Universitatea Transilvania din Brașov

INTERDISCIPLINARY DOCTORAL SCHOOL

Faculty of Silviculture and Forest engineering

Gabriel Osei FORKUO

Application of Deep Learning and Computer Vision Techniques for Postural Classification of Forest Operations Using the OWAS Method

SUMMARY

Scientific supervisor

Prof.Dr. Eng. Stelian Alexandru BORZ

BRAȘOV, 2025

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

1. AC: Action Category
2. AI: Artificial Intelligence
3. APM: Analysis of Posture and Motion
4. CNN: Convolutional Neural Network
5. CV: Computer Vision
6. DCNN: Deep Convolutional Neural Network
7. DL: Deep Learning
8. DT: Decision Trees
9. EMG: Electromyography
10. HMM: Hidden Markov Model
11. HOG: Histogram of Oriented Gradients
12. HRNet: High-Resolution Network
13. IMU: Inertial Measurement Unit
14. KNN: K-Nearest Neighbors
15. LCNN: Lightweight Convolutional Neural Network
16. MDS: Multi-dimensional Scaling
17. ML: Machine Learning
18. MSD: Musculoskeletal Disorder
19. MSDs: Musculoskeletal Disorders
20. OWAS: Ovako Working Posture Analysing System
21. PO: Picture Only
22. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

23. PS: Picture With Skeleton

24. REBA: Rapid Entire Body Assessment

25. RF: Random Forest

26. RM: Ratings of the DL model

27. RNN: Recurrent Neural Network

28. RULA: Rapid Upper Limb Assessment

29. SIFT: Scale-Invariant Feature Transform

30. SO: Skeleton Only

31. SVM: Support Vector Machine

32. TL: Transfer Learning

33. YOLO: You Only Look Once

INTRODUCTION

Relevance of smart solutions for postural assessment in forest operations

Forest operations encompass a wide array of activities related to the management, cultivation, and harvesting of forests and timber resources (Rummer et al., 2002). These operations are vital for the sustainable utilization of forest ecosystems, aiming to reconcile economic needs with the fundamental objective of reducing waste and emissions while minimizing impacts on the environment's structures and functions (Heinimann, 2007; Marchi et al., 2018). However, forestry remains one of the most physically demanding and hazardous occupations. Workers are frequently required to perform repetitive tasks (Sibiya et al., 2021), engage in heavy lifting (Paini et al., 2020), and maintain prolonged awkward postures (Yovi and Prajawati, 2015). These factors significantly increase the risk of developing musculoskeletal disorders (MSDs) (Calvo, 2009; Punnett and Wegman, 2004), which include a wide range of inflammatory and degenerative conditions affecting muscles, tendons, ligaments, and joints. According to Da Costa and Vieira (2010), poor ergonomic practices and a lack of attention to postural health are primary contributors to MSDs, which represent one of the most common occupational health issues globally and a significant cause of work-related disability (Bevan, 2015; Enez and Nalbantoğlu, 2019; Pascual and Naqvi, 2008). Consequently, the need for effective, accurate, and efficient postural assessment in forest operations is both urgent and critical to mitigate these risks and prescribe appropriate interventions.

For decades, the field of ergonomics has relied on traditional assessment methods to identify and quantify postural risks. Methodologies such as the Rapid Upper Limb Assessment (RULA) (McAtamney and Corlett, 1993), the Rapid Entire Body Assessment (REBA) (Hignett and McAtamney, 2000), and the Ovako Working Posture Analysing System (OWAS) (Karhu et al., 1977) have been widely used. These methods are endorsed for their ability to systematically classify postures and assign risk levels. However, they rely heavily on observational techniques that are often subjective, time-consuming, and impractical for large-scale or continuous implementation (Gomez-Galan et al., 2017; Takala et al., 2010). Their static, snapshot-based nature often fails to capture the dynamic, real-time postural changes inherent in demanding environments like forestry (Bačić et al., 2024). While modern technologies like wearable sensors and motion capture systems offer more quantitative and objective data, they are often expensive and require controlled environments, limiting their applicability in the rugged, mobile, and outdoor settings of forest operations (Nadeem et al., 2021).

Recent advancements in machine learning (ML), deep learning (DL), and computer vision (CV) offer transformative opportunities to overcome these challenges (Roggio et al., 2024; Yang et al., 2024). These "smart solutions" can analyze postures from video or image data, enabling non-invasive, scalable, and cost-effective assessment. DL models, particularly Convolutional Neural Networks (CNNs) like ResNet-50 (He et al., 2016), have demonstrated remarkable success in image classification and human pose estimation, allowing for the precise tracking of body joints and angles (Cao et al., 2017). This capability is highly relevant for ergonomic risk assessment, where DL can identify posture deviations and predict potential risks from visual data. Similarly, CV technologies, powered by algorithms like OpenPose (Cao et al., 2019) and advanced architectures like the High-Resolution Network (HRNet) (Sun et al., 2019), can now detect and classify human postures with high accuracy, even in complex and dynamic environments. The integration of these advanced technologies into ergonomic assessment workflows promises to enhance the efficiency, accuracy, and real-time applicability of postural analysis. This thesis sought to develop and evaluate such smart solutions tailored for postural assessment in forest operations, addressing the critical role of ergonomics in minimizing work-related risks by leveraging the immense potential of these advanced technologies.

Aim and objectives

The aim of this thesis was to improve the existing postural assessment methods by selecting, testing, tailoring, and validating innovative ML, DL and CV techniques so as to help prevent injuries and improve work ergonomics, save resources, and enable big data analysis in forest operations postural assessment. By integrating established ergonomic methods with ML and CV technologies, this research will bridge gaps in existing methods and provide a novel framework for effective postural ergonomic assessment. The specific objectives of the thesis were as follows:

O1. To perform a comprehensive review of existing ergonomic assessment methods to identify current gaps and limitations;

O2. To collect baseline/representative data on the types of manual, motor-manual and partly mechanized forest operations and the variability in operational factors that affect the postural condition in the form of media files;

O3. To select and test the best candidates of ML and CV classification algorithms by integrating them in the conceptual framework of postural assessment methods for manual, motor-manual and partly mechanized operations implemented especially in Europe;

O4. To identify and optimize the most promising algorithms of ML and CV in terms of performance by fine-tuning of their operating hyperparameters;

O5. To disseminate the results.

Organization of the thesis

The PhD thesis is organized into manuscript-style chapters, each concentrating on a specific topic:

Chapter 1: A systematic survey of conventional and new postural assessment methods. This chapter presents a systematic literature review that identifies gaps in existing postural ergonomic assessment methods, laying the groundwork for future research. It examines contemporary ergonomic assessment techniques, particularly in forestry, contrasting traditional methods with modern approaches utilizing Machine Learning and Computer Vision.

Chapter 2: Development and evaluation of automated postural classification models in forest operations using deep learning-based computer vision. This chapter focuses on the development and evaluation of DL models for automating ergonomic assessments using the OWAS method. It details the creation of a large, annotated image dataset and the comparative performance of four pre-trained CNN models, culminating in the selection and fine-tuning of the best-performing model.

Chapter 3: Approaching full accuracy by deep learning and computer vision in OWAS postural classification. This chapter explores how enhancing conventional 2D images with computer-generated body keypoints can improve the accuracy of DL-based postural classification. It compares the performance of a ResNet-50 model on three different image analysis approaches: images only, skeletons only, and images combined with skeletons.

Chapter 4: Postural classification by image embedding and transfer learning. This chapter investigates the use of image embedding and transfer learning to automate the OWAS method for evaluating postures in motor-manual cross-cutting work. It analyzes the performance of Google's Inception V3 and SqueezeNet models and discusses the challenges of domain differences and unseen data.

Chapter 5: Human and machine reliability in postural assessment of forest operations by OWAS method. This chapter evaluates the reliability of postural assessments by comparing the ratings of three human experts with the predictions of a deep learning model. It analyzes intra- and inter-rater reliability, agreement with the model's "ground truth," and the time efficiency of both human and machine assessments.

Chapter 6: Conclusions. Original contributions. Dissemination of results. This final chapter synthesizes the findings from the preceding chapters, outlines the original contributions of the thesis to the field, and provides a list of scientific publications and presentations that have disseminated the research.

CHAPTER 1. A SYSTEMATIC SURVEY OF CONVENTIONAL AND NEW POSTURAL ASSESSMENT METHODS

1.1. MATERIALS AND METHODS

The purpose of this study was to explore the existing research on ML and CV applications in postural assessment. This review follows a systematic methodology based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2015; Page et al., 2021) to identify, select, appraise, synthesize, and report relevant studies from various databases. A protocol was developed that outlined the objectives, scope, and methodology of the review. Appropriate academic databases (e.g., PubMed, Scopus, ScienceDirect, Google Scholar, Elsevier, Web of Science, IEEE Xplore, Springer Link) were chosen, and a comprehensive list of keywords and synonyms related to the research topic were developed (e.g., musculoskeletal disorders, OWAS, RULA, REBA, ML, DL, CV, postural assessment). The search was limited to articles published in English without any year range.

A total of 182 articles were selected for inclusion in this review, based on their relevance to the topic and methodological quality. Inclusion criteria considered studies that were published on postural assessment at any time, including research conducted with humans, related to ML and CV applications in postural assessment, and were available in English. To extract and analyze relevant information, a data collection framework was created to systematically document key information from selected articles, such as research objectives, methodology, findings, limitations, and gaps. The PRISMA flow diagram in **Figure 1.1** illustrates the methodological phases of the literature review.

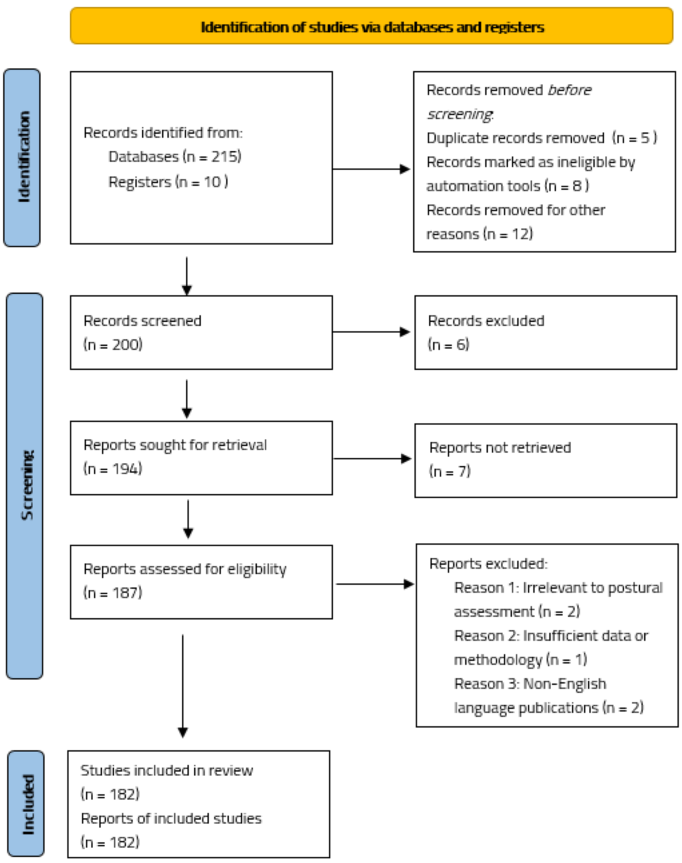


Figure 1.1. The PRISMA flow diagram of the methodological phases of the literature review for this study.

1.2. RESULTS AND DISCUSSION

A wide variety of methodologies used in postural assessment have been identified and grouped into three categories: direct measurements, observational approaches, and self-reports (David, 2005; Gomez-Galan et al., 2017; Li and Buckle, 1999). Worker self-reports (e.g., diaries, interviews, questionnaires) are valuable for gathering information on exposure to work-related hazards but can be vague and unreliable (David, 2005; Kolgiri et al., 2017). Observational methods are essential for assessing the external exposure of workers and can be subjective (e.g., body maps, rating scales) or systematic (e.g., recording postures and movements) (Lorenzini et al., 2023). Simpler observational techniques like OWAS, REBA, and RULA use pro-forma sheets to record workplace exposure and prioritize interventions. Advanced methods use computerized data collection and analysis to provide a more objective assessment of postural risk factors. Direct measurement of exposure variables is achieved by attaching sensors to the subject's workspace, including hand-held instruments like electronic goniometers, or systems that use optical, acoustic, or electromagnetic markers to track body movements (David, 2005; Kolgiri et al., 2017). While these methods provide highly accurate data, they can be costly, require technical expertise, and may be uncomfortable for the subject.

Algorithms form the core of ML and CV systems. Recent advances in ML and DL algorithms have enabled the development of efficient and accurate CV methods for assessing posture from visual cues, such as images and videos (Debnath et al., 2022; Jiang et al., 2023). Common traditional ML algorithms used in CV for postural assessment include Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG), Support Vector Machine (SVM), and Hidden Markov Model (HMM) (Ding et al., 2020; Jiang et al., 2023). Other classification algorithms include Random Forest (RF), Decision Trees (DT), and k-Nearest Neighbors (KNN). In recent years, researchers have proposed several Convolutional Neural Network (CNN) algorithms and architectures, such as stacked hourglass networks, multi-stage pose estimation networks, convolutional pose machines (CPM), and high-resolution nets (Jiang et al., 2023). DL-based models like ResNet-50, YOLO, and OpenPose have demonstrated great success in postural assessment and identification, achieving high accuracy by automatically extracting complex features from images.

The use of ML and CV tools provides several benefits for data analysis, but there are also drawbacks, including computational demands, the need for high-quality data, and the cost of software (Alpaydin, 2020; Imbeault-Nepton et al., 2022). Deep learning is a new trend that involves training algorithms using big data sets and neural networks. However, CV algorithms still face challenges such as occlusion, variations in lighting conditions, scalability, and processing time (Russakovsky et al., 2015). Future research aims to develop new network architectures, refine and adapt ML/CV techniques for specialized scenarios, enhance model interpretability, and address challenges related to training and deploying large-scale models. A promising area is the application of sophisticated ML algorithms to improve the accuracy and reliability of postural assessments in forest operations.

1.3. CONCLUSIONS

Postural assessment is a key aspect of ergonomics. Both conventional and new methods can be used and can be effective in improving ergonomic conditions. ML algorithms are useful as they can learn from their environment and adapt to changes, while CV is a rapidly evolving field for analyzing and interpreting visual data. As ML and CV continue to advance, they hold immense potential to transform numerous industries. Future research should focus on refining these models, enhancing training protocols, expanding annotated datasets, and optimizing data preprocessing techniques to improve accuracy and applicability across diverse environments.

CHAPTER 2. DEVELOPMENT AND EVALUATION OF AUTOMATED POSTURAL CLASSIFICATION MODELS IN FOREST OPERATIONS USING DEEP LEARNING-BASED COMPUTER VISION

2.1. MATERIALS AND METHODS

To the best of the Authors' knowledge, no industry-specific annotated datasets were available at the time of this study. A field campaign was carried out between February and April 2024 to collect video footage of forest workers. This was complemented by videos from previous projects, resulting in 157 video files and 174,231 still images. After a manual similarity removal operation, the dataset covered 229 of the 252 possible OWAS classes. To generate images for the remaining 23 classes, a text-to-photo app (Freepik) was used. A final subset of exactly 23,000 images was retained for model development, including 115 artificially generated images.

The development and evaluation of the models involved several steps carried out using MATLAB R2023b software. The workflow, shown in **Figure 2.1**, included data pre-processing, image annotation, creation of an image datastore, data partition and augmentation, selection of pre-trained CNN models, training, validation and testing.

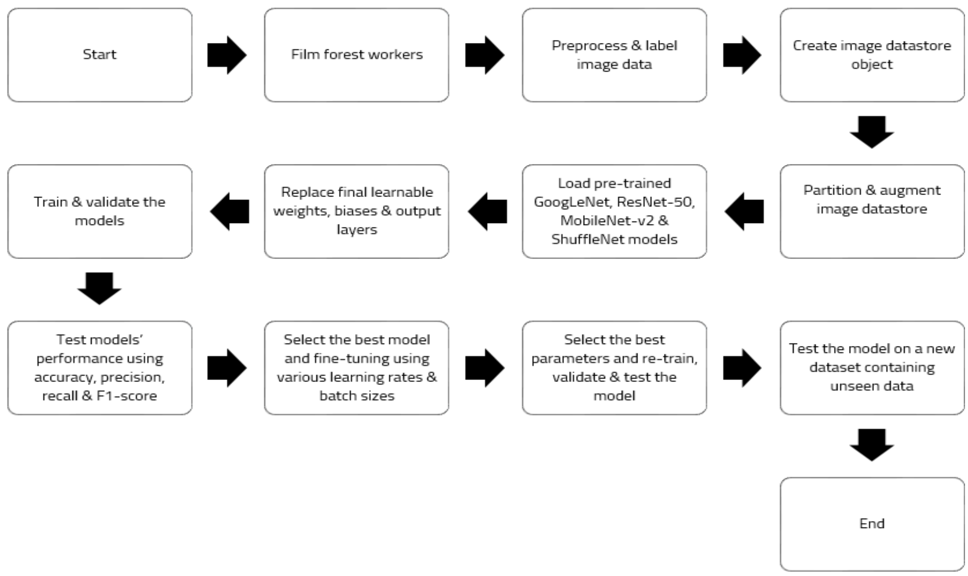


Figure 2.1. Workflow used to develop the smart OWAS classification model

Label definitions were formulated using OWAS posture codes, resulting in 252 image data labels. The data was partitioned into training, validation, and test sets at a ratio of 70:15:15%. Four CNN models—GoogLeNet, ResNet-50, MobileNet-v2, and ShuffleNet—were selected for the study (**Table 2.1**). Transfer learning was used to tune the models to the 252 classes in the dataset.

After evaluating the four pre-trained models, the one with the highest classification accuracy and F1-score (ResNet-50) was selected for fine-tuning. The model's performance was optimized by testing three sets of training options for learning rate (0.01, 0.001, 0.00001) and batch size (32, 64, 128). The final model was selected based on a comprehensive evaluation of performance metrics across various fine-tuning configurations.

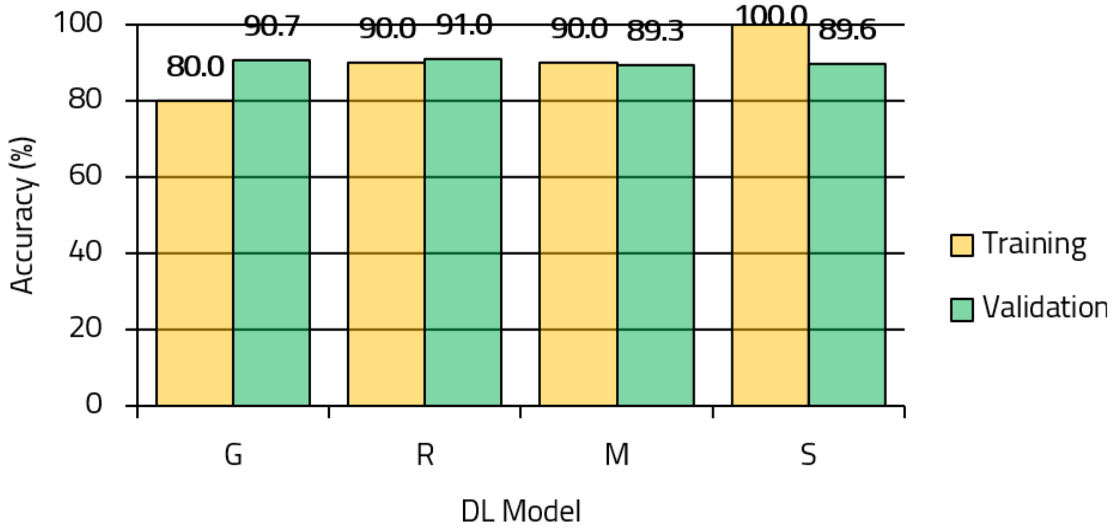
A separate data set of 20 unseen images was used to check the quality of classification and generalization ability of the final model. The classify function of MATLAB was used to produce labels and probability scores for the images.

Table 2. 1. Features of the pre-trained networks selected for the study. Source: MathWorks, Inc., 2024b

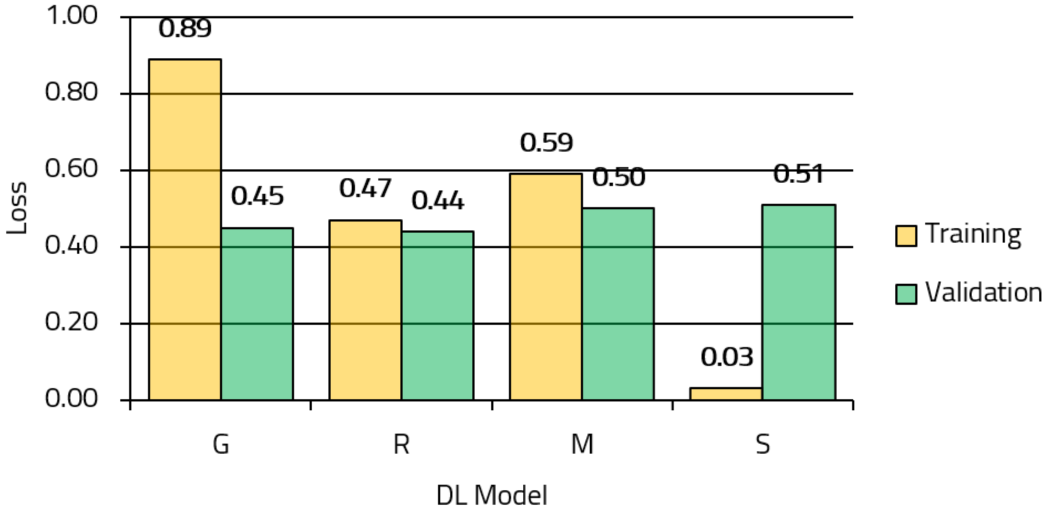
Pre-trained network (Reference)	No. of layers	Depth	Size (MB)	Parameters (M)	Total Learnables (M)	Input Size
GoogLeNet (Szegedy et al., 2015)	144	22	27.0	7.00	6.200	224×224×3
ResNet-50 (He et al., 2016)	177	50	96.0	25.60	24.000	224×224×3
MobileNet-v2 (Sandler et al., 2018)	155	53	13.0	3.50	2.500	224×224×3
ShuffleNet (Zhang et al., 2018)	50	50	5.20	1.40	0.999	224×224×3

2.2. RESULTS

Figure 2.2 summarizes the main results regarding the training and validation of the models. The best-performing model was ResNet-50, with an accuracy of 90.0% and 91.0% during training and validation, respectively. All models showed lower losses during validation compared to training, indicating effective generalization. **Figure 2.3** summarizes the performance metrics of the testing phase. In terms of classification performance, ResNet-50 achieved the highest accuracy of 91.2% and an F1-score of 77.9%.



a



b

Figure 2.2. Classification accuracy (a) and loss (b) of the models during training and validation phases. Legend: G – GoogLeNet, R – ResNet-50, M – MobileNet-v2, S – ShuffleNet

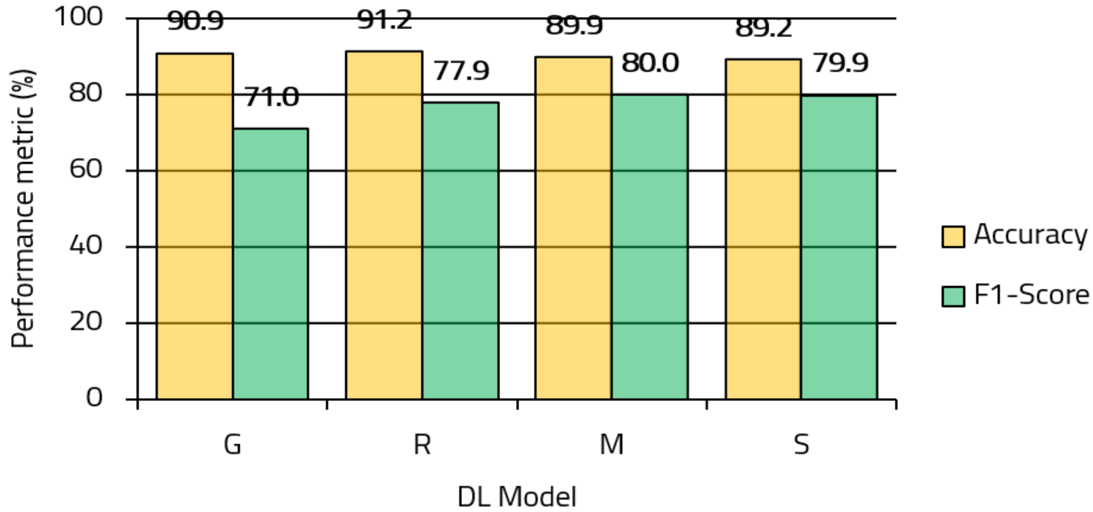


Figure 2.3. Classification performance during testing for the four pre-trained models. Legend: G – GoogLeNet, R – ResNet-50, M – MobileNet-v2, S – ShuffleNet

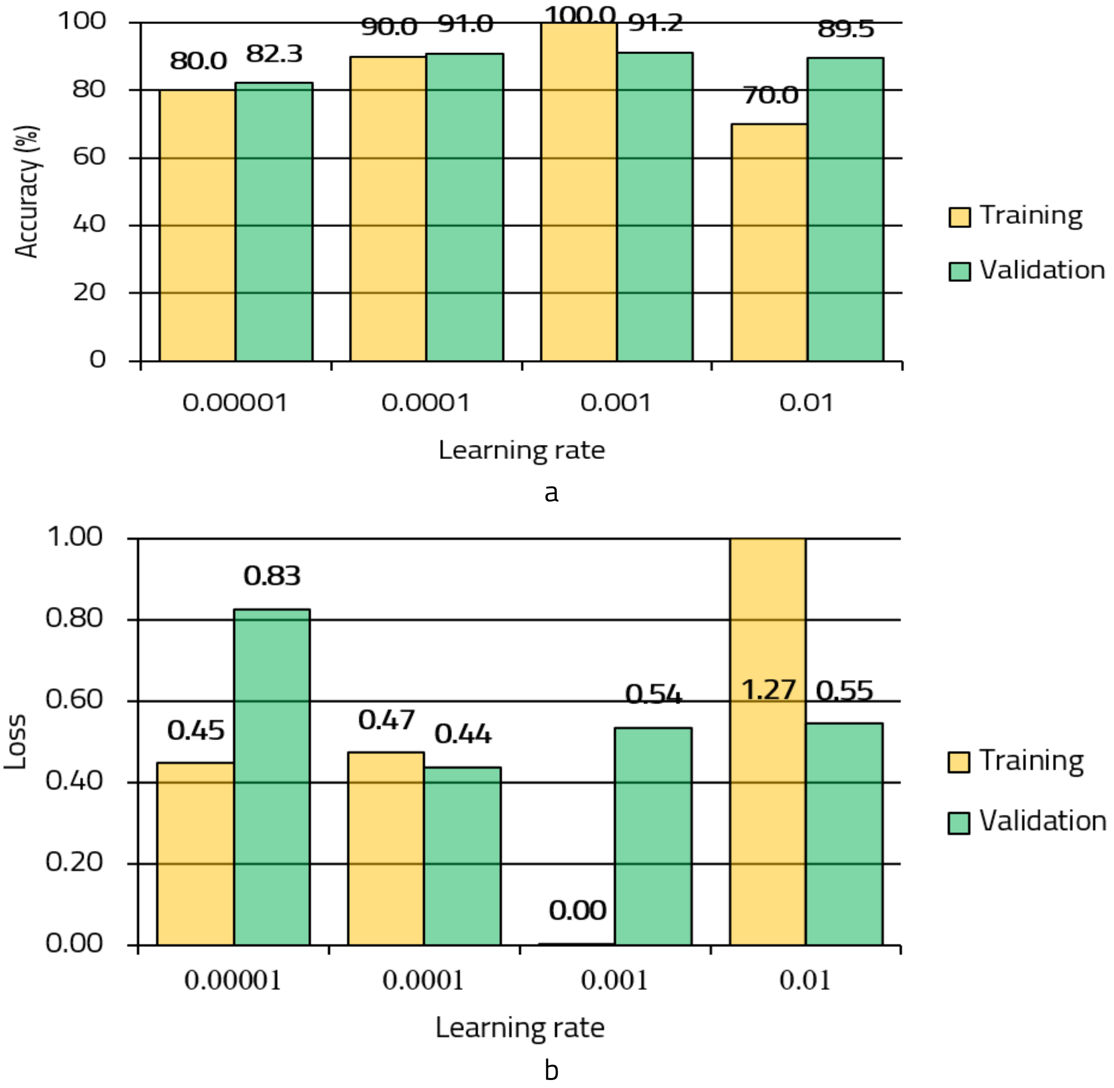


Figure 2.4. Classification accuracy (a) and loss (b) of the ResNet-50 model as a function of the learning rate during training and validation phases. Note: batch size was kept at 10.

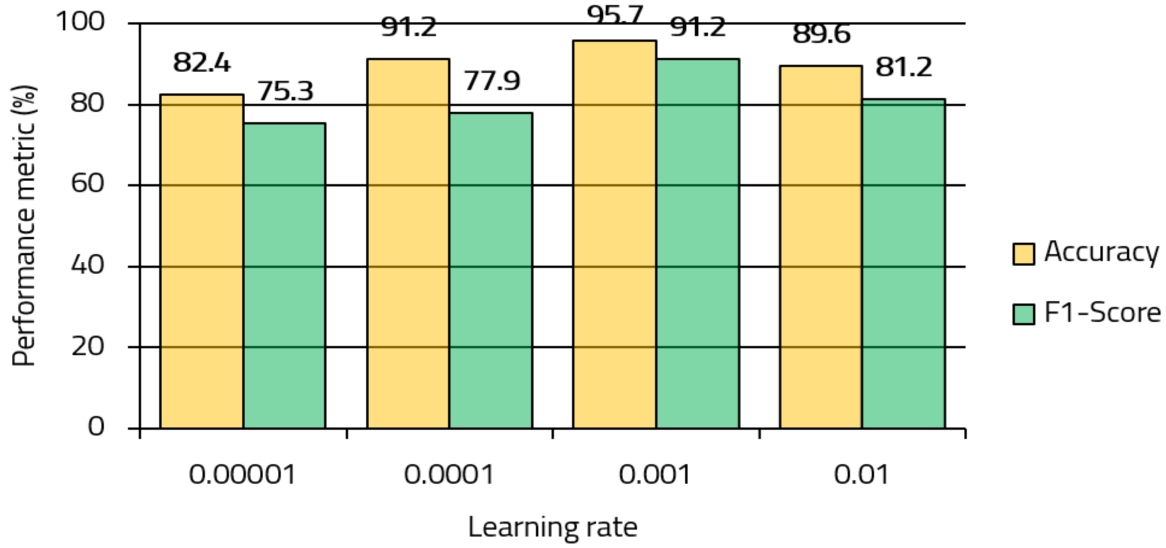
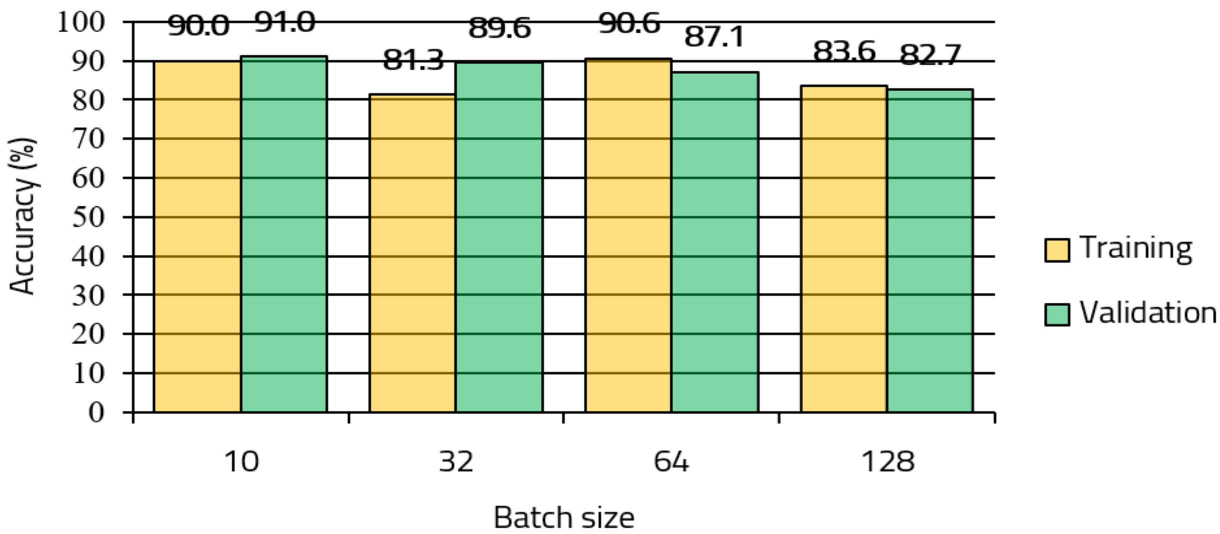
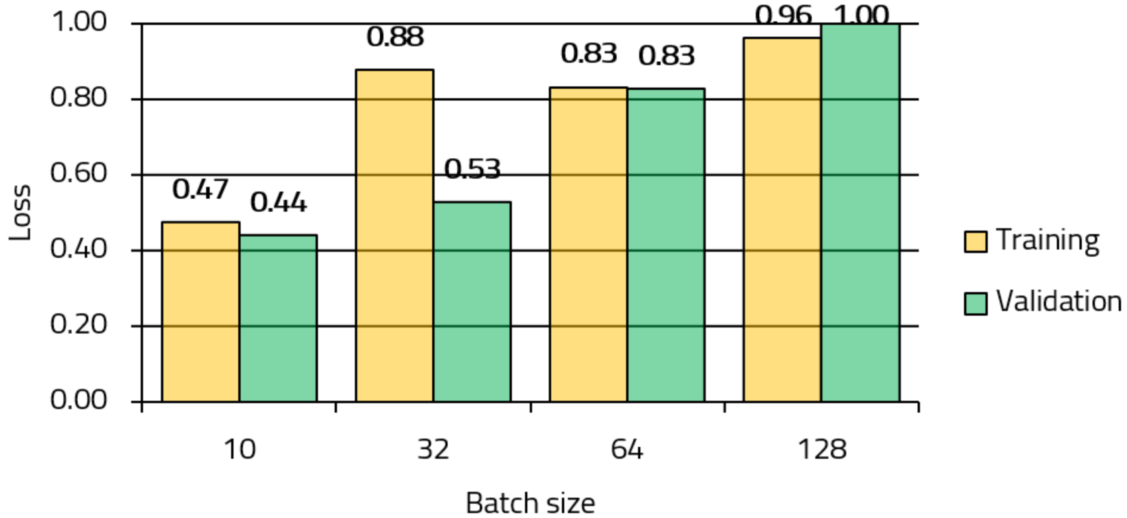


Figure 2.5. Classification performance of the ResNet-50 model during testing. Note: batch size was kept at 10.



a



b

Figure 2.6. Classification accuracy (a) and loss (b) of the ResNet-50 model as a function of the batch size during training and validation phases. Note: learning rate was kept at 0.0001.

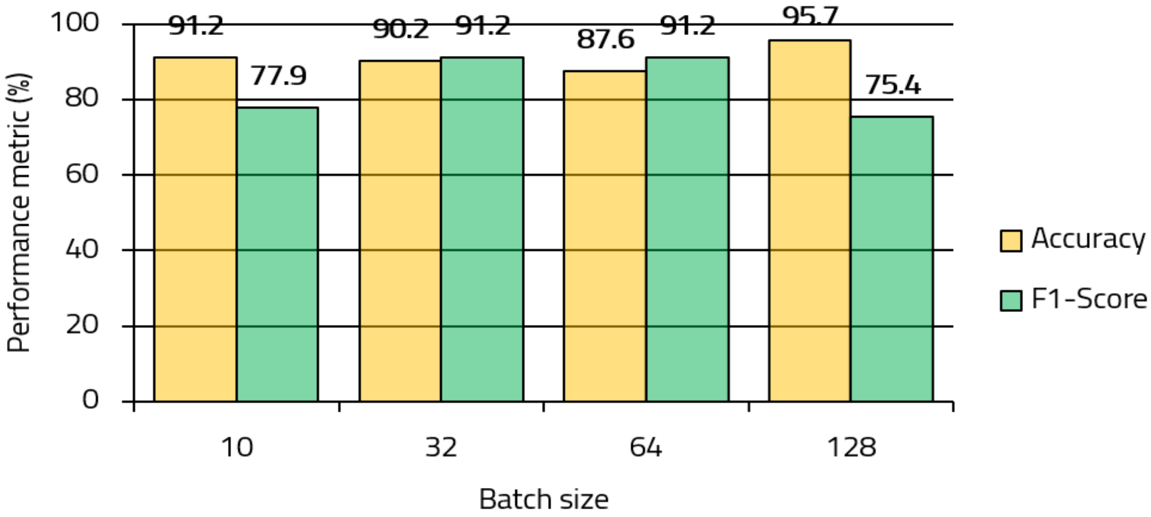
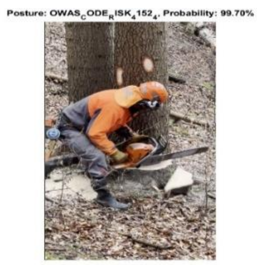


Figure 2.7. Classification performance of the ResNet-50 model during testing. Note: learning rate was kept at 0.0001.

Figure 2.8 shows prediction scores on several examples of unseen data. The model achieved very high probabilities for several images across all postural classes, indicating strong confidence in its predictions. The model also performed well on AI-generated images, with overall probabilities exceeding 50%.

Prediction: postural class = 4142, action category = 4
Probability = 98.94%



Prediction: postural class = 4152, action category = 4
Probability = 99.70%

a

Prediction: postural class = 3111, action category = 1
Probability = 99.97%

Prediction: postural class = 4171, action category = 2
Probability = 99.03%

b

Prediction: postural class = 4171, action category = 2
Probability = 99.94%

Prediction: postural class = 2121, action category = 2
Probability = 99.81%

c

Prediction: postural class = 4263, action category = 4
Probability = 96.50%

Prediction: postural class = 2373, action category = 4
Probability = 71.38%

d

Figure 2.8. Prediction accuracy on unseen data: a – predictions on classes with dominance in number of images, b – predictions on classes with a medium number of images, c – predictions on classes with a low number of images, d – predictions on classes with AI-generated images.

2.3. DISCUSSION

The goal of this study was to develop a smart OWAS model that uses deep learning-based CV to classify working postures. ResNet-50, a DCNN model, emerged as the most robust model in terms of superior performance. The fine-tuning process resulted in a model that achieved an accuracy of over 96%, demonstrating the effectiveness of the approach. The predictions on unseen data reflect the model's confidence levels, with high probabilities indicating robust training. However, variability in probabilities suggests the importance of data quantity and quality. The study also had limitations, including class imbalance and the computationally intensive nature of DL techniques. Future research could explore ensemble learning and address the subjectivity in rating postures from still images.

2.4. CONCLUSIONS

OWAS is a widely used method for classifying working postures. Automating it using deep learning-based CV techniques offers many benefits. The results of this study demonstrate that four commonly used deep learning models were highly accurate in postural classification based on transfer learning. After fine-tuning, ResNet-50 achieved impressive results with over 96% in classification accuracy and F1-score. Potential improvements in classification performance may be achieved using ensemble learning, as well as using recurrent neural networks.

CHAPTER 3. APPROACHING FULL ACCURACY BY DEEP LEARNING AND COMPUTER VISION IN OWAS POSTURAL CLASSIFICATION: AN EXAMPLE ON HOW COMPUTER-GENERATED BODY KEYPOINTS CAN IMPROVE DEEP LEARNING BASED ON CONVENTIONAL 2D DATA

3.1. MATERIALS AND METHODS

A field campaign was executed between February and April 2024 across various sites in Romania (**Figure 3.1**) to document the postures and movements of forest workers. The dataset was extended with videos from prior field campaigns, resulting in a refined set of 23,000 images for model development and testing.

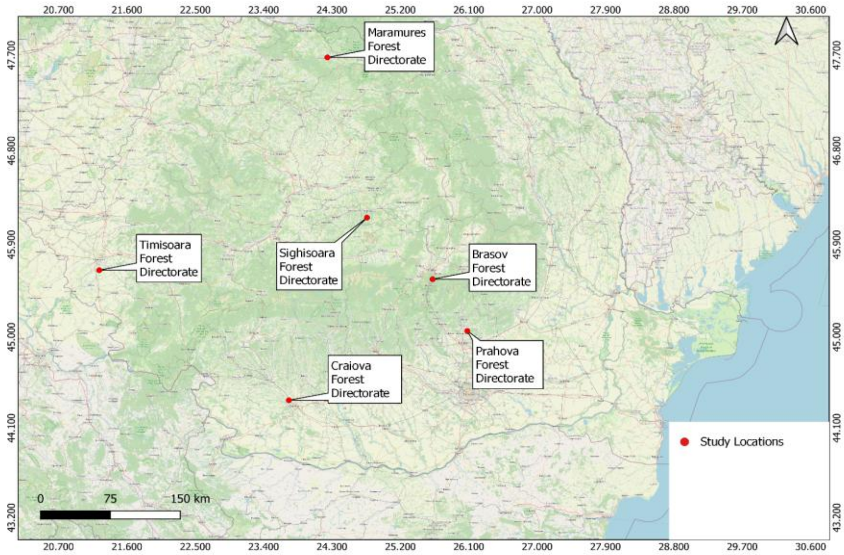


Figure 3.1. Map of Romania showing the sites of field data collection. Note: Developed in QGIS 3.32.1 'Lima' using QuickMapServices, OSM Standard.

Three datasets were used in this study: PictureOnly (PO), which contained the original pictures; PictureWithSkeleton (PS), which contained pictures with the body keypoints and their connectors overlaid on the original image; and SkeletonOnly (SO), which contained the body keypoints and their connectors only. A subset of 1,260 images was set aside for model development. The YOLOv4 object detector was used to detect bounding boxes around persons, and the HRNet object keypoint detector was used to identify keypoints.

The models were developed on a desktop computer using the PyCharm Community Edition Python IDE and TensorFlow. The workflow is shown in **Figure 3.2**. For each model, the ResNet-50 architecture pre-trained on the ImageNet dataset was used as the base model. The model was compiled using Adam optimization, categorical cross-entropy loss, and metrics including accuracy, precision, and recall.

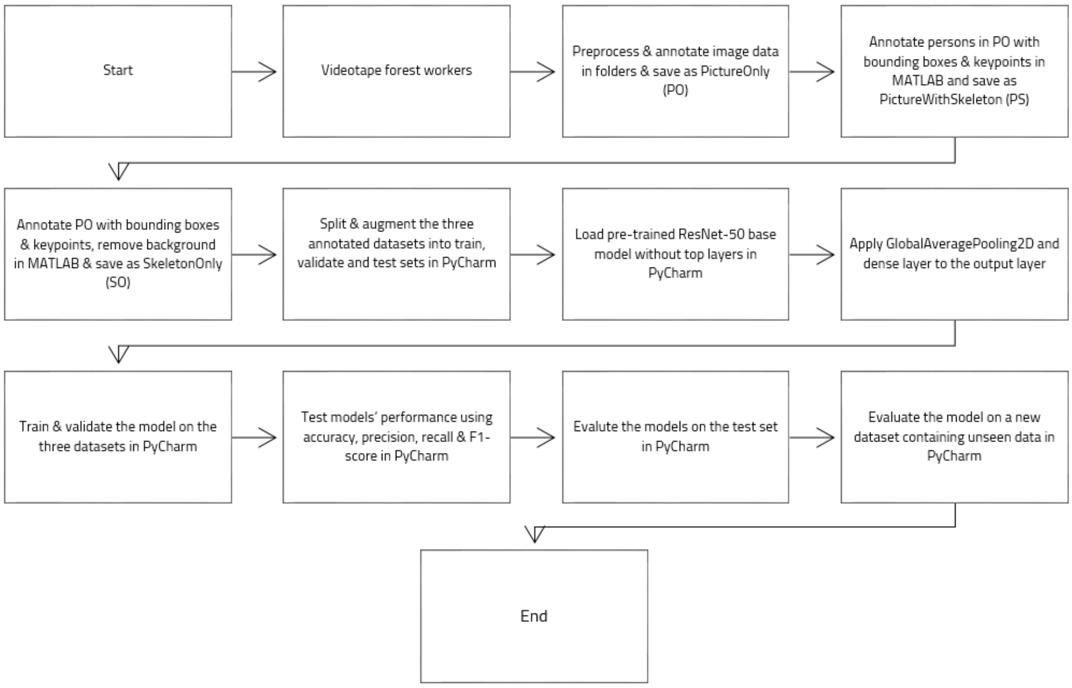
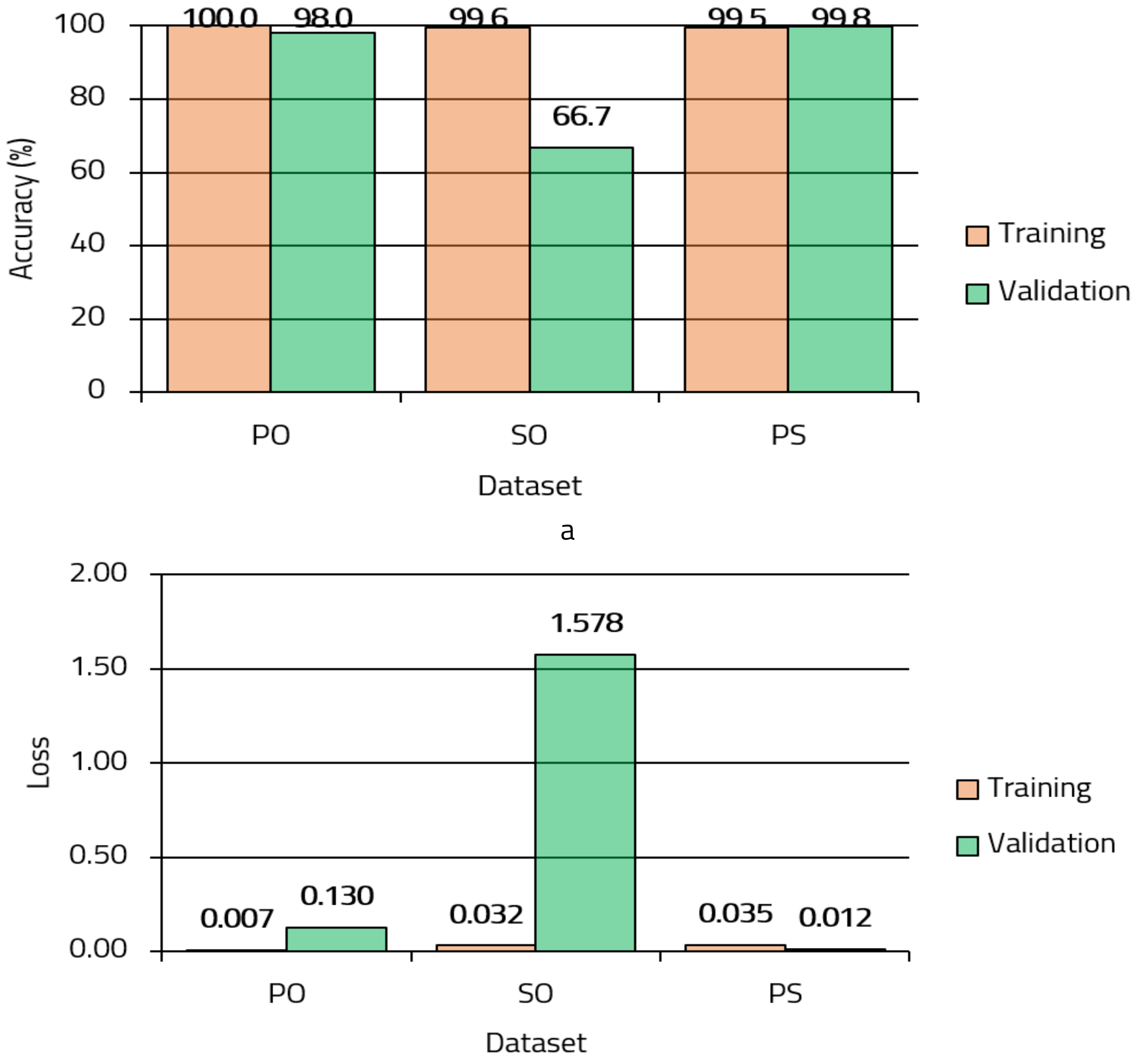


Figure 3.2. Flowchart showing the steps of data collection, annotation, and models' development and evaluation.

The evaluation was done in two steps. First, performance on testing data was evaluated using metrics like accuracy, precision, recall, and F1-score. Second, the evaluation was conducted using a new dataset of 200 unseen images.

3.2. RESULTS

Figure 3.3 summarizes the key findings related to the training and validation of the ResNet-50 models. The model based on the PS dataset performed consistently well, with 99.5% accuracy in training and 99.8% in validation. In contrast, the model with SO exhibited a significant drop in performance from training (99.6%) to validation (66.7%). The model with PO maintained a high performance (100.0% in training to 98.0% in validation). **Figure 3.4** shows the performance metrics during the testing phase. The model with PS achieved a final accuracy and F1-score of 99.8%. The model with PO performed well with an accuracy of 97.8% and an F1-score of 97.30%. The model with SO had a significantly lower performance, with an accuracy of 66.2% and an F1-score of 64.3%.



b

Figure 3.3. Accuracy (a) and loss (b) of the pretrained ResNet-50 models during training and validation phases. Legend: PO - PictureOnly, SO - SkeletonOnly, PS – PictureWithSkeleton.

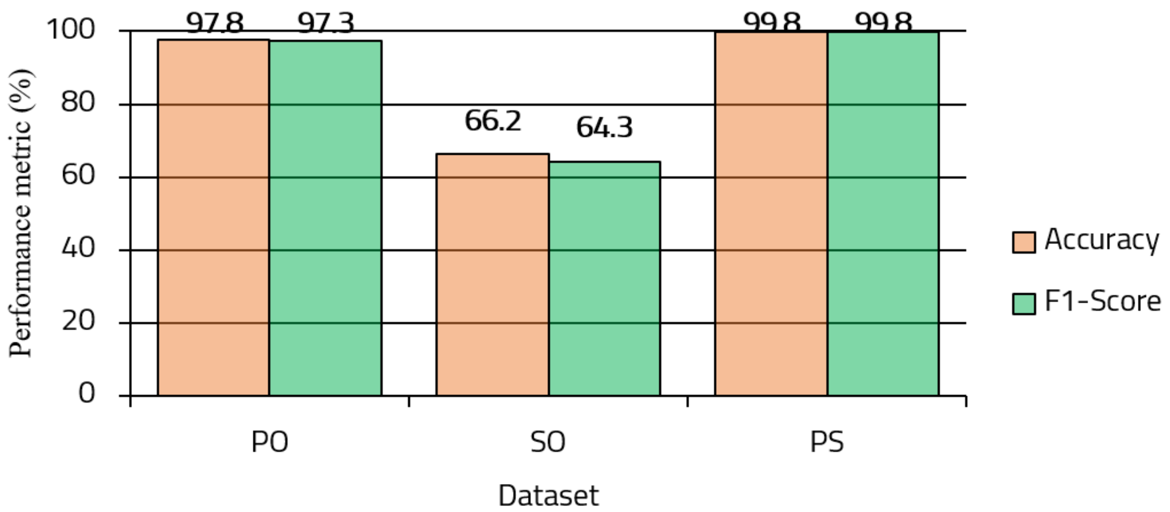


Figure 3.4. Classification performance during testing for the pre-trained ResNet-50 models on the three datasets. Legend: PO - PictureOnly, SO - SkeletonOnly, PS – PictureWithSkeleton.

Figures 3.5, 3.6, and 3.7 show the probabilities of the models on unseen data. The models displayed very high probabilities in several images across all postural classes and forest operations, indicating strong confidence. The probabilities generally exceeded 55%.

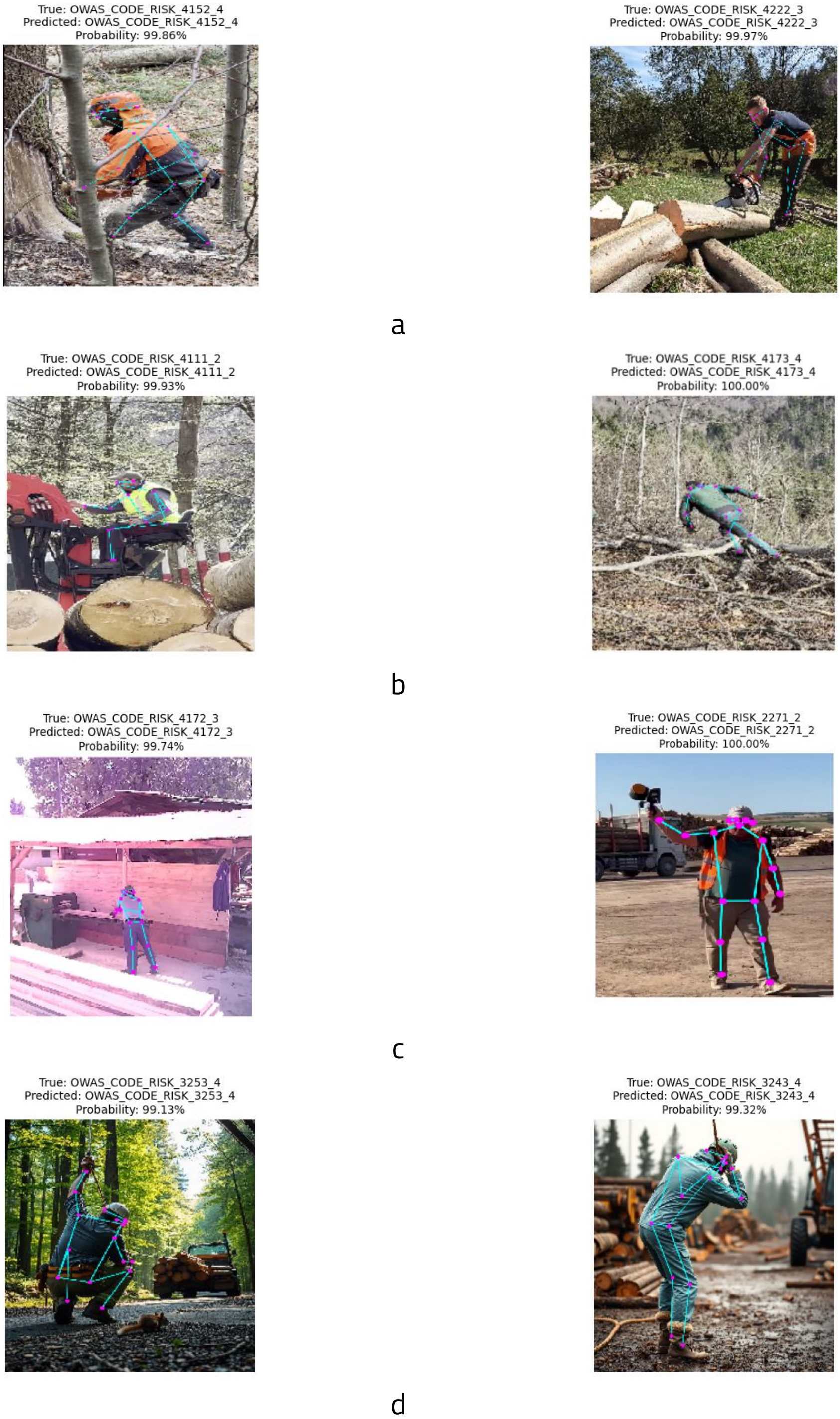


Figure 3.5. Prediction accuracy of the model developed with PS on unseen data: a - predictions on operations with dominance in number of images, b - predictions on operations with a medium number of images, c - predictions on operations with a low number of images, d - predictions on classes with AI-generated images.

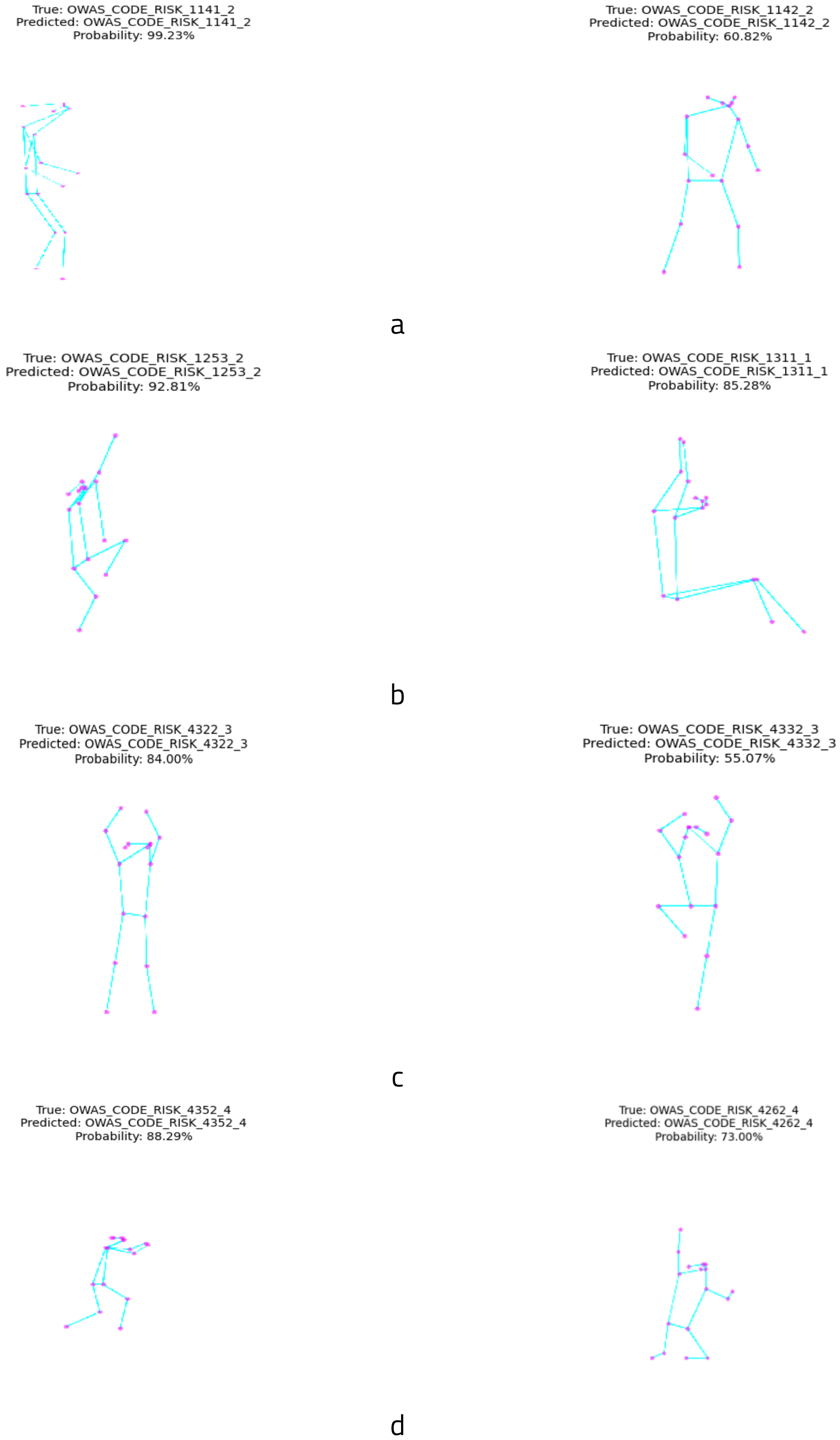


Figure 3.6. Prediction accuracy of the model developed with SO on unseen data: a - predictions on operations with dominance in number of images, b - predictions on operations with a medium number of images, c - predictions on operations with a low number of images, d - predictions on classes with AI-generated images.

Figure 3.7. Prediction accuracy of the model developed with PO on unseen data: a - predictions on operations with dominance in number of images, b - predictions on operations with a medium number of images, c - predictions on operations with a low number of images, d - predictions on classes with AI-generated images.

3.3. DISCUSSION

The ResNet-50 architecture has demonstrated exceptional performance across various CV tasks. In this study, the model was evaluated on three distinct datasets. The results indicate that the ResNet-50 model trained on the PO dataset consistently outperformed those trained on the SO dataset. Notably, the model developed using PS achieved the highest classification performance. The presence of comprehensive, background features in the PO aids in capturing additional contextual information. The high probability scores from model predictions provide important insights into model confidence and reliability. The study also had limitations, including the small size of the datasets and the underrepresentation of some posture categories.

3.4. CONCLUSIONS

This study demonstrates that the ResNet-50 model trained on the PS dataset achieved the highest performance of 99.8% accuracy and F1-score during the testing phase. This underscores the importance of comprehensive skeleton and contextual information in training robust models. The integration of OWAS with deep learning is a promising alternative for a more effective assessment and mitigation of ergonomic risks.

CHAPTER 4. POSTURAL CLASSIFICATION BY IMAGE EMBEDDING AND TRANSFER LEARNING: AN EXAMPLE OF USING THE OWAS METHOD IN MOTOR-MANUAL WORK TO AUTOMATE THE PROCESS AND SAVE RESOURCES

4.1. MATERIALS AND METHODS

The field survey included the collection of media footage documenting the motor-manual crosscutting of teak (*Tectona grandis L.f.*) in Kanchanaburi province, Thailand (**Figure 4.1**). A total of 14 videos were collected, resulting in more than 5000 still images deemed valid for detailed analysis.

Figure 4.1. An example of an image used for modeling (left) and the location of the study (right).

All images were visually assessed according to the OWAS postural classification system. Following the visual assessment, the images were stored in folders according to the four-digit codes (Posture dataset) and corresponding action categories (Action dataset). The final dataset for analysis comprised 5001 images. **Table 4.1** shows the distribution of images across action categories.

Table 4.1. Frequency of the action categories identified in the dataset used for machine learning.

Action Category Code	Description	Absolute Frequency (n)	Relative Frequency (n/N×100)
1	No corrective action is needed	80	1.6
2	Corrective actions are needed in the near future	1849	36.9
3	Corrective actions are needed as soon as possible	125	2.5
4	Corrective actions are required immediately	2951	59.0
Total (N)	-	5001	100

Note: n represents the absolute frequency of the action category, and N represents the size of the data sample.

Orange Visual Programming software was used to develop, train, and test the machine learning models. Two image embedders, Google's Inception V3 and SqueezeNet, were used to obtain the vector representation of the images. An artificial neural network was used as the local classifier. The models were trained and tested using a range of architectures (1 to 10 layers) and neuron counts (10, 100, 1000). A separate set of 406 unseen images was prepared for testing.

4.2. RESULTS

The main results of the classification accuracy metrics are shown in **Figures 4.2–4.5**. The best-performing models for the Inception V3 image embedder achieved a maximum classification accuracy of 0.836. For SqueezeNet, the top classification accuracy was 0.820. In general, neural network architectures containing 100 and 1000 neurons performed better.

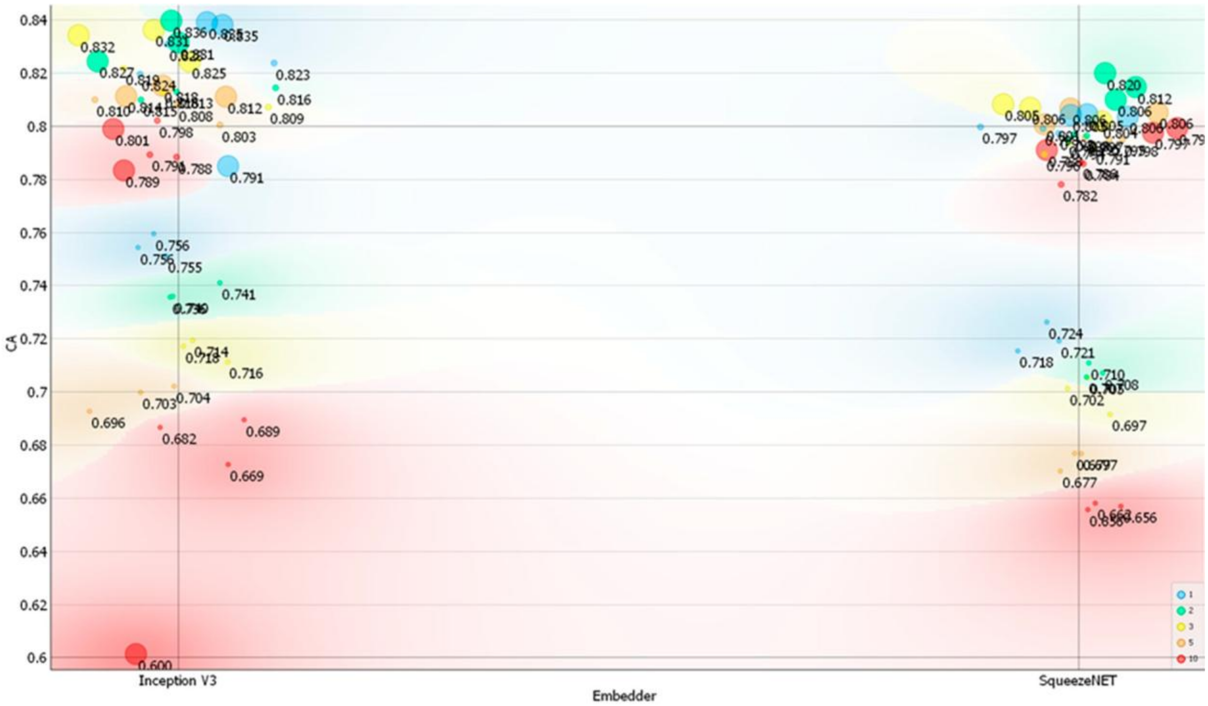


Figure 4.2. Classification accuracy (CA) on the Posture dataset depending on the embedder used, the number of layers, and the number of neurons per layer. Legend: size of the data points represents the number of neurons used (small—10 neurons, medium—100 neurons, and large—1000 neurons) and color represents the number of layers used (red—ten, orange—five, yellow—three, greentwo, and blue—one).

Figure 4.3. F1 score (F1) on the Posture dataset depending on the embedder used, the number of layers, and the number of neurons per layer. Legend: size of the data points represents the number of neurons used (small—10 neurons, medium—100 neurons, and large—1000 neurons) and color

represents the number of layers used (red—ten, orange—five, yellow—three, green—two, and blue—one).

Figure 4.4. Precision (PREC) on the Posture dataset depending on the embedder used, the number of layers, and the number of neurons per layer. Legend: size of the data points represents the number of neurons used (small—10 neurons, medium—100 neurons, and large—1000 neurons) and color represents the number of layers used (red—ten, orange—five, yellow—three, green—two, and blue—one).

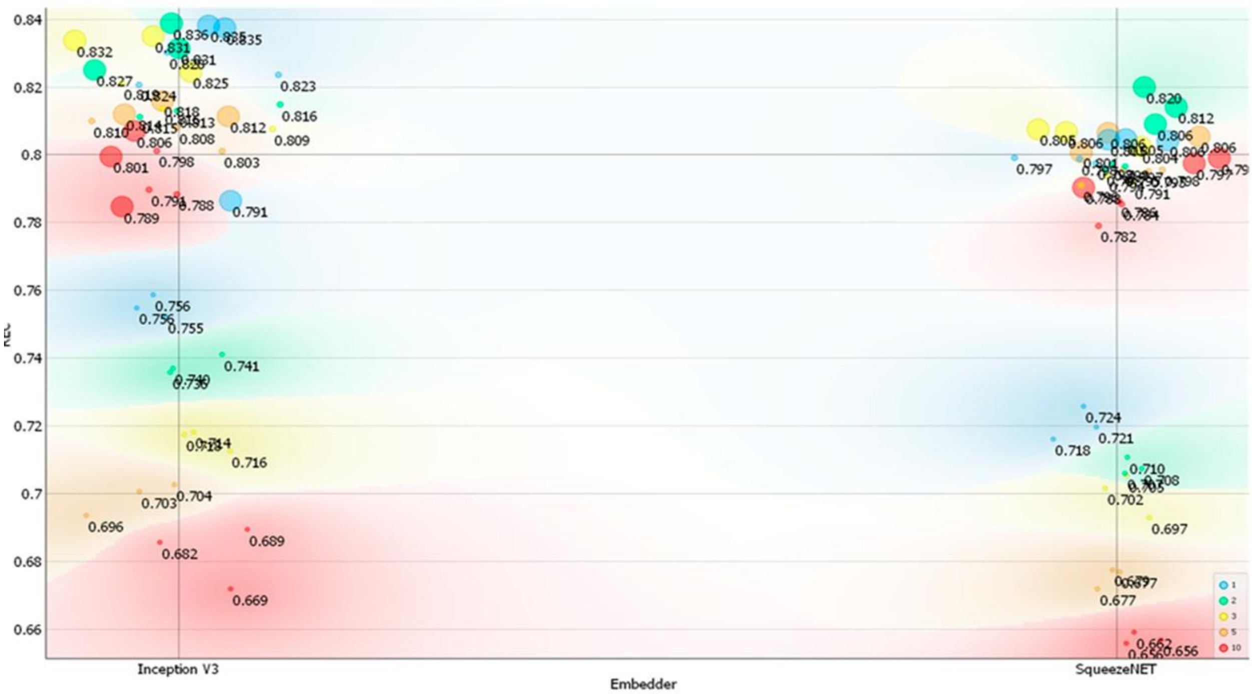


Figure 4.5. Recall (REC) on the Posture dataset depending on the embedder used, the number of layers, and the number of neurons per layer. Legend: size of the data points represents the number of neurons used (small—10 neurons, medium—100 neurons, and large—1000 neurons) and color represents the number of layers used (red—ten, orange—five, yellow—three, green—two, and blue—one).

The classification accuracies of the models on the Action dataset are presented in **Figures 4.6–4.9**. The two image embedders performed similarly, achieving top classification accuracies and recalls of 0.888–0.889 and a F1 score and precision of 0.886. The improved classification accuracy on the Action dataset may be attributed to its lower classification complexity.

Figure 4.6. Classification accuracy (CA) on the Action dataset depending on the embedder used, the number of layers, and the number of neurons per layer. Legend: size of the data points represents the number of neurons used (small—10 neurons, medium—100 neurons, and large—1000 neurons) and color represents the number of layers used (red—ten, orange—five, yellow—three, green—two, and blue—one).

Figure 4.7. F1 score (F1) on the Action dataset depending on the embedder used, the number of layers, and the number of neurons per layer. Legend: size of the data points represents the number of neurons used (small—10 neurons, medium—100 neurons, and large—1000 neurons) and color represents the number of layers used (red—ten, orange—five, yellow—three, green—two, and blue—one).

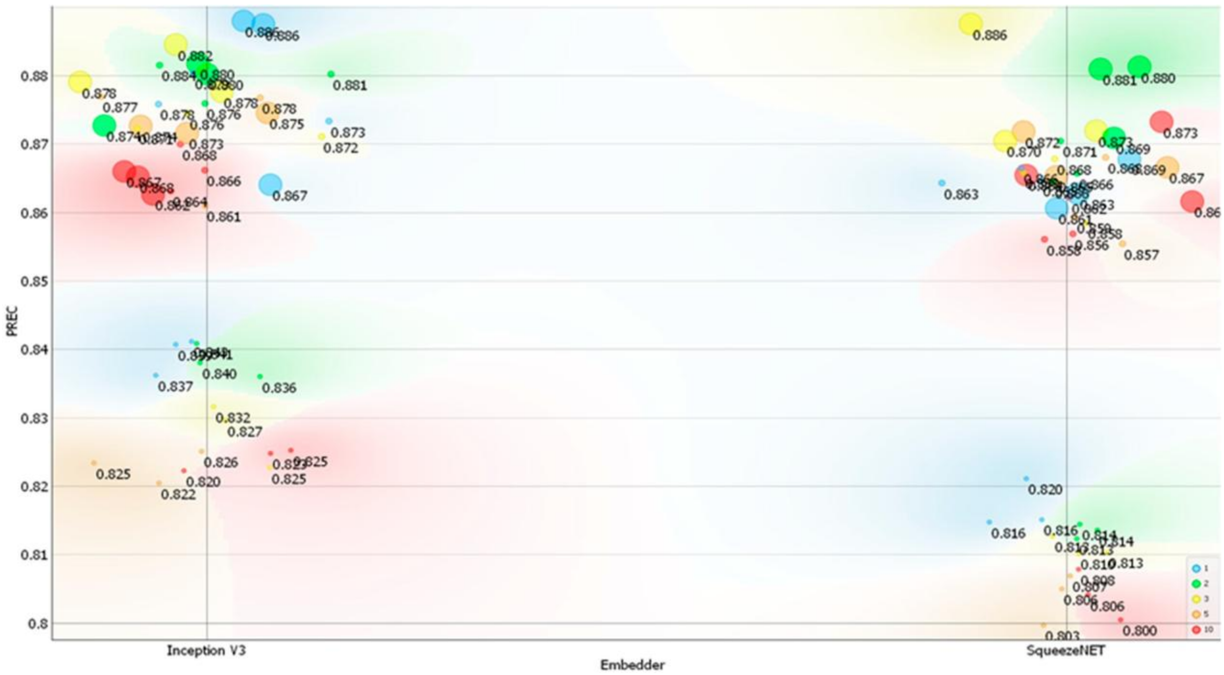


Figure 4.8. Precision (PREC) on the Action dataset depending on the embedder used, the number of layers, and the number of neurons per layer. Legend: size of the data points represents the number of neurons used (small—10 neurons, medium—100 neurons, and large—1000 neurons) and color represents the number of layers used (red—ten, orange—five, yellow—three, green—two, and blue—one).

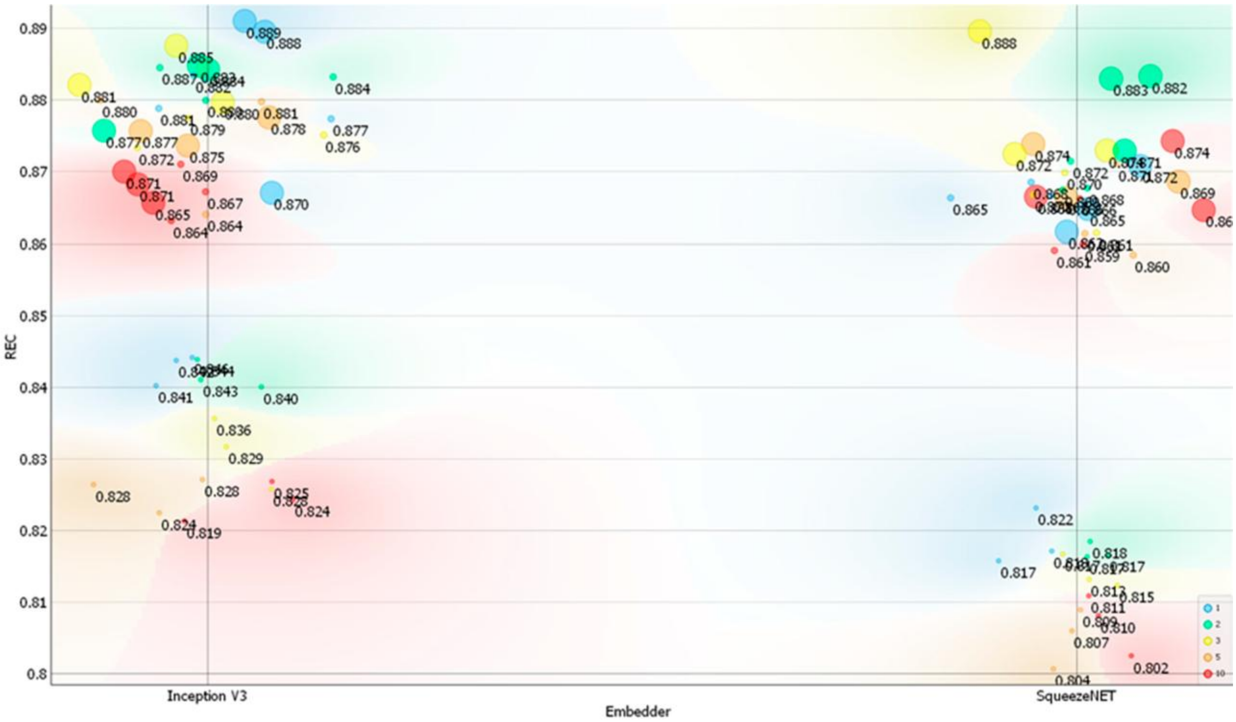


Figure 4.9. Recall (REC) on the Action dataset depending on the embedder used, the number of layers, and the number of neurons per layer. Legend: size of the data points represents the number of neurons used (small—10 neurons, medium—100 neurons, and large—1000 neurons) and color represents the number of layers used (red—ten, orange—five, yellow—three, green—two, and blue—one).

Table 4.2 presents the results of applying the trained models to unseen data. The classification performance decreased, with predicted classification accuracies ranging from 49% to 52% for the posture data and from 51% to 60% for the action category data. This suggests that the training and validation datasets may not have been fully representative of the real-world data distribution.

Table 4.2. Performance of the models on unseen data.

Model	Description & main parameters	Posture or Action category	Number of instances	Correct predictions	Classification accuracy
1	Inception V3, Postural data, 2 hidden layers, 1000 neurons each, and α = 0.0001	1131	3	0	0.0
		2141	7	2	28.6
		2171	76	4	5.3
		2271	11	0	0.0
		3121	4	0	0.0
		3141	14	0	0.0
		3171	8	0	0.0
		4131	20	0	0.0
		4141	203	174	85.7
		4151	21	0	0.0
		4171	30	15	50.0
		Overall	397	195	49.1

Model	Description & main parameters	Posture or Action category	Number of instances	Correct predictions	Classification accuracy
2	SqueezeNet, Postural data, 2 hidden layers, 1000 neurons each, and α = 0.001	1131	3	0	0.0
		2141	7	0	0.0
		2171	76	19	25.0
		2271	11	0	0.0
		3121	4	0	0.0
		3141	14	0	0.0
		3171	8	0	0.0
		4131	20	0	0.0
		4141	203	161	79.3
		4151	21	6	28.6
		4171	30	21	30.0
		Overall	397	207	52.1
3	Inception V3, Action data, 1 hidden layer, 1000 neurons each, and α = 0.001	1	24	0	0.0
		2	137	44	32.1
		3	21	0	0.0
		4	224	161	71.9
		Overall	406	205	50.5
4	SqueezeNet, Action data, 3 hidden layers, 1000 neurons each, and α = 0.001	1	24	0	0.0
		2	137	102	74.5
		3	21	2	1.0
		4	224	139	62.1
		Overall	406	243	59.9

4.3. DISCUSSION

This study evaluated the effectiveness of image embedding and transfer learning in facilitating precise postural classification. The performance of the models on unseen data showed a classification accuracy range of 49% to 52% for posture data and 51% to 60% for action category data. Previous studies have highlighted the strengths of deep learning models. A comparison of these studies with the current study reveals lower performance outcomes, highlighting areas for potential improvement. The findings emphasize the importance of dataset size and diversity in enhancing machine learning model efficiency.

4.4. CONCLUSIONS

This study proposes a novel method to automate the process of postural classification in motor-manual work via the OWAS while saving resources. The findings prove that i) classifying complex problems such as those of postural assessment can be performed with remarkable accuracy (84%–89%), ii) it is possible to reconfigure deep learning networks with less effort, and iii) the learned image representations may be less effective on unseen data (50%–60%). The proposed method can potentially reduce the cost of ergonomic assessments.

CHAPTER 5. HUMAN AND MACHINE RELIABILITY IN POSTURAL ASSESSMENT OF FOREST OPERATIONS BY OWAS METHOD: LEVEL OF AGREEMENT AND TIME RESOURCES

5.1. MATERIALS AND METHODS

The ResNet-50 model (He et al., 2016), a deep convolutional neural network, was utilized as a reference for this study due to its proven effectiveness in image classification tasks. Its selection was based on prior experimental results (Forkuo and Borz, 2024) that demonstrated its superior classification accuracy and favorable balance with computational efficiency compared to other models like GoogLeNet, MobileNet-v2, and ShuffleNet.

A separate dataset of 100 images was compiled to accurately reflect the diverse postures and movements of forest workers. Three expert human raters (R1, R2, and R3) were selected to evaluate these images using the OWAS method (**Table 5.1**). To ensure reliability and mitigate recall bias, the rating process was conducted in two separate replications (r1 and r2), with a one-month interval between them.

Table 5.1. Description of the OWAS codes and categories used in the study

Feature	Abbreviation in the study	Number of categories according to OWAS	Description
Back	B	4	Describes the posture of the back starting from a neutral straight posture and ending with the back being bent and twisted
Arms	A	3	Describes the posture of the arms starting from a neutral posture with both arms below shouder level and ending with both arms being at or above the shoulder level
Legs	L	7	Describes the posture of the legs by seven categories starting from a neutral sitting posture and ending with legs being engaged in walking or moving
Force exertion	F	3	Describes the level of force exertion starting with handling loads or exerting forces less than 10 kg and ending with handling loads or exerting forces over 20 kg
Action category	AC	4	Indicates the level of postural risk by the urgency of the ergonomic interventions required, starting from no intervention required and ending with intervention required immediately

Reliability was assessed using several datasets as detailed in **Table 5.2**. Intra-rater reliability was determined by comparing a rater's two replications (e.g., R1r1 vs. R1r2). Inter-rater reliability involved pairwise and overall comparisons between different raters for the same replication. The deep learning model was used to generate a reference dataset (RM), considered the ground truth, against which human rater reliability was also assessed.

Table 5.2. Description of the datasets used in the assessment

Rater No.	Replication No.	Abbreviation of the dataset	Description of the dataset
R1	r1	R1r1	Ratings of the first rater in the first replication
R1	r2	R1r2	Ratings of the first rater in the second replication
R2	r1	R2r1	Ratings of the second rater in the first replication
R2	r2	R2r2	Ratings of the second rater in the second replication
R3	r1	R3r1	Ratings of the third rater in the first replication
R3	r2	R3r2	Ratings of the third rater in the second replication
RM	-	RM	Rating of the deep learning model

Cohen's kappa (Cohen, 1960) and Fleiss' kappa (Fleiss, 1971) were the primary metrics for assessing reliability. Time efficiency was evaluated by recording the time taken by each human rater per image and comparing it to the programmatic assessment time of the DL model. Statistical analyses, including multi-dimensional scaling (MDS) for visual agreement and non-parametric tests for time consumption data, were performed using Orange Visual Programming software and Python.

5.2. RESULTS AND DISCUSSION

The MDS analysis (**Figure 5.1**) revealed a degree of agreement among human raters, though with considerable dispersion. When the DL model's ratings were included (**Figure 5.2**), a higher level of disagreement was observed, with the model's data points often positioned distinctly from the human raters', indicating a different classification pattern.

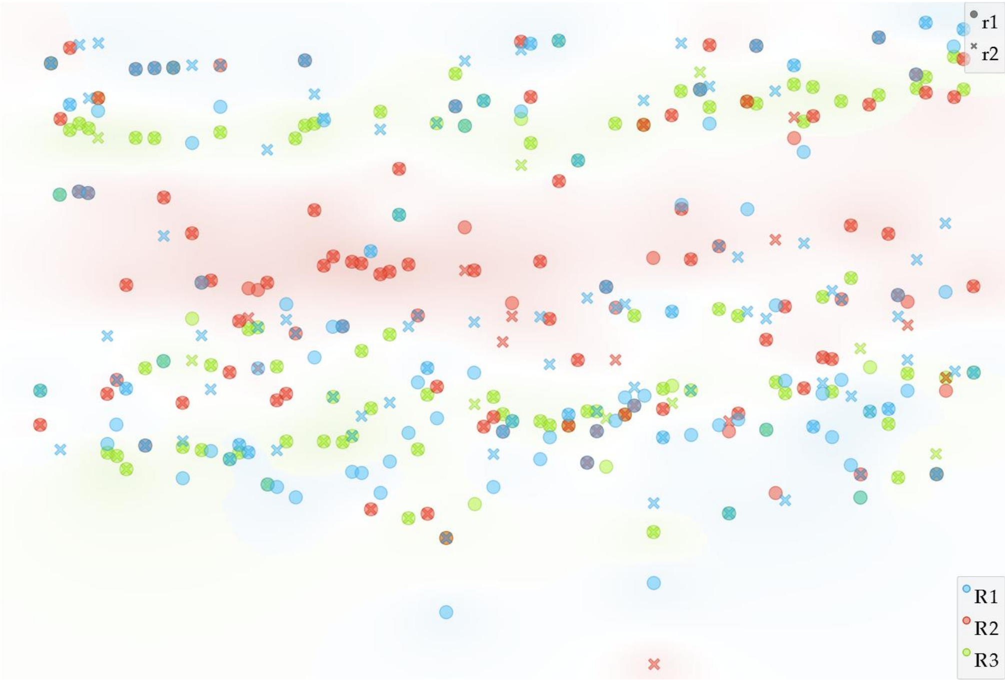


Figure 5.1. Results of multi-dimensional scaling concerning human rater agreement. Legend: R1 – rater 1, R2 – rater 2, R3 – rater 3, r1 – data from the first replication, r2 – data from the second replication.

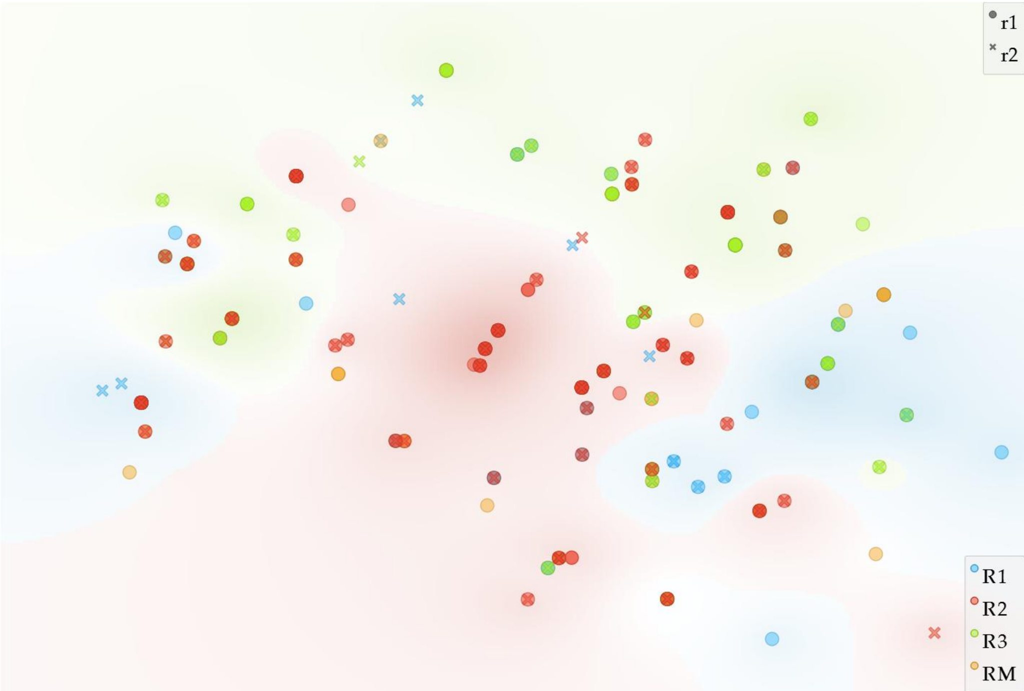


Figure 5.2. Results of multi-dimensional scaling concerning human raters and model agreement. Legend: R1 – rater 1, R2 – rater 2, R3 – rater 3, RM – rating of the deep learning model, r1 – data from the first replication, r2 – data from the second replication. Note: for RM a single rating was used.

Intra-rater agreement (**Table 5.3**) was high, with Cohen's kappa values ranging from 0.48 to 1.00, indicating moderate to almost perfect consistency for individual raters over time.

Table 5.3. Results of intra-rater reliability for the three human raters

Compared datasets		# Ratings	*Po*	*Pe*	*k*	%Agreement	Interpretation of kappa
BR1r1	BR1r2	100	0.69	0.29	0.56	69	Moderate agreement
AR1r1	AR1r2	100	0.93	0.71	0.76	93	Substantial agreement
LR1r1	LR1r2	100	0.68	0.26	0.57	68	Moderate agreement
FR1r1	FR1r2	100	0.90	0.62	0.74	90	Substantial agreement
ACR1r1	ACR1r2	100	0.61	0.25	0.48	61	Moderate agreement
BR2r1	BR2r2	100	0.97	0.33	0.96	97	Almost perfect agreement
AR2r1	AR2r2	100	1.00	0.73	1.00	100	Almost perfect agreement
LR2r1	LR2r2	97	0.99	0.25	0.99	99	Almost perfect agreement
FR2r1	FR2r2	100	0.95	0.51	0.90	95	Almost perfect agreement
ACR2r1	ACR2r2	97	0.95	0.26	0.93	95	Almost perfect agreement
BR3r1	BR3r2	100	0.96	0.39	0.93	96	Almost perfect agreement
AR3r1	AR3r2	100	0.98	0.84	0.88	98	Almost perfect agreement
LR3r1	LR3r2	100	0.99	0.32	0.99	99	Almost perfect agreement
FR3r1	FR3r2	100	0.98	0.48	0.96	98	Almost perfect agreement
ACR3r1	ACR3r2	100	0.96	0.32	0.94	96	Almost perfect agreement

Note: ***Po*** denotes observed agreement; ***Pe*** denotes expected agreement by chance; ***k*** denotes Cohen's kappa statistic, B denotes the posture of the back, A denotes the posture of the arms, L denotes the posture of the legs, F denotes the level of force exertion, AC denotes the action category. The full abbreviations were composed by using the type of feature under assessment (B, A, L or AC, **Table 5.2**) and the datasets presented in **Table 5.1.**

In contrast, pairwise inter-rater reliability among human experts (**Table 5.4**) was significantly lower and more variable, with kappa values from 0.02 (slight) to 0.64 (substantial), highlighting the inherent subjectivity of manual assessment.

Table 5.4. Results of inter-rater reliability among the three human raters

Compared datasets		# Ratings	*Po*	*Pe*	*k*	%Agreement	Interpretation of kappa
BR1r1	BR2r1	100	0.46	0.24	0.29	46	Fair agreement
BR1r1	BR3r1	100	0.62	0.36	0.41	62	Moderate agreement
BR2r1	BR3r1	100	0.34	0.29	0.07	34	Slight agreement
AR1r1	AR2r1	100	0.91	0.70	0.70	91	Substantial agreement
AR1r1	AR3r1	100	0.89	0.75	0.56	89	Moderate agreement
AR2r1	AR3r1	100	0.88	0.78	0.46	88	Moderate agreement
LR1r1	LR2r1	97	0.57	0.21	0.45	57	Moderate agreement
LR1r1	LR3r1	100	0.64	0.26	0.52	64	Moderate agreement
LR2r1	LR3r1	100	0.60	0.25	0.46	60	Moderate agreement
FR1r1	FR2r1	100	0.74	0.52	0.46	74	Moderate agreement
FR1r1	FR3r1	100	0.70	0.53	0.37	70	Fair agreement
FR2r1	FR3r1	100	0.72	0.48	0.46	72	Moderate agreement
ACR1r1	ACR2r1	100	0.54	0.24	0.40	54	Fair agreement

Compared datasets		# Ratings	*Po*	*Pe*	*k*	%Agreement	Interpretation of kappa
ACR1r1	ACR3r1	100	0.52	0.27	0.34	52	Fair agreement
ACR2r1	ACR3r1	97	0.40	0.23	0.22	40	Fair agreement
BR1r2	BR2r2	100	0.58	0.28	0.41	58	Moderate agreement
BR1r2	BR3r2	100	0.41	0.30	0.15	41	Slight agreement
BR2r2	BR3r2	100	0.32	0.30	0.02	32	Slight agreement
AR1r2	AR2r2	100	0.90	0.73	0.62	90	Substantial agreement
AR1r2	AR3r2	100	0.92	0.79	0.63	92	Substantial agreement
AR2r2	AR3r2	100	0.86	0.78	0.37	86	Fair agreement
LR1r2	LR2r2	100	0.56	0.24	0.42	56	Moderate agreement
LR1r2	LR3r2	100	0.75	0.31	0.64	75	Substantial agreement
LR2r2	LR3r2	100	0.58	0.25	0.44	58	Moderate agreement
FR1r2	FR2r2	100	0.79	0.55	0.53	79	Moderate agreement
FR1r2	FR3r2	100	0.73	0.55	0.40	73	Fair agreement
FR2r2	FR3r2	100	0.75	0.48	0.52	75	Moderate agreement
ACR1r2	ACR2r2	100	0.56	0.25	0.42	56	Moderate agreement
ACR1r2	ACR3r2	100	0.41	0.25	0.22	41	Fair agreement
ACR2r2	ACR3r2	100	0.40	0.23	0.22	40	Fair agreement

Note: ***Po*** denotes observed agreement; ***Pe*** denotes expected agreement by chance; ***k*** denotes Cohen's kappa statistic, B denotes the posture of the back, A denotes the posture of the arms, L denotes the posture of the legs, F denotes the level of force exertion, AC denotes the action category. The full abbreviations were composed by using the type of feature under assessment (B, A, L or AC, **Table 5.2**) and the datasets presented in **Table 5.1**.

Agreement between human raters and the DL model as the ground truth (**Table 5.5**) was generally poor to fair, with kappa values ranging from -0.03 to 0.34. This suggests a systematic difference between individual human interpretation and the data-driven patterns learned by the model.

Table 5.5. Results of pair-based agreement between the human raters and the deep learning model

Ratings Under Comparison		# Ratings	*Po*	*Pe*	*k*	%Agreement	Interpretation of kappa
BR1r1	BRM	100	0.43	0.34	0.13	43	Slight agreement
BR1r2	BRM	100	0.34	0.30	0.06	34	Slight agreement
BR2r1	BRM	100	0.32	0.30	0.03	32	Slight agreement
BR2r2	BRM	100	0.30	0.30	0.00	30	Poor agreement
BR3r1	BRM	100	0.57	0.37	0.32	57	Fair agreement
BR3r2	BRM	100	0.57	0.38	0.31	57	Fair agreement
AR1r1	ARM	100	0.75	0.76	-0.03	75	Poor agreement
AR1r2	ARM	100	0.79	0.79	-0.02	79	Poor agreement
AR2r1	ARM	100	0.78	0.78	-0.02	78	Poor agreement
AR2r2	ARM	100	0.78	0.78	-0.02	78	Poor agreement

AR3r1	ARM	100	0.85	0.84	0.04	85	Slight agreement
AR3r2	ARM	100	0.85	0.84	0.04	85	Slight agreement
LR1r1	LRM	100	0.38	0.24	0.18	38	Slight agreement
LR1r2	LRM	100	0.46	0.28	0.25	46	Fair agreement
LR2r1	LRM	97	0.44	0.25	0.26	44	Fair agreement
LR2r2	LRM	100	0.43	0.24	0.25	43	Fair agreement
LR3r1	LRM	100	0.50	0.29	0.29	50	Fair agreement
LR3r2	LRM	100	0.49	0.30	0.28	49	Fair agreement
FR1r1	FRM	100	0.60	0.47	0.24	60	Fair agreement
FR1r2	FRM	100	0.59	0.49	0.20	59	Slight agreement
FR2R1	FRM	100	0.53	0.44	0.16	53	Slight agreement
FR2r2	FRM	100	0.56	0.44	0.21	56	Fair agreement
FR3r1	FRM	100	0.61	0.44	0.31	61	Fair agreement
FR3r2	FRM	100	0.63	0.44	0.34	63	Fair agreement
ACR1r1	ACRM	100	0.32	0.26	0.08	32	Slight agreement
ACR1r2	ACRM	100	0.38	0.25	0.18	38	Slight agreement
ACR2r1	ACRM	97	0.35	0.24	0.15	35	Slight agreement
ACR2r2	ACRM	100	0.36	0.24	0.16	36	Slight agreement
ACR3r1	ACRM	100	0.50	0.29	0.29	50	Fair agreement
ACR3r2	ACRM	100	0.51	0.30	0.30	51	Fair agreement

Note: ***Po*** denotes observed agreement; ***Pe*** denotes expected agreement by chance; ***k*** denotes Cohen's kappa statistic, B denotes the posture of the back, A denotes the posture of the arms, L denotes the posture of the legs, F denotes the level of force exertion, AC denotes the action category. The full abbreviations were composed by using the type of feature under assessment (B, A, L or AC, **Table 5.2**) and the datasets presented in **Table 5.1.**

The overall agreement, when considering all human raters against the DL model using Fleiss' kappa (**Table 5.6**), improved to fair to moderate levels (kappa = 0.28–0.49). This indicates that while individual raters may diverge, their collective assessment is more aligned with the model's predictions.

Table 5.6. Results of overall agreement among the three human raters and the ResNet-50 model

Ratings Under Comparison				# Ratings	*Po*	*Pe*	*k*	%Agreement	Interpretation of kappa
BR1R1	BR2R1	BR3R1	BRM	100	0.53	0.34	0.28	53	Fair agreement
AR1R1	AR2R1	AR3R1	ARM	100	0.88	0.77	0.49	88	Moderate agreement
LR1R1	LR2R1	LR3R1	LRM	97	0.52	0.23	0.37	52	Fair agreement
FR1R1	FR2R1	FR3R1	FRM	100	0.66	0.47	0.37	66	Fair agreement
ACR1R1	ACR2R1	ACR2R1	ACRM	97	0.52	0.26	0.35	52	Fair agreement
BR1R2	BR2R2	BR3R2	BRM	100	0.49	0.31	0.26	49	Fair agreement
AR1R2	AR2R2	AR3R2	ARM	100	0.89	0.79	0.47	89	Moderate agreement
LR1R2	LR2R2	LR3R2	LRM	100	0.53	0.25	0.38	53	Fair agreement

FR1R2	FR2R2	FR3R2	FRM	100	0.68	0.47	0.37	68	Fair agreement
ACR1R2	ACR2R2	ACR2R2	ACRM	100	0.51	0.27	0.33	51	Fair agreement

Note: ***Po*** denotes observed agreement; ***Pe*** denotes expected agreement by chance; ***k*** denotes Fleiss's kappa statistic, B denotes the posture of the back, A denotes the posture of the arms, L denotes the posture of the legs, F denotes the level of force exertion, AC denotes the action category. The full abbreviations were composed by using the type of feature under assessment (B, A, L or AC, **Table 5.2**) and the datasets presented in **Table 5.1.**

In terms of time efficiency (**Table 5.7**), the DL model was vastly superior, performing assessments 19 to 53 times faster than human raters on average. Significant variability in assessment time was also observed among the human raters, further underscoring the efficiency and consistency benefits of the automated approach.

Table 5.7. Results of comparison tests for time consumption data

Variables under comparison	Median values (s)	Results of normality test[1]	Results of comparison test[2]
TR1r1-TR1r2	30.0 – 24.0	No, p < 0.001-No, p < 0.001	Yes, p < 0.001
TR2r1-TR2r2	52.5 – 44.0	No, p < 0.001-No, p < 0.001	Yes, p < 0.001
TR3r1-TR3r2	19.0 – 20.0	No, p < 0.001-No, p < 0.001	No, p = 0.608
TR1r1-TR2r1	30.0 – 52.5	No, p < 0.001-No, p < 0.001	Yes, p < 0.001
TR1r1-TR3r1	30.0 – 19.0	No, p < 0.001-No, p < 0.001	Yes, p < 0.001
TR2r1-TR3r1	52.5 – 19.0	No, p < 0.001-No, p < 0.001	Yes, p < 0.001
TR1r2-TR2r2	24.0 – 44.0	No, p < 0.001-No, p < 0.001	Yes, p < 0.001
TR1r2-TR3r2	30.0 – 20.0	No, p < 0.001-No, p < 0.001	Yes, p = 0.003
TR2r2-TR3r2	44.0 – 20.0	No, p < 0.001-No, p < 0.001	Yes, p < 0.001

Note: 1 – According to Shapiro-Wilk test; 2 – significant differences according to Mann-Whitney two-tailed nonparametric test, T stands for the time consumption dataset

5.3. CONCLUSIONS

This study shows that DL models present significant advantages for conducting OWAS-based postural assessments, offering remarkable speed enhancements while achieving comparable levels of reliability to traditional human-rater methods. The findings showed that while human raters exhibited moderate to almost perfect intra-rater reliability, their inter-rater agreement was considerably lower. The DL model serves not only as a highly resource-efficient alternative but also as a stable reference point for evaluating OWAS assessments.

CHAPTER 6. CONCLUSIONS. ORIGINAL CONTRIBUTIONS. DISSEMINATION OF RESULTS

6.1. Conclusions

This PhD thesis underscores the importance of postural assessment in ergonomics, particularly within forest operations, where identifying high-risk postures is essential for developing effective interventions. The research demonstrates that both conventional and innovative methods, including machine learning (ML) and computer vision (CV), significantly enhance the accuracy and efficiency of postural classification. Automating the OWAS method using deep learning-based CV techniques yields impressive classification performance, with the ResNet-50 model emerging as the most effective option in real-world applications. Furthermore, the integration of comprehensive skeletal and contextual information is shown to optimize model reliability and effectiveness. The introduction of a novel approach leveraging image embedding and transfer learning allows for accurate postural classification with reduced reliance on computer programming expertise. Additionally, the findings reveal that deep learning models offer substantial advantages over traditional human assessments, providing resource-efficient solutions that decrease assessment time and enhance consistency. Despite challenges such as variability in human ratings and limitations with unseen data, this research highlights the transformative potential of ML and CV in advancing postural assessment methods. Future research should prioritize refining these models, enhancing algorithm training protocols, expanding annotated datasets, and optimizing data preprocessing techniques to further improve accuracy and applicability across diverse environments. Through these efforts, the ultimate goal of enhancing worker health, safety, and operational efficiency in forestry and beyond can be achieved.

6.2. Original contributions

This PhD thesis has made significant contributions to the field of ergonomic assessment in forest operations through the development and evaluation of novel methods and extensive datasets. Key original contributions include:

Creation of a Comprehensive and Novel Dataset: A dataset of 23,000 annotated images for OWAS classification in forestry was created. Its innovative integration of field-collected data with AI-generated images ensures complete coverage of all 252 OWAS postural combinations, addressing the persistent challenge of data scarcity.

Development of an Automated Workflow: A novel workflow for automated OWAS classification was established, identifying optimal hyperparameters (learning rate of 0.001, batch size of 32) through systematic fine-tuning, creating a robust framework for future applications.

Enhancement of Input Data with Skeletal Keypoints: An innovative strategy was implemented that enhanced conventional input images with computer-generated body keypoints, which was proven to significantly improve classification accuracy by 2-2.5% and achieve a near-perfect accuracy of 99.8%.

First Evaluation of DL for OWAS in Forestry: This thesis represents the first development and evaluation of deep learning-based computer vision techniques for automating OWAS postural classification in forestry, testing four pre-trained CNN models and setting new performance standards, with the fine-tuned ResNet-50 model achieving over 96% classification accuracy.

Advancement of Automation via Image Embedding: The automation of posture assessment was advanced through the novel application of image embedding and transfer learning, achieving accuracy rates between 84% and 89% without the need for specialized sensors or extensive programming expertise.

Pioneering Analysis of Human vs. Machine Reliability: A pioneering analysis of human and machine-based postural assessments was conducted, quantifying intra- and inter-rater reliability and establishing a comparison with the ResNet-50 model as a ground truth, highlighting the superior speed and consistency of the DL model.

6.3. Dissemination of results

6.3.1. Scientific publications based on this PhD thesis

A. Papers published in BDI journals

Forkuo, **G.O.**, 2023. A systematic survey of conventional and new postural assessment methods. *Revista Pădurilor, 138*(3), 34p. Available online at: http://revistapadurilor.com/wp-content/uploads/2024/01/RP_138-3-2023.-BT.pdf.

B. Papers published in journals indexed by Clarivate Analytics (former ISI Web of Science)

Forkuo, G.O., Borz, S.A., Kaakkurivaara, T., Kaakkurivaara, N., 2025. Postural classification by image embedding and transfer learning: An example of using the OWAS method in motor-manual work to automate the process and save resources. *Forests*, *16*(3), 492. https://doi.org/10.3390/f16030492

Forkuo, G.O., Marcu, M.V., Kaakkurivaara, N., Kaakkurivaara, T., Borz, S.A., 2025. Human and machine reliability in postural assessment of forest operations by OWAS method: Level of agreement and time resources. *Forests*, *16*(5), 759. https://doi.org/10.3390/f16050759.

6.3.1.1. Papers presented at national scientific conferences

Forkuo, **G.O.**, 2023. A systematic survey of conventional and new postural assessment methods. In: *Book of Abstracts, Proceedings of the 6th Edition of the Integrated Management of Environmental Resources Conference Suceava – Romania, 23-24 November 2023*. Available online at: https://silvic.usv.ro/imer2023/proceedings_imer_2023.pdf (accessed 4 March 2025)

Forkuo, G.O., Borz, S.A., 2024. Evaluation and development of smart ovako working posture analysis system (OWAS) solutions for postural classification in forest operations using deep learning-based computer vision. *Graduates in Front of Companies (AFCO) 2024.* Available online at: https://afco.unitbv.ro/2024/images/Documente/AFCO_2024_Lucrari_inscrise_.pdf (accessed 5 March 2025).

Forkuo, G.O., Borz, S.A., 2024. Evaluation and development of smart ovako working posture analysis system (owas) solutions for postural classification in forest operations using deep learning-based computer vision. *DoCo2024*. Available online at: https://www.unitbv.ro/documente/cercetare/Detailed_Programme.pdf (accessed 2 March 2025).

6.3.1.2. Papers presented at international scientific conferences

Forkuo, G.O., 2023: A systematic survey of conventional and new postural assessment methods. In: *Book of Abstracts, Proceedings of the 6th Edition of the Integrated Management of Environmental Resources Conference Suceava - Romania. Session: Forest Ecosystems and Climate*. Available online at: https://ibn.idsi.md/sites/default/files/imag_file/Book_of_abstracts_IMER_23_Suceava.pdf (accessed 1 April 2025).

Forkuo, G.O., Borz, S.A., 2024. Development and evaluation of automated postural classification models in forest operations using deep learning-based computer vision. In: *Book of Abstracts, Proceedings of the 11th International Symposium on Forest and Sustainable Development, FSD 2024*. Available online at: https://silvic.unitbv.ro/images/conferinte/fsd2024/Book_of_abstracts_FSD_2024c.pdf. (accessed 1 April 2025).

6.3.2. Scientific publications based on the results produced by participation in research teams external to the PhD thesis scope

A. Papers published on preprint servers

Forkuo, G.O., Borz, S.A., Proto, A.R.,2024. Accuracy of low-cost mobile lidar technology in estimating the severity and extent of soil disturbance in forest operations. SSRN Preprint SSRN-4685980. http://dx.doi.org/10.2139/ssrn.4685980. Submitted to Croatian Journal of Forest Engineering. Status: Accepted.

Forkuo, G.O., Borz, S.A., 2024. Intra-and inter-rater reliability in log volume estimation based on lidar data and shape reconstruction algorithms: A case study on poplar logs. SSRN Preprint SSRN-4948247. http://dx.doi.org/10.2139/ssrn.4948247. Submitted to Frontiers in Remote Sensing. Status: Under Review.

B. Papers published in journals indexed by Clarivate Analytics (former ISI Web of Science)

Borz, S.A., **Forkuo, G.O.**, Oprea-Sorescu, O., Proto, A.R., 2022. Development of a robust machine learning model to monitor the operational performance of fixed-post multi-blade vertical sawing machines. Forests, 13(7), 1115. https://doi.org/10.3390/f13071115

Forkuo, G.O., Borz, S.A., 2023. Accuracy and inter-cloud precision of low-cost mobile LiDAR technology in estimating soil disturbance in forest operations. Front. For. Glob. Change, 6, 1224575. https://doi.org/10.3389/ffgc.2023.1224575

Forkuo, G.O., Marcu, M.V., Iordache, E., Borz, S.A., 2024. Timber extraction by farm tractors in low-removal-intensity continuous cover forestry: a simulation of operational performance and fuel consumption. Forests, 15(8), 1422. https://doi.org/10.3390/f15081422

Borz, S.A., Morocho Toaza, J.M**., Forkuo, G.O**., Marcu, M.V., 2022. Potential of measure app in estimating log biometrics: A comparison with conventional log measurement. Forests, 13(7), 1028. https://doi.org/10.3390/f13071028

Presecan, M.F., **Forkuo, G.O.,** Borz, S.A., 2024. Soil compaction induced by three timber extraction options: A controlled experiment on penetration resistance on silty-loamy soils. Appl. Sci., 14(12), 5117. https://doi.org/10.3390/app14125117

6.3.2.1. Papers presented at national scientific conferences

Forkuo, G.O., Borz, S.A., 2023. Accuracy and inter-cloud precision of low-cost mobile LiDAR technology in estimating soil disturbance in forest operations. In: *Book of Abstracts, Proceedings of the VIIIth edition of the National Session of Student Scientific Communications on Forestry and Environmental Protection, May 26-27, 2023, Suceava - Vatra Dornei*. Available online at: http://silvic.usv.ro/avizier/2023_mai_simoz_national.pdf (accessed 4 March 2025).

6.3.2.2. Papers presented at international scientific conferences

Forkuo, G.O., Elias, M., Borz, S.A., 2023. Forwarding by farm tractors in low removal intensity, continuous cover forestry: A simulation of operational performance and fuel consumption. In: *Book of Abstracts, Proceedings of the 55th International Symposium on Forest Mechanization (FORMEC) and the 7th Forest Engineering Conference (FEC), Florence, Italy, September 20-22, 2023.* Available online at: https://www.formec.org/images/proceedings/2023/Book_of_abstract_FORMEC_FEC_23.pdf (accessed 4 March 2025).

Forkuo, G.O., Borz, S.A., 2024. Intra-and inter-rater reliability in log volume estimation based on lidar data and shape reconstruction algorithms: A case study on poplar logs. In: *Book of Abstracts, Proceedings of the 56th International Symposium on Forest Mechanization (FORMEC), Gdańsk, Poland, June 11 - 14, 2024.* Available online at: https://www.formec.org/images/proceedings/2024/Proceedings_FORMEC_2024_Gdansk_02.pdf (accessed 4 March 2025).

Selective Bibliography (Full list is included in the PhD thesis)

1. Alpaydin, E., 2020. Introduction to Machine Learning, 4th ed. MIT Press, Cambridge, MA.
2. Bačić, M., Landekić, M., Pandur, Z., Šušnjar, M., Šporčić, M., Nevečerel, H., Lepoglavec, K., 2024. Forestry ergonomics publications in the last decade: A review. Forests, 15(4), 616.
3. Bevan, S., 2015. Economic impact of musculoskeletal disorders (MSDs) on work in Europe. Best Pract. Res. Clin. Rheumatol. 29, 356-373.
4. Calvo, A., 2009. Musculoskeletal disorders (MSD) risks in forestry: A case study to propose an analysis method. Agric. Eng. Int. CIGR J., 11, 1-9.
5. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y., 2017. Realtime multi-person 2d pose estimation using part affinity fields, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7291-7299.
6. Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., Sheikh, Y., 2019. OpenPose: realtime multi-person 2d pose estimation using part affinity fields, arXiv preprint arXiv:1812.08008.
7. Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas., 20(1), 37-46.
8. Da Costa, B.R., Vieira, E.R., 2010. Risk factors for work-related musculoskeletal disorders: a systematic review of recent longitudinal studies. Am. J. Ind. Med., 53(3), 285-323.
9. David, G.C., 2005. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. Occup. Med., 55(3), 190-199.
10. Debnath B., O'Brien M., Yamaguchi M., Behera A., 2022. A review of computer vision-based approaches for physical rehabilitation and assessment. Multimed. Syst., 28(1), 209-239.
11. Ding, Z., Li, W., Yang, J., Ogunbona, P., Qin, L., 2024. An attention-based CNN for automatic whole-body postural assessment. Expert Syst. Appl., 238, 122391.
12. Enez, K., Nalbantoğlu, S.S., 2019. Comparison of ergonomic risk assessment outputs from OWAS and REBA in forestry timber harvesting. Int. J. Ind. Ergon., 70, 51-57.
13. Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. Psychol. Bull., 76(5), 378-382.
14. Forkuo, G.O., Borz, S.A., 2024. Development and evaluation of automated postural classification models in forest operations using deep learning-based computer vision. Preprint SSRN 4875562.
15. Gomez-Galan, M., Perez-Alonso, J., Callejón-Ferre, Á.J., López-Martínez, J., 2017. Musculoskeletal disorders: OWAS review. Ind. Health, 55(4), 314-337.
16. He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778.
17. Heinimann, H.R., 2007. Forest operations engineering and management-the ways behind and ahead of a scientific discipline. Croat. J. For. Eng., 28(1), 107-121.
18. Hignett, S., McAtamney, L., 2000. Rapid entire body assessment (REBA). Appl. Ergon., 31(2), 201-205.

19. Imbeault-Nepton T., Maitre J., Bouchard K., Gaboury S., 2022. Filtering data bins of UWB radars for activity recognition with random forest. Procedia Comput. Sci., 201, 48-55.
20. Jiang X., Hu Z., Wang S., Zhang Y., 2023. A survey on artificial intelligence in posture recognition. CMES-Comp. Model. Eng. Sci., 137(1), 35.
21. Karhu, O., Kansi, P., Kuorinka, I., 1977. Correcting working postures in industry: A practical method for analysis. Appl. Ergon., 8(4), 199-201.
22. Kolgiri S., Bansode S., Hiremath R., 2017. A literature review on work-related body part discomfort with respect of postures and movements. IAETSD J. Adv. Res. Appl. Sci., 4(5).
23. Li, G., Buckle, P., 1999. Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. Ergonomics, 42(5), 674-695.
24. Lorenzini, M., Lagomarsino, M., Fortini, L., Gholami, S., Ajoudani, A., 2023. Ergonomic human-robot collaboration in industry: a review. Front. Robot. AI, 9, 262.
25. Marchi, E., Chung, W., Visser, R., Abbas, D., Nordfjell, T., Mederski, P.S., McEwan, A., Brink, M., Laschi, A., 2018. Sustainable Forest Operations (SFO): A new paradigm in a changing world and climate. Sci. Total Environ., 634, 1385-1397.
26. McAtamney, L., Corlett, E.N., 1993. RULA: a survey method for the investigation of work-related upper limb disorders. Appl. Ergon., 24(2), 91-99.
27. Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst. Rev., 4(1), 1-9.
28. Nadeem, M., Elbasi, E., Zreikat, A.I., Sharsheer, M., 2021. A review of wearable technology for posture monitoring. In: 2021 international conference on information technology (ICIT), pp. 848-853.
29. Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int. J. Surg., 88, 105906.
30. Paini, A.D.C., Lopes, E.D.S., Fiedler, N.C., de Souza, A.P., Rodrigues, C.K., 2020. Postural analysis and repetitive motions of skidder operators in wood extraction. Sci. For., 48(125), e3095.
31. Pascual, S. A., Naqvi, S., 2008. An investigation of ergonomics analysis tools used in industry in the identification of work-related musculoskeletal disorders. Int. J. Occup. Saf. Ergon., 14, 237-245.
32. Punnett, L., Wegman, D.H., 2004. Work-related musculoskeletal disorders: The epidemiologic evidence and the debate. J. Electromyogr. Kinesiol., 14(1), 13-23.
33. Roggio, F., Trovato, B., Sortino, M., Musumeci, G., 2024. A comprehensive analysis of the machine learning pose estimation models used in human movement and posture analyses: A narrative review. Heliyon, 10(21), e39977.
34. Rummer, B., Wear, D.N., Greis, J., 2002. Forest operations technology. Southern Forest Resource Assessment. General Technical Report. SRS-53, USDA-Forest Service, Southern Research Station, Asheville, NC, pp. 341-353.

35. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis., 115, 211-252.
36. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. MobileNetV2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4510-4520. https://doi.org/10.48550/arXiv.1801.04381
37. Sibiya, Z., Ackerman, P., Ntinga, N., Ackerman, S., 2021. An ergonomics risk assessment of manual and motor-manual pruning of pinus patula at two different lifts. For. Eng. Family-Growing Forward Together, p.12.
38. Sun, K., Xiao, B., Liu, D., Wang, J., 2019. Deep high-resolution representation learning for human pose estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5693-5703.
39. Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594
40. Takala, E.P., Pehkonen, I., Forsman, M., Hansson, G.Å., Mathiassen, S.E., Neumann, W.P., Sjøgaard, G., Veiersted, K.B., Westgaard, R.H., Winkel, J., 2010. Systematic evaluation of observational methods assessing biomechanical exposures at work. Scand. J. Work Environ. Health, 3-24.
41. Yovi, E.Y., Prajawati, W., 2015. High risk posture on motor-manual short wood logging system in Acacia mangium plantation. J. Manaj. Hutan Tropika, 21(1), 11-18.
42. Zhang, X., Zhou, X., Lin, M., Sun, J., 2018. ShuffleNet: An extremely efficient convolutional neural network for mobile devices, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6848–6856.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Eng. Stelian Alexandru Borz, for his invaluable guidance, insightful feedback, and steadfast support throughout the course of this research. His mentorship was instrumental in shaping the direction and depth of this study. I am also deeply indebted to the members of my Supervisory Committee: Prof. Dr. Eng. Petrițan Ion Cătălin, Assoc. Prof. Dr. Marcu Marina Viorela, Assoc. Prof. Dr. Eng. Iordache Eugen, and Assoc. Prof. Dr. Eng. Nicuşor Boja, for their constructive feedback and thought-provoking questions that enriched this work. This research was made possible through financial support from the Transilvania University of Brașov's PhD Thesis Project Competition (2023 Edition) and was partly funded by the Romanian Ministry of Education and Research through CNCS – UEFISCDI. Finally, I extend my gratitude to the Faculty of Silviculture and Forest Engineering for providing the necessary equipment, and to the forest companies, administrators, and research participants who provided logistical support and generously contributed their time and insights to this project.