the challenges associated with posture classification in busy hospital settings, we propose leveraging pre-trained models as a means of enhancing this process. Specifically, our study aims to optimize these pre-trained models using image augmentation techniques on new datasets. This approach enables us to minimize the time and computational resources required for training, thereby improving the efficiency and accuracy of posture classification.

By utilizing this approach, we can develop robust models capable of accurately distinguishing between sitting and standing postures, ultimately contributing to safer healthcare work environments and improved work-health outcomes for healthcare professionals. Our solution provides a scalable means of optimizing posture classification, ensuring more reliable ergonomic evaluations and enabling healthcare organizations to better support the well-being of their staff.

**Research Objectives**

1. To develop an algorithm for evaluating human posture using artificial intelligence.
2. To compare transfer learning models using images as a dataset for training models.

**Literature Review**

In developing machine learning models for image classification, two critical factors are key: training performance and duration. Traditional CNN-based approaches typically involve training from scratch, which can be computationally intensive and time-consuming due to the complexity of high-dimensional image data. In contrast, transfer learning offers a more efficient approach by utilizing pre-trained models that have already extracted valuable features from large-scale datasets. This literature review examines the effectiveness and efficiency of developing models for medical image classification using architectures such as MobileNet

(Mobile Network) (Sandler, et al., 2018; Howard, et al., 2019), DenseNet (Densely Connected Convolutional Network) (Rybiałek & Jeleń, 2020), ResNet (Residual Network) (He, et al., 2015; Jaruenpunyasak, et al., 2025), and VGG (Visual Geometry Group Network) (Deepthi, et al., 2024).

First of all, MobileNet models have gained popularity in mobile application devices due to their compact size and impressive performance in real-time scenarios. Its structure is based on depthwise separable convolutions, which separate spatial and depth filtering into two efficient operations (Howard, et al., 2017) As a result, the MobileNetV3 excels in terms of precision and recall across various complex pose classifications such as yoga posture classification (Rajendran & Sethuraman, 2024). Considering efficiency, MobileNetV2 in medical monitoring on an IoT device used only 44% CPU and 2.1 FPS for inference time (Nguyen Huu, et al., 2022). However, these models still face the challenge of balancing model size reduction with accuracy preservation (Iandola, et al., 2016). In summary, MobileNet models are well-suited for real-time applications due to their compact size, but they require fine-tuning to achieve optimal performance on specific datasets.

DenseNet differs from MobileNet in its approach to addressing the vanishing gradient problem. By introducing dense connectivity between layers, DenseNet allows each layer to receive input from all preceding layers and pass its own feature maps to all subsequent layers. This architecture has been widely adopted in health-oriented applications such as pneumonia classification (Bundea & Danciu, 2024), and breast cancer classification (Rybiałek & Jeleń, 2020), and various classification tasks (Zhou, et al., 2022). Nonetheless, Siddiqui notes that DenseNet still faces challenges related to redundant skip connections, which can result in increased memory usage and computational inefficiency. In short, DenseNet offers strong feature reuse and mitigates vanishing gradients for medical image classification, but its efficiency can be improved by addressing redundant skip connections for resource-constrained environments.

The ResNet model, introduced by He, et al. (2015), utilizes residual layers with shortcut connections to facilitate the training of very deep networks and overcome the degradation problem. By allowing gradients to flow through identity mappings, this innovation enables ResNet to not only be deeper but also more accurate and efficient in convergence. In case of its effectiveness, ResNet has demonstrated strong performance in various medical applications such as sperm morphology assessment (Jaruenpunyasak, et al., 2025) and pancreatic tumor detection in early-stage diagnosis applications (Deepthi, et al., 2024). Nevertheless, ResNet's computational complexity poses challenges for deployment in embedded or mobile systems (Qawasmeh, et al., 2025). Briefly, ResNet's use of residual connections are well-suited for medical image analysis, but deployment in low-resource environments still requires further optimization to achieve efficient and effective results.

VGG, introduced by Simonyan and Zisserman (2014), boasts a straightforward and uniform architecture built with 3×3 convolutional filters and 2×2 max-pooling layers. Its design showcases that deep networks with small filters can excel in large-scale image recognition tasks (Deepthi, et al., 2024). VGG is widely used in medical imaging due to its simplicity and strong classification performance as skin cancer classification when combined with transfer learning (Jalili, et al., 2024). However, VGG falls short regarding efficiency compared to MobileNet, DenseNet, and ResNet due to its larger size and higher computational cost, which can limit its suitability for real-time or mobile applications (Wu, et al., 2020). To sum up, VGG's uniform architecture contributes to its effectiveness in classification tasks and ease of implementation. Nonetheless, improving its efficiency through model optimization remains crucial for real implementation

To conclude, although transfer learning has advanced medical classification, research on human pose classification remains limited. This study aims to evaluate multiple pre-trained models for human posture classification, focusing on model parameters, training efficiency, and performance metrics.

**Research Methodology**

This section presents an overview of the dataset used in this study, including a description of the methods utilized for image augmentation, feature extraction, model development, and experimental testing. The research process consists of four sequential steps: data acquisition, image augmentation, transfer learning approach, and evaluation.

1. Data Acquisition

Our research on human sitting and standing classification in hospital scenarios was supported by a diverse dataset of human foreground images sourced from publicly available online sources (Nambair, 2021; Maurya, 2024). To ensure data quality, we applied specific filters to retain only images with the entire body visible, resulting in 471 sitting and 337 standing images. Additionally, we randomly selected 200 samples from the dataset with balanced classes. For hospital background images, we chose 30 wide-angled view images that included the floor from an online dataset (Ahmad, 2019). By combining these foreground and background images through data augmentation processes, we created a comprehensive dataset for our research.

2. Image Augmentation

To enhance our human sitting and standing dataset in hospital scenarios, we employed two primary methods for data augmentation. Firstly, we used a pre-trained DeepLabV3 model (Chen, 2017) to segment human images from the original dataset. This resulted in transparent foreground images that were then resized to match the background images. This approach yielded a dataset of 6,000 images (Asati, 2021). Secondly, we applied general image augmentation techniques, including rotation, shift, shear, and flip transformations with varying ranges (Shorten, 2019; Islam, 2024). The results were 60,000 images, which were then randomly sampled to create a new dataset (N = 4,000) to reduce overfitting. Moreover, the parameters of each process are illustrated in Table 1. Furthermore, the examples of these sitting and standing images were displayed in Figure 1 and Figure 2, respectively.

Table 1 The parameters of the image augmentation process

| Method | Parameters | Number of Outcome Samples |
|---|---|---|
| Human Segmentation | DeepLabv3 model with a ResNet50 backbone | 200 |
| Resize | Foreground to 224 x 224<br>Background to 224 x 224 | 200<br>30 |
| Merge | Foreground (N = 200) and Background (N = 30) | 6,000 |
| Image Augmentation (10 times) | Rotation range ±10°<br>Shift range ±10%<br>Shear range ±20%<br>Horizontal flip | 60,000 |
| Random Selected | Balanced classes | 4,000 |

3. Transfer Learning Approach

In this experiment, we leveraged six pre-trained convolutional neural network (CNN) models such as MobileNetV2, DenseNet121, DenseNet201, ResNet50, VGG16, and VGG19. The classification head for each model is a single linear layer adapted to match the number of target classes. We utilized the PyTorch library to load the default weights of these pre-trained models, capitalizing on the knowledge gained from large-scale datasets such as ImageNet. The dataset consisted of 4,000 images, which were subsequently divided into training, validation, and test sets with an 80%, 10%, and 10% ratio respectively. For data preprocessing, the images were resized to 224x224 pixels, converted into tensors, and standardized based on the mean and standard deviation values of ImageNet. The models were trained using the training dataset, with a learning rate of 0.0001, 50 epochs, and the Adam optimizer. We also froze the convolutional layers of the pre-trained models to prevent overfitting, allowing only the custom classification layers to be trained on the posture dataset.

Figure 1: Example of human sitting images,
Source: modified from Nambair (2021), Maurya (2024), and Ahmad (2019)

Figure 2: Example of human standing images,
Source: modified from Nambair (2021), Maurya (2024), and Ahmad (2019)

4. Evaluation

Evaluation of the transfer learning models involves three key aspects: model complexity, training and inference efficiency, and classification performance.

Firstly, we examine the number of trainable parameters for each model, providing insight into its complexity, computational requirements, and potential for overfitting or underfitting. This metric enables us to assess the resource intensity of each model and its expected performance on smaller or larger datasets.

Secondly, we evaluate training and inference time, as well as analyze loss and accuracy curves for both the training and validation datasets. We measure the total training time and inference time per image to gauge the efficiency of each model, while plotting loss and accuracy curves can support us assess how the model converges during training and generalizes effectively.

Lastly, we evaluate the classification performance based on the confusion matrix, as shown in Table 2. The performance metrics for each model, including accuracy, precision, recall, and F1-score, were calculated using Equations (1) – (4). These metrics are essential for understanding the model's ability to classify sitting and standing postures accurately.

Table 2 Confusion matrix

| | Actual Sitting | Actual Standing |
|---|---|---|
| Predicted Sitting | True Positive (TP) | False Positive (FP) |
| Predicted Standing | False Negative (FN) | True Negative (TN) |

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

$$\text{Recall} = TP / (TP + FP) \quad (3)$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

**Research Results**

Table 3 presents the number of parameters, model size, and performance metrics of six pre-trained CNN models for posture classification: MobileNetV2, DenseNet121, DenseNet201, ResNet50, VGG16, and VGG19. First of all, VGG19 achieved perfect scores across all metrics (accuracy, precision, recall, and F1-score = 1.0000), followed closely by VGG16 with 99.75% accuracy. However, both models are considerably large, with sizes exceeding 500MB and over 138 million parameters. In contrast, MobileNetV2, the smallest model with only 3.5 million parameters and 9MB in size, delivered a strong performance with 98% accuracy. DenseNet201 also offered a good balance between size and performance, achieving 97.75% accuracy with a moderate model size. These findings indicate that while larger models yield higher accuracy, lightweight architectures like MobileNetV2 can offer competitive performance with significantly fewer parameters and smaller model size.

Table 3 The result of this experimental

| Model | $N_P$ (Million) | Size (MB) | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|---|
| MobileNetV2 | 3.5 | 9 | 0.9800 | 0.9800 | 0.9800 | 0.9800 |
| DenseNet121 | 8.0 | 28 | 0.9625 | 0.9626 | 0.9625 | 0.9625 |
| DenseNet201 | 20.0 | 72 | 0.9775 | 0.9775 | 0.9775 | 0.9775 |
| ResNet50 | 25.6 | 92 | 0.9650 | 0.9657 | 0.9650 | 0.9650 |
| VGG16 | 138.4 | 545 | 0.9975 | 0.9975 | 0.9975 | 0.9975 |
| VGG19 | 143.7 | 524 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

Note: $N_P$ stands for number of parameters

**Conclusion and Discussion**

The development of sustainable health monitoring technologies is crucial for optimizing patient care and mitigating occupational health risks among healthcare workers, particularly in preventing musculoskeletal disorders caused by poor posture. This study leverages transfer learning to investigate the application of deep learning models for posture classification that represents sitting and standing postures in hospital scenarios. The evaluated

pre-trained models exhibit exceptional classification performance, with accuracy, precision, recall, and F1-scores consistently above 0.96. Notably, MobileNetV2 emerges as the most efficient option that achieves 98% accuracy while boasting only 3.5 million parameters and a mere 9 MB model size. This model may be an ideal choice for real-time and resource-constrained healthcare applications. In contrast, VGG16 and VGG19 offer superior performance, but they require significant memory and computational resources, which limit their practicality in embedded systems. These findings underscore the potential of lightweight deep learning models to support sustainable and real-time posture monitoring solutions in healthcare environments that enhance occupational safety and efficient patient care delivery. Future research will focus on expanding the dataset with diverse postures and hospital contexts, as well as developing real-time monitoring systems integrated with edge devices to assess practical deployment performance.

**Acknowledgement**

We would like to express our sincere gratitude to the Faculty of Medicine, Prince of Songkla University, for their financial support of this master's degree program. Additionally, we would like to extend our appreciation to the "AI innovator · AI engineer · AI researcher - Super AI Engineer Season 5" project organized by the Artificial Intelligence Association of Thailand for their contributions.

**References**

Ahmad, M. (2019). *MIT indoor scenes* (Version 1). Kaggle. https://www.kaggle.com/datasets/itsahmad/indoor-scenes-cvpr-2019/data

Asati, M., Kraisittipong, W., & Miyachi, T. (2021). Extract and merge: Merging extracted humans from different images. In *Advances in computer, communication and computational sciences: Proceedings of IC4S 2019* (pp. 1023–1033). Springer Singapore.

Bundea, M., & Danciu, G. M. (2024). Pneumonia image classification using DenseNet architecture. *Information, 15*(10), 611. https://doi.org/10.3390/info15100611

Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv:1706.05587*. https://arxiv.org/abs/1706.05587

Deepthi, G., Anusha Bamini, A. M., & Praveen, Y. J. (2024). A comparative analysis of pancreatic tumor detection using VGG16, ResNet, and DenseNet. In *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 1–8). IEEE. https://doi.org/10.1109/icaaic60222.2024.10575802

Gill, T. K., Mittinty, M. M., March, L. M., Steinmetz, J. D., Culbreth, G. T., Cross, M., ... & Vasankari, T. J. (2023). Global, regional, and national burden of other musculoskeletal disorders, 1990–2020, and projections to 2050: A systematic analysis of the Global Burden of Disease Study 2021. *The Lancet Rheumatology, 5*(11), e670–e682.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv:1512.03385*. https://arxiv.org/abs/1512.03385

Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., … Le, Q. (2019). Searching for MobileNetV3. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 1314–1324). IEEE. https://doi.org/10.1109/iccv.2019.00140

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., & Adam, H. (2017). *MobileNets: Efficient convolutional neural networks for mobile vision applications* (arXiv:1704.04861). arXiv. https://arxiv.org/abs/1704.04861

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). *SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5MB model size* (arXiv:1602.07360). arXiv. https://arxiv.org/abs/1602.07360

Islam, T., Hafiz, M. S., Jim, J. R., Kabir, M. M., & Mridha, M. F. (2024). A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions. *Healthcare Analytics*, 100340. https://doi.org/10.1016/j.health.2024.100340

Jalili, A., Sajedi, H., Tabrizchi, H., & Mosavi, A. (2024). Skin cancer classification using DenseNet. In *2024 IEEE 22nd Jubilee International Symposium on Intelligent Systems and Informatics (SISY)* (pp. 333–340). IEEE. https://doi.org/10.1109/sisy62279.2024.10737516

Jaruenpunyasak, J., Maneelert, P., Nawae, M., & Choksuchat, C. (2025). Artificial intelligence model for the assessment of unstained live sperm morphology. *Reproduction and Fertility, 6*(2). https://doi.org/10.1530/RAF-25-0014

Maurya, A. (2024). *Human-activity-image-data* (Version 1). Kaggle. https://www.kaggle.com/datasets/ayush5556/human-activity-image-data

Nambair, J. (2021). *Human activity detection dataset* (Version 3). Kaggle. https://www.kaggle.com/datasets/jithinnambiarj/human-activity-detection-dataset

Nguyen Huu, P., Nguyen Thi, N., & Ngoc, T. P. (2022). Proposing posture recognition system combining MobileNetV2 and LSTM for medical surveillance. *IEEE Access, 10*, 1839–1849. https://doi.org/10.1109/access.2021.3138778

Ou, Y.-K., Liu, Y., Chang, Y.-P., & Lee, B.-O. (2021). Relationship between musculoskeletal disorders and work performance of nursing staff: A comparison of hospital nursing departments. *International Journal of Environmental Research and Public Health, 18*(13), 7085. https://doi.org/10.3390/ijerph18137085

Qawasmeh, B., Oh, J.-S., & Kwigizile, V. (2025). Comparative analysis of AlexNet, ResNet-50, and VGG-19 performance for automated feature recognition in pedestrian crash diagrams. *Applied Sciences, 15*(6), 2928. https://doi.org/10.3390/app15062928

Rajendran, A. K., & Sethuraman, S. C. (2024). Transfer learning-based yogic posture recognition using deep pre-trained features. *SN Computer Science, 5*(6). https://doi.org/10.1007/s42979-024-03086-8

Rybiałek, A., & Jeleń, Ł. (2020). Application of DenseNets for classification of breast cancer mammograms. In *Lecture Notes in Computer Science: Computer Information Systems and Industrial Management* (pp. 266–277). Springer. https://doi.org/10.1007/978-3-030-47679-3_23

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4510–4520). IEEE. https://doi.org/10.1109/cvpr.2018.00474

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*(1), 1–48. https://doi.org/10.1186/s40537-019-0197-0

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition* (arXiv:1409.1556). arXiv. https://arxiv.org/abs/1409.1556

Sirajudeen, M. S., Pillai, P. S., & Vali, G. M. (2013). Assessment of knowledge of ergonomics among information technology professionals in India. *Age (Years), 20*(29), 135.

Wu, M., Ma, W., Li, Y., & Zhao, X. (2020). The optimization method of knowledge distillation based on model pruning. In *2020 Chinese Automation Congress (CAC)*. IEEE. https://doi.org/10.1109/cac51589.2020.9327625

Zhou, T., Ye, X., Lu, H., Zheng, X., Qiu, S., & Liu, Y. (2022). Dense convolutional network and its application in medical image analysis. *BioMed Research International, 2022*, 2384830. https://doi.org/10.1155/2022/2384830