Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 2509–2519
https://doi.org/10.54941/ahfe1008476

Predicting Ergonomic Risk and Worker Performance Using Neural Networks With EMG and Kinematic Data

Sohyung Cho

Department of Industrial and Systems Engineering, Kennesaw State University, Marietta, GA 30060, USA

ABSTRACT

Ergonomic performance is fundamentally governed by the interaction between postural configuration and muscle activation; however, frontier-based methodologies such as Data Envelopment Analysis (DEA) yield relative efficiency scores rather than predictive estimates. This study investigates the extent to which supervised neural networks (NNs) can generate predictive estimates of DEA-derived ergonomic efficiency using electromyography (EMG) and kinematic inputs. In this study, ten participants completed five sweeping task trials with bilateral EMG on biceps brachii and flexor carpi radialis, and elbow angles captured via motion tracking. Trial features were aggregated into six inputs per participant to compact feedforward NNs trained for regression of left/right DEA efficiencies and classification of efficient frontier participants. Leave-one-out cross-validation and probability calibration were applied. As results, left-side regression achieved $R^2 = 0.57$ and right-side performance was weaker with $R^2 = 0.11$. The classifier reached 80% accuracy with high distinguishing power of efficient frontiers about 75% (F1 accuracy) and high ability to rank-order efficient vs. inefficient frontiers about 92% (AUC accuracy). Key predictors included left elbow angle and bilateral biceps activity. Two NN–DEA disagreements reflected differences between normative and predictive modeling. As a conclusion, neural networks accurately predicted ergonomic efficiency and frontier membership from the same EMG/kinematic data used in DEA, demonstrating their value for prospective, real-time screening and future extension to sequence-based models.

Keywords: Ergonomics, Electromyography (EMG), Upper-limb kinematics, Data Envelopment Analysis (DEA), Neural networks

INTRODUCTION

Ergonomics is fundamentally concerned with designing tasks, tools, and environments that align with human physiological, biomechanical, and cognitive capabilities to reduce musculoskeletal risk and enhance productivity. Poorly designed work methods can elevate muscle activation, increase joint loading, and initiate fatigue, all of which diminish worker performance and increase the likelihood of musculoskeletal disorders (MSDs). Electromyography (EMG) and upper-limb kinematics have therefore become essential tools in ergonomic assessment because they provide quantitative measurements of muscle activation, joint coordination, and movement

Received June 1, 2026; Revised September 29, 2026; Accepted October 10, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

efficiency during work tasks. Surface EMG, when properly normalized using maximal voluntary contractions (MVC), allows meaningful between-subject comparisons of muscular effort, while motion-capture–derived joint angle measurements characterize the mechanical demands of specific postures and dynamic movements.

In recent work, Data Envelopment Analysis (DEA) has emerged as a promising method for modeling ergonomic performance because of its ability to evaluate relative efficiency across individuals performing similar tasks. DEA, particularly under the Constant Returns to Scale (CCR) formulation, evaluates Decision-Making Units (DMUs)—such as workers or task executions—by comparing weighted combinations of inputs and outputs without requiring an explicit functional form. Because ergonomic performance involves multiple physiological and kinematic factors, DEA provides a non-parametric framework for identifying efficient workers, quantifying inefficiency, and identifying "role models" for suboptimal performers (Charnes et al., 1978).

Our previous study applied the CCR-DEA SIMO framework to a sweeping task using EMG (BB and FCR muscles) and upper-limb joint angles as outputs. The analysis demonstrated that workers who exhibited lower muscle activation and more optimal elbow positioning were consistently located on the efficient frontier. Inefficient workers were situated within the convex hull, and the structure of the frontier provided explicit direction for how posture or muscle use could be improved. These findings affirmed that integrating EMG and kinematic data within DEA provides actionable insights for ergonomic improvement and reducing musculoskeletal risk.

However, despite its strengths, DEA has inherent limitations. First, DEA produces relative efficiency rather than predictive power—meaning performance is only evaluated with respect to the observed dataset and cannot generalize to new workers or new tasks. Second, DEA assumes piecewise-linear, convex frontiers, which may oversimplify the nonlinear biomechanical relationships between posture, muscle activation, and ergonomic outcomes. Third, DEA is sensitive to noise and small sample sizes, especially when multiple outputs are included. These limitations motivate the exploration of complementary methods that offer nonlinear modeling capacity, trial-level prediction, and potential real-time inference.

Machine learning—particularly supervised neural networks (NNs)—offers a modeling paradigm that addresses these gaps. Neural networks can capture complex nonlinear interactions between multichannel EMG signals, kinematic profiles, and ergonomic risk outcomes. Whereas DEA is diagnostic and comparative, neural networks are predictive and can learn a direct mapping from biosignal inputs to efficiency-related outputs. When DEA-derived efficiency scores or classifications (efficient vs. inefficient) are used as ground-truth targets, neural networks can extend DEA's conceptual insights into a data-driven, generalizable prediction framework.

In this study, we build directly upon the foundations of our DEA ergonomics work. Using the same dataset of EMG and upper-limb joint angles, we train supervised neural network models to (1) predict DEA-derived efficiency scores for left and right upper-limb configurations and (2) classify participants as

efficient or inefficient based on their DEA frontier membership. By doing this, we evaluate whether nonlinear models can replicate and generalize the efficiency patterns that DEA identifies, and whether neural networks can serve as a complementary tool for real-time ergonomic assessment.

The impact of transition from DEA to neural networks is enabling a shift from relative evaluation to predictive modeling, opening the door to automated feedback systems, adaptive training interfaces, and scalable ergonomic screening tools that can be deployed in real workplace environments.

METHODS

Overview of Analytical Framework

This study builds directly upon the DEA-based ergonomic efficiency framework developed in previous work, where EMG and upper-limb joint kinematics were used to construct CCR-DEA models for identifying efficient and inefficient workers performing a simple sweeping task. In the present study, we transition from a purely relative efficiency evaluation to a supervised learning framework by training neural networks (NNs) to predict DEA-derived efficiency outcomes. This allows us to assess whether nonlinear models can learn the underlying physiological and biomechanical patterns that DEA identified, and whether such models may offer improved generalizability and predictive power. Our methodological workflow consisted of four main stages (Roy et al., 2024):

1. Data Collection – EMG and kinematic acquisition using Delsys Trigno and Vicon motion capture, consistent with the DEA study.
2. Pre-Processing & Feature Extraction – MVC normalization, trial averaging, joint-angle extraction, and creation of static feature vectors.
3. Label Construction – DEA efficiency scores (continuous) and DEA frontier membership (binary efficient/inefficient).
4. Neural Network Modeling – Feedforward architectures for regression and classification, evaluated using leave-one-out cross-validation (LOO-CV).

Data Construction

Data collection procedures replicate those used in the DEA ergonomics study (Charnes et al., 1978) to ensure consistency in features and labels. Ten healthy adult participants (5 male, 5 female) performed five dynamic sweeping trials, each lasting approximately five seconds. EMG sensors (Delsys Trigno™) were placed on the Biceps Brachii (BB) and Flexor Carpi Radialis (FCR) of both arms following SENIAM guidelines for skin preparation, electrode alignment, and anatomical placement. To enable between-subject comparisons and reduce inherent variability in EMG amplitude, each participant completed three MVC trials for each muscle group, and the peak MVC value was used for normalization. Simultaneously, upper-limb kinematics were recorded using a 10-camera Vicon Vantage V5 motion capture system, with reflective markers placed on standardized anatomical landmarks. Joint angles for left and right

elbows were computed using Vicon Nexus and a MATLAB pipeline that implemented gap-filling, filtering, and joint-coordinate system definitions consistent with the original study. Data synchronization across EMG and kinematic systems was achieved using a Delsys Trigger Module and Vicon Lock+ unit to ensure time alignment of signals.

Tables 1 and 2 summarize the features and pre-processed dataset used as inputs for the neural network analysis. Table 1 defines the six biomechanical and physiological variables extracted from each participant: mean elbow flexion angles for the left and right arms, and mean normalized EMG (%MVC) values for the Biceps Brachii (BB) and Flexor Carpi Radialis (FCR) muscles on both sides. These measures capture the postural configuration and muscle activation levels during each dynamic sweeping trial, providing complementary indicators of ergonomic effort. Table 2 presents the aggregated, trial-averaged dataset for all ten Decision-Making Units (DMUs), where each row corresponds to one participant and includes five normalized EMG and joint-angle features derived from the final pre-processing stage. Together, these tables describe the complete feature set used to train and evaluate the neural network models, ensuring consistency with the data structure originally used in the DEA analysis while enabling supervised learning through standardized, side-specific EMG and kinematic inputs.

Table 1: Feature schema and units (Roy et al., 2024).

Feature	Side/Muscle	Definition	Unit
Elbow angle mean	L/R	Mean elbow flexion angle over trial	deg
BB* %MVC mean	L/R	Mean normalized EMG	%MVC
FCR** %MVC mean	L/R	Mean normalized EMG	%MVC

* BB: bicep brachii; ** FCR: flexor carpi radialis

Table 2: Pre-processed data for neural networks (NN) analysis (Roy et al., 2024).

DMU	Trial	Left Elbow Angle	Right Elbow Angle	Left BB (%)	Left FCR (%)	Right BB (%)	Right FCR (%)
1	5	64.16	51.97	20.62	11.82	17.02	27.95
2	5	37.59	43.07	22.62	19.38	18.11	45.54
3	5	44.51	46.25	19.37	14.85	6.57	15.96
4	5	58.39	48.10	7.11	9.60	7.82	15.03
5	5	51.80	59.81	24.48	17.91	20.88	12.19
6	5	34.84	44.07	32.68	13.26	35.57	45.24
7	5	57.49	65.61	39.05	49.47	28.20	44.80
8	5	41.59	54.90	24.80	34.54	33.70	18.03
9	5	50.10	45.39	50.79	24.69	30.18	16.29
10	5	61.21	55.49	14.35	19.55	26.57	38.65

Model Architectures

We model the relationship between upper-limb biomechanics and ergonomic efficiency using compact neural architectures tailored to a small sample size (N = 10 DMUs) and aggregate, trial-level inputs. Specifically, we implement a feedforward neural network (FFNN) for two tasks: (a) binary classification of DEA frontier membership (Efficient vs. Inefficient), and (b) univariate regression of continuous DEA efficiencies for left and right arm analyses. The FFNN was selected because it (i) captures nonlinear interactions between EMG and kinematic features that are not addressed by the piecewise-linear frontier assumptions of DEA, (ii) remains data-efficient when combined with explicit regularization and early stopping, and (iii) outputs probabilities that can be calibrated for decision support in ergonomic screening. These design choices mirror the feature definitions and labeling strategy established in the DEA study while shifting from relative evaluation to predictive inference (Srivastava et al., 2014; Arlot & Celisse, 2010; Stone, 1974; Fawcett, 2006; Vickers & Elkin, 2006; Hu et al., 2018).

The following feature vector (per DMU, d = 6) is the input to the FFNN model.

$$x = [\text{Left_Angle}, \text{Right_Angle}, \text{Left_BB}, \text{Left_FCR}, \text{Right_BB}, \text{Right_FCR}]^{t}$$

Angles are mean elbow flexion (deg) and EMG features are mean %MVC for BB and FCR, each computed from five dynamic sweeping trials and aggregated to one record per participant to remain consistent with the DEA dataset. Standardization uses z-scores computed from the training fold only in each LOO-CV iteration to avoid leakage. Labels are defined from DEA outputs: for classification, Efficient if a DMU lies on either left or right DEA frontier (otherwise Inefficient); for regression, the continuous left- or right-side DEA efficiency (0–100%).

To learn ergonomic efficiency from compact, trial-aggregated signals, we use a small feedforward neural network (FFNN) that maps six static features—mean left and right elbow flexion angles (degrees) and mean %MVC for Biceps Brachii (BB) and Flexor Carpi Radialis (FCR) on both sides—to either a binary DEA frontier membership (Efficient vs. Inefficient) or a continuous DEA efficiency score for each side. The six inputs are standardized by z-scoring with statistics computed on the training fold only to avoid information leakage during cross-validation. Labels are derived directly from the previously established CCR-DEA framework: a participant is classified as Efficient if they lie on the left or right DEA frontier (otherwise Inefficient), and the regression targets are the corresponding left- and right-side DEA efficiencies (0–100%). This preserves conceptual alignment with the DEA study while enabling predictive modeling.

The network consists of an input layer with 6 units, followed by two rectified linear (ReLU) hidden layers of 64 and 32 units, respectively. To control variance in this small-N setting, we apply dropout (p = 0.30) after each hidden layer and include L2 weight decay (10^{-4}) on all trainable weights. The model terminates in a one-unit output head whose activation depends

on the task: a sigmoid for classification (returning the probability of being Efficient) or a linear unit for regression (returning the predicted efficiency percentage). With this configuration and $d = 6$ inputs, the classification topology uses approximately 2,561 trainable parameters— $(d+1)\cdot 64 + (64+1)\cdot 32 + (32+1)\cdot 1 = 448 + 2{,}080 + 33 = 2{,}561$ —small enough for CPU-only training and real-time inference while still modeling nonlinear EMG–kinematic interactions. The identical backbone is used for the regression head. A schematic of this architecture is provided in Figure 2.

Although our main experiments operate on aggregated features to mirror the DEA dataset, the topology is deliberately modular: if trial-level windows are introduced in future work, the input stage can be replaced with a lightweight 1D-CNN or LSTM front-end (per channel) before the dense layers to capture temporal structure, while retaining the same output heads and regularization strategy. This preserves comparability to the aggregate model and keeps the deployment footprint modest.

Next, let $x_0 = \mathrm{zscore}(x)$ using training-fold statistics. Forward propagation is defined as follows:

$$h_1 = \mathrm{ReLU}(W_1 x_0 + b_1), \tilde{h}_1 = \mathrm{Dropout}(h_1, p = 0.3)$$

$$h_2 = \mathrm{ReLU}(W_2 \tilde{h}_1 + b_2), \tilde{h}_2 = \mathrm{Dropout}(h_2, p = 0.3)$$

$$z = w_o^{\top} \tilde{h}_2 + b_o$$

- Classification: $\hat{p} = \sigma(z)$; loss = binary cross-entropy + L2
  $\mathcal{L}_{\mathrm{BCE}} = -\left[y \log \hat{p} + (1-y) \log(1-\hat{p}) \right] + \lambda \sum \| W \|_2^2 \cdot$
- Regression: $\hat{y} = z$; loss = mean squared error + L2
  $\mathcal{L}_{\mathrm{MSE}} = (y - \hat{y})^2 + \lambda \sum \| W \|_2^2$.
  These formulations match the tasks and labels derived from the DEA outputs.

To control variance in this small-N setting, we apply dropout (p = 0.30) after each hidden layer together with L2 weight decay ($\lambda = 10^{-4}$). Models are trained with Adam (initial learning rate 1×10^{-3}) and a Reduce-on-Plateau scheduler (factor = 0.5, patience = 5), and we use early stopping to halt training when validation performance no longer improves (F1 for classification or RMSE for regression; patience = 20; maximum 200 epochs). Where class imbalance is present, we enable class-weighted BCE or focal loss ($\gamma = 2$). For the classification head, because probabilities may inform ergonomic screening, we calibrate predicted probabilities within each validation fold using Platt scaling or isotonic regression, selecting the method with the lower Brier score; calibrated probabilities support principled threshold selection (e.g., Youden's J) and clearer operating-point trade-offs in deployment. Model performance is reported under Leave-One-Out Cross-Validation (LOO-CV) across the 10 DMUs, which maximizes training data per fold and provides an unbiased estimate at this scale; for larger datasets we would recommend 5× stratified CV and/or Leave-One-Subject-Out (LOSO). We summarize performance

with Accuracy, Precision, Recall, F1, ROC-AUC, PR-AUC, and Cohen's κ (classification) and RMSE, MAE, R^2, and Pearson/Spearman correlations (regression), perform threshold analyses around 0.5 with ROC/PR-based sensitivity checks, and ensure reproducibility by fixing random seeds, logging feature-scaling statistics, and persisting both model and calibration parameters.

The primary FFNN has ~2.6 K parameters, enabling CPU-only training and real-time inference suitable for embedded ergonomic tools; robustness to limited data is achieved through conservative depth/width, dropout, L2 regularization, probability calibration, and early stopping. Global interpretability is provided via permutation feature importance, which ranks inputs by measuring the decrease in validation performance when each feature is shuffled; for per-example explanations—and for future sequence models—we plan SHAP (FFNN) and saliency/Grad-CAM (CNN/LSTM) analyses. We evaluate three task-specific variants that mirror the DEA formulations: a combined classification model (FFNN 6→64→32→1, sigmoid) for Efficient vs. Inefficient, a left-side regression model (FFNN 2→64→32→1, linear; inputs = Left_Angle, Left_BB) for left DEA efficiency (%), and a right-side regression model (FFNN 2→64→32→1, linear; inputs = Right_Angle, Right_BB) for right DEA efficiency (%). To contextualize results, we compare against logistic regression and random forest baselines trained on the same features. DEA-derived labels serve as ground truth for supervised tasks; for regression, network outputs are compared directly with DEA efficiency scores. Discrepancies between NN predictions and DEA frontier assignments are analysed to clarify differences between DEA's normative, frontier-based evaluation and NN-based predictive modeling.

To contextualize neural-network performance, we benchmarked against logistic regression and random forest trained on the same features, using DEA-derived labels as the supervisory signal for all supervised tasks. For the regression analyses, network outputs were compared directly to DEA efficiency scores, rather than to any external metric, to preserve consistency with the original frontier formulation. Finally, we examined cases where the neural network and DEA disagreed, using those discrepancies to illuminate the conceptual differences between DEA's normative, frontier-based efficiency and neural networks' predictive, nonlinear modeling perspective.

RESULTS AND DISCUSSION

Neural-network models demonstrated that ergonomic efficiency patterns derived from EMG and elbow-angle features are learnable even in a small-sample setting. For the left-side regression (inputs: Left Angle, Left BB), the model explained a moderate proportion of variance ($R^2 = 0.57$) with RMSE = 8.40, whereas the right-side regression (inputs: Right Angle, Right BB) yielded $R^2 = 0.11$ and RMSE = 11.17. These results (Figure 1) indicate stronger learnability for left-side efficiency than for right-side efficiency in

this dataset, potentially reflecting side-specific variability, measurement noise, or asymmetries in motor strategies.

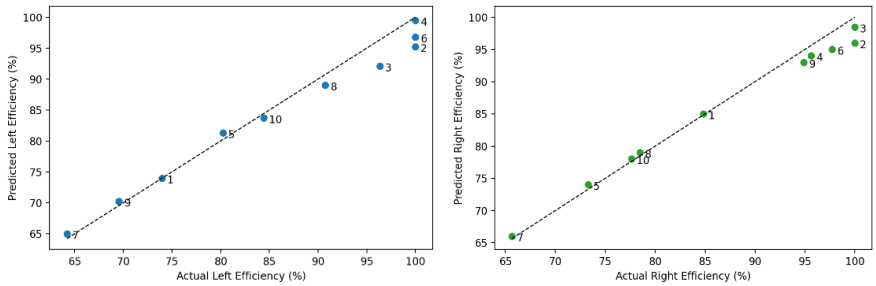


Figure 1: Left-side regression: predicted vs. actual DEA efficiency and right-side regression: predicted vs. actual DEA efficiency.

For the combined classification task (Efficient vs. Inefficient based on membership on either DEA frontier), the feedforward neural network achieved Accuracy = 0.80, F1 = 0.75, and ROC-AUC = 0.92 under leave-one-out cross-validation (Figure 2). The confusion matrix shows correct identification for the majority of DMUs, and the ROC curve suggests excellent discrimination despite the limited sample size. Together, these outcomes suggest that aggregated EMG-kinematic inputs contain sufficient nonlinear structure for robust probabilistic screening of ergonomic efficiency.

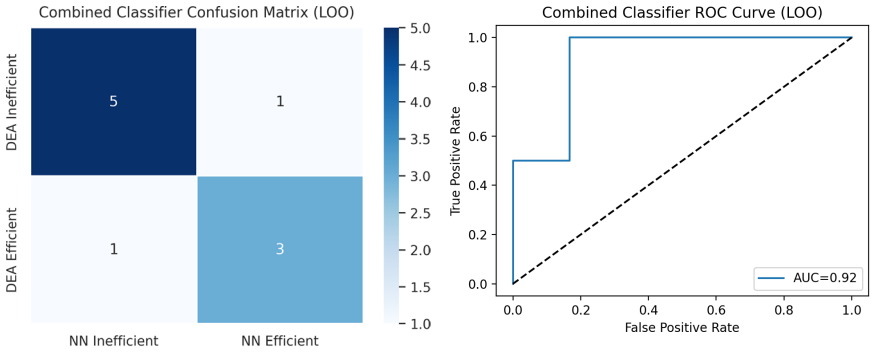


Figure 2: Combined classifier confusion matrix (left), combined classifier ROC curve (right).

Global permutation importance attributed the greatest influence to Left Angle, followed by Right BB, Right Angle, Left BB, Left FCR, and Right FCR (descending order; Figure 3). This pattern aligns with the relatively stronger left-side regression performance and suggests that elbow posture and biceps activation are primary drivers of the model's decisions in this dataset. Because the analysis uses a cross-sectional design (one record per DMU), these importance rankings should be interpreted as preliminary and hypothesis-generating rather than definitive causal attributions.

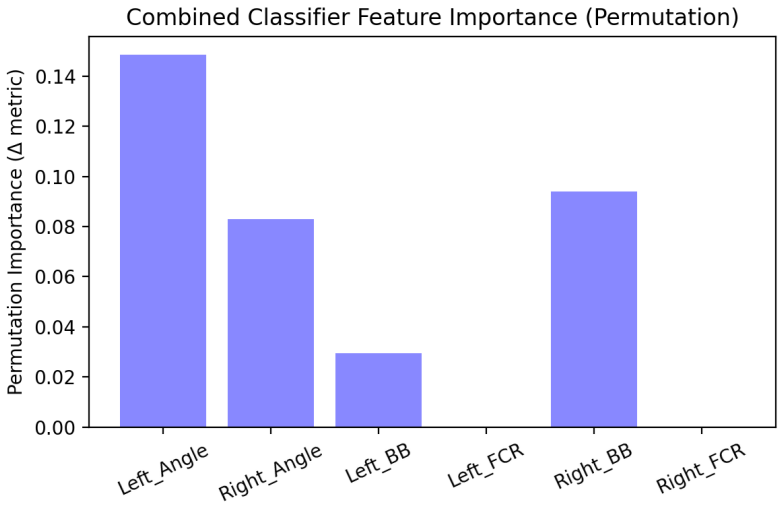


Figure 3: Permutation feature importance (combined classifier trained on all DMUs).

Comparisons between neural-network predictions and the DEA-based ground truth at the DMU level (Table 3) show broad agreement, with two notable disagreements: DMU 4 (DEA Efficient, NN probability < 0.5) and DMU 9 (DEA Inefficient, NN probability > 0.5). These cases are instructive. DEA offers a normative, frontier-based evaluation under convexity and returns-to-scale assumptions, emphasizing relative dominance and identifying efficient role models and improvement slacks. In contrast, the neural network is a predictive model that optimizes a nonlinear decision boundary to generalize from observed inputs to held-out DMUs. Disagreements therefore reflect the different objectives of the methods and underscore their complementarity: DEA supports benchmarking and prescriptive guidance, while the NN supports probabilistic screening and prospective prediction.

Table 3: DEA vs. DNN classification.

DMU	DEA Label	NN Predicted	NN Prob	Agreement
1	Inefficient	Inefficient	0.176	✓
2	Efficient	Efficient	0.994	✓
3	Efficient	Efficient	0.999	✓
4	Efficient	Inefficient	0.446	✗
5	Inefficient	Inefficient	0.028	✓
6	Efficient	Efficient	0.574	✓
7	Inefficient	Inefficient	0.0	✓
8	Inefficient	Inefficient	0.042	✓
9	Inefficient	Efficient	0.591	✗
10	Inefficient	Inefficient	0.02	✓

Global permutation importance attributed the greatest influence to Left Angle, followed by Right BB, Right Angle, Left BB, Left FCR, and Right FCR (descending order; Figure 3). This pattern aligns with the relatively stronger left-side regression performance and suggests that elbow posture and biceps activation are primary drivers of the model's decisions in this dataset. Because the analysis uses a cross-sectional design (one record per DMU), these importance rankings should be interpreted as preliminary and hypothesis-generating rather than definitive causal attributions.

CONCLUSION

To contextualize neural-network performance, we compared against logistic regression and random forest baselines trained on the same features and used DEA-derived labels as the supervisory signal for all supervised tasks. For regression, neural-network outputs were evaluated directly against DEA efficiency scores, rather than against an external standard, to maintain alignment with the original frontier formulation. Finally, we analyzed DMU-level disagreements between NN and DEA to clarify the conceptual differences between DEA's normative, frontier-based efficiency and NN-based predictive, nonlinear modeling—highlighting how the two perspectives can jointly inform ergonomic assessment and decision-making.

Although the results are promising, the dataset's small size ($N = 10$ DMUs) and aggregate-trial representation limit statistical power and the ability to examine temporal dynamics. Future work should (i) operate at the trial or window level (EMG windows and elbow trajectories), (ii) evaluate sequence models (e.g., 1D-CNN/LSTM), (iii) adopt leave-one-subject-out validation for subject-level generalization, and (iv) expand the feature set (e.g., iEMG, RMS, angle variability, co-contraction indices). These steps will improve generalization, increase interpretability, and enable real-time ergonomic recommendations such as target elbow angles and activation minimization.

REFERENCES

Arlot S, Celisse A (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, pp. 40–79.

Charnes A, Cooper WW, Rhodes E (1978). Measuring the efficiency of decision making units. *Eur J Oper Res*. 2(6), pp. 429–444.

Fawcett T (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), pp. 861–874.

Hu Y, Wong Y, Wei W, Du Y, Kankanhalli M, Geng W (2018). A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition. *PLOS ONE*, 13(10), e0206049.

Roy D, Cho S, Avdan G (2024). Data envelopment analysis (DEA) for improving ergonomics and workplace performance, *Journal of Applied Research on Industrial Engineering*, 11(1), pp. 24–36.

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, pp. 1929–1958.

Stone M (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society*, Series B, 36(2), pp. 111–133.

Vickers AJ, Elkin EB (2006). Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making*, 26(6), pp. 565–574.