*Proceedings of the International Conference on Industrial Engineering and Operations Management*

Publisher: IEOM Society International, USA
Published: February 18, 2025

DOI: 10.46254/AN15.20250361

# A Video-Based Deep Learning Model to Automate Lifting Stage Identification for Ergonomic Risk Assessment

**Listia Anjani**
Master's Degree Student in Industrial Engineering
Universitas Gadjah Mada
Indonesia
Listia.anjani@mail.ugm.ac.id

**Kung-Jeng Wang**
Distinguished Professor, Department of Industrial Management
School of Management
National Taiwan University of Science and Technology
Taiwan
kjwang@mail.ntust.edu.tw

**Hilya Mudrika Arini**
Associate Professor, Department of Mechanical and Industrial Engineering
Universitas Gadjah Mada
Indonesia
hilya.mudrika@ugm.ac.id

## Abstract

Manual lifting tasks are a leading cause of workplace injuries, necessitating reliable methods to evaluate and mitigate associated risks. Accurately identifying lifting stages is a critical step in ergonomic risk assessment tools such as the Revised NIOSH Lifting Equation (RNLE), which requires precise determination of the lifting action's origin and destination points. This study addresses the challenge of lifting stage identification by developing a hybrid deep learning model that integrates Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. Using a dataset of 117 videos, representing single lifting actions, 75,920 labeled frames were generated and classified into four stages: "ready," "start," "processing," and "end." Data augmentation and class balancing techniques were employed to handle label imbalances. The proposed model achieved a robust 99% accuracy in 5-fold cross-validation, effectively distinguishing between lifting stages, even in asymmetric lifting scenarios. Misclassification primarily occurred between visually similar stages such as "ready" and "processing". The results highlight the model's potential to automate lifting stage identification, a key requirement for implementing ergonomic assessments like RNLE. By providing accurate and scalable predictions, this approach offers a practical solution for improving workplace safety. Future work could expand this model's applicability to diverse lifting conditions and real-time analysis.

## Keywords
Lifting Stages, Deep Learning, CNN-LSTM, Ergonomic Risk Assessment, Revised NIOSH Lifting Equation (RNLE).

# 1. Introduction

Manual lifting tasks are a major contributor to workplace injuries, particularly musculoskeletal disorders (MSDs), which can lead to significant physical, economic, and operational consequences. In industries such as manufacturing, logistics, and construction, workers frequently engage in repetitive lifting actions, exposing them to risks that necessitate proper assessment and preventive measures(Sekkay 2023). According to the U.S. Bureau of Labor Statistics, MSDs account for a substantial portion of nonfatal workplace injuries (Li et al. 2020), emphasizing the importance of reliable ergonomic risk assessment tools to ensure worker safety and reduce injury-related costs.

One of the most widely used tools for evaluating manual lifting tasks is the Revised NIOSH Lifting Equation (RNLE) (Fox et al. 2019). The RNLE provides a systematic approach to assess the potential risk of lifting actions by calculating the Recommended Weight Limit (RWL) and Lifting Index (LI) (Waters et al. 1994). However, the accurate application of the RNLE depends on correctly identifying the lifting action's origin (starting point) and destination (ending point). In this study we classify the stages as "ready," "start," "processing," and "end" to identify the origin and destination of the lifting action by differentiate it with the process in between these stages. Misidentification of these stages can result in inaccurate RNLE calculations, undermining the tool's effectiveness in mitigating risks.

Traditional methods for identifying lifting stages rely on manual observation or wearable sensors, which are time-consuming, prone to human error, and limited in scalability (Jacob et al. 2012; Hernandez et al. 2021). With advances in computer vision and deep learning, automated solutions have emerged as a promising alternative (Li et al. 2020). By leveraging video data, machine learning models can process lifting actions to accurately determine lifting stages, reducing dependency on manual intervention and enhancing consistency in ergonomic evaluations.

This study develops and evaluates a hybrid deep learning model that combines Convolutional Neural Networks (CNN) for spatial feature extraction (Roopa et al. 2024)and Long Short-Term Memory (LSTM) networks for temporal sequence learning. The model is trained on a dataset of 117 videos, covering both symmetric and asymmetric lifting actions. The main objective of this study is to develop a reliable and scalable method for identifying lifting stages, to handle a crucial requirement for accurate risk assessments using the RNLE.

## 1.1 Objectives

This research aims to develop an automated method for accurately identifying lifting stages in manual lifting tasks to support ergonomic risk assessments like the Revised NIOSH Lifting Equation (RNLE). The study focuses on designing a hybrid CNN-LSTM deep learning model capable of classifying lifting stages ("ready," "start," "processing," and "end") with high accuracy.

To ensure the model's robustness and reliability, techniques such as data augmentation and class balancing are employed to address challenges like label imbalances and visual ambiguities between stages. The model is evaluated using rigorous cross-validation to validate its performance across various scenarios, including both symmetric and asymmetric lifting actions.

By automating the identification of lifting stages, this research demonstrates how the developed model enhances the accuracy and scalability of ergonomic risk assessments. The study highlights its contributions to improving RNLE applications and lays the groundwork for integrating advanced deep learning techniques into future real-time ergonomic risk analysis.

# 2. Literature Review

Ergonomic risk assessment methods are essential for identifying workplace hazards and ensuring safety during manual tasks. These methods can be categorized into three main types: basic, observational, and direct measurement methods (Lowe et al. 2019). Basic methods use simple tools like tape measures, stopwatches, or thermometers to evaluate risks. Observational methods, such as the Rapid Upper Limb Assessment (RULA), Rapid Entire Body Assessment (REBA), and the Revised NIOSH Lifting Equation (RNLE), involve structured observations to assess task-related hazards systematically. Direct measurement methods rely on advanced tools, such as heart rate monitors or goniometers, to provide highly precise measurements.

Among these, Lowe et al. (2019) mentioned that the RNLE stands out as a widely adopted tool for assessing manual lifting tasks, particularly in the USA and UK. It calculates recommended weight limits (RWL) to reduce the risk of low back injuries, incorporating parameters such as horizontal and vertical distances, lifting angles, and load weights (Waters et al. 1994). Traditionally, identifying these parameters has been performed manually or with wearable devices, but these methods are time-intensive and prone to errors. This has created a demand for automated solutions that streamline the process while maintaining accuracy.

Machine learning has emerged as a powerful tool for automating ergonomic risk assessments, particularly in lifting stage identification. Neural networks, such as Convolutional Neural Networks (CNNs), have been widely applied to extract spatial features from data, while hybrid models like CNN-LSTM integrate temporal dynamics, making them ideal for analyzing motion sequences. For instance, Li et al. (2020) used CNNs to detect body joint positions from image data with an accuracy of 93%. Similarly, Fernández et al. (2020) applied CNNs with OpenPose coordinates to classify ergonomic risks, reporting strong performance with a Cohen's kappa statistic above 0.6. Lu et al. (2020) demonstrated the use of random forests on IMU sensor data to evaluate RNLE and ACGIH TLV parameters, achieving strong correlations with ergonomic indices.

Video-based methods have also shown significant promise in ergonomic risk assessments. MediaPipe pose landmarks have been integrated with CNNs for spatial feature extraction, as demonstrated by Jeong and Kook (2023) in their REBA-based analysis. Hybrid models like 1D CNN-LSTM, as utilized by Thomas et al. (2022), achieved 93.7% accuracy by combining spatial and temporal data from IMU sensors. Carlos et al. (2024) further advanced video-based modeling by introducing the TimeDistributed Conv2D layer, which preserves spatial and temporal relationships across video frames. This technique is particularly useful for lifting stage identification, where both spatial features and temporal sequences are critical.

This research improves the previous studies by introducing an innovative approach to ergonomic risk assessment that combines deep learning model with video analysis for lifting stage identification. The method utilizes a hybrid 2D CNN-LSTM model to analyze video frames and accurately classify lifting stages into "ready," "start," "start_A," "processing," and "end." Unlike traditional methods reliant on manual labeling or wearable sensors, this approach automates the process, reduces human intervention, and ensures consistent accuracy.

## 3. Methods

The research framework in Figure 1 outlines an approach to build a deep learning model for lifting stages identification in manual lifting tasks. The process begins with video data collection, where video recordings of manual lifting actions are gathered. These videos are then extracted into individual frames, which are manually labeled with one of the predefined lifting stages: "ready," "start," "start_A," "processing," or "end." This labeled dataset forms the foundation for training the deep learning model, ensuring the accurate representation of lifting stages.

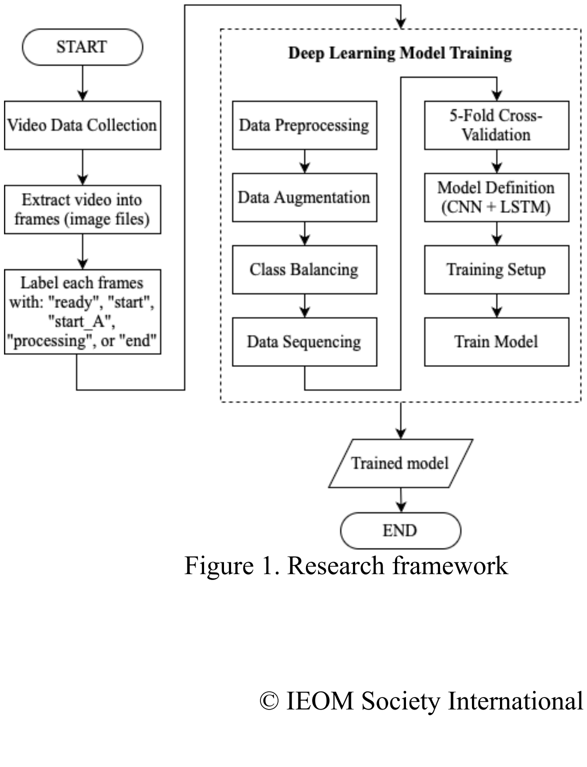

START
Video Data Collection
Extract video into frames (image files)
Label each frames with: "ready", "start", "start_A", "processing", or "end"
Deep Learning Model Training
Data Preprocessing
Data Augmentation
Class Balancing
Data Sequencing
5-Fold Cross-Validation
Model Definition (CNN + LSTM)
Training Setup
Train Model
Trained model
END

Figure 1. Research framework

In the Deep Learning Model Training phase, the labeled frames undergo several preparatory steps to ensure they are suitable for training. The process includes Data Preprocessing to clean and format the data, Data Augmentation to increase dataset diversity, and Class Balancing to address any class imbalances and ensure fair representation of all lifting stages. The frames are then organized into sequences to capture the temporal dependencies inherent in lifting tasks. A hybrid CNN-LSTM model is then developed and trained using 5-Fold Cross-Validation, ensuring robustness and reliability in stage classification.

The output of this study is a trained deep learning model, which is ready for deployment to classify lifting stages in new datasets. This structured process provides an automated and scalable solution for identifying lifting stages, supporting ergonomic risk assessments such as the Revised NIOSH Lifting Equation (RNLE).

## 4. Data Collection

Table 1 below provides a detailed description of the labeled lifting stages used in this study. These stages are critical for accurately analyzing and classifying lifting actions, which form the foundation for ergonomic risk assessments such as the Revised NIOSH Lifting Equation (RNLE). The stages are defined based on observable movements, ensuring clarity and consistency in labeling frames extracted from video data.

Table 1. Lifting stage definition

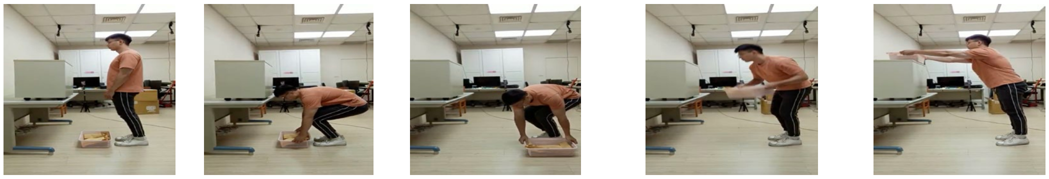
| **Stages (labels)** | **Definition** |
|---|---|
| ready | The first frame of the video until the operator touches the box. |
| Start and start_A | When the operator touches the box until the box lift a little. |
| processing | All the frames after the start and start_A stages until the box is fully placed in the lifting destination position. |
| end | When the box fully placed at the destination until the operator's hand leave the box. |

The video dataset used to train the model was gathered from a real-world experiment. Each video captures a single lifting action. The recordings were taken at six distinct locations and involved various operators. Each operator performed two types of lifting actions: symmetric and asymmetric. In total, 117 videos were used for training, with the distribution of types detailed in Table 2. The videos were converted into individual frames, transforming the video dataset into an image dataset to be used as input for the training process. Each frame was labeled according to the defined lifting stage classifications. This study categorizes the lifting process into four stages: ready, start, processing, and end. To distinguish between symmetric and asymmetric lifting actions, the start stage of an asymmetric action is specifically labeled as "start_A."

Table 2. Number of videos for training

| **Location** | **Symmetric (NA)** | **Asymmetric (A)** |
|---|---|---|
| Loc_1 | 18 | 9 |
| Loc_2 | 9 | 9 |
| Loc_3 | 9 | 9 |
| Loc_4 | 9 | 9 |
| Loc_5 | 9 | 9 |
| Loc_6 | 9 | 9 |
| **Total** | **117 videos** | |

Figure 2 illustrates five labeled stages of the manual lifting process as defined in this study: (a) Ready, (b) Start, (c) Start_A, (d) Processing, and (e) End. Each image represents a specific phase in the lifting action, captured from a single lifting task.



(a) Ready (b) Start (c) Start_A (d) Processing (e) End

Figure 2. Lifting stages image dataset

## 5. Results and Discussion

### 5.1 Class Balancing

Each video in the dataset represents a single lifting action and is divided into four stages: "ready," "start" or "start_A," "processing," and "end." When frames are extracted from the videos, some stages naturally have more frames than others. For example, the "ready" and "processing" stages consistently have a higher number of frames compared to the "start" stage. This imbalance results in an uneven class distribution, which can negatively impact the model's ability to effectively learn from the underrepresented stages, potentially leading to biased predictions.

To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) is applied to balance the dataset. SMOTE generates synthetic data for the underrepresented classes to equalize the number of frames across all stages (Kamalov 2024). Table 3 compares the number of frames for each stage before and after applying SMOTE. Initially, the dataset contains 75,920 frames, with a significant disparity between stages. After applying SMOTE, each class is balanced with 44,444 frames, resulting in a total of 222,220 frames. This class balancing ensures a fair representation of all stages, improving the model's learning process and prediction accuracy.

Table 3. Class distribution before and after balancing

| **Label code** | **Label** | **Number of Image Data (Frame)** | |
|---|---|---|---|
| | | **Before SMOTE** | **After SMOTE** |
| 0 | Ready | 44,444 | 44,444 |
| 1 | Start | 2,104 | 44,444 |
| 2 | Start_A | 1,916 | 44,444 |
| 3 | Processing | 19,924 | 44,444 |
| 4 | End | 7,532 | 44,444 |
| **Total** | | **75,920 Frames** | **222,220 Frames** |

### 5.2 Model Architecture

This research utilizes a deep learning approach that combines convolutional and recurrent neural networks. The proposed model architecture includes a 2D Convolutional Neural Network (2D CNN) layer and a Long Short-Term Memory (LSTM) layer. The 2D CNN extracts spatial features from the input data, which are then processed by the LSTM layer to capture temporal dependencies (Vrskova et al. 2022).

Figure 3 illustrates the diagram of the proposed model architecture, which is adapted from Carlos et al. (2024). Details of the architecture are provided in Table 4. The input layer accepts sequences of frames, each with a sequence length of 5 and an image size of 64 x 64 pixels (height x width), processed in grayscale. The use of grayscale reduces computational complexity without significantly affecting prediction accuracy, as the dataset consists of uniform video types depicting manual lifting actions.

The TimeDistributed Conv2D layer applies 2D convolution with 8 filters and a kernel size of 3 x 3. The ReLU activation function introduces non-linearity, enabling the model to learn spatial features from each frame in the sequence. The output of this layer is passed through a MaxPooling2D layer, which reduces the spatial dimensions of each feature map from 62 x 62 to 31 x 31. This pooling operation not only controls overfitting but also reduces computational load. The resulting feature maps are flattened using a Flatten layer, which converts the 2D feature maps into 1D feature vectors, making them suitable for further processing in the LSTM layer.

The TimeDistributed wrapper ensures that operations are independently applied to each frame in the sequence (Jin et al. 2022), preserving the temporal structure of the data. In the convolutional layers, this wrapper extracts spatial features, such as edges or textures, from each frame without mixing data across frames.

A LSTM layer with 32 units processes the sequence of flattened feature vectors to learn temporal relationships between frames. The output of the LSTM layer represents the temporal information extracted from the entire sequence and is passed to the subsequent layers.

The model includes a Dense layer with 16 units and ReLU activation, designed to capture more complex relationships between features (Javid et al. 2021). To prevent overfitting, a Dropout layer with a rate of 0.3 randomly sets 30% of the neurons to zero during training (Wu and Gu 2015), ensuring the model does not become overly reliant on specific neurons. The final layer is a Dense layer with softmax activation, used for multi-class classification. The number of units in this layer corresponds to the number of output categories, which are "ready," "start" or "start_A," "processing," and "end."

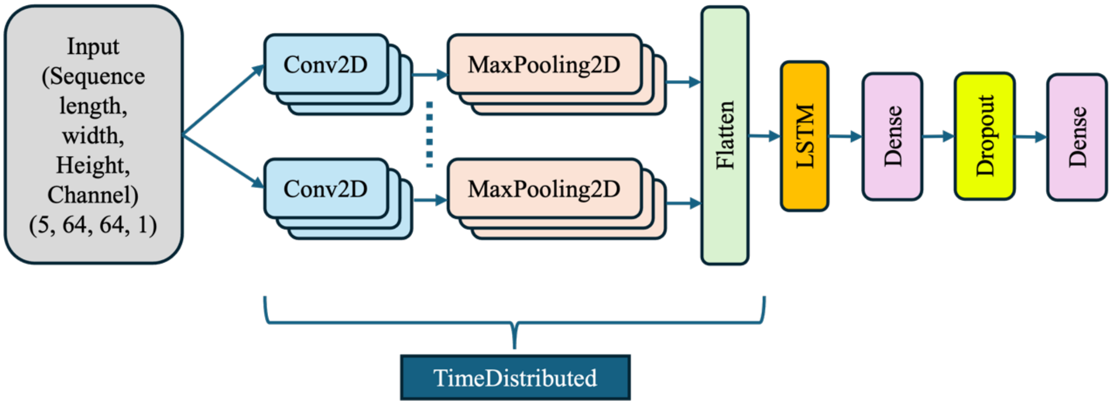


Figure 3. Model architecture diagram

Table 4. Model architecture

| Layer | Shape | Activation |
|---|---|---|
| Input | (5, 64, 64, 1) | N/A |
| TimeDistributed Conv2D (3x3) | (5, 62, 62, 8) | ReLU |
| TimeDistributed MaxPooling2D (2x2) | (5, 31, 31, 8) | None |
| TimeDistributed Flatten | (5, 7696) | None |
| LSTM | (32) | tanh |
| Dense | (16) | ReLU |
| Dropout | (16) | None |
| Output (Dense) | (5) | Softmax |

## 5.3 Hyperparameters

The hyperparameters utilized for training process includes the learning rate, batch size, sequence length, and number of epochs. The Adam optimizer was employed with an initial learning rate of 0.001 to ensure stability and facilitate fast convergence. To minimize computational complexity, a batch size of 32 and a sequence length of 5 were used. The training was configured to run for up to 50 epochs, with an EarlyStopping function implemented to halt the process if the validation loss failed to improve for 3 consecutive epochs. Under this hyperparameter configuration, the total training time was 5.2 hours.

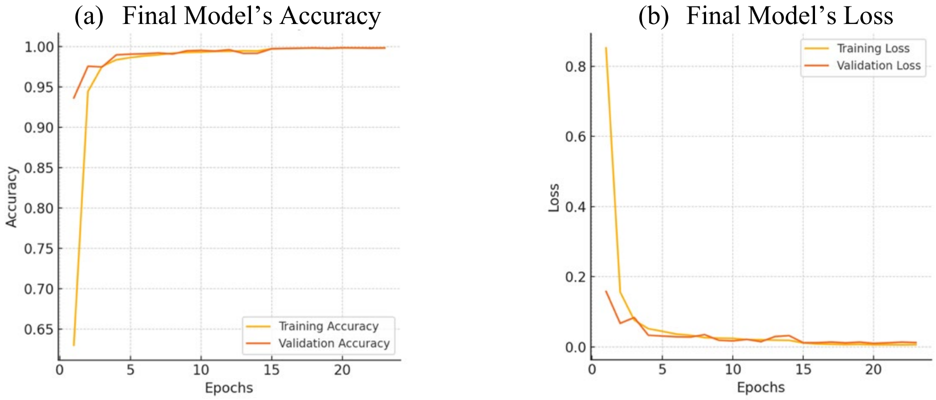
## 5.4 Cross-Validation Training Result

The cross-validation results demonstrate the model's robustness and strong performance across all five folds. As summarized in Table 5, the training, validation, and test accuracies for each fold are consistently high, averaging around 99%, while the loss values remain low. The number of epochs required to achieve similar accuracy varies between folds, indicating that the early stopping mechanism effectively reduces overall training time by halting the process once the model converges.

Table 5. Accuracy and Loss Comparison Between Folds

| Fold | Number of Epochs | Final Train Accuracy | Final Validation Accuracy | Final Train Loss | Final Validation Loss | Test Accuracy | Test Loss |
|---|---|---|---|---|---|---|---|
| Fold 1 | 23 | 0.998 | 0.998 | 0.006 | 0.009 | 0.998 | 0.008 |
| Fold 2 | 15 | 0.997 | 0.997 | 0.008 | 0.012 | 0.997 | 0.011 |
| Fold 3 | 13 | 0.998 | 0.998 | 0.008 | 0.012 | 0.998 | 0.009 |
| Fold 4 | 15 | 0.995 | 0.998 | 0.009 | 0.012 | 0.998 | 0.010 |
| Fold 5 | 23 | 0.998 | 0.998 | 0.006 | 0.012 | 0.998 | 0.010 |
| Mean | | 0.997 | 0.998 | 0.007 | 0.011 | 0.998 | 0.010 |
| Standard Deviation | | 0.001 | 0.000 | 0.001 | 0.001 | 0.000 | 0.001 |

Figure 4. displays the training and validation results for the final model of the suggested deep learning model. The figure's part (a) demonstrates that the training accuracy begins at about 65% and rises after 5–10 epochs, followed by the validation accuracy, which has a comparable value. Training and validation accuracy declined concurrently until 10 epochs, as seen by the final model's loss in the part (b) graph. Between 10 and 15 epochs, the validation loss varies, but for the remaining epochs, it remains low and is comparable to the training loss.


(a) Final Model's Accuracy

(b) Final Model's Loss

Figure 4. Model's Accuracy and Loss Plot

The accuracy and loss graphs indicate that the training process was effective, with the model learning efficiently and showing minimal risk of overfitting. All five folds completed training in fewer than 50 epochs, typically stopping around 20 epochs, thanks to the EarlyStopping mechanism implemented in the training process. Detailed results for each fold are discussed in the following section.

The confusion matrix for the overall 5-fold cross-validation, shown in Figure 5, demonstrates the strong performance of the model in classifying lifting stages. The high values along the diagonal indicate that the model accurately predicts the true labels for most cases. In contrast, the off-diagonal values represent instances of misclassification. For the true label "ready," the model often misclassifies frames as "processing," and similarly, "processing" frames are occasionally misclassified as "ready." These errors are likely due to the similar features shared between these two stages. On the other hand, the “start” and “start_A” labels show the least misclassification, which is particularly important as these stages are critical for identifying lifting actions (Figure 5).

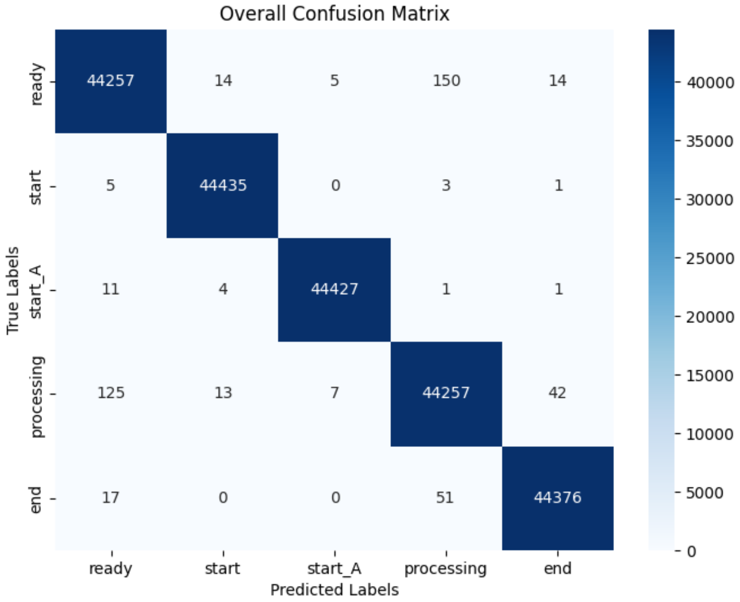


Figure 5. Overall Confusion Matrix from Cross-Validation

## 6. Conclusion

This study developed a video-based deep learning model to automate the identification of lifting stages in manual lifting tasks, supporting ergonomic risk assessments like the Revised NIOSH Lifting Equation (RNLE). The hybrid CNN-LSTM model demonstrated high accuracy, achieving an average of 99% across training, validation, and test datasets during 5-fold cross-validation. The use of SMOTE effectively addressed class imbalances, ensuring reliable classification of the five lifting stages: "ready," "start," "start_A," "processing," and "end."

This research highlights the potential of deep learning models to improve the scalability and accuracy of ergonomic risk assessments by automating complex processes like lifting stage identification. The integration of temporal and spatial features in the hybrid CNN-LSTM architecture provides a scalable solution that minimizes human intervention and enhances consistency in ergonomic evaluations.

Future work could explore applying the model to diverse lifting conditions and scenarios. Additionally, incorporating advanced techniques for feature extraction and dataset expansion could further enhance the model’s performance and adaptability, making it an even more robust tool for workplace safety and health assessments.

## References

Carlos, W. C., Copetti, A., Bertini, L., Moreira, L. B. and Gomes, O. de S. M., Human activity recognition: an approach 2D CNN-LSTM to sequential image representation and processing of inertial sensor data. *AIMS Bioengineering* [online], 11 (4), 527–560, 2024. Available from: http://www.aimspress.com/article/doi/10.3934/bioeng.2024024.

Fernández, M. M., Fernández, J. Á., Bajo, J. M. and Delrieux, C. A., Ergonomic risk assessment based on computer vision and machine learning. *Computers and Industrial Engineering*, 149. 2020.

Fox, R. R., Lu, M. L., Occhipinti, E. and Jaeger, M., Understanding outcome metrics of the revised NIOSH lifting equation. *Applied Ergonomics*, 81. 2019.

Hernandez, J., Valarezo, G., Cobos, R., Kim, J. W., Palacios, R. and Abad, A. G., Hierarchical Human Action Recognition to Measure the Performance of Manual Labor. *IEEE Access*, 9, 103110–103119. 2021.

Jacob, V., Bhasi, M. and Gopikakumari, R., Effect of work related variables on human errors in measurement. *Journal of Electrical and Electronics Engineering Research*, 4 (1). 2012.

Javid, A. M., Das, S., Skoglund, M. and Chatterjee, S., A relu dense layer to improve the performance of neural networks. *In*: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. Institute of Electrical and Electronics Engineers Inc., 2810–2814. 2021.

Jeong, S. O. and Kook, J., CREBAS: Computer-Based REBA Evaluation System for Wood Manufacturers Using MediaPipe. *Applied Sciences (Switzerland)*, 13 (2). 2023.

Jin, B., Peng, Y., Kuang, X., Zhang, Z., Lian, Z. and Wang, B., Robust Dynamic Hand Gesture Recognition Based on Millimeter Wave Radar Using Atten-TsNN. *IEEE Sensors Journal*, 22 (11), 10861–10869. 2022.

Kamalov, F., ASYMPTOTIC BEHAVIOR OF SMOTE-GENERATED SAMPLES USING ORDER STATISTICS. *Gulf Journal of Mathematics*, 17 (2), 327–336. 2024.

Li, L., Martin, T. and Xu, X., A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders. *Applied Ergonomics*, 87. 2020.

Lowe, B. D., Dempsey, P. G. and Jones, E. M., Ergonomics assessment methods used by ergonomics professionals. *Applied Ergonomics*, 81. 2019.

Lu, M.-L., Barim, M. S., Feng, S., Hughes, G., Hayden, M. and Dwight, W., Development of a Wearable IMU System for Automatically Assessing Lifting Risk Factors. *Posture, Motion and Health: 11th International Conference*, (Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management.), 194–213. 2020.

Roopa, B. S., Prema, K. N. and Smitha, S. M., Brief Study on Convolutional Neural Networks. *International Journal For Multidisciplinary Research*, 6 (5). 2024.

Sekkay, F., Prevention of Work-Related Musculoskeletal Disorders supported by Artificial Intelligence. *In*: *Human Interaction and Emerging Technologies (IHIET-AI 2023): Artificial Intelligence and Future Applications*. AHFE International. 2023.

Thomas, B., Lu, M. L., Jha, R. and Bertrand, J., Machine Learning for Detection and Risk Assessment of Lifting Action. *IEEE Transactions on Human-Machine Systems*, 52 (6), 1196–1204. 2022.

Vrskova, R., Hudec, R., Kamencay, P. and Sykora, P., A New Approach for Abnormal Human Activities Recognition Based on ConvLSTM Architecture. *Sensors*, 22 (8). 2022.

Waters, T. R., Putz-Anderson, V. and Garg, A., Application Manuals For The Revised NIOSH Lifting Equation. Ohio: DHHS (NIOSH). 1994.

Wu, H. and Gu, X., Towards dropout training for convolutional neural networks. *Neural Networks*, 71, 1–10. 2015.

## Biographies

**Listia Anjani** is a master’s dual degree student in Industrial Engineering of Universitas Gadjah Mada, Indonesia, and Industrial Management of National Taiwan University of Science and Technology, Taiwan. She was graduated from the same major, Industrial Engineering of Universitas Gadjah Mada in 2020. Before starting her master’s degree, she worked at the multinational retail company, IKEA, from 2021-2022. Her current research area is application of deep learning approach in ergonomic risk assessment.

**Kung-Jeng Wang** is a Distinguished Professor in Department of Industrial Management, School of Management, National Taiwan University of Science and Technology. With expertise in supply chain management, technology and operations management, and smart manufacturing, he has led over 40 research projects sponsored by NSC, Taiwan Tech, and industry partners. He has received numerous accolades, such as the NTUST Outstanding Research Award and Excellence Teaching Award. He earned his PhD in Industrial Engineering from the University of Wisconsin-Madison and has also been a visiting scholar at Dresden University of Technology and a trainee at Harvard Business School.

**Hilya Mudrika Arini** is an Associate Professor in Industrial Engineering, Department of Mechanical and Industrial Engineering at Universitas Gadjah Mada, Indonesia. Her research focuses on system dynamics, communication strategies, and evidence-based policies to address global challenges such as renewable energy adoption, climate policy communication, and disaster management. With experience in collaborative projects across Australia, Europe, and Asia, she works to bridge academia, industry, and government to develop sustainable solutions. Dr. Arini holds a PhD in Management Science from the University of Strathclyde, United Kingdom.