ISSN (Print) : 0972-4397
ISSN (Online) : 0974-0805
Journal of Ecophysiology and Occupational Health, *Vol 17(1&2), DOI : 10.18311/jeoh/2017/v17i1&2/17926 57-62, January- June 2017*

Data Mining Approach to Pre-Screening of Manual Material Handlers Based on Hand Grip Strength

Lokesh Kumar Sharma* and Joydeep Majumder

ICMR-National Institute of Occupational Health, Ahmedabad- 380016, Gujarat, India

Abstract

The general incidence of musculoskeletal disorders (MSD) is one the major causes of mobility and concerns about financial and physical losses. The more objective pre-employment selection and employee placement test methods can reduce the work related injuries and Employers in general prefer for post-offer pre-placement (POPP) testing as a tool for evaluating worker's physical proficiency towards essential functions within a job. This functions as an effective screening technique intended to identify vulnerable workers who may be at greater risk of developing future musculoskeletal injuries. In this study, data mining approaches to estimate the proper work group for a pre-defined specialized job is used. In the first phase, the cluster analysis for the grouping the young workers according to their physical parameters was applied and according to the characteristic of each cluster, particularly with increasing handgrip strength values, the clusters were coined as very poor, poor, fair, good and excellent. In the second phase, supervised learning methods are utilized to classify the worker based on the anthropometric variables.

Keywords: Bayesian Networks, Cluster Analysis, Data Mining, Handgrip Strength, K-Means, Multilayer Perceptron, Musculoskeletal Disorder Anthropometry, Supervised Learning

1 . Introduction

Manual material handling jobs are labour intensive jobs with indefinite work time schedule. Jobs like carpentry, agriculture, construction, porters, etc. involve heavy muscular loading, resulting in musculoskeletal disorders. BoLSta (2013) reports musculoskeletal disorders as a concern to several employers which accounts for around one third of all injuries and illnesses. A study conducted by ICMR has reported that the prevalence of MSDs in Delhi is 7.08%, in Dibrugarh 11.52% and in Jodhpur 9.53% (Sharma, 2012). The basic reason may be the fitment of a worker into the prevailing work situation without judging the capacity and physical pertinence of the worker. Although at some workplaces, employers in general settle for post-offer pre-placement (POPP) testing as a tool for evaluating worker's physical proficiency towards essential functions within a job. (Chaffin et al, (1978); Rosenblum et al, (2006). Jenkins et al. (2016) suggested that pre-employment physical capacity testing scores might be predictive of subsequent musculoskeletal injury in paramedics. This also functions as an effective screening technique intended to identify vulnerable workers who may be at greater risk of developing future musculoskeletal injuries as well as affecting the desired productivity from the work. Screening tests generally include evaluating worker's physical strength and flexibility before advancing to performance testing of the work-related essential functions. The pre - employment examination usually results in the Low risk of injury or disease, at risk of injury or disease, at high risk of injury or disease, Serra et al, (2007); Schaafsma et al., (2016).

**Author for correspondence*

Data mining techniques are applied in the various anthropometry applications such as body sizing, Lin et al.; (2008); Niu et al., (2010); Paquet et al., (2011) and Majumder and Sharma, (2015). Data mining is a mechanism to explore and explicate data as contained in the database and focuses on ascertaining new relationships that cannot be found using standard data retrieval tools. The popular techniques of data mining are descriptive such as cluster analysis and predictive modelling. Cluster analysis is an important research area of data mining; it is used to identify homogeneous groups of objects called clusters. Objects in a specific cluster share many characteristics, but are very dissimilar to objects not belonging to that cluster. The data mining approach in anthropometrics based on clustering method, shall guide in coining significant clusters and their subsequent archetypes. The benefits of clustering approach remains within the possibility of being applied in the detection of obvious groups, i.e., clusters based on different body types. The identified clusters and epitomes thus becomes significant variables while designing "better fit" clothes or aid in deriving human body models on which industrial designs can come up. The predictive modelling is a technique used to develop a model to relate a dependent variable with a set of independent variables. It is also considered as a supervised learning or classification task. The appropriate goal is to predict group membership of new records based on their characteristics (independent variables), Wilson et al., (2004).

To estimate the proper work group for a pre-defined specialized job, cluster analysis is an active technique in data mining. Components in a specific cluster share many similar characteristics, but can also have many disparate features not belonging to that cluster. Researchers have been using data mining techniques on static as well as dynamic anthropometrics for excavating sizing systems using stature and bust/waist circumference following principal component analysis, Zakaria et al., (2008); Sharma and Majumder, (2013); Majumder and Sharma, (2014). In the present study, the clusters attributes have been used with the presumption for understanding the strength limits of the workers, as an appropriate tool towards post-offer pre-placement testing on the workers' front. Once the data arranged in the respective groups, the supervised learning method was applied to train the model and predict the class for unknown data.

The rest of the paper is organized as follows. The methodology is presented in the section 2. The section 3 provides the result and discussion. Finally, the work is concluded in the **section 4.**

2. Methodology

Anthropometric variables from 347 young Indian men (17.4±3.0 years) which include height, weight, body mass index (BMI), body surface area (BSA), body fat percentage (BF), right second: fourth digit ratio (RH2D:4D), left second: fourth digit ratio (LH2D:4D), right handgrip strength (RHTG) and left handgrip strength (LHTG) and age were taken as attributes. Height was measured on a stadio meter. Weights of the subjects were measured on an electronic balance (Rossmax, Swiss Gmbh). The thickness of skin folds were measured with skin fold calliper (Holtain Ltd., Crymych, UK) at biceps, triceps, sub scapular and suprailiac sites, following methodology reported in Gite et al. (2009). Body mass index (BMI, kg/m^2), an indicator of content of body fat, is defined as the individual's body mass divided by the square of one's body height. BMI ranging between 18.5-25.0 is considered as normal. The value less than 18.5 denotes chronic under nutrition, BMI in the range of 25.0-30.0 is considered as overweight and score above 30 indicates obese. The body surface area (BSA) is the measured or calculated surface area of a human body. BSA is an indicator of metabolic mass than body weight because it is less affected by abnormal adipose mass. Body fat consists of both essential body fat and storage body fat. Body fat percentage of a person is the overall content of fat in the body, with reference one's body weight. This is a direct indicator of body composition, irrespective of body height or body weight of a person. It may be noted BMI is a measure derived from body heights and weights. Since the body composition differs under different conditions, BMI may not necessarily be a precise indicator of body fat. Anthropometric techniques are applied in determining body fat from the measurement of double fold thicknesses of skin (subcutaneous adipose tissue) of different parts of body, and applying prediction equation to arrive at body density (Durnin and Womersley, 1974) and body fat (Siri, 1961), from skinfold thicknesses of bicep, tricep, subscapular and suprailiac sites. Lean body mass (LBM) is the measure of the mass of the body minus the absolute body fat. Although, the limitation of determining body fat from skinfold thicknesses is that since two individuals having similar skinfold thickness might differ in their body fat content due to differences in deposits of fat in the

abdominal cavity. It is also likely that skinfold measurement might confuse in estimating body fat for elderly and children due to variation in skin elasticity. The right hand impression was outlined on a white paper and 2D and 4D length were measured. The inter-digital web-spaces on the second and fourth digits were considered as the base line. The distal most part of each digit on the outline was considered as tip line of the finger. The 2D length was measured as the length between the midpoints of base line to the midpoint of tipline of index finger. In the same way, 4D length was measured for the ring finger. Hand grip strength is the strength applied by each hand in its neutral position on the hand grip dynamometer. The data is measured in kg of strength applied. The protocol is to apply the maximum voluntary strength for two seconds and sustain it for another three seconds. This measurement is done thrice and the average of the three values is considered. Whilst the anthropometric attributes being measured in variable units, initially normalization process of the attributes was performed. Thereafter, cluster analysis was applied in Weka 3.7 (Hall et al., 2009). For estimating the opposite number of clusters, Expectation Maximization (EM) technique (Ghosh and Liu, 2009) was used. Step (E) estimates probability and Step (M) acquires an approximation mixture model. The k-means, an iterative clustering algorithm was utilized to form homogenous clusters from the considered attributes. The functional algorithm was performed in two steps: (a) clustering all objects in the dataset based on the distance between each object and its nearest cluster representative and (b) cluster representatives by re-estimating. As the number of clusters were estimated, k-means algorithm was used to form homogenous clusters. Further formed cluster quality was evaluated through statistical differences in SPSS.

The clustering technique distributes the data on the different group depends on its similarity. In the real scenario, every time new workers join and it is not possible to form a new cluster. Therefore, a supervised learning method can apply. In the first step of learning known data or training data is utilized to train the model. In the second step, the model predicts the class for the unknown data. In this paper, the result of clustering method is utilized for the model training. There are different supervised learning methods are available in data mining or machine learning and it is effective applied in different applications. In this paper, we consider Instance-based (IB), Naïve Bayesian networks (BN), Multilayer Perceptron Learning (MLP), Support Vector Machine (SVM) and Radial Basis Function (RBF) classifier for the verification acceptability and accuracy of above illustrated the anthropometric variables.

Instance-based (IB) learning algorithms was used as this process require less computation time during the training phase than other algorithms such as BN, although it requires more computation time during the classification process (Kotsiantis, 2007).

Naïve Bayesian networks (BN) is simple Bayesian networks, a combination of trees with only one parent, characterised by unseen node and several children resembling to seen nodes with a strong supposition of independence among child nodes in relation to parents. Thus, this independence model is based on estimating probabilities. The higher probability specifies that the class label value resemble more close to the actual label. BNs have been proven to be a strong tool to discover the relationships between variables that attempts to separate out direct and indirect dependencies, Fuster-Parra et al., (2016).

Multilayer Perceptron Learning (MLP) is an artificial neural network supervised learning method. It is capable to classify the non-linearly input data. It uses extended gradient-descent based delta rule-learning rule common known as back propagation. It consists of a large number of units (neurons) joined in a pattern of connections. The classification task is performed by supplying the input data on neural network and the activation function is used to determine the output value. Each input unit has an activation value representing some feature external to the net, Naraei et al., (2016).

Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for classification. SVM identifies a decision boundary with the maximum possible margin between the data points of each class, Razzaghi et al., (2015).

Radial Basis Function (RBF) classifier has been also widely applied in many science and engineering fields). An RBF network is a three-layer feedback network, in which each hidden unit implements a radial activation function and each output unit implements a weighted sum of hidden unit outputs. Its training procedure is usually divided into two stages. First, the centres and widths of the hidden layer are determined by clustering algorithms. Second, the weights connecting the hidden layer with the output layer are determined by Singular Value Decomposition (SVD) or Least Mean Squared (LMS) algorithms.

Table 1. Distribution of Data among Different Clusters

Attribute	Cluster 1 Very Poor	Cluster 2 Poor	Cluster 3 Fair	Cluster 4 Good	Cluster 5 Excellent
N	80	31	63	97	76
Age	14.12±0.88	15.41±1.4	20.83±1.9	16.402±1.6	20.11±1.63
Height	153.05±7.7	163.3±6.9	166.66±8.01	165.56±5.4	172.08±5.8
Weight	38.18±5.3	64.4±13.9	49.88±4.8	47.06±4.7	62.28±7.1
BMI	16.26±1.7	24.1±4.8	18.02±1.99	17.18±1.7	21.02±2.0
BSA	1.29±0.11	1.68±0.17	1.54±0.10	1.49±0.08	1.73±0.11
BF	20.58±4.6	31.32±4.4	18.89±2.7	18.60±2.86	24.78±4.3
RH2D:4D	0.96±0.035	0.94±0.025	0.97±0.049	0.94±0.035	0.94±0.043
LH2D:4D	0.95±0.033	0.95±0.031	0.98±0.047	0.94±0.035	0.97±0.0473
RHTG	51.65±9.7	65.29±11.8	71.89±8.5	75.38±9.4	85.41±9.6
LHTG	50.45±10.4	59.9±10.7	68.16±12.1	73.08±8.8	80.42±10.3

Table 2. Comparative Result of Different Supervised Learning Methods

	IB	BN	MLP	SVM	RBF
Correctly Classified Instances %	92.5	93.37	95.10	95.10	89.91
Incorrectly Classified Instances %	7.4	6.62	4.89	4.89	10.08
Kappa statistic %	90.3	91.4	93.7	93.7	87.06
Mean absolute error	0.03	0.04	0.02	0.24	0.04
Relative absolute error %	11.06	13.79	8.6	77.53	14.9
Total Instances	347	347	347	347	347

3. Result and Discussion

Five subsets of volunteers were distributed as five clusters. It was observed that clusters were formed with accumulative raise in right handgrip strength (51.65±9.7 kg, 65.29±11.8 kg, 71.89±8.5 kg, 75.38±9.4 kg, 85.41±9.6 kg) and left handgrip strength (50.45±10.4 kg, 59.9±10.7 kg, 68.16±12.1 kg, 73.08±8.8 kg, 80.42±10.3 kg). As observed that, right hand 2D:4D followed a pattern of diminishing trend with the progressive cluster number except for Cluster 3 formed. It is iterated that 2D:4D is a marker of exposure to pre-natal testosterone wherein 2D:4D is inversely related to pre-natal testosterone exposure and hence higher strength profile. Also, research revealed that digit ratio aid in comprehending body composition markers in terms of positive correlation between lower 2D:4D and male type pattern of anthropometric indicators (Majumder and Bagepally, 2015). The distribution of cluster result is shown the **Table 1.** Further, each cluster revealed distinct inter-cluster differences, but resemblance exists within each cluster. Therefore, according to the characteristic of each cluster, particularly with increasing handgrip strength values, the clusters were coined as very poor, poor, fair, good and excellent. Further, these data was utilized for the training the supervised learning methods. The comparative results of IB, BN, MLP, SVM, RB classifier were reported in the **Table 2.** It can be observed that accuracy 89.91% of RBF is lowest among these algorithms. The performance of IB and BN is comparable similar with accuracy 92.5% and 93.37%. The MLP and SVM give high accuracy 95.10% for the both, but it is noted that the relative absolute error of SVM is more than MLP. Therefore, the performance of MLP can be consider as best one among these algorithms.

4. Conclusion

In this study, data mining techniques cluster analysis and supervised learnings were applied for grouping the

young workers according to their physical anthropometric parameters. The results were further utilized for the model construction and predict the class. The rule of classification of the human handgrip strength can aid in classifying the class for subsequent unlabelled cases. The reported results would guide in prospects of pre-screening of proper work group for a pre-defined specialized job, anthropometrics, as well as application in occupational health.

5. References

1. BoLSta (2013) Nonfatal Occupational Injuries and Illnesses Requiring Days away from Work, Bureau of Labour Statistics, United States Department of Labour.
2. Chaffin, D.B., Herrin, G.D. and Keyserling, W.M. (1978) Pre-Employment Strength Testing: An Updated Position, J Occup, 20, 403–408.
3. Durnin, J.V.G.A. and Womersley, J. (1974). Body Fat Assessed from the Total Body Density and its Estimation From Skinfold Thickness: Measurements on 481 Men and Women Aged from 16 to 72 Years. British Journal of Nutrition, 32, 77-97. Crossref PMid:4843734
4. Fuster-Parra, P., Tauler, P., Bennasar-Veny, M., Ligêza, A., López-González, A.A. and Aguiló, A. (2016) Bayesian Network Modeling: A Case Study of an Epidemiologic System Analysis of Cardiovascular Risk', Computer Methods and Programs in Biomedicine, 126, pp. 128-142. Crossref PMid:26777431
5. Ghosh, J. and Liu, A. (2009) The Top Ten Algorithms in Data Mining, Taylor & Francis Group, LLC: pp. 21.
6. Gite LP, Majumder J, Mehta CR and Khadatkar A (2009). Anthropometric and Strength Data of Indian Agricultural Workers for Farm Equipment Design, CIAE Bhopal (ISBN 978-81-909305-0-5).
7. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. H. and Witten, I. H. (2009) The WEKA Data Mining Software: An Update, ACM SIGKDD Explorations, 11, 10-18. Crossref
8. Jenkin N. et al. (2016). Pre-Employment Physical Capacity Testing as a Predictor for Musculoskeletal Injury in Paramedics: A Review of the Literature, Work, 55, 565-575. Crossref PMid:27792024
9. Kotsiantis, S. B. (2007) Supervised Machine Learning: A Review of Classification Techniques, Informatica, 31, 249-268.
10. Lin, H. F., Hsu, C. H., Wang, M. J. and Lin, Y. C. (2008) An Application of Data Mining Technique in Developing Sizing System for Army Soldiers in Taiwan, WSEAS Tran on Computers, 4, 245-252.
11. Majumder, J. and Bagepally, B. S. (2015). The Right Hand Second to Fourth Digit Ratio (2D:4D) and its Relationship with Body Composition Indicators among Young Population, Asian Journal of Medical Sciences, 6, 78–84.
12. Majumder, J. and Sharma, L.K. (2014) Application of Data Mining Techniques to Audiometric Data among Professionals in India', Journal of Scientific Research & Reports 3, 2860-2971. Crossref PMid:28537663
13. Majumder, J. and Sharma, L.K. (2015) Identifying Body Size Group Clusters from Anthropometric Body Composition Indicators', J. Ecophysiol. Occup. Hlth. 15, 81-88.
14. Naraei, P., Abhari, A. and Sadeghian, A. (2016) 'Application of Multilayer Perceptron Neural Networks and Support Vector Machines in Classification of Healthcare Data', IEEE Future Technologies Conference (FTC), 838-847. Crossref
15. Niu, J., He, Y., Li, M., Zhang, X. and Ran, L., Chao, C. and Zhang, B. (2010) A Comparative Study on Application of Data Mining Technique in Human Shape Clustering: Principal Component Analysis vs. Factor Analysis, the 5th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2014-2018.
16. Paquet, E., Pena, I. and Viktor, H. L. (2011) 'From Anthropometric Measurements to three Dimentional Shape', Indian Journal of Fiber and Textile Research, 36, pp. 336-343.
17. Razzaghi, T., Roderick, O., Safro, I. and Marko, N. (2016) Multilevel Weighted Support Vector Machine for Classification on Healthcare Data with Missing Values, PLoS ONE 11, Rosenblum, K.E. and Shankar, A. (2006) A Study of the Effects of Isokinetic Pre-Employment Physical Capability Screening in the Reduction of Musculoskeletal Disorders in a Labor-Intensive Work Environment', Work (Reading, Mass), 26, 215–228.
18. Schaafsma, F.G., Mahmud, N., Reneman, M.F., Fassier, J.B. and Jungbauer, F.H.W.(2016) Pre-Employment Examinations for Preventing Injury, Disease and Sick Leave in Workers',Cochrane Database of Systematic Reviews, 1, Art. No.: CD008881. Crossref PMid:26755127
19. Serra, C., Rodriguez, M. C., Delclos, G. L., Plana, M., Lopez, L. I. and Benavides, F. G.(2007) Criteria and Methods Used for the Assessment of Fitness for Work: A Systematic Review, Occupational and Environmental Medicine, 64, 304–312. Crossref PMid:17095547 PMCid:PMC2092557
20. Sharma, L. K. and Majumder, J. (2013) Application of Artificial Neural Network on Body Somatotype Analysis Among Indian Population', In proceedings of International Conference of HWWE-2013, 124-131.
21. Sharma, R. (2012) Epidemiology of Musculoskeletal Conditions in India, Indian Council of Medical Research (ICMR), New Delhi, India.
22. Siri H.E. (1961). Body Composition from Fluid Space and Density. In: Brozek, J., Hanschel, A. (Eds.), Techniques for Measuring Body Composition. National Academy of Science, Washington, DC, 223-244.

23. Wilson, A. M., Thabane, L. and Holbrook, A. (2004) Application of Data Mining Techniques in Pharmacovigilance, Br J Clinpharmacol, 57, pp.127-134. Crossref PMid:14748811 PMCid:PMC1884444

24. Zakaria, N., Mohd, J. S., Taib, N. , Tan, Y. Y. and Wah, Y. B. (2008) Using Data Mining Technique to Explore Anthropometric Data towards the Development of Sizing System, International Symposium on Information Technology, 2, 1-7. Crossref