Machine learning methods for electromyography error detection in field research: An application in full-shift field assessment of shoulder muscle activity in apple harvesting workers

## Abstract

This study presented an alternative technique for processing electromyography (EMG) data with sporadic errors due to challenges associated with the field collection of EMG data. The application of this technique was used to detect errors, clean and optimize EMG data in order characterize and compare shoulder muscular load in farmworkers during apple harvesting in a trellised orchard. Surface EMG was used to take measurements from twenty-four participants in an actual field work environment. Under challenging field settings, data from twenty participants were retained. Anomalies in the EMG data were detected and removed with a customized algorithm using principal component analysis, interquartile range cut-off and unsupervised cluster analysis. This study found significantly greater upper trapezius muscle activity in farmworkers who used a ladder as compared to the alternative platform-based method where a team of mobile platform workers harvested apples from the tree tops and a second separate team of ground workers harvested apples from the tree bottoms. By comparing the unprocessed and the processed, anomaly-free EMG data, the robustness of our proposed method was demonstrated.

**Keywords:** anomaly detection; electromyography; horticulture; k-means clustering; principle component analysis

## 1. Introduction

**1.1 Horticultural industry context**

It has been nearly three decades since Latino migrant farmworkers have been seasonally employed to fill the gaps in agricultural labour shortage in the United States (Blank 1998). Musculoskeletal pains are highly prevalent in this population (Kearney et al., 2016), particularly among women (Ramos et al., 2020), with one common source repetitive and awkward work postures (Xiao et al., 2013), and most commonly during the harvesting season when the job demands are high.

Horticulture in North America is moving towards trellised architecture, where the growth of the trees is controlled to facilitate the use of semi-automated, mobile orchard platforms. Such platforms can semi-autonomously move and transport workers through tree rows while they perform their regular work activities such as thinning, pruning and harvesting fruits. The use of the platforms is expected to reduce risk of musculoskeletal and fall injuries, which are common in this type of work (Fathallah 2010), and to enable more diverse populations to work in the orchards. However, this technology might also introduce other unknown musculoskeletal risk factors due to the work repetition and prolonged static work postures, especially in the shoulders due to the work often being above shoulder level.

**1.2 Electromyography field data collection**

Surface electromyography (EMG) is a non-invasive technique for measuring electrical activity resulting from muscle contraction and relaxation. The amplitude of EMG can be used as an estimator of muscular load. Surface EMG has been shown to have utility for estimating muscular loads during work. For example, Freivalds et al. (2006) simulated the task of carrying a heavy fruit bucket while reaching over and down to pick fruits in laboratory environment, and measured upper trapezius EMG. Meyer and Radwin (2007) also used surface EMG to measure muscle activity in trapezius as well as in erector spinae during simulated harvesting tasks on the ground, and made a comparison between stoop and prone postures. Additionally, Jin et al. (2009) used similar methods to measure low back muscle activity during simulated harvesting tasks in kneeling, squatting, and stooping postures in laboratory.

Despite the prior successful use in more controlled settings, the use of surface EMG in real, outdoor agricultural environments is challenging; due to the fact that equipment including data logger, electrodes and cables may interfere with tasks and/or catch on equipment or plants; and that perspiration can lead to the electrodes slipping, losing skin contact or falling off. Thus, to date, there has been only a few outdoor, agricultural field studies where EMG has been successfully used. Among those studies, Douphrate et al. (2017) used surface EMG to evaluate muscle activity in upper trapezius, anterior deltoid, forearm flexors and forearm extensors in order to compare different designs of large-scale cow milking operations. In another observational study, surface EMG was used for estimating muscle activity in multiple muscle groups during various agricultural tasks (Fethke et al., 2020). Lastly, farmers’ erector spinae EMG were measured to evaluate muscle activity changes by the use of assistive device (Thamsuwan et al., 2020a).

**1.3 Electromyography signal processing**

The basic EMG signal processing in ergonomics include the signal amplitudes, describing muscle activity with relationship with force, and the spectral parameters, used for detecting frequency changes in muscle activity that are thought to be associated with muscle fatigue. In occupational ergonomics, EMG amplitudes are derived for the percentiles of the Amplitude Probability Distribution Frequency (APDF) to define static, median and peak muscle activities (Jonsson 1982). Also, a shift in mean or median power frequency towards lower frequency have been thought to be associated with muscle fatigue in the workplace (Herberts et al., 1980). Established EMG parameters may be used to identify and remove anomalies in EMG data beyond typical errors known to be associated EMG instrumentation, ambient noise, and/or motion artifacts.

Artificial intelligence could enable error detection and removal in EMG signals; thus, EMG data that may have been previously thought to be unusable could be retained further analysis. Various machine learning techniques have previously been developed for detecting gestures (Cheron et al., 1996; Mukaeda and Shima, 2017; Jaramillo and Benalcazar, 2017; Baghdadi et al., 2018; Anil and Sreletha, 2018); thus, it may be possible to apply such techniques to also detect and remove EMG errors. Cheron et al. (1996) applied dynamic recurrent neural network to establish relationship between EMG and arm kinematics, and obtained a classification accuracy of 84%. Mukaeda and Shima (2017) developed a new hidden Markov model to recognize hand gesture patterns based on EMG data and yielded 91% accuracy. Baghdadi et al. (2018) used support vector machine to differentiate fatigued and non-fatigued EMG with the accuracy of 90%. Still, those previously-mentioned techniques are supervised learning, which require data labels of correct gestures. In other words, if the algorithms were to be applied to error detection, they would need to know in advance whether the data point is an error or not.

Clustering, as used in unsupervised machine learning, on the other hand, could be applied in case of unknown anomalies. K-means and hierarchical clustering techniques were used for detecting muscle fatigue in EMG signals and for selecting car seats (Atieh et al., 2005). Still, there are several EMG parameters in both time and frequency domains, i.e. muscle activity levels associated with force estimation and median frequencies of muscle activity associated with muscle fatigue, where anomalies in these parameters could alter the clustering. Therefore, principal component analysis (PCA) should be also used for dimensionality reduction to minimize the influence of outliers in one dimension. Chu et al. (2006) proposed the extension of PCA to EMG data processing for hand pattern recognition. In addition, Ijaz and Choi (2018) applied PCA together with hierarchical clustering for anomaly detection of EMG signals in gait analysis.

**1.4 Objectives of this study**

Given the feasibility and challenges of using surface EMG technique in field settings, this study attempted to, firstly, apply existing EMG and data analysis methods to measure and characterize the upper trapezius muscle activity of farmworkers while harvesting apples with ladders and mobile orchard platforms. Then this study aimed to develop a machine learning algorithm to detect and remove sporadic errors in the collected EMG signals in order to enable a comparison across harvesting methods: (1.) harvesting apples from the full tree with a ladder, (2.) standing on the mobile orchard platforms and harvesting apples from the upper level of the trees, and (3.) harvesting apples from the lower part of the trees without using a ladder or a platform.

## 2. Methods

### 2.1 Study site, participants and work tasks

This study was carried out in a trellised orchard inn Washington State. In trellised orchards, the trees are force or trained to grow a certain way, to facilitate tree grow and harvesting.

Twenty-four farmworkers from one trellised orchard were recruited to participate in this study. They were divided equally into three groups; that is, eight different farmworkers were assigned into each group. The first group used a ladder and harvesting bags with shoulder straps to pick apples from tree-top to tree-bottom (the full trees). They also carried loads of apples in their bag from the trees to large, centrally-located, apple bins placed between the tree rows. The second group stood on a mobile elevated orchard platform while harvesting apples from the upper level of the trees. The large apple bin resided on the floor in the middle of the platform. The third group were ground workers who only picked apples from the lower level of the trees; their work did not require the use of ladders or platforms. Similar to the ladder group, they used the harvesting bags with shoulder straps to carry the apples from the trees to the larger, centrally located, apple bins. This last group of workers is denoted as ‘ground’ workers.

The study protocols were approved by the university’s Human Subject’s Division Institutional Review Board. All the participants provided informed consent to participate in the study.

### 2.2 Instrumentation and protocol

EMG in right and left upper trapezius muscles was collected at 1,000 Hz for an entire 8-hour work day with a 30-minute break. As shown in Figure 1, differential electrode pairs were placed one centimetre distally from the mid points between the spine level C7 and the acromions. The bi-polar electrodes with a 1 centimetre diameter were placed next to each other with the inter-electrode distance of two centimetres according to the recommendation by Jensen et al. (1993). Ground electrodes were placed over the bone and tendons of the acromions.

[Insert Figure 1 here]

Prior to attaching the electrode, the skin-electrode interfaces were prepared by cleaning the skin with alcohol wipe to remove dead cells and applying tincture of benzoin, i.e., a sweat-resistant adhesive, over the area where the adhesive portion of the electrodes would contact the skin. There was no adhesive over the area where the circular active electrode gel interface contacted the skin. Single-use disposable pre-gelled electrodes (*Blue Sensor N; Ambu; Ballerup, Denmark*) were attached and secured to the participants with breathable medical tape (Transpore™; 3M, St. Paul; MN), and via wires with EMG pre-amplifiers, the electrodes were connected to a battery-powered portable data logger (*Biomonitor ME6000; Mega Electronics Ltd.; Kuopio, Finland*). The data were recorded to 2MB compact flash memory card in the data logger. The participants used a waist strap containing a small pocket to carry the data logger with them over the course of their work day.

Since individual factors such as skin impedance, adipose tissue and the cross-sectional area of muscle fibres can impact the measured EMG voltages, the EMG signals had to be normalized, i.e., calibrated to a reference. Three submaximal reference voluntary contraction (RVC) trials were performed by each participant at the beginning of their work day. During the RVC, the participants stood upright, held their arms straight out in front of their body and parallel to the ground and carried a 0.91-kg dumbbell in each hand. The RVC trials lasted for 30 seconds with at least 5-second breaks between each trial.

An RVC activity was chosen in this study instead of maximal voluntary contraction (MVC) in order to minimize the risk of shoulder injuries prior to the data collection. In addition, the researchers did not use MVC because they could have caused fatigue in the participants (Oberg 1995; Minning et al., 2007) and affected the subsequent full-shift measurements. In contrast, other researchers believe RVCs have greater reliability and are more repeatable (Attebrant et al., 1995; Jackson et al., 2009). Like the RVC proponents, our RVC load was uniform across all participants. In other words, all the workers’ muscular loads were calibrated to a similar and specific muscular demand, rather than against their own muscular capability like with the MVC, which can vary depending on individual.

### 2.3 Data processing

First, raw EMG signals were filtered with a second-order Butterworth dual-pass band-pass between 10-350 Hz. The filtered signals were partitioned into 125-millisecond overlapping windows of which root-mean-square (RMS) amplitudes were calculated as Equation 1. Then EMG data were summarized for each one second with the following parameters:

1. mean power frequency (MNF in Equation 2)
2. median power frequency (MDF in Equation 3)
3. mean of RMS amplitude
4. standard deviation of RMS amplitude
5. skewness of RMS amplitude
6. kurtosis of RMS amplitude
7. 0th percentile (or minimum) of RMS amplitude
8. 1st percentile of RMS amplitude
9. 2.5th percentile of RMS amplitude
10. 5th percentile of RMS amplitude
11. 10th percentile of RMS amplitude
12. 25th percentile (or 1st quartile) of RMS amplitude
13. 50th percentile (or median) of RMS amplitude
14. 75th percentile (or 3rd quartile) of RMS amplitude
15. 90th percentile of RMS amplitude
16. 95th percentile of RMS amplitude
17. 97.5th percentile of RMS amplitude
18. 99th percentile of RMS amplitude
19. 100th percentile (or maximum) of RMS amplitude

Equation 1:

where S is the window length and f(s) is the data at each time point s.

Equation 2:

Equation 3:

where is the frequency value at frequency bin j, is the EMG power spectrum at frequency bin j, and M is the length of the frequency bin.

The nineteen parameters obtained through the data pre-processing were considered to be input for detecting various types of anomalies, most notably the loss of contact between skins and surface electrodes. PCA were conducted muscle by muscle to reduce the dimension of input features from 19 to k. The first three principal components, i.e. the uncorrelated linear combinations of the 19 initial components, corresponding to highest eigenvalues could explain over 88-94% of the variances in both sides of upper trapezius in all participants; thus, they were used for identifying the potential EMG anomalies.

Two different techniques were employed for identifying anomalies. One was to consider data points that were outliers, i.e. smaller than Q1 – 1.5\*IQR or larger than Q3 + 1.5\*IQR, of any three principal components. The other way was to use k-means clustering with k = 2, 3, or 4 to separate errors from good data. EMG errors were expected to be in cluster(s) other than the cluster of good data. Then visualization of the clusters were also used for verification. An example of anomalies detected using the two techniques are shown in Figure 2. Once possible errors were detected and verified, the specific portions of data corresponding to these apparent EMG errors were removed.

[Insert Figure 2 here]

Finally, the remaining EMG data without the errors/anomalies were as analyzed using multiple 10-minute intervals throughout the participants’ work period. The static, median, and peak muscle activity were extracted from each interval as the 10th, 50th and 90th percentile of the RMS amplitudes, respectively. These three parameters were divided by the mean of RMS amplitude during the RVC. The median values among the three RVC were used as the denominators. That is, the 10th, 50th and 90th percentile of the RMS amplitudes were normalized as %RVC.

### 2.4 Statistical analysis

In this study, the effect of harvesting method were to be tested in *non-dominant* and *dominant* upper trapezius. That is, the dependent variables were static, median and peak muscle activity, and the independent variables included the fixed effects of work method (*Ladder, Platform, Ground*), muscle side, and work time, and the random effect of the subjects.

The muscle activity parameters were not normally distributed and when log-transformed they were still were not normally distributed, according to the Shapiro-Wilk tests, and had non-uniform variances; therefore, the data did not meet the assumptions for using ANOVA methods. Since we wanted to include subject as a random effect, non-parametric tests like Wilcoxon signed-rank test and Kruskal-Wallis test could not be used.

Instead, it was deemed most appropriate to use generalized additive mixed models (Wood 2004, 2006a, 2006b). This method extended from the generalized linear mixed model, which considers random effects and does not restrict to only normally-distributed response variables. At the same time, it allows the relationship between the predictors and response variable to be non-linear. Since the dependent variables, i.e., the 10th, 50th and 90th percentile of the RMS amplitudes, were originally skewed right, they were log-transformed to obtain heavy-tailed distributions that otherwise could have been modelled as Gaussian. Then the log-transformed data were modelled with the scaled t distributions.

To further investigate the effect of time on muscle activity in each side of upper trapezius and each harvesting method, non-parametric Friedman rank sum tests for repeated measures in a complete block design were conducted. In each harvesting method and muscle, subjects were the same and, therefore, served as a blocking factor.

All the statistical analysis was conducted using R programming language. The tests were considered statistically significant at p < 0.05.

## 3. Results

Due to instrumental errors during the field data acquisition, EMG data were dropped from three ladder workers and from one ground worker. The statistical analysis comparing the three groups indicated that there were no significant differences in worker demographics. The workers had an average age of 27.8 (SD = 2.6) years, weight of 77.1 (SD = 4.0) kg, height of 172.9 (SD = 2.7) cm and a BMI of 25.8 (SD = 1.6).

Muscle activity normalized as %RVC is presented in Figure 3. Generally, all the static, median and peak muscle activities had similar pattern where ladder workers had wider variation than the platform- and ground-based groups. In addition, Figure 3 suggested that the muscle activity in the dominant side of upper trapezius of ladder workers could be higher than ones in the non-dominant side, and ones in the other participant groups.

[Insert Figure 3 here]

Based on the generalized additive mixed models, harvesting method had a significant effect on the muscle activity. Specifically, with negative coefficients of model estimates as shown in Table 1-3, participants harvesting apples on the mobile elevated platform had less static, median and peak muscle activity than those using ladders. Moreover, workers who harvested apples from the lower level of trees without using a ladder or standing on the platform had even lower less static, median and peak muscle activity than those using ladders. Consider overlapping confidence intervals of the muscle activity estimates, the difference in muscle activity between the ground and platform workers were not statistically significant.

[Insert Table 1, Table 2, and Table 3 here]

Another significant factor affecting the muscle activity was the length of time into the work shift. The negative coefficients shown in Table 1-3 suggested that the muscle activity was not stable over time. When the data were separated and analyzed by harvesting method and muscle side, some different phenomena were observed. As shown in Figure 4, at the end of the work shift, there was a slight increase in muscle activity of the dominant upper trapezius among workers harvesting on the platform and no apparent change in the ground workers. During the same time period, there was a bilateral decrease in trapezius muscle activity in the ladder workers. Additionally, the ladder workers had higher variation, i.e. wider confidence intervals, of muscle activities.

[Insert Figure 4 here]

Regarding the time effect, follow-up analysis were also conducted separately for each muscle and each work method. According to the results from the Friedman test (Table 4), time of work had a significant effect on muscle activity in the ladder and ground workers but no significant effect across the platform workers. In addition, there were no difference in the trends of significance between the dominant and non-dominant sides. These results were all similar for static, median and peak muscle activity.

[Insert Table 4 here]

Finally, according to the comparison between processed and unprocessed data (Table 5), the new technique reduced the variability and allowed the inclusion of data from most participants in contrast to the unprocessed data where from many of the participants had to be excluded due to errors.

[Insert Table 5 here]

## 4. Discussions

### 4.1 Instrumentation, error detection and removal

This study measured EMG in a challenging environment. Perspiration and the interference of apple bags and ladders on the shoulder caused the EMG electrodes to come off participants’ skin. The issue of electrodes losing the contact to the skin was discovered at the end of work shift so it is important to analyze the EMG signals in the way that can find the time when this happened to be able to retain the rest of the data.

The traditional method in EMG error removal is a band-pass filter between 10 or 20 Hz and 350 Hz (De Luca, 1997), which can eliminate low-frequency movement artifacts and high-frequency noise. This technique does not always capture incidences when EMG electrodes are not in contact with participant’s skin or when physical damage or electrical short circuits develop in any components in the EMG data collection system. In previous studies (Duphrate et al., 2020; Thamsuwan et al., 2020a), raw EMG signals in time domain were visually inspected for any signal drift indicating the loss in electrode-skin interface, and the data from certain participants or tasks were removed. In the present study, if the same paradigm was to be applied, a great amount of data would have lost (46%). On a contrary, this study applied machine learning techniques to systematically detect errors and removed only half as much of what was thought to be unusable EMG data (21%).

To summarize, this study introduced the application of unsupervised machine learning, i.e. clustering, and PCA described in the method section to identify specific errors in all participants; thus, it eliminated the need to drop the entire data set from five participants. This results also reflect lower variation (IQR in Table 5) in general. The more usable data we have, the more accurate the comparison could be achieved.

### 4.2 Interpretation of findings, and implication to industry

These preliminary findings indicate that relative to ladder harvesting, these relatively new, mobile platforms may have the a potential to improve work conditions. This current study was a complementary to the previous studies that have already showed that the platforms reduce non-neutral work postures (Thamsuwan et al., 2020b) and repetitive motions (Thamsuwan et al., 2020c). The higher muscle activity was often the consequence of the more prolonged exertion posture and the more frequent work repetition, which were already shown. The results from this study confirmed the platform’s benefits; that is, muscular load demand in harvesting fruits with a ladder could be reduced by converting harvesting operations to mobile platforms for the tree-tops and ground harvesting for the tree-bottoms. This could ultimately lead to the reduction of the musculoskeletal risks known to contribute to shoulder injuries.

Additionally, as this study also found that time throughout the work shift significantly affected the farmworkers’ shoulder muscle activity especially in the ladder and ground workers (Table 4), orchard managers may consider this information when designing the schedule of work shift. In addition, from the results shown in Figure 4, muscle activity seemed to be stable during the first two and half hours of work. Afterwards, muscle activity measured from the ladder workers decreased; this could have been related to or a result of some form of physical and/or muscle fatigue.

### 4.3 Limitations and lessons learned

First and foremost, the ladder workers presented the greatest challenges with respect to our use of surface EMG. One work modification we requested of the ladder workers was to avoid their usual practice of carrying their ladder on their shoulder, in order to keep the electrodes intact. This requested adaptation could have changed the nature of their work. However, based on a time motion study, moving the ladder only accounted for only 7% of total work time (Zhang et al., 2019) and our own observations was the ladder movement was infrequent. Thus, the change in how the participants carried the ladder was expected to only marginally affect the muscular load.

Secondly, although all the participants were recruited from the same orchard and at the same time, one constraint for the selection of ladder workers was their experience harvesting with ladders. At the study site, only the platform was used in the season and several of the migrant workers never had experience working with ladders so they could only be assigned to the platform or ground group. Still, their differences in terms of demographic and anthropometric characteristics were not statistically significant.

Lastly, the presence of researchers in the field during regular work hour could have created a some for of a systematic bias by their presence. The participants might have worked differently from the way they normally did. However, for the purpose of comparison across different work method, this bias was anticipated to be similar across the participant groups.

## 5. Conclusion

Exposures to risk factors of work-related musculoskeletal disorders are unique in horticulture due to the diversity of work tasks; thus, the methodologies used to assess such physical or musculoskeletal risk factors may have to be customized and optimized. In this study, surface EMG was used to evaluate shoulder muscular load in the upper trapezius. Upon removing EMG anomalies due to challenges related to full-day, field measurements, EMG data suggested that static, median and peak muscle activity in workers using ladders were higher than alternative the ground and platform-based workers. This finding supports the adoption of the mobile elevated orchard platform as it could reduce the risk of harvesting-related, shoulder injuries that may be associated with the more strenuous ladder work.

## References

Anil, N. & Srelatha, S. H. (2018) EMG Based Gesture Recognition Using Machine Learning, 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1560-1564, doi: 10.1109/ICCONS.2018.8662987.

Atieh, M., Younes, R., Khalil, M., & Akdag., H. (2005) Classification of the Car Seats by Detecting the Muscular Fatigue in the EMG Signal. *International Journal of Computational Cognition*, 3 (4), 48-54. https://hal.archives-ouvertes.fr/hal-01170983

Attebrant, M., Mathiassen, S. E., & Winkel, J. (1995). Normalizing upper trapezius EMG amplitude: Comparison of ramp and constant force procedures. *Journal of Electromyography and Kinesiology*, 5(4), 245-250. https://doi.org/10.1016/1050-6411(95)00005-4

Baghdadadi, A., Megahed, F. M., Esfahani, E. T. & Cavuoto, L. A. (2018). A machine learning approach to detect changes in gait parameters following a fatiguing occupational task. *Ergonomics*, 61(8), 1116-1129. https://doi.org/10.1080/00140139.2018.1442936

Blank, S. C. (1998). *The end of agriculture in American portfolio*, Quorum Books, Westport, CT.

Cheron, G., Draye, J., Bourgeios, M., & Libert, G. (1996). A dynamic neural network identification of electromyography and arm trajectory relationship during complex movements. IEEE Transactions on Biomedical Engineering, 43, 552-558. https://doi.org/10.1109/10.488803

Chu, J.-U., Moon, I., & Mun, M.-S. (2006). A Real-Time EMG Pattern Recognition System Based on Linear-Nonlinear Feature Projection for a Multifunction Myoelectric Hand, *IEEE Transactions on Biomedical Engineering*, 53(11), 2232-2239. https://doi.org/10.1109/TBME.2006.883695.

De Luca, C. J. (1997). The use of surface electromyography in biomechanics. *Journal of Applied Biomechanics*, 13(2), 135-163. https://doi.org/10.1123/jab.13.2.135

Douphrate, D. I., Fethke, N. B., Nonnenmann, M. W., Rodriguez, A., Hagevoort, R., & Gimeno Ruiz de Porras, D. (2017). Full-shift and task-specific upper extremity muscle activity among US large-herd dairy parlour workers. *Ergonomics*, 60(8), 1042-1054. https://doi.org/10.1080/00140139.2016.1262464

Fathallah, F. A. (2010). Musculoskeletal disorders in labor-intensive agriculture. *Applied Ergonomics*, 41(6), 738-743. https://doi.org/10.1016/j.apergo.2010.03.003

Fethke, N. B., Shall Jr, M. C., Chen, H., Branch, C. A., & Merlino, L. A., (2020). Biomechanical factors during common agricultural activities: Results of on-farm exposure assessments using direct measurement methods. *Journal of Occupational and Environmental Hygiene*, 17(2-3), 85-96. https://doi.org/10.1080/15459624.2020.1717502

Freivalds, A., Park, S., Lee, C., Earle-Richardson, G., Mason, C., & May, J. J. (2006). Effect of belt/bucket interface in apple harvesting. *International Journal of Industrial Ergonomics*, 36(11), 1005-1010. https://doi.org/10.1016/j.ergon.2006.08.005

Herberts, P., Kadefors, R., Broman, H. (1980) Arm positioning in manual tasks: An electromyographic study of localized muscle fatigue. *Ergonomics*, 23(7), 655-665. https://doi.org/10.1080/00140138008924780

Ijaz, A., & Choi, J. (2018). Anomaly Detection of Electromyographic Signals. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, *26*(4), 770–779. https://doi.org/10.1109/TNSRE.2018.2813421

Jackson, J. A., Mathiassen, S. E., & Dempsey, P. G. 2009. Methodological variance associated with normalization of occupational upper trapezius EMG using sub-maximal reference contractions. *Journal of Electromyography and Kinesiology,* 19(3), 416-427. https://doi.org/10.1016/j.jelekin.2007.11.004

Jaramillo, A. G. & Benalcázar, M. E. (2017) Real-time hand gesture recognition with EMG using machine learning, IEEE Second Ecuador Technical Chapters Meeting (ETCM), pp. 1-5, doi: 10.1109/ETCM.2017.8247487.

Jensen, C., Vasselijen, O., & Westgaard, R. H. (1993). The influence of electrode position on bipolar surface electromyogram recordings of the upper trapezius muscle. *European Journal of Applied Physiology and Occupational Physiology*, 67(3), 266-273. https://doi.org/10.1007/BF00864227

Jin, S., McCulloch, R., & Mirka, G. A. (2009). Biomechanical evaluation of postures assumed when harvesting from bush crops. *International Journal of Industrial Ergonomics*, 39(2), 347-352. https://doi.org/10.1016/j.ergon.2008.07.005

Jonsson, B. (1982) Measurement and evaluation of local muscular strain on the shoulder during constrained work, *Journal of Human Ergology*, 11, 73-88. https://doi.org/10.11183/jhe1972.11.73

Kearney, G. D., Allen, D. L., Balanay, J., & Barry, P. (2016). A Descriptive Study of Body Pain and Work-Related Musculoskeletal Disorders Among Latino Farmworkers Working on Sweet Potato Farms in Eastern North Carolina. Journal of Agromedicine, 21, 234-243. https://doi.org/10.1080/1059924X.2016.1178613

Meyer, R. H., & Radwin, R. G. (2007). Comparison of stoop versus prone postures for a simulated agricultural harvesting task. Applied Ergonomics, 38(5), 549-555. https://doi.org/10.1016/j.apergo.2006.08.005

Minning, S., Eliot, C. A., Uhl, T. L., & Malone, T. R. (2007). EMG analysis of shoulder muscle fatigue during resisted isometric shoulder elevation. *Journal of Electromyography and Kinesiology*, 17(2), 153-159. https://doi.org/10.1016/j.jelekin.2007.11.004

Öberg, T., (1995). Muscle fatigue and calibration of EMG measurements. *Journal of Electromyography and Kinesiology*, 5(4), 239-243. https://doi.org/10.1016/1050-6411(96)85582-9

Ramos, A., Carvajal-Suarez, M., Trinidad, N., Michaud, T. L., Grimm, B., Levan, T. D., & Siahpush, M. (2020). A Cross-sectional Study of Gender-related Differences in Reporting Fatigue and Pain among Latino/A Migrant Farmworkers. Journal of Agromedicine, 25, 319-329. https://doi.org/10.1080/1059924X.2020.1713272

Thamsuwan, O., Milosavljevic, S., Srinivasan, D., & Trask, C. (2020a). Potential exoskeleton uses for reducing low back muscular activity during farm tasks. American Journal of Industrial Medicine, 63(11), 1017-1028. https://doi.org/10.1002/ajim.23180

Thamsuwan, O., Galvin, K., Tchong-French, M., Aulck, L., Boyle, L., Ching, R. P., McQuade, K. J., & Johnson, P. W. (2020b). Comparisons of physical exposure between workers harvesting apples on mobile orchard platforms and ladders, part 1: Back and upper arm postures. Applied Ergonomics, 89, 103193. https://doi.org/10.1016/j.apergo.2020.103193

Thamsuwan, O., Galvin, K., Tchong-French, M., Aulck, L., Boyle, L., Ching, R. P., McQuade, K. J., & Johnson, P. W. (2020c). Comparisons of physical exposure between workers harvesting apples on mobile orchard platforms and ladders, part 2: Repetitive upper arm motions. Applied Ergonomics, 89, 103192. https://doi.org/10.1016/j.apergo.2020.103192

Wood, S.N. (2004) Stable and efficient multiple smoothing parameter estimation for generalized additive models. Journal of the American Statistical Association. 99:673-686. https://doi.org/10.1198/016214504000000980

Wood, S.N. (2006a) Low rank scale invariant tensor product smooths for generalized additive mixed models. Biometrics 62(4):1025-1036 https://doi.org/10.1111/j.1541-0420.2006.00574.x

Wood S.N. (2006b) Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC Press.

Xiao, H., McCurdy, S. A., Stoecklin-Marois, M. T., Li, C. S., Schenker, M. B. (2013). Agricultural work and chronic musculoskeletal pain among Latino farm workers: the MICASA study. American Journal of Industrial Medicine, 56(2), 216-225. https://doi.org/10.1002/ajim.22118

Zhang, Z., Lu, R., Igathinathane, C. (2019). A time and motion study for evaluation of apple harvest processes with different harvest methods. *Transactions of the ASABE*, 63(6), 1957-1967. https://doi.org/10.13031/trans.14144

Table 1: Parameter estimates from the generalized additive mixed models where dependent variable (y) is the log-transformed of the 10th percentile of the RMS amplitude. Note that the intercept reflected the “Work method: Ladder”, “Muscle: Right/dominant” and Work Time at the first period.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t-value | p-value |
| (Intercept) | 4.891 | 0.071 | 68.880 | < 0.0001 \* |
| Work Method: Platform | - 0.213 | 0.059 | - 3.582 | 0.0003 \* |
| Work Method: Ground | - 0.326 | 0.073 | - 4.481 | < 0.0001 \* |
| Muscle: Left/non-dominant | - 0.131 | 0.051 | - 2.600 | 0.0093 |
| Work Time | - 0.00057 | 0.000242 | - 2.340 | 0.019 \* |

Table 2: Parameter estimates from the generalized additive mixed models where dependent variable (y) is the log-transformed of the 50th percentile of the RMS amplitude. Note that the intercept reflected the “Work method: Ladder”, “Muscle: Right/dominant” and Work Time at the first period.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t-value | p-value |
| (Intercept) | 5.397 | 0.071 | 76.172 | < 0.0001 \* |
| Work Method: Platform | - 0.215 | 0.059 | - 3.615 | 0.0002 \* |
| Work Method: Ground | - 0.330 | 0.073 | - 4.550 | < 0.0001 \* |
| Muscle: Left/non-dominant | - 0.061 | 0.050 | - 0.208 | 0.23 |
| Work Time | - 0.00064 | 0.00024 | - 2.648 | 0.0081 \* |

Table 3: Parameter estimates from the generalized additive mixed models where dependent variable (y) is the log-transformed of the 90th percentile of the RMS amplitude. Note that the intercept reflected the “Work method: Ladder”, “Muscle: Right/dominant” and Work Time at the first period.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t-value | p-value |
| (Intercept) | 5.879 | 0.071 | 82.533 | < 0.0001 \* |
| Work Method: Platform | - 0.222 | 0.059 | - 3.718 | 0.0002 \* |
| Work Method: Ground | - 0.323 | 0.073 | - 4.430 | < 0.0001 \* |
| Muscle: Left/non-dominant | - 0.005 | 0.050 | - 0.100 | 0.92 |
| Work Time | - 0.00065 | 0.00024 | - 2.669 | 0.0076 \* |

Table 4: Friedman rank sum tests of the 10th, 50th and 90th percentile for differences across the different work time segments where subjects were considered as block (random effect) and the results of each work method was tested separately.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Work |  | Non-dominant |  |  | Dominant |  |  |
| Method | Percentile | Chi-squared | df | p-value | Chi-squared | df | p-value |
| Ladder | 10th percentile | 64.117 | 35 | 0.0019 \* | 73.171 | 35 | 0.00017 \* |
| 50th percentile | 71.104 | 35 | 0.0030 \* | 73.626 | 35 | 0.00015 \* |
| 90th percentile | 69.351 | 35 | 0.00048 \* | 72.559 | 35 | 0.00020 \* |
| Platform | 10th percentile | 31.368 | 35 | 0.64 | 42.928 | 35 | 0.17 |
| 50th percentile | 36.524 | 35 | 0.40 | 42.396 | 35 | 0.18 |
| 90th percentile | 41.486 | 35 | 0.21 | 38.156 | 35 | 0.33 |
| Ground | 10th percentile | 27.766 | 17 | 0.048 \* | 43.216 | 17 | 0.00045 \* |
| 50th percentile | 27.895 | 17 | 0.046 \* | 48.468 | 17 | < 0.0001\* |
| 90th percentile | 30.690 | 17 | 0.022 \* | 49.485 | 17 | < 0.0001 \* |

Table 5 Median (interquartile range, IQR) comparing muscle activity (normalized as % RVC) and Median Power Frequency between the processed data (new method using principle component analysis and k-means clustering) and unprocessed data (dropping entire subjects) by harvesting method.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Work |  | Non-dominant | | Dominant | |
| Method | Percentile | Processed | Unprocessed | Processed | Unprocessed |
| Ladder | 10th percentile | 96 (348) | 98 (380) | 186 (264) | 215 (170) |
| 50th percentile | 167 (573) | 179 (668) | 167 (461) | 178 (528) |
| 90th percentile | 272 (939) | 281 (1030) | 535 (840) | 547 (932) |
| Median Power Frequency | 56.7 (7.1) | 56.7 (7.6) | 54.4 (5.4) | 54.4 (6.6) |
| Platform | 10th percentile | 84 (48) | 85 (89) | 82 (100) | 77 (123) |
| 50th percentile | 142 (123) | 148 (123) | 134 (160) | 134 (194) |
| 90th percentile | 247 (179) | 250 (171) | 217 (131) | 211 (236) |
| Median Power Frequency | 60.8 (11.0) | 59.3 (11.6) | 56.9 (11.9) | 56.8 (12.0) |
| Ground | 10th percentile | 71 (66) | 72 (76) | 85 (111) | 89 (96) |
| 50th percentile | 127 (97) | 135 (106) | 137 (110) | 147 (96) |
| 90th percentile | 228 (173) | 242 (240) | 217 (257) | 241 (274) |
| Median Power Frequency | 53.4 (7.4) | 54.9 (7.6) | 56.3 (7.3) | 57.4 (8.1) |