

Selection of hand features based on Random Forest algorithm and hand shape recognition

DOI: 10.35530/IT.075.03.202351

XIN LI
XIAO-JUN DING
ZHOU-YAN PENG

XI-YAN LIN
FENG-YUAN ZOU

ABSTRACT – REZUMAT

Selection of hand features based on Random Forest algorithm and hand shape recognition

To obtain effective features applicable to hand morphology recognition, the method of obtaining effective feature indicators based on the Random Forest (RF) algorithm to downscale hand morphology parameters is proposed. Firstly, 232 female university students collected three-dimensional hand information, constructed auxiliary point, line, and surface standardised measurement methods, obtained 33 characteristic parts of human dimensions, and used k-means clustering for hand morphology subdivision. Hand morphology can be divided into three categories: short, slender and broad. The RF algorithm is used for feature index importance assessment and hand shape recognition model. The accuracy of the feature metrics determined by the RF algorithm, PCA, and VC method applied to hand shape recognition is compared and analysed to verify the effectiveness of the dimensionality reduction of the RF algorithm. The results showed that the feature indexes used for hand shape recognition were five items: hand length, thickness at the metacarpal, thenar width, the distance between the thumb and index finger, and distance from the root of the little finger to the centre of the wrist. Using the RF algorithm to reduce the dimensionality is more effective; the average recognition accuracy of the four hand shape recognition models is 93.78% on average, compared with PCA and VC reduction methods, the average accuracy of hand shape recognition models is increased by 19.17%, and 14.86% respectively. The study's results can provide methodological references for the objective selection of characteristic indicators and morphological recognition of human body parts.

Keywords: 3D scanning, hand shape classification, Random Forest algorithm, feature selection, hand shape recognition

Selectarea caracteristicilor mâinii pe baza algoritmului Random Forest și recunoașterea formei mâinii

Pentru a obține caracteristici eficiente aplicabile recunoașterii morfologiei mâinii, este propusă metoda de obținere a indicatorilor eficienți bazați pe algoritmul Random Forest (RF) pentru a reduce scalarea parametrilor morfologiei mâinii. În primul rând, 232 de studenți au colectat informații tridimensionale ale mâinii, au construit metode de măsurare standardizate a punctelor auxiliare, liniilor și suprafeței, au obținut 33 de părți caracteristice ale dimensiunilor umane și au folosit gruparea k-means pentru subdiviziunea morfologiei mâinii. Morfologia mâinii poate fi împărțită în trei categorii: scurtă, îngustă și lată. Algoritmul RF este utilizat pentru evaluarea importanței indicelui caracteristicilor și modelul de recunoaștere a formei mâinii. Precizia parametrilor caracteristicilor determinate de algoritmul RF, PCA și metoda VC aplicată recunoașterii formei mâinii sunt comparate și analizate pentru a verifica eficacitatea reducerii dimensionalității algoritmului RF. Rezultatele au arătat că indicii de caracteristici utilizați pentru recunoașterea formei mâinii au fost cinci itemi: lungimea mâinii, grosimea la nivelul metacarpianului, lățimea palmei, distanța dintre degetul mare și cel arătător și distanța de la baza degetului mic până la centrul încheieturii mâinii. Utilizarea algoritmului RF pentru a reduce dimensionalitatea este mai eficientă; precizia medie de recunoaștere a modelelor de recunoaștere a formei mâinii este de 93,78% în medie; în comparație cu metodele de reducere PCA și VC, precizia medie a modelelor de recunoaștere a formei mâinii crește cu 19,17% și, respectiv, 14,86%. Rezultatele studiului pot oferi referințe metodologice pentru selecția obiectivă a indicatorilor caracteristici și recunoașterea morfologică a părților corpului uman.

Cuvinte-cheie: scanare 3D, clasificarea formei mâinii, algoritmul Random Forest, selecție de caracteristici, recunoașterea formei mâinii

INTRODUCTION

Studying human hand characteristics and classification is essential for size development, hand product design, fit research, and pattern optimization [1]. Hand size acquisition is the basis of hand feature and classification research. There are mainly manual contact measurements, non-contact 3D scanning, and picture measurement methods to acquire hand size.

With the evolution of scientific instruments, non-contact 3D body scanning is more standardized and faster than a traditional tape measure. Since the scanners can measure more parameters, such as the angle, thickness, width, height, volume, and curvature of the human body accurately [2–4], this method is believed to be the most suitable method to obtain anthropometric data in the shortest time and at the cheapest cost [5]. However, 3D body scanning also

presents the problem that the complexity of parameter dimensionality while acquiring multi-dimensional parameters makes it difficult to determine the main feature parameters effectively. To improve the interpretability and accuracy of hand morphology recognition, extracting key feature indicators from multidimensional hand parameters is a critical issue in hand morphology research.

Currently, research on hand morphology classification recognition is gradually increasing, mainly in terms of PCA (Principal Component Analysis). Jee et al. used factor analysis to identify three main factors from 21 hand dimensions, hand width, palm length, and finger length, giving descriptive statistics of hand dimensions in Korea [6]. Liu et al. used the PCA method to select five hand characteristics factors from 37 hand measurements for the hand characteristics study [7]. Fan et al. measured 24 hand dimensions and used PCA to dimensionalize the hand measurements and perform K-means cluster analysis to classify the hand morphology of 328 samples into five categories: small, short and fat, standard, long and slender, and wide, to provide a reference for hand product design [1]. All of the above methods in hand morphology classification and identification research idea to use PCA first to reduce the dimensionality of hand measurements and determine feature indicators for classification and identification, in which the cumulative contribution of principal component factors are all below 80%, indicating that at most 80% of hand, features can be explained. Since PCA can eliminate correlations between variables, provided that such correlations are assumed linear, good results are not obtained for nonlinear dependencies [8]. Zhang et al. to reduce the high-dimensional feature index for chest morphology characterization, CV (Coefficient of Variation) was used to eliminate the influence of measurement scale and magnitude, parameters with CV values greater than 10% were selected for chest morphology analysis [9]. Zhang et al. measured 22 characteristic parameters related to neck and shoulder morphology in men and then selected anterior tilt angle, dorsal entry angle, oblique shoulder angle, neck-shoulder width ratio, and transverse sagittal neck diameter ratio as cluster analysis variables based on CV analysis to classify

the neck and shoulder morphology and summarize the discriminant rules [10]. The above methods extract variables from multidimensional parameters but require multiple screening and comparison of the extracted feature indicators. Zhou et al. used an RF algorithm to quantify the importance of female breast feature indicators and determine the best breast features for classification and identification. The results showed that dimensionality reduction by RF algorithm is more effective than traditional dimensionality reduction methods such as PCA [5]. CV and PCA methods require multiple screening when extracting feature metrics, and this study expects to determine the metrics of effective hand features directly through dimensionality reduction by RF algorithm.

Therefore, to determine the effective hand features applicable to hand morphology recognition, this paper will start from hand 3D point cloud data, use K-Means for clustering analysis, RF algorithm to quantify the importance of hand feature indexes, and at the same time, build BP (backpropagation), RBF (Radial Basis Function), SVM (Support Vector Machines), RF four supervised machine learning mathematical models, hand model recognition accuracy as a benchmark, compare and analyse the accuracy of hand shape recognition with feature indicators determined by RF algorithm, PCA, VC method, further verify the RF algorithm dimensionality reduction effectiveness, provide reference for hand feature indicator selection and hand shape recognition.

MATERIALS AND METHODS

Participants and measurement

To ensure the representativeness and independence of the sample, the experiment randomly sampled 240 female college students aged 18–25 without physical activities, diseases, and other particular hand morphology as the study subjects. To exclude abnormal data from the analysis, values outside the range of “the mean difference $\pm 3\sigma$ ” (σ : SD of differences) were considered weird and excluded from the study [11]. After excluding 3.33% of the samples, the adequate sample size was determined to be 232 people. Refer to GB/T 16252-1996 “Adult Hand Size”, GB/T 5703-2010 “Anthropometric Basic Items for Technical



Fig. 1. Schematic diagram of hand measurement items

Design”, 33 hand measurement items, including hand length, were determined, as shown in figure 1.

Measuring

The results of manual contact measurements are susceptible to subjective factors of the tester and the muscle tissue undergoes different degrees of extension and contraction movements during the measurement process, resulting in measurement errors [12]. Therefore, this study used the non-contact 3D scanning method to measure the hand dimensions. According to GB/T 23698-2009 “General Requirements for 3D Scanning Anthropometric Methods”, the temperature of the experimental environment was $27\pm 3^{\circ}\text{C}$, and the relative humidity was $(60\pm 10)\%$. A multifunctional handheld 3D scanner (EinScan, Xianlin, China, measurement accuracy 0.04 mm) was used to capture the 3D point cloud of the hand. Through point cloud smoothing, rounding, and filling, solve the noise and missing point clouds generated by slight shaking during scanning, and get smooth and complete point cloud data. In this paper, the measurement method of handheld 3D scanning is divided into the following three steps:

Auxiliary point, line and surface definitions: To ensure the stability of repeated measurements, the standardization of auxiliary points, lines, and surfaces before determining the characteristic points, calibration points, lines, and surfaces as shown in figure 2 (finger part to the middle finger as an example).

Hand feature point determination: After determining the required auxiliary points, lines, and surfaces according to the above steps, the hand feature points are determined by combining GB/T 16252-1996 and GB/T 5703-2010. To better analyse the local morphology, such as fingers, the coordinate system needs to be reconstructed for the finger part. Take the middle finger for example. The feature points are calibrated as shown in figure 2. Similarly, other hand width and circumference feature points can be determined.

Hand size measurements: A handheld 3D scanner scans the hand, and 33 hand feature parameters are measured based on the processed hand 3D point cloud data by the step-by-step method described above.

Extraction of hand features

To extract key feature indicators from multidimensional hand parameters for hand morphology recognition and consider the discrete and nonlinear hand feature part size, this paper uses the RF algorithm to determine the indicators of practical hand features. The RF algorithm can quantify the importance of hand features and simplify the calculation process, improving the efficiency and objectivity of the results [5]. RF trains different base classifiers according to each subset and determines the final classification result by the results of the base classifier voting. Gini (Gini index) or OOB (out-of-bag) error rate was used as a metric to evaluate the contribution of each feature in the training process [13]. To determine the hand feature indicators for hand shape recognition, this study uses Gini to evaluate the contribution of feature indicators to the decision tree and the process of scoring the importance of 33 hand feature indicators by RF algorithm including Gini result output calculation, feature indicator importance calculation and feature indicator importance score normalization three stages, the result output expression is as follows

1. Gini results output calculation:

$$GI_q^i = 1 - \sum_{k=1}^K P_{qk}^2 \quad (1)$$

2. The importance of feature indicator X_j at node q of the i decision tree is:

$$VIM_{iq}^{G(i)} = GI_q^i - GI_{l_q}^i - GI_{r_q}^i \quad (2)$$

Suppose there are l decision trees in the RF and the importance of the feature indicators X_j is:

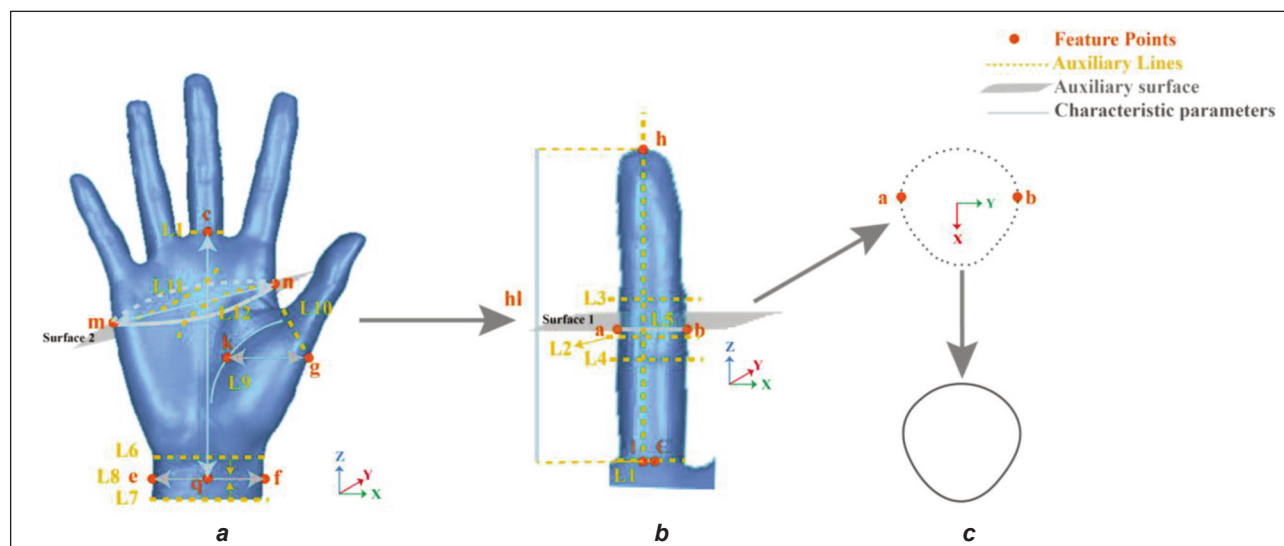


Fig. 2. Determine the hand characteristic parameters: a – palm measurement method; b – middle finger measurement method, c – middle finger perimeter section curve fitting

$$VIM_j^{GI} = \sum_{i=1}^I VIM_j^{GI(i)} \quad (3)$$

3. Normalization of importance scores of feature indicators:

$$VIM_j = \frac{VIM_j^{GI}}{\sum_{j=1}^{33} VIM_j^{GI}} \quad (4)$$

Recognition models for hand shapes

Supervised learning finds the mapping relationship between input and output from the already labelled training set. It applies this mapping relationship to the unknown data for classification and prediction [14]. Therefore, this study used the BP neural network, RBF neural network, SVM support vector machine, and RF algorithm to construct four hand morphology recognition models. The model parameters were set as shown in table 1.

RESULT AND DISCUSSIONS

k-Means clustering analysis

To explore the hand morphology classification, the optimal number of clusters needs to be determined. The hypothesis of normal distribution was first tested on the sample, and the data of 33 variables were tested to be approximately normally distributed. Take the length of the hand as an example, a histogram of Gaussian distribution is obtained, as shown in figure 3, a.

In this study, a mixed F-statistic was used to determine the number of clusters. The mixed F statistic is abbreviated as, and the larger the value calculated means that all variables are more closely related within classes and more dispersed between classes, so the number of categories corresponding to the largest is the optimal number of categories [15].

By calculating the FMixed value corresponding to each classification number obtained, it can be seen from figure 3, b that when the FMixed value is the largest, the corresponding C = 3, so the best classification of 232 female college students' hands is in 3 categories.

Due to the large amount of data obtained from hand measurements, the fast sample clustering (K-Means) method was applied to perform fast clustering analysis on hand data in order to facilitate the exploration of hand morphology classification [16]. The hand morphology and characteristics were analysed based on the clustering index. After ten iterations, the clustering centre converged. Finally, the hand morphology was divided into three categories and named according to the characteristics of each type of hand shape, which were wide and large, slender and short, and fat, as shown in table 2.

To analyse the clustering results more clearly, the clustering centres of the three classes of hand types are described with the corresponding shapes. Among them, the three types of hand morphology contour line pairs are shown in figure 4, a. The three types of hand morphology intermediate models and intermediate values are shown in figure 4, b.

Table 1

MATHEMATICAL MODEL PARAMETER SETTING					
Models	Parameter	Value	Models	Parameter	Value
BP	Hidden nodes	11	SVM	Penalty Factor	10
	Input transfer function	Tansig	RBF	Layer	3
	Output transfer function	Purelin		Number of neurons	190
	Layer	3		Kernel functions	newrbe
	Expectation error	0.01	RF	Number of decision trees	500
SVM	Kernel functions	rbf		Minimum number of leaves	5

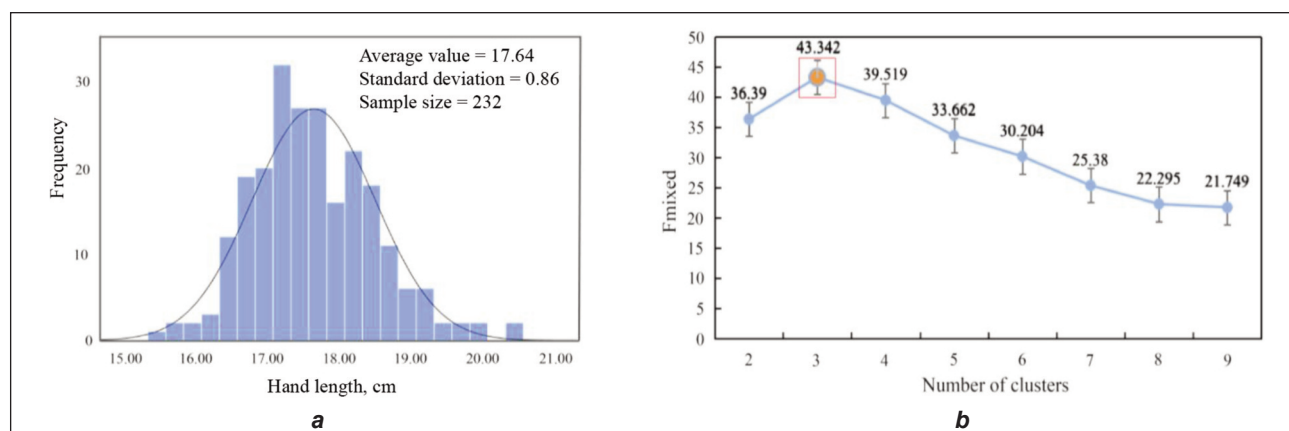


Fig. 3. Graphs of: a – histogram of Gaussian distribution; b – the mixed F-statistic

HAND SHAPE CHARACTERISTICS AND PROPORTION		
Name	Characteristics	Proportion (%)
Wide large	The palm length is 60% of the hand length, the palm is relatively wide, and the overall size is large.	22.5
Slender type	The hand length, width, and circumference dimensions are smaller, the length dimension is larger, the palm is flatter, and the hand morphology is slender.	40.2
Short fat type	The hand features shorter length dimensions but larger circumference, width, and palm thickness and a short, fat hand morphology.	37.3

Importance assessment of the hand features

The RF algorithm performs feature metric importance quantification to evaluate the importance of hand feature metrics. It mainly includes three steps of RF algorithm parameter setting, feature index importance ranking, feature index determination, ranking feature index importance, and dividing the samples into a training set and test set in the ratio of 8:2, which are represented as follows:

1. Decision tree, number of branches determined

Combined with the RF algorithm decision tree's minimum tree in other machine learning datasets, this study keeps the number of branch variables constant. It determines the decision tree forest size $N_{trees} = \{50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000\}$. The RF algorithm prediction results were compared with the classification results, and the best decision tree value of 500 was finally determined, and the results are shown in figure 5, a. Determine the number of branches using the square root of the characteristic index [17]. Therefore, $d = 33$ and $M = 5$

to 6 in this study. The number of branches was determined to be five by changing the size of the number of branches and comparing the predicted results of the RF algorithm with the classification results. The results are shown in figure 5, b.

2. Characteristic index determination

The RF algorithm calculated the hand feature index importance score; the results are shown in figure 5, c. The importance scores of five indicators, namely, hand length, the distance between the thumb and index finger, thenar width, thickness at metacarpal, and distance between the thumb and index finger, were more significant than 0.4. The five indicators reflected information on hand length characteristics, thumb characteristics, and palm characteristics. There were 15, 7, and 5 hand feature indicators with importance scores in the range of $(0.0, 0.2]$, $(0.2, 0.4]$, and $(0.4, 1.4]$, respectively. The top 15% (top 5), 36% (top 12), and 82% (top 27) sets of hand feature indicators were determined based on the feature indicator importance scores.

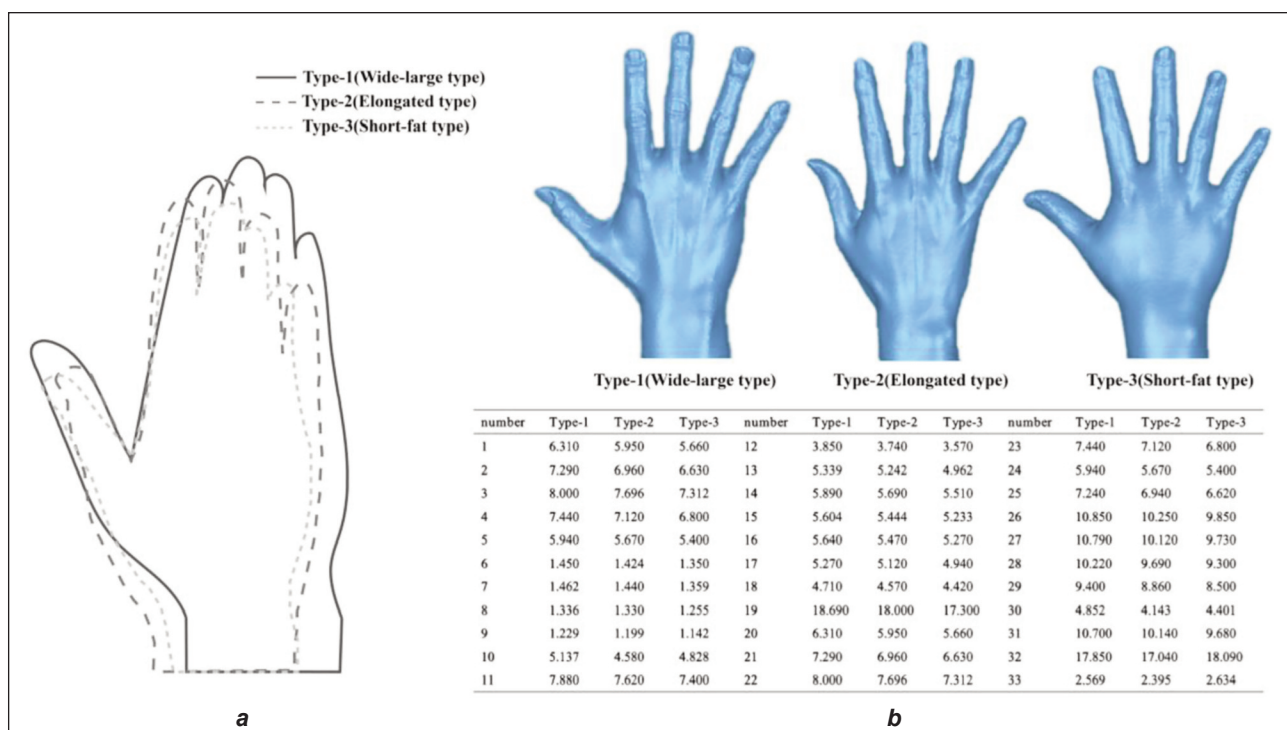


Fig. 4. Comparison of three types of hand morphology: a – shape comparison of three hand types; b – three types of subdivided hand shape

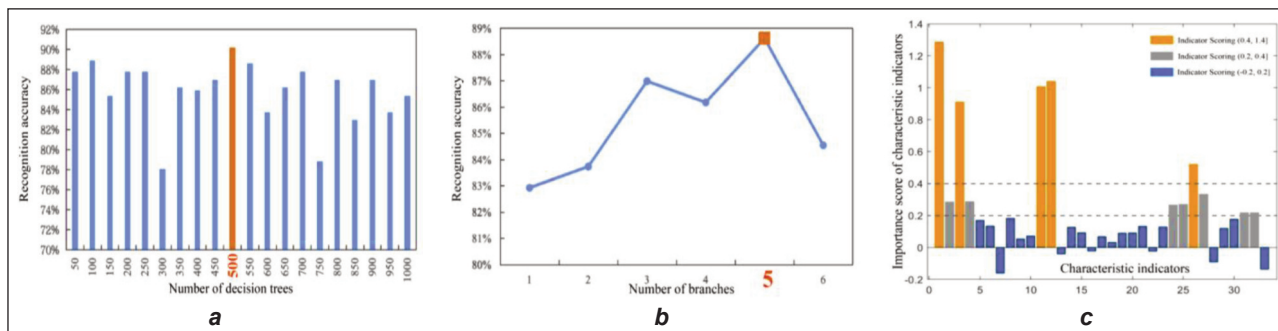


Fig. 5. Importance assessment of the hand features: *a* – recognition accuracy with different numbers of decision trees; *b* – recognition accuracy with different numbers of branches; *c* – index importance score

Validation of the hand feature selection

To verify the best feature index dimension of the RF algorithm for dimensionality reduction, this study used the BP neural network, RBF neural network, RF algorithm, and SVM algorithm to construct four hand shape recognition models, respectively, and the hand shape recognition model accuracy results are shown in table 3 and figure 6.

Combined with t-SNE algorithm 3D visualization to evaluate the dimensionality of feature indicators and RF and other three dimensionality reduction methods (RF, PCA, CV) for hand morphology recognition and analyse the degree of sample dispersion and crossover. t-SNE is a nonlinear dimensionality reduction

algorithm that reduces multidimensional data to 2D and 3D visualization more effectively [18]. Analysis of figure 7 shows that when five hand feature indicators are obtained by the RF algorithm for hand shape recognition through dimensionality reduction, the 3 class hand shape boundaries are the most obvious, the distance between the cluster centres is far, and the sample crossover is the smallest, indicating that hand shape recognition is the most effective. Due to the discrete and nonlinear nature of hand size data, the RF algorithm quantifies the contribution value of each feature metric on the decision tree when performing feature metric importance assessment. The importance ranking of the feature indicators is derived by comparing and analysing the

Table 3

HAND SHAPE RECOGNITION ACCURACY OF DIFFERENT FEATURE INDEX DIMENSIONS					
Characteristic index dimension	Recognition accuracy (%)				Average recognition accuracy (%)
	BP	SVM	RBF	RF	
5 (RF)	97.43	90	95	92.68	93.78
12 (RF)	87.80	82.50	77.50	92.68	85.12
27 (RF)	82.94	67.50	58.30	78.86	71.90
33	90.23	70	52.50	78	72.68

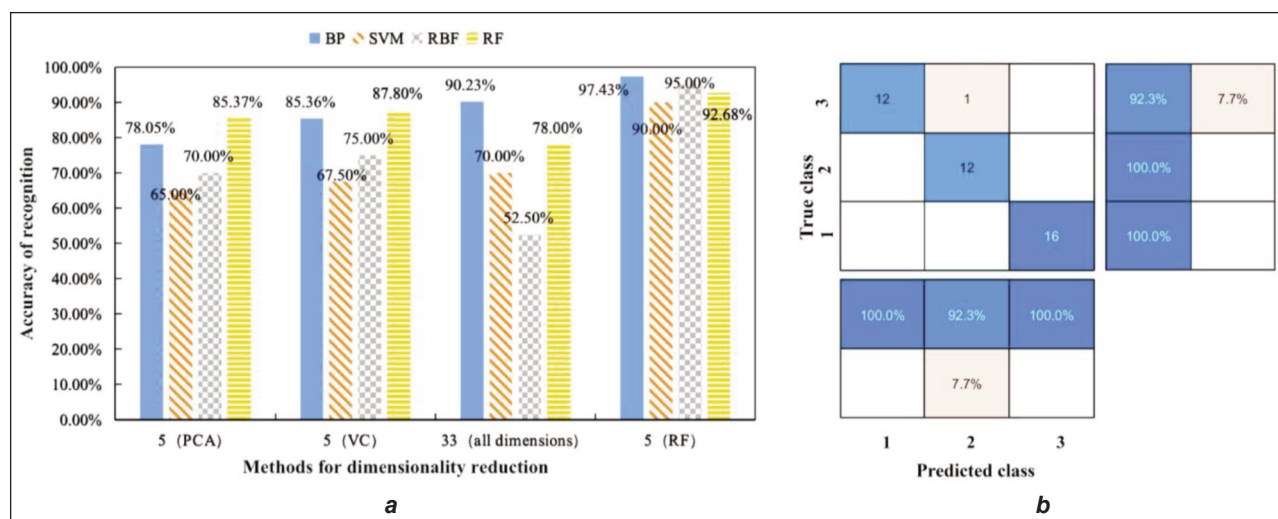


Fig. 6. The accuracy of hand shape recognition by neural networks: *a* – hand shape recognition accuracy with different feature indexes; *b* – confusion matrix of BP hand shape recognition model

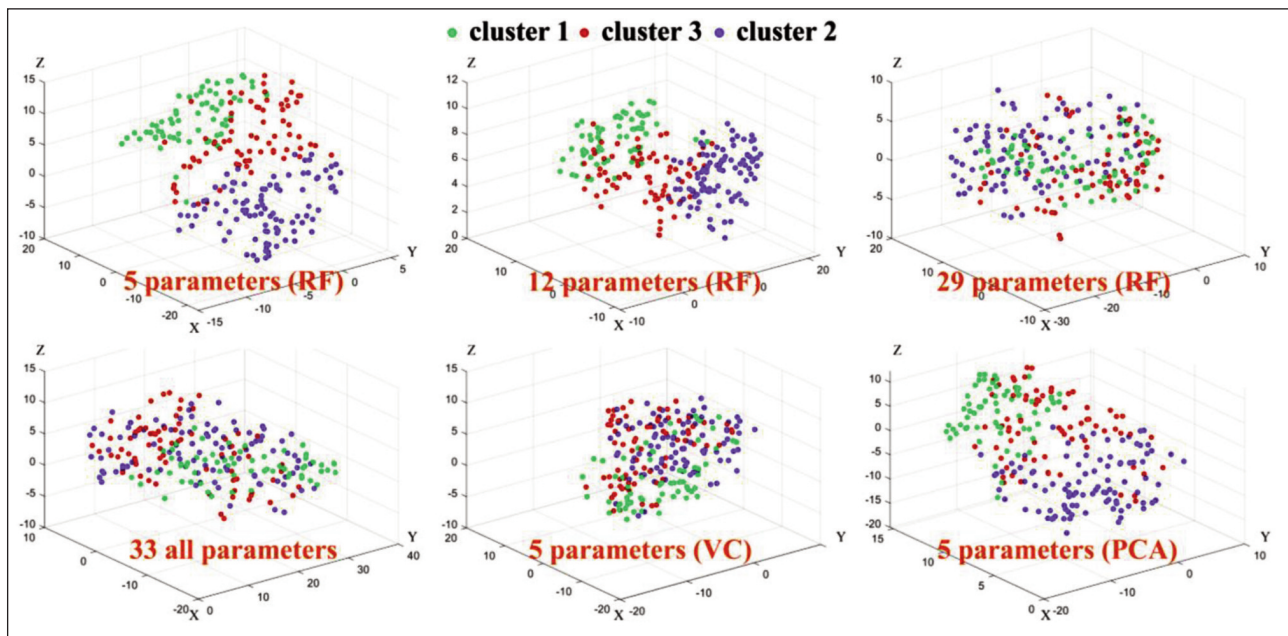


Fig. 7. The accuracy of hand shape recognition by neural networks

magnitude of the contribution value of each feature indicator, so the extracted feature indicators are more objective.

CONCLUSION

To improve the effectiveness of hand feature index selection and hand shape recognition accuracy, this study uses an RF algorithm to evaluate the importance of hand feature indexes. The main conclusions drawn from this study are:

1. The optimal number of clusters was determined to be 3 using the F mixed statistic, and three categories of hand morphology were determined by the K-Means clustering method: short and fat, slender and broad. Among them, the short fat type hand features shorter length dimensions but larger circumference, width, and palm thickness; the slender type hand has smaller width and circumference dimensions and larger length dimensions; the wide large palm is flatter palm is wider and has larger overall dimensions.
2. Dimensionality reduction of the hand feature indicators by RF algorithm to determine the feature indicator dataset as 5, 12, 27; Four hand shape recognition models of BP, RBF, SVM, and RF were constructed, and the best feature index dataset was determined by comparing and analysing the

feature index dataset 5 (hand length, distance between the thumb and index finger, thenar width, thickness at metacarpal, and distance between the thumb and index finger). The four models achieved the highest average recognition accuracy of 93.78%.

3. Comparing and analysing the accuracy of RF, PCA, and VC methods to determine feature indicators and 33 total indicators for hand shape recognition, the highest accuracy of hand shape recognition was achieved using the RF algorithm for dimensionality reduction. By comparing and analysing the recognition accuracy of four recognition models, it can be seen that the highest recognition accuracy of the BP neural network hand shape recognition model is 97.43%, which is 2.43%, 7.43%, and 4.75% higher than RBF, SVM, and RF hand shape recognition models, respectively.

The purpose of this paper is to improve the effectiveness of hand characteristic index selection and hand shape recognition accuracy, but due to different regions, different physical movements, and other factors are prone to cause hand morphology differences, further explore the main factors affecting hand morphology differences, expanding the applicability of the method.

REFERENCES

- [1] Fan, D.Y., Ma, X.M., Wang, L.J., *Hand measurement and analysis based on image and mark watershed algorithm*, In: International Journal of Clothing Science and Technology, 2021, 33, 5, 822–842
- [2] Zheng, R., Yu, W., Fan, J., *Development of a new Chinese bra sizing system based on breast anthropometric measurements*, In: International Journal of Industrial Ergonomics, 2007, 37, 8, 697–705
- [3] Pei, J., Fan, J., Ashdown, S.P., *A novel method to assess breast shape and breast asymmetry*, In: Journal of the Textile Institute, 2018, 110, 8, 1129–1240
- [4] Pandarum, R., Yu, W., Hunter, L., *3-D breast anthropometry of plus-sized women in South Africa*, In: Ergonomics, 2011, 54, 9, 866–875

- [5] Zhou, J., Mao, Q., Zhang, J., Newman, M.L., Chen, J.M., *Selection of breast features for young women in northwestern China based on the random forest algorithm*, In: Textile Research Journal, 2022, 92, 24, 5095–5105
- [6] Jee, S., Yun, M.H., *An anthropometric survey of Korean hand and hand shape types*, In: International Journal of Industrial Ergonomics, 2016, 53, 10–18
- [7] Liu, P., Fu, D.N., Jiang, S.Q., Chen, H., Sheng, S.J., *A survey of Chinese young medical students' hand anthropometry Chinese*, In: Journal of Ergonomics, 2017, 23, 3, 60–66
- [8] Olive, D.J., *Robust principal component analysis*, In: Springer, Cham, 2017
- [9] Zhang, B.B., Sun, Y., Zhong, Z., Zhong, Z.J., Xia, Y.P., Jin, S.N., Li, Y.Q., *Breast shape classification and discrimination driven by local features-focusing on Chinese women in their 20s*, In: International Journal of Industrial Ergonomics, 2022, 90, 103304
- [10] Zhang, J., Xu, K.Y., Zhao, X.L., Gu, B.F., *Classification and recognition of young males' neck-shoulder shape based on 2-D photos*, In: Journal of Textile Research, 2022, 43, 5, 143–149
- [11] Jin, J.F., Yang, Y.C., Zou, F.Y., *Developing an intelligent model for the construction a hip shape recognition system based on 3D body measurement*, In: Fibres & Textiles in Eastern Europe, 2016, 5, 119, 110–118
- [12] Chi, C., Zeng, X., Bruniaux, P., Guillaume, T., *A study on segmentation and refinement of key human body parts by integrating manual measurements*, In: Ergonomics, 2022, 65, 1, 60–77
- [13] Xu, S.C., Zhang, H.C., *Weighted feature selection algorithm based on random forest*, In: Statistics & Decision, 2018, 34, 18, 25–28
- [14] Li, P., Wang, H., Bhm, C., Shao, J., *Online Semi-supervised Multi-label Classification with Label Compression and Local Smooth Regression*, In: Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence IJCAI-PRICAI-20, 2020
- [15] Yin, L., Xia, L., Xu, C., *Female body shape prediction based on random forest*, In: Journal of Textile Research, 2014, 35, 5, 113–117
- [16] Jing, L., Ng, M.K., Huang, J.Z., *An Entropy Weighting k-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data*, In: IEEE Transactions on Knowledge & Data Engineering, 2007, 19, 8, 1026–1041
- [17] Domb, M., Bonchek-Dokow, E., Leshem, G., *Lightweight adaptive Random-Forest for IoT rule generation and execution*, In: Journal of Information Security and Applications, 2017, 34, 2, 218–224
- [18] Rauber, P.E., Falc, A.X., Telea, A.C., *Visualizing time-dependent data using dynamic t-SNE*, In: Eurographics. Eurographics Association, 2016

Authors:

XIN LI¹, XIAO-JUN DING^{1,2,3}, ZHOU-YAN PENG¹, XI-YAN LIN¹, FENG-YUAN ZOU^{1,2,3}

¹Fashion College, Zhejiang Sci-Tech University, Hangzhou, 310018, China

²Key Laboratory of Silk Culture Inheriting and Products Design Digital Technology, Ministry of Culture and Tourism, Zhejiang Sci-Tech University, 310018, China

³Clothing Engineering Research Center of Zhejiang Province, Zhejiang Sci-Tech University, 310018, China

Corresponding author:

FENG-YUAN ZOU
e-mail: zfy166@zstu.edu.cn