J. Ecophysiol. Occup. Hlth. 15 (3 & 4), 2015, 81–88
© 2015 The Academy of Environmental Biology, India
DOI : 10.15512/joeoh/2015/v15i3-4/121588

Identifying Body Size Group Clusters from Anthropometric Body Composition Indicators

Joydeep Majumder and Lokesh Kumar Sharma*

National Institute of Occupational Health (ICMR), Ahmedabad-380016, India

Abstract: *Mining of anthropometric data by clustering technique would categorically classify homogenous body size group. The objective of this study is to classify homogenous human body size according to anthropometric body composition indicators. Anthropometric data was measured from 382 men and 391 women of Orissa, India. Percent body fat was calculated from the skinfold parameters. Cluster analysis was then applied with self reported age, stature, weight and percent body fat. The clusters formed were tested statistically for their validity of formation. Multivariate analysis considering age, stature, weight and percent body fat revealed significant differences among men and women (p<0.001). Expectation Maximization (EM) estimated five clusters for men and women by age, stature, weight and percent body fat. Consequently, k-means cluster algorithm was used and it formed five clusters with cumulative increment in stature, weight and percent body fat, for both men and women. However, age individually, did not influence the body size indicators. The clusters were named small, medium, large, X-large and XX-large. Silhouette plot validation of clusters reveals that for both men and women, 95.5% and 98.7% of data, respectively were well-clustered. These cluster results further can generate classification rules to categorize subsequent unseen cases, and may aid in anthropometric database creation, nutritional status, body growth research, etc.*

Keywords: Anthropometry, Cluster Analysis, K-means, Silhouette Plot, Body Sizing.

Introduction

Relevance of human body dimensions was recognized decades back for partial fitting in equipment design among Germans, French, Italians, Japanese, Thais and Vietnamese (Abeysekera and Shahnavaz, 1989). It has also been noted that differences in anthropometric data of people from different regions in India exists (Saha, 1985; Dewangan *et al*., 2005; Gite *et al.,* 2009). This indicates varying anthropometric dimensions among population and need for its quantification and classification. The quantification of anthropometric data would build up a methodical body size group (Yu, 2004; Aldrich, 2007). As a part of the country-wide anthropometric database for use in various research and application, including health and fitness research, evaluation of body composition, study of nutritional status and disorders, monitoring of body growth, postural analysis and skeletal disorders/deformities, body kinetics and performance analysis, body size grouping of the population would be of immense help. Mining of anthropometric data by clustering technique (Abdali *et al*., 2004) would classify homogenous body size group, which will create meaningful clusters, and their corresponding archetype(s) would aid our understanding of the distinguishing physical characteristics of the population.

Cluster analysis is an active research area of data mining (Xu and Wunsch, 2005; Gosh and Liu, 2009) and an easily replicable way of constructing classifications (Aldenderfer and Blashfield, 1984). Objects in a specific cluster share many characteristics, but are very dissimilar to objects not belonging to that cluster. Clustering approach allows detection of natural groups, in the form of clusters, as based on different body types. Further, rule

*Email: sharma.lk@gov.in

of classification of body types would classify class for subsequent unseen cases, as they are added to the database. Mahalanobis *et al.* (1949) used this method in an anthropometric survey of the united province. Vasulu and Pal (1989) studied the relationship between anthropometric differentiation and cultural diversity in the Yanadi tribe in different regions of India. Rao *et al.* (2013) used the technique on anthropometric data of 10096 children in 54 districts of Uttar Pradesh and classified the children in 4 clusters according to their average height and weight.

Data mining technique has also been used on anthropometric measurements for the development of sizing system using stature and bust/waist circumference as determined by principal component analysis (Zakaria *et al.,* 2008; Bagherzadeh *et al.,* 2010). These studies were focussed on development of sizing system for clothing industry, in particular, and not on the overall body composition characteristics of the population. However, in our study, age, stature, weight and percent body fat are considered as the input attributes for cluster analysis, so as to look into the body size in terms of anthropometric body composition indicators. The results can further generate classification rules to categorize subsequent unseen cases. The clusters and archetypes as identified may be of use for nutritional status, anthropological classifications, as well as derive human body models for industrial designs (Abdali *et al.,* 2004; Hsu *et al.,* 2009).

Materials and Methods

The study was a cross-sectional survey on a community based rural population in Odisha, India (19.5^0 N, 84.5^0 E). The study was approved by the Ethics Committee of the Institute. The volunteers were 382 men and 391 women of variable age groups (16-85 years). All the subjects were healthy without any physical disabilities. Before the measurement, the subjects were informed about the purpose of the study and the measurement procedures. Each participant had the opportunity to ask questions before agreeing to participate in the study and signing an informed consent form. The anthropometric measurements were then recorded by two trained researchers for the entire sample. Stature was taken on a flat base with stadiometer (Bioplus, India) attached to the wall. Weight was measured with volunteers barefoot, on an electronic balance (Rossmax, Swiss Gmbh) accurate to 0.1 kg. Four skinfold measurements were taken with skinfold calliper (Holtain Ltd., Crosswell, Crymych, UK) at biceps, triceps, sub scapular, and supra iliac sites as per the methodology described in Gite *et al.* (2009). The body mass index (BMI), body surface area (BSA) and percent body fat was calculated from the anthropometric parameters.

Data analysis was performed in SPSS 16.0 for Windows. The descriptive statistics were reported. Skewness was measured as the measure of the asymmetry of the probability distribution and kurtosis, as measure of the peakedness. Statistical hypothesis testing using univariate and multivariate analysis were done to examine the significance of differences between men and women volunteers as well as between the formed clusters. A p value < 0.05 was considered significant. Expectation Maximization (EM), a probabilistic optimization technique (Ghosh and Liu, 2009) is a two-step iterative optimization. Step (E) estimates probability and Step (M) finds an approximation mixture model. EM is utilized to estimate the appropriate number of clusters. k-means, an iterative clustering algorithm that partitions a given dataset into a user-specified number of clusters (Ghosh and Liu, 2009). The algorithm works by iterating over two steps: (a) clustering all points in the dataset based on the distance between each point and its closest cluster representative and (b) re-estimating the cluster representatives. The cluster analysis was implemented in this study with self reported age, measured anthropometric attributes as stature and weight, and computed percent body fat from the four skinfold measurement, in

Waikato Environment for Knowledge Analysis software (Weka 3.7), developed at University of Waikato, Hamilton, New Zealand (Hall *et al.*, 2009). As the anthropometric attributes were measured in various units, attributes were normalized initially. The number of clusters was estimated by EM technique. On determination of the number of clusters, k-means algorithm was used to form homogenous clusters.

The anthropometric variables and the calculated attributes for the men and women volunteers are presented in Table 1 and 2, respectively. As the data were non-normally distributed, non-parametric Mann-Whitney U test was applied. Among the volunteers studied, women tend to have higher values of body composition characteristics. However, the distribution of BMI indicated that there is no significant difference between men and women studied ($p = 0.430$).

The distribution of subcutaneous body fat deposition, as measured in four different sites of the body was also higher in women than men, with a deviation in the subscapular site, where the skin fold thickness is marginally more in men. Further, the cumulative skinfold thickness of the four sites was 19.6% more in the women than that of the men ($p < 0.001$). The BSA of women is less than that of the men ($p < 0.001$). Absolute body fat of men and women though have no significant difference ($p = 0.497$, NS), percent body fat in women was observed to be significantly higher than that of the men ($p < 0.001$). This indicates that although the body surface area of men and women does not have any significant variation, the deposition of subcutaneous fat among women is higher than that of the men studied (Flegal *et al.,* 2010; Joseph *et al.,* 2011). Multivariate analysis considering age, stature, weight and percent body fat also revealed significant difference exist among men and women (Wilks' Lambda = 0.505, $p < 0.001$).

For identification of homogenous groups of men and women, cluster analysis for both sexes was applied separately. Age, stature, weight and percent body fat were taken into consideration for the cluster analysis, which estimated five clusters for both men and women. The distribution statistics are presented in Table 3 and 4 for men and women, respectively.

It is noticed that with clusters were formed with cumulative increment in weight and percent body fat, in case of men. However, the clusters formation was not as per the incremental trend of stature. Non-parametric Kruskal-Wallis

Table 1 Descriptive statistics of anthropometric and derived attributes among men (N = 382).

Attributes/Indices	Mean	SD	Range	Skew-ness	Kurtosis
Age (years)	44.0	17.2	16-85	0.094	-1.017
Stature (m)	1.63	0.072	1.35-1.79	-0.485	0.567
Weight (kg)	56.6	10.6	35-94	0.557	0.195
Bicep skin fold thickness (m)	0.0045	0.002	0.0018-0.016	2.272	6.915
Tricep skin fold thickness (m)	0.0088	0.005	0.0022-0.0264	1.207	1.273
Subscapular skin fold thickness (m)	0.011	0.004	0.0028-0.0272	1.071	0.979
Supra iliac skin fold thickness (m)	0.0089	0.005	0.0032-0.0288	1.125	0.718
% Body Fat	19.9	5.6	8.92-33.7	0.280	-0.824
Body Mass Index	21.4	3.4	14.3-35.2	0.723	0.663
BSA (m^2)	1.6	0.2	1.2-2.0	-0.108	-0.276

Table 2 Descriptive statistics of anthropometric and derived attributes among women (N = 391).

Attributes/Indices	Mean	SD	Range	Skew-ness	Kurtosis
Age (years)	45.2	15.2	16-80	0.016	-0.957
Stature (m)	1.52	0.058	1.36-1.76	0.203	0.330
Weight (kg)	49.1	9.6	26-89	0.625	0.256
Bicep skin fold thickness (m)	0.0063	0.003	0.0018-0.017	1.206	1.239
Tricep skin fold thickness (m)	0.0123	0.005	0.002-0.025	0.447	-0.519
Subscapular skin fold thickness (m)	0.0104	0.004	0.0032-0.026	1.025	0.861
Supra iliac skin fold thickness (m)	0.0123	0.006	0.0024-0.029	0.509	-0.332
% Body Fat	23.1	5.6	4.9-35.3	-0.216	-0.465
Body Mass Index	21.2	3.6	11.3-34.8	0.654	0.468
BSA (m^2)	1.4	0.1	1.1-1.9	0.330	-0.204

Table 3 Distribution statistics of the cluster formation for men (N = 382).

Visualization Colour	Age (yrs)	Stature (cm)	Weight (kg)	Absolute body fat (kg)	Indicator Size	N (%)
Purple	32.2±10.5	158.4±6.1	47.7±6.1	15.0±2.8	S	73(19.1)
Blue	61.7±9.1	160.4±6.3	49.7±5.8	16.8±3.8	M	104(27.2)
Orange	26.7±8.1	169.4±4.4	49.8±5.1	18.3±3.1	L	73(19.1)
Green	34.9±10.1	158.8±6.9	59.2±7.5	25.7±2.8	XL	56(14.7)
Red	54.7±10.1	166.1±5.6	70.0±8.9	26.5±3.3	XXL	76 (19.9)
Chi-square value	271.09*	145.76*	232.94*	253.04*		

* p<0.001

Table 4 Distribution statistics of the cluster formation for women (N = 391).

Visualization Colour	Age (yrs)	Stature (cm)	Weight (kg)	Absolute body fat (kg)	Indicator Size	N (%)
Purple	30.7±8.9	150.3±5.6	40.9±3.7	17.5±3.4	S	76 (19.4)
Blue	60.1±8.3	148±4.7	41.5±4.8	19±3.7	M	100 (25.6)
Green	32.2±7.5	153.2±5.1	52±5.4	25.4±2.8	L	84 (21.5)
Orange	57.5±7.3	154.4±5	53.5±6.3	26.3±3	XL	78 (19.9)
Red	40.4±8	155.8±5.3	64.1±6.6	30.5±2.2	XXL	53 (13.6)
Chi-square value	291.05*	96.83*	277.33*	282.16*		

* p<0.001

test was applied to examine the statistical difference among the parameters between the clusters formed. The results reveal that significant difference in age (Chi-Square = 271.09, p<0.001), stature (Chi-Square = 145.76, p<0.001), weight (Chi-Square =

232.94, p<0.001) and percent body fat (Chi-Square = 253.04, p<0.001) exist between the five clusters. In case of women, clusters were formed with cumulative increment in stature, weight and percent body fat, except a deviation in stature for two clusters (purple and blue). Non-parametric Kruskal-Wallis test reveal that age (Chi-Square = 291.05, p<0.001), stature (Chi-Square = 96.83, p<0.001), weight (Chi-Square = 277.33, p<0.001) and percent body fat (Chi-Square = 282.16, p<0.001) significantly vary between the five clusters formed. Accordingly, the clusters were coined as small (S), medium (M), large (L), X-large (XL) and XX-large (XXL). The cluster results for men and women are visualized in Fig. 1 and 2. For both figures, the positions of the cluster points are on the basis of the parameters taken into consideration. However, the visualization of point size has been done on the basis of percent body fat. Accordingly, for men, the clusters formed were of red, green, orange, blue and purple colour, in terms of descending percent body fat values. For women, the clusters formed were of red, orange, green, blue and purple colour, in terms of descending percent body fat values.

In order to validate the quality of clusters formed, Silhouette plot was used to graphically represent how well observations were accrued within each cluster based on the input parameters. Silhouette value ranges between -1 and +1, and positive values indicate strong clustering, while negative values indicate weak clustering. Fig. 3 and 4 shows the Silhouette plot for men and women respectively. 95.5% and 98.7% of the data for men and women, respectively, were above zero, indicative of the fact that the clusters were well structured, with most observations seeming to belong to the cluster.

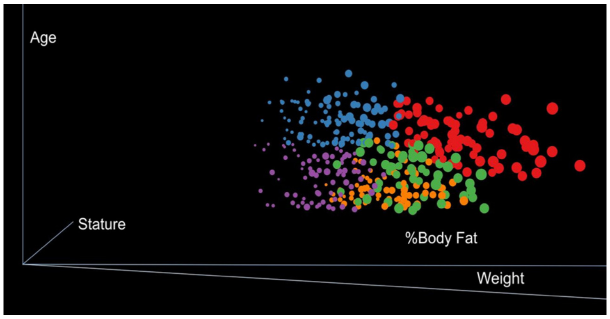

Fig. 1 Visualization of Cluster results for men with age, stature, weight and percent body fat. Positions of the cluster points are on the basis of age, stature, weight and percent body fat. Point size is on the basis of percent body fat. Clusters formed (red, green, orange, blue and purple) are on the basis of descending percent body fat values.

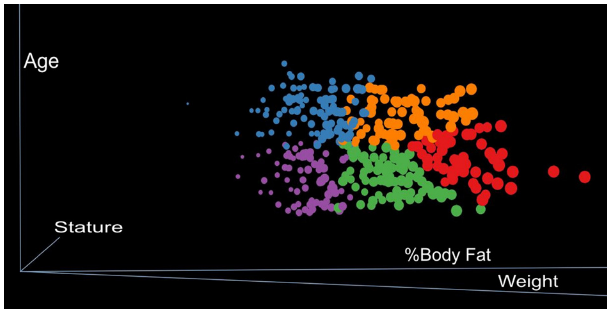

Fig. 2 Visualization of Cluster results for women with age, stature, weight and percent body fat. Positions of the cluster points are on the basis of age, stature, weight and percent body fat. Point size is on the basis of percent body fat. Clusters formed (red, orange, green, blue and purple) are on the basis of descending percent body fat values.

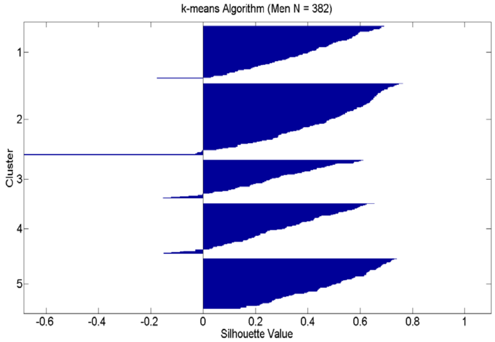

Fig. 3 Silhouette plot for men. It shows that most points in each of the five clusters have a silhouette value greater than 0, indicating that the clusters are separated from each other.

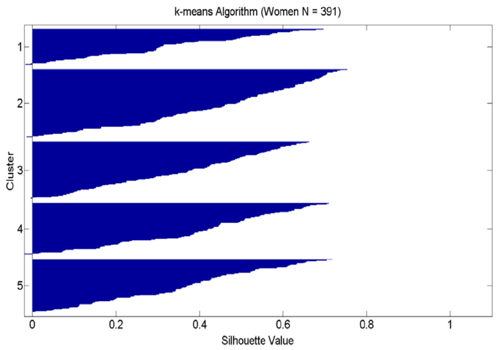

Fig. 4 Silhouette plot for women. It shows that most points in each of the five clusters have a silhouette value greater than 0 (98.7%), indicating that the clusters are separated from each other.

Results and Discussion

As evident from Table 1 and 2, anthropometric data, as recorded for men and women were positively skewed, with values concentrated on left of the mean, and extreme values to the right. The values had a platykurtic distribution, flatter than a normal distribution and the values are wider spread around the mean. The results are in concordance with the earlier study on agricultural workers of both sexes. The stature recorded in our study is similar to that of the agricultural workers reported, however, the weight was observed around 13% elevated for both men and women than that the agricultural workers (Gite *et al.*, 2009). This suggests the trend of the population towards turning endomorphic (Ellis, 2001). In the present study, the mean body fat (19.9% in men and 23.1% in women**)** also advocates the endomorphic trend i.e. characterization of body in terms of increased fat storage.

Literature has emphasised on clustering of anthropometric attributes for development of sizing system for clothing industry, though sizing in terms of body composition indicators elucidate even the characteristics of the population. Perhaps, anthropometric studies on sizing system for apparels attract the science and technological advances towards appearance and fit (Aldrich, 2007; Abdali *et al.*, 2004; Paquet *et al.*, 2011; Jeyasingh and Appavoo, 2012). The present study, therefore considered age, stature and weight as the attributes for the cluster analysis because they were actual measured data (Paquet, *et al.*, 2011). Percent body fat was considered because it was the computed cumulative observation of four skinfold measurements (Gite *et al.*, 2009) and studies also reveal that percent body fat is probably an appropriate and important measure for defining obesity in Asian populations (Pongchaiyakul *et al.*, 2005).

The clusters coined as S, M, L, XL and XXL, is in concordance to other cluster analysis studies on anthropometric data (Paquet *et al.*, 2011; Jeyasingh and Appavoo, 2012; Brantley, 2000; Chung *et al.*, 2007). Although, the attributes in these studies differ from each other, the distribution of the morphological features in the coined clusters complement with our study. As seen in Table 3 and 4, stature, weight and percent body fat among the volunteers vary significantly from each cluster formed. Also, weight and percent body fat followed the incremental curve with the increment in coined clusters. However, age individually was found not influencing the body size indicator. The results are in concordance with the study of Paquet *et al.* (2011). The results indicate that k-means algorithm computed well-defined clusters. Silhouette plot (Fig. 3 and 4) also indicates that cluster points are very distant from neighbouring clusters (Rousseeuw, 1987; Brun *et al.*, 2007). This technique computed the silhouette width for each data point, average silhouette width for each cluster and overall average silhouette width for the total data set, indicative of the fact that the clusters were compact, well-structured and separated clusters, with most observations seeming to belong to the cluster.

In the present study, the clustering was done with the measured anthropometric parameters as well as the calculated attribute (percent body fat). The morphological shape of the volunteers as determined from somatotype of the individual was not considered. Although, studies have been carried out with body shape as a component along with the measured anthropometric parameters, however all those studies were on the sizing system for clothing design (Paquet *et al.*, 2011; Jeyasingh and Appavoo, 2012). As clothing design is dependent on both anthropometric parameters as well as shape of the body, somatotype information is crucial to determine a close fit. This may be a limitation of the present study; however the purpose of this study was to identify the body size from the point of view of body composition.

Identifying body size group from anthropometric body composition indicators is crucial from the point of view of nutritional status, body growth research. The results of this study indicate that although the body surface area of men and women are quite similar, women were having high deposition of subcutaneous fat than men. The rural population in Odisha, India are classified into five discrete body size clusters on the basis of age, stature, weight and percent body fat. Each cluster showed distinct inter-cluster differences but similarity exist within each cluster, as confirmed by Silhouette plot. The results further can be utilized to generate classification rules. The rule of classification of the body types extracted can help to classify the class for subsequent unseen cases. These results, as reported in this paper would aid in prospects of anthropometric database creation, application in health, nutrition and body growth research.

Acknowledgement

The authors are grateful to the field team for data collection. The authors also acknowledge the whole-hearted participation of the volunteers in the study.

References

Abdali, O., Viktor, H.L., Paquet, E. and Rioux, M. (2004) Exploring anthropometric data through cluster analysis; In: *SAE Technical Paper*, (DOI: 10.4271/2004-01-2187).

Abeysekera, J.D.A. and Shahnavaz, H. (1989) Body Size variability between people in developed and developing countries and its impact on the use of imported goods. *Int. J. Ind. Ergonom*., **4,** 139–149.

Aldenderfer N.S. and Blashfield, R.K. (1984) Cluster analysis, (1st ed). Beverly hills, CA: Sage Publication, pp. 7–17.

Aldrich, W. (2007) History of sizing systems and ready to wear garments. In: Ashdown SP, editors. Sizing in clothing: Developing effective sizing systems for ready to wear clothing. England: Woodhead publishing, pp. 43.

Bagherzadeh, R., Latifi, M. and Faramarzi, A.R. (2010) Employing a three-stage data mining procedure to develop sizing system. *WASJ.*, **8,** 923–929.

Brantley, J.D. (2000) Field evaluation of the sizing and tariff of the U.S. marine corps interceptor body armor. U.S. Army Soldier and Biological Chemical Command Soldier Systems Center, Natick, Massachusetts 01760-5050, (Technical Report Imatick /TR-00/014).

Brun, M., Sima C., Hua, J., Lowey. J., Carroll, B, Suh, E. and Dougherty, E.R. (2007). Model-based evaluation of clustering validation measures. *Pattern Recogn.,* **40,** 807–824.

Chung, M., Lin, H. and Wang, M. (2007) The development of sizing systems for Taiwanese elementary and high-school students. *Int. J. Ind. Ergonom.,* **37,** 707–716.

Dewangan, K.N., Prasanna Kumar, G.V, Suja, P.L. and Choudhury M.D. (2005) Anthropometric dimensions of farm youth of the north eastern region of India. *Int. J. Ind. Ergonom.,* **35,** 979–989.

Ellis, K.J. (2001) Selected body composition methods can be used in field studies. *J. Nutr.*, **131,** 1589S–1595S

Flegal, K.M., Carroll, M.D., Ogden, C.L. and Curtin, L.R. (2010) Prevalence and trends in obesity among US adults, 1999-2008. *J. Am. Med. Assoc.*, **303,** 235–241.

Ghosh, J. and Liu, A. (2009) k-means. In: Wu X, Kumar V. (Eds.). The top ten algorithms in data mining. Taylor and Francis Group, LLC, pp. 21–36.

Gite, L.P., Majumder, J., Mehta, C.R. and Khadatkar, A., (Eds.). (2009) Anthropometric and strength data of Indian agricultural workers for Farm equipment design. CIAE Bhopal.

Hall, M., et al. (2009) The WEKA data mining software: An update. *ACM SIGKDD Explorations.,* **11,** 10–18.

Hsu, C.H., Lee, T.Y., Kuo and H.M. (2009) Mining the body features to develop sizing systems to improve business logistics and marketing using fuzzy clustering data mining. *WSEAS Transactions on computers.*, **7,** 1215–1224.

Jeyasingh, M.M. and Appavoo, K. (2012) Mining the shirt sizes for Indian men by clustered classification. *Int. J. Info. Technol. Comp. Sci.,* **6,** 12–17.

Joseph, L., et al. (2011) Appropriate values of adiposity and lean body mass indices to detect cardiovascular risk factors in Asian Indians. *Diabetes Technol. Ther.*, **13,** 899–906.

Mahalanobis, P.C., Majumdar, D.N. and Rao, C.R. (1949) Anthropometric survey of the united provinces, 1941. *A statistical study, Sankhya.,* **9,** 89–324.

Paquet, E., Pena. I. and Viktor, H.L. (2001) From anthropometric measurements to three-dimensional shape. *Indian J. Fibre Text.***, 36,** 336–343.

Pongchaiyakul, C., et al. (2005) Prediction of percentage body fat in rural Thai population using simple anthropometric measurements. *Obes. Res.,* **13,** 729–738.

Rao, M.V.V., Kumar, S. and Brahmam, G.N.V. (2013) A study of the geographical clustering of districts in Uttar Pradesh using nutritional anthropometric data of preschool children. *Indian J. Med. Res.*, **137,** 73–81.

Rousseeuw, P.J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.,* **20,** 53–65.

Saha, P.N. (1985) Anthropometric characteristics among industrial workers in India. Proceedings of International Symposium on Ergonomics in Developing Countries. Jakarta, Indonesia, pp. 158–161.

Vasulu, T.S. and Pal, M. (1989) Size and shape components of anthropometric differences among the Yanadis. *Ann. Hum. Biol.,* **16,** 449–462.

Xu, R. and Wunsch, D. (2005) Survey of Clustering Algorithm. *IEEE Trans Neural Netw,* **16,** 645–678.

Yu, W. (2004) Human anthropometrics and sizing systems. In: Fan J, Yu W, Hunter L. Clothing appearance and fit: Science and technology. Boston, CRC Press, pp. 169–193.

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. *Int. Symp. Inform. Technol.*, **2,** 1–7.