Final Report:
Classification of Occupational Risk During Repetitive Lifting Tasks Using EMG and IMU Sensor Data

Isabel Andaya Evan Chang

June 13, 2025

Abstract

Abstract

This project investigates the use of wearable sensor data, including surface electromyography(EMG) and inertial measurement unit(IMU) signal, for binary categorization of biomechanical risk during repetitive lifting tasks. During controlled lifting trials, data were gathered from four EMG sensors(biceps and deltoids) and IMUs. Our goal is to create classifications models that can differentiate between low and high risk conditions utilizing EMG and IMU separately, with the ultimate goal of combining both modalities to increase predictive accuracy. After preprocessing and extracting characteristics, we trained and tested models like Logistic and Random Forest. The finding show that some models outperform others in risk classification and that specific variables from each sensor type contribute differently to prediction performance depending on the modeling strategy.

Contents

1 Introduction

Manual labor tasks are common in physically demanding industries such as logistics, construction, and manufacturing, and pose a significant physical danger to workers. Prolonged exposure to repetitive motions, poor posture, and overexertion can result in work-related musculoskeletal disorders(WMSDs) which are one of the leading causes of occupational injury and long-term impairment. Despite increased awareness of these concerns, manual labor work environments lack the objective tools for assessing physical effort and detecting dangerous movements. Wearable sensor technology provides a possible solution to collect continuous data on muscle activity and body motion in natural work settings allowing for data-driven decisions to prevent injuries.

Surface electromyography(EMG) and inertial measurement units(IMU) are two common sensor types utilized in this context. EMG sensors measure the electrical signals produced by active muscle fibers. These signals indicate how much a muscle is engaged, how much muscle effort required, and how rapidly it fatigues. An EMG sensor is essential for quantifying workload in upper-body muscle groups. In contrast, IMUs are made up of accelerometers and gyroscopes that gather movement information. They give exact measures of posture, joint angles, and movement dynamics. An IMU sensor enables biomechanical analysis of dynamic tasks involving lifting and reaching.

This allows users to test both EMG and IMU data to develop a binary classification algorithm that can distinguish between low and high risk lifting operations. Using a machine learning approach, we hope to develop a system that can detect dangerous movement patterns. Such a tool could be used to monitor physical workload in real time and prevent future injuries to support occupations in labour-heavy environments.

- **Problem Statement:** The risk of WMSD while performing certain tasks is currently unknown and must be predicted to prevent long-term injury.
- **Objective / Research Question:** The objective of this project develop a machine learning model that uses EMG and IMU data to predict occupational risk during repetitive lifting tasks. We aim to answer the question: can a supervised model accurately distinguish high- and low-risk tasks based on these features?

2 Data Understanding and Preprocessing

2.1 Dataset Description

The dataset used in this study was collected from two human subjects during a repetitive lifting experiment. The participants consisted of one male (Participant 1) and one female (Participant 2), each performing identical lifting tasks under two conditions: High risk (load greater than twenty pounds) and Low risk (load less than twenty pounds). This resulted in a total of four time-series data files:

- P1_High.csv — male subject high risk
- P1_Low.csv — male subject low risk
- P2_High.csv — female subject high risk
- P2_Low.csv — female subject low risk

Each trial involved repetitive box lifting using a standardized technique, simulating occupational lifting in an industrial setting. Each session lasted a few minutes and included a sufficient number of lifting repetitions to assess fatigue-related and ergonomic patterns.

The data includes signals from the following wearable sensors:

- **EMG Sensors:** Four surface EMG sensors were placed bilaterally on the upper body:
  - Sensor 1 — Right Bicep
  - Sensor 2 — Right Deltoid
  - Sensor 3 — Left Bicep
  - Sensor 4 — Left Deltoid

  EMG data were sampled at 1259 Hz to capture high-frequency muscle activation dynamics.

- **IMU Sensors:** Each EMG sensor was accompanied by an inertial measurement unit (IMU) that recorded:
  - Triaxial acceleration (X, Y, Z)
  - Triaxial gyroscope (angular velocity in X, Y, Z)

  IMU data were sampled at 148 Hz.

Each CSV file includes 56 columns representing raw signal data and multiple timestamp columns (some of which were redundant and removed during preprocessing). The dataset serves as the foundation for supervised classification tasks aimed at predicting whether a lifting instance was performed with or without exoskeleton assistance based on extracted physiological and kinematic features.

2.2 Data Cleaning

Effective data cleaning ensures that the input data is reliable, consistent, and conducive to accurate model training.

Importance of Data Cleaning

- Cleaning is essential to reduce irreducible error caused by noise, faulty sensors, or measurement anomalies.
- Many machine learning algorithms cannot handle missing values or extreme outliers, leading to unreliable predictions.
- Reducing noise and removing anomalies from the training dataset also mitigates overfitting and enhances model generalizability.
- Preprocessing decisions such as how to handle missing values or filter noise should be treated as hyperparameters and validated using cross-validation or grid search.

Table 1: Summary of Data Cleaning Techniques

Step	Methodology and Rationale
Missing value imputation	Missing values were filled using median imputation.
Outlier Removal	Drop any raw sample beyond ±3 from the mean to remove extreme values from EMG and IMU data.
Standardization	

Cleaning Methods Applied in Our Project

Effective data cleaning is a fundamental and often underestimated step in preparing

2.3 Feature Engineering

To transform raw time-series signals from EMG and IMU sensors into machine learning inputs, we implemented a windowed feature extraction method with physiologically relevant metrics.

Windowing Strategy: We applied a sliding window of 100 samples with a step size of 30 across synchronized EMG and IMU signals. This enabled high-resolution, continuous feature extraction across time.

EMG Features:

- Mean, maximum, minimum, standard deviation, RMS

IMU Features:

- Acceleration peak and range: Physical exertion indicators
- Gyroscope peak and range: Angular movement intensity

Labeling: Each windowed feature set was tagged with a binary label (High or low risk) and participant number. These labels supported supervised classification.

2.4 Example Table for Feature Statistics

Table 2: Summary Statistics of Key Features

Feature	Mean	Standard Deviation	Range
EMG	0.008371	0.1204	5.4628
IMU Acceleration	0.7005	20.8550	182.2948

3 Modeling

3.1 Model Selection

For this study, we adopted a classification approach to distinguish betwen lifting tasks associated with high or low biomechanical risk. Given that the task involves assigning discrete lables of either "high risk" or "low risk" for each instance, the problem is inherently binary and label driven. This makes classification a more suitable methodology than regression, which is typically used for predicting continuous outcomes/movements/ By framing the task as a supervised classification problem, we were able to leverage labled training data to develop models that learn patterns in muscle activity and motion features to predict risk level.

We considered several machine learning models that are well-established in time-series and wearable sensor analysis:

- **Logistic Regression**: A suitable choice for binary classification task since it explicitly models the chance of an instance being high or low risk. Its simplicity and interpretability make it an excellent baseline for learning how individual EMG and IMU variables contribute to risk classification. Furthermore, it performs efficiently with well-engineered features and has a lower risk of overfitting, which is useful given the moderate size of the experimental data.

- **Random Forest**: A powerful and flexible algorithm suited for binary classification. Unlike linear models, random forest detec complicated nonlinear correlations between characteristics and risk labels. It works as a collection of decision tress, reducing the risk of overfitting while enhancing generalization by predictions from numerous tress. Random forest also provide feature importance scores which offer insights into which sensor derived metrics are most predictive of risk level. Its ability to handle noisy/outlier data, mixed feature types, and complexity between classes makes it a strong candidate for modeling biomechanical risk.

Models were selected to balance between **interpretability, performance, and suitability for temporal sensor data**. Evaluation of each model's accuracy, precision, and recall allowed us to assess which architecture most effectively differentiates between high and low biomechanical risk conditions and contribute meaningfully to risk assessments.

Table 3: EMG - Logistic Regression Hyperparameters

Hyperparameter	**Description**	**Default**
`C`	Inverse of regularization strength. Smaller values imply stronger regularization (L2 by default).	`1.0`
`penalty`	Type of regularization (`'l2'`, `'l1'`, `'elasticnet'`, or `'none'`). Helps prevent overfitting.	`'l2'`
`solver`	Optimization algorithm. liblinear support binary classification.	`'liblinear'`
`multi_class`	`'ovr'` (one-vs-rest) or `'multinomial'` (softmax). Use `'multinomial'` with `'lbfgs'`.	`'ovr'`
`max_iter`	Maximum iterations for solver convergence.	`100`
`tol`	Tolerance for stopping criteria. Lower values make training more precise but slower.	`1e-4`

3.2 Model Training

Detail the process of training your model. Explain how you split the dataset, tuned hyperparameters, and optimized model performance.

EMG Based Classification

Table 4: EMG - Random Forest Hyperparameters

Hyperparameter	Description	Default
n_estimators	Number of trees in the forest. More trees improve performance but increase computational cost.	100
max_depth	Maximum depth of each tree. Limits tree growth and helps prevent overfitting.	None
min_samples_split	Minimum number of samples required to split an internal node.	2
min_samples_leaf	Minimum number of samples required to be at a leaf node.	1
min_weight_fraction_leaf	Minimum weighted fraction of total samples required at a leaf node.	0.0
max_leaf_nodes	Maximum number of leaf nodes. Limits model complexity.	None
max_features	Number of features considered at each split. Introduces randomness and reduces overfitting.	"auto"
bootstrap	Whether bootstrap samples are used when building trees. If False, pasting is used.	True
n_jobs	Number of CPU cores to use for parallel training. -1 uses all cores.	None
criterion	Function to measure split quality: "gini" or "entropy".	"gini"

Table 5: IMU - Random Forest Hyperparameters

Hyperparameter	Description	Default
n_estimators	Number of decision trees in the forest. Larger values improve performance but increase computation.	100
max_depth	Maximum depth of each tree. Shallower trees reduce overfitting.	None
min_samples_split	Minimum number of samples required to split an internal node. Higher values simplify trees.	2
min_samples_leaf	Minimum number of samples required to be at a leaf node. Promotes more balanced splits.	1
min_weight_fraction_leaf	Minimum weighted fraction of total samples required to be at a leaf node. Useful for imbalanced datasets.	0.0
max_leaf_nodes	Maximum number of leaf nodes per tree. Forces simpler tree structures.	None
max_features	Number of features considered for the best split at each node. Reduces variance and overfitting.	"auto"
bootstrap	Whether bootstrap samples are used when building trees. If False, pasting is used instead.	True
n_jobs	Number of CPU cores to use during training and prediction. -1 means all cores.	None
random_state	Controls the randomness of the estimator for reproducibility.	None
criterion	Function to measure the quality of a split: "gini" or "entropy".	"gini"

Surface Electromyogrpahy(EMG) data gives information about muscle activation by recording the electrical signal generated during muscular contractions. In this work, we collected data from four EMG sensors placed on the biceps and deltoids of both arm. To clean the raw signals, missing values were filled with column means and severe outliers were removed using z-score clipping at plus/minus 3 standard deviations. The raw data were then standardized with StandardScaler to ensure that all data had a zero mean and unit variance. This step helps improves model convergence and stability when utilizing feature scale sensitive model.

For feature extraction, a function extract_rolling_emg_features() were created to extract features necessary for the model. It begins by defining a rolling window mechanism across all the csv. For each window size, we used a window size = 100, meaning each subset of the data will have 100 data selected. The parameter step size = 30 is set to generate an overlapping segment of 30 data for each subset of data. This approach provide great resolution while also allowing more testing data samples sets which proves critical for identifying activity changes during lifting actions. We then add a Lebel column(0 = Low Risk, 1 = High Risk) and appends this labeled feature DataFrame to the list. We ccombined all the labeled feature and extract the datframes into a single dataset. Lastly, we split into features and labels seperately into X, which contains the numerical features used for training, and y, which contains the binary classification labels.

For the model optimization, we chose to select logistic regression and random forest classifier. For logistic regression, we selected penalty = 'l2', solver = 'liblinear', and random state = 42. Penalty = 'l2', also known as ridge regularization, helps prevent overfitting by adding a penalty to large coefficient values in the model. Solver = 'liblinear' is useful for a small to medium sized binary datasets and support the l2 penalty. Random state = 42 is the standard for random state and ensures reproducibility. This means every time you run your code, you get the same results which helps with debugging and scientific reporting. For random forest classifier, our parameter includes n estimator =

100 and random state = 42. N_estimator specifies the number of decision trees in the ensemble/forest and 100 is the common default value. Similiarly to the logistic regression, we chose random_state = 42 because it is the standard for random state.

IMU Based Classification

Inertial measurement units(IMU) provide a view of biomechanical behavior by tracking motion through acceleration and angular velocity. For this project, we used IMU data recorded from accelerometer and gyroscope sensors along three axes X, Y, and Z. Similarly to EMG, raw data was processed for accelerometer using a sliding window method where each window produced derived metrics such as peak acceleration, mean acceleration, total acceleration, and range of acceleration magnitude. Similar features was extracted for gyroscope signals, capturing rotational dynamics during lifting task.

The features were extracted from each trail were labeled as either low-risk or high-risk and all segments were compiled into a complete dataset. A function extract_rolling_imu_features() was created with the same parameter of window size = 100 and step size = 30 to represent a similar test model to EMG test data. Unlike EMG data, the IMU features did not undergo standardization since the feature scales were already interpretable and relatively consistent.

A random forest classifier was trained on the IMU feature set using a 70/30 stratified train-test split. Prior to training, missing values were handled by imputing the mean of each feature from the training data. To maintain consistency with the EMG model, the classifier's hyperparameter were set to n_estimator = 100 and random_state = 42.

3.3 Evaluation Metrics

Explain the metrics used to evaluate model performance (e.g., accuracy, F1-score for classification; R^2, RMSE for regression). Present performance results comparing models for EMG and IMU.

- **Accuracy**: Proportion of correctly identified instance, both high and low risk, out of the total number of predictions made. High accuracy indicates the model is generally reliable in distinguishing between low and high risk lifting task.

$$\text{F1} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \tag{1}$$

- **Precision**: Proportion of correctly identified high risk cases among all instances the model label as high risk. High precision means the model is good at avoiding false alarms.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2}$$

- **Recall**: Proportion of actual high-risk tasks that were correctly identified by the model. High recall means the model effectively detects most risky lifting instance.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3}$$

- **F1-Score**: Evaluation of the performnace of a classification model and calculate the mean of precision and recall, providing a balanced view of model's ability to correctly identify positive instances while minimizing false positve.

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

	Precision	Recall	F-1 Score	Support
Low Risk	0.58	0.52	0.55	1981
High Risk	0.57	0.63	0.60	2002
Accuracy:			58%	
Mean Squared Error:			0.25	
Root Mean Squared Error:			0.5	
R2 Score:			0.02	
Confusion Matrix				
1030		951		
734		1268		

Table 6: EMG - Logistic Regression Evaluations

- **Mean Square Error(MSE)**: Measures the average of the squared differences between predicted values and actual labels. MSE help assess how well the model's confidence scores align with reality. A lower MSE means the predicted probabilities are close to actual outcomes.

$$\text{MSE}(y, \hat{y}) = \frac{\sum_{i=0}^{N-1}(y_i - \hat{y}_i)^2}{N}$$

- **Root Mean Squared Error(RMSE)**: Square root of MSE. Brings the error metrics back to the original scale of the target variable. RMSE helps assess the model's confidence calibration, on average how far the model's predicted risk probability is from the true risk label.

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1}(y_i - \hat{y}_i)^2}{N}}$$

- **R2 Score**: Coefficient of determination. Measurement to evaluate how well a model explains the variability of the target variable. It ranges from 0 to 1, with 1 being the best and 0 being the lowest. It can be negative if the model is worse than simply predicting the mean.

$$\text{R2 Score} = 1 - \frac{\sum_{i=0}^{N-1}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{N-1}(y_i' - y_i)^2}$$

Based on the overall results, the random forest model with EMG data demonstrated superior performance. Random forest using EMG data achieved the highest recall rate, which is important in a risk classification model to ensure that high risk cases are correctly identified and prioritizing safety by minimizing the chance of missing dangerous risk. Additionally, IMU data did not contribute meaningful results to classification performance across various window sizes and preprocessing strategies.

	Precision	Recall	F-1 Score	Support
Low Risk	0.82	0.85	0.83	1981
High Risk	0.84	0.81	0.83	2002
Accuracy:			83%	
Mean Squared Error:			0.17	
Root Mean Squared Error:			0.41	
R2 Score:			0.32	
Confusion Matrix				
1682		299		
374		1628		

Table 7: EMG - Random Forest Evaluations

	Precision	Recall	F-1 Score	Support
Low Risk	0.80	0.08	0.15	2134
High Risk	0.52	0.98	0.68	2148
Accuracy:			53%	
Mean Squared Error:			0.47	
Root Mean Squared Error:			0.68	
R2 Score:			-0.87	
Confusion Matrix				
176		1958		
43		2105		

Table 8: IMU Random Forest Evaluations

4 Results and Discussion

4.1 Results

4.2 Figure for Model Performance

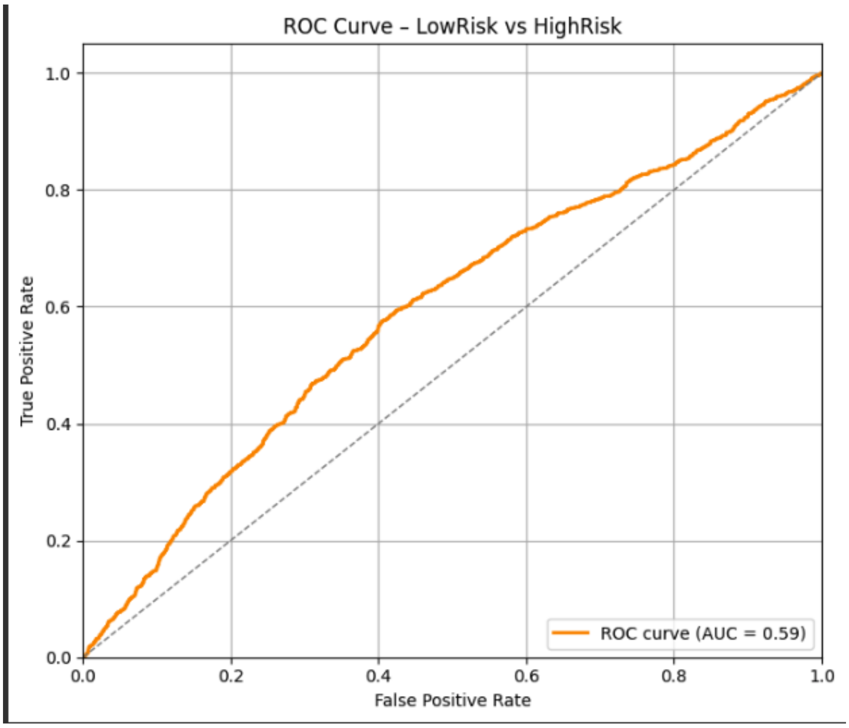

Figure 1: Logistic Regression ROC Curve Visualization for EMG data set

According to the ROC Curve findings, the logistic regression model performs only slightly better than random chance, as indicated by the position of the curve being slightly above the dashed diagonal reference line. The gradual rise of the curve suggests only moderate sensitivity across various false positive rates. However, with an AUC of only 0.59, indicating that only a 59% chance model assigns a higher probability to a randomly chosen HighRisk, the model demonstrates limited discriminative ability. Since the model did not meet the common default standard of AUC = +0.7. To have a better understanding the model's limitations and identify potential misclassification, we wanted to incorporate another visual representation that can reveal distribution of data points and highlight position where the model struggles to distinguish between the classes.

This PCA scatter plot visualize the logistic regression model's prediction. Each color and shape represents the predicted label.

Correctly Predicted Low Risk - Blue circle

Correctly Predicted High-Risk - Red X

From the plot, we can observe a significant overlap between the two classes and no clear boundary separating the high and low risk tasks. Many predictions are visibly misaligned with the true label, especially toward the middle. This supports the earlier AUC = 0.59 and concludes that the model is not sufficiently effective for reliable risk classification.

This generated visual display is a decision tree model for the random forest classifier used to distinguish between high and low risk tasks. The tree starts by splitting on

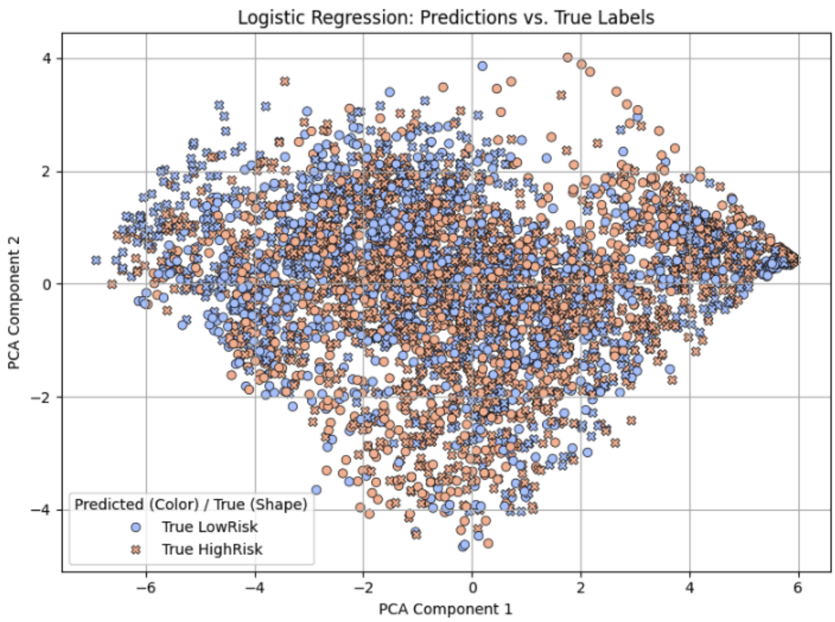

Figure 2: Logistic Regression Scatter Plot Visualization for EMG Data Set

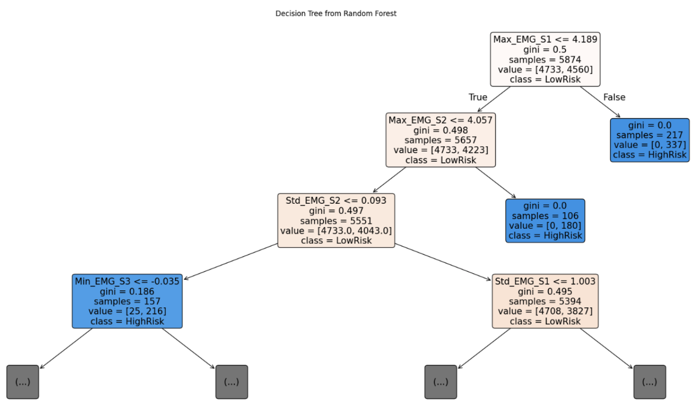

Figure 3: Random Forest Classifier Decision Tree Visualization for EMG Data Set

Max_EMG_S1 less than or equal to 4.189. This means that throughout the training, Sensor 1's Max EMG is the strongest indicator in the initial decision. If this condition is true, the model continues down the left sub tree; otherwise, it is classified as high risk with perfect certainty(Gini = 0.0). The tree continues to display different criteria, highlighting the strength of Random Forest in capturing non linear patterns in our low and high risk classification model.

4.3 Discussion

This study explored the effectiveness of using wearable EMG and IMU sensors to classify biomechanical risk in repetitive lifting tasks. Our objective was to establish a reliable supervised model capable of distinguishing between high-risk and low-risk conditions based on muscle activation and kinematic patterns.

The random forest classifier trained on EMG-derived features demonstrated the strongest overall performance, with precision, recall, and F1-scores (approximately 0.83). This suggests EMG signals effectively captured essential differences in muscle activation intensity and fatigue, key indicators of biomechanical risk. The ability of EMG signals to reflect physiological stress through muscle activity makes them robust predictors in ergonomic assessments.

The logistic regression model using EMG data demonstrated limited discriminatory power, as evidenced by a relatively low AUC of 0.59. The linear assumptions underlying logistic regression appear insufficient to capture the complex and potentially nonlinear interactions inherent in muscle activation patterns during dynamic movements. This limitation is further highlighted in the PCA scatter plot, which revealed significant overlap between predicted classes, indicating logistic regression's difficulty in differentiating subtle distinctions in risk.

The findings emphasize the superiority of random forest models in capturing intricate patterns within EMG data due to their flexibility in handling nonlinear relationships. The random forest algorithm's ensemble-based approach effectively mitigated overfitting, enabling better generalization across lifting scenarios and participants. This aligns with existing research emphasizing random forest's effectiveness in physiological and biomechanical data analysis.

IMU-based features consistently demonstrated weaker predictive performance across models and preprocessing strategies. Although IMUs theoretically provide detailed kinematic data, the random forest classifier trained on this data exhibited poor precision and recall, alongside a notably negative R^2 score (-0.87). Examination of the confusion matrix indicated a consistent misclassification bias toward high-risk predictions, suggesting a lack of distinctive biomechanical patterns differentiating risk conditions in the IMU data. One potential experimental improvement could involve structuring tasks to intentionally exaggerate IMU-based biomechanical differences, for example by having participants perform lifting repetitions to failure, potentially inducing measurable deviations in posture and movement that clearly delineate high-risk conditions.

Implementing fatigue-inducing protocols introduces ethical considerations. There is a risk of causing undue physical stress or injury to participants, particularly those with pre-existing musculoskeletal vulnerabilities. Ethical guidelines necessitate thorough informed consent procedures, clear communication of potential risks, appropriate safety measures, and careful monitoring during the experiment to ensure participant well-being. Researchers must balance the scientific benefit of clearly differentiated risk conditions against the ethical responsibility to protect participants from potential harm.

Challenges encountered during this project include data variability stemming from participant-specific muscle activation patterns and the limited size of the dataset, which constrained the extent of model generalization and increased the potential for overfitting. In addition, future studies would benefit from a larger and more diverse participant pool to enhance the robustness and applicability of predictive models.

Exploring advanced feature extraction techniques, such as frequency-domain analyses for EMG signals or more sophisticated motion segmentation methods for IMU data, might enhance predictive accuracy. Future research could also investigate sensor fusion techniques, combining EMG and IMU data within integrated machine-learning frameworks, potentially capturing complementary information that may enhance classification performance.

5 Conclusion

This study demonstrates the viability of using a machine learning model with random forest classifiers to accurately distinguish between high-risk and low-risk conditions in repetitive lifting tasks. EMG-derived features effectively captured meaningful variations in muscle activation, highlighting their potential in ergonomic risk assessment and injury prevention in occupational settings. Conversely, IMU-based features did not provide sufficient differentiation between risk categories, underscoring limitations in capturing distinct biomechanical patterns within the experimental protocol.

For occupational risk analysis, these findings imply that EMG sensors and machine learning models could form the basis of real-time monitoring systems, enabling proactive risk mitigation in physically demanding jobs. Future improvements should include expanding participant diversity and sample sizes to enhance model robustness and exploring advanced feature extraction methods, particularly for IMU data. Additionally, ethical and practical implications of inducing fatigue for clearer biomechanical differentiation must be carefully balanced against participant safety.

6 References

1. Hudgins, B., Parker, P., Scott, R. N. (2000). A new strategy for multifunction myoelectric control *IEEE Transactions on Biomedical Engineering*, 22(24), 9713. `jneuroengrehab.biomedcentral.com`

2. Phinyomark, A., Khushaba, R. N., Scheme, E. (2018). Feature extraction and selection for myoelectric control. *IEEE Reviews in Biomedical Engineering*. `researchgate.net`

3. Hume, P. A., Groll, J. H. (2025). Decision Trees and IMU Sensors for Risk Prediction in Manual Material Handling.*Springer*. `link.springer.com`

4. Hossain, S., et al. (2023). Hybrid Learning Models for IMU-Based HAR with Feature Analysis *Sensors, 23(18), 7802*. `mdpi.com`

5. Di Nardo, F., et al. (2024). Biomechanical Risk Classification in Repetitive Lifting Using Multi-Sensor Data.*Biosensors, 15(2), 84*. `mdpi.com`

6. OpenAI. (2024). *ChatGPT* (GPT-4) [Large Language Model]. `https://chat.openai.com`

7. Microsoft. (2024). *GitHub Copilot* [AI Code Assistant]. GitHub. `https://github.com/features/copilot`