Body Shape Recognition and Prototype Construction Based on LightGBM Algorithm

Lanmin WANG[a,c], Hongmin WANG[d], Li LIU[a, b,1], and Hao LIU[a,c,2]

[a] *School of Textile Science and Engineering, Tiangong University, Tianjin, China*
[b] *School of Fashion Art and Engineering, Beijing Institute of Fashion Technology, Beijing, China*
[c] *Institute of Smart Wearable Electronic Textiles, Tiangong University, Tianjin, China*
[d] *State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou, China*

Abstract. The regional characteristics of middle-aged women in Xinjiang are obvious, and the phenomenon of incompatible clothing is prominent. In order to avoid the information loss of principal component analysis or factor analysis during body type classification, this study uses the LightGBM algorithm to establish a body type recognition model based on the results of K-Means clustering, and compares it with random forest and linear regression recognition models. It is found that the LightGBM model has the best body type recognition effect. it is good. Through correlation analysis and regression analysis, regression equations between variables are obtained. Based on the new cultural prototype, Liu Ruipu women's clothing prototype, and Donghua prototype, a clothing prototype suitable for middle-aged women in Xinjiang is established. Through virtual fitting evaluation and actual fitting, the experimental prototypes established based on the Donghua prototype and the Liu Ruipu prototype can meet the needs of middle-aged women for the comfort and aesthetics of clothing. The experimental prototype established by the Donghua prototype has the best effect.

Keywords. Middle-aged women, K-Means clustering, body shape recognition, LightGBM algorithm, clothing prototype

1. Introduction

After entering middle-age women, due to reduced metabolism and fertility, subcutaneous fat continues to accumulate, and Xinjiang is affected by geographical environment and eating habits. As a result, middle-aged women in Xinjiang have a larger carcass than other parts of China, and their body shape is regionally characteristic. obvious. In recent years, some scholars have conducted extensive measurement and body type classification research on different genders and age groups in different regions of China, but the research on the body types of middle-aged women in Xinjiang is still at a blank

---

stage.

Body type classification has a great influence on clothing plate making and clothing size coverage, and the accuracy of body type classification method affects clothing fit. There are various body type classification methods. Cluster analysis is a commonly used method in body type research, but some information will be lost when principal component analysis or factor analysis and the highest feature of the load matrix in the principal component are selected, which makes the evaluation lack of objectivity. Moreover, the interpretation of principal components is usually vague, not as clear and precise as the meaning of the original variables. This is the price that must be paid for dimensionality reduction.

In recent years, the continuous update and iteration of machine learning algorithms has promoted the rapid development of artificial intelligence technology [1]. Machine learning algorithms can quickly process large-scale data, have high computational efficiency and excellent model effects. They have been widely used in various fields to solve real-life and academic problems [2-4].

LightGBM belongs to the Boosting algorithm and is an improved version of the GBDT (Gradient Boosting Decision Tree) algorithm framework [5]. The LightGBM algorithm has the advantages of high training efficiency, low memory usage, high accuracy, support for parallel learning and processing large-scale data [6]. This article uses the LightGBM algorithm to explore the body shape of middle-aged women in Xinjiang and establish a body shape recognition model. On the basis of body type classification, a clothing prototype suitable for middle-aged women in Xinjiang is established.

2.   Material Preparation

2.1.   *Anthropometric Experiment*

The contact measurement was conducted on middle-aged women aged 45-59 in Xinjiang by random sampling. The experimental locations were Urumqi, Altay, Aksu, and Ili Kazakh Autonomous Region. Measuring instruments include altimeter, weight gauge, soft ruler, Martin gauge, angle gauge, rod gauge, etc. [7]. Refer to GB/T5703-2010 "Anthropometric Basic Projects Designed Using Technology" [8] to determine the experimental measurement items. According to the formula 1, the sample size required for the largest measurement site is 486.20 people. Considering noise factors such as errors, it is reasonable to determine the total number of samples as 537 people.

$$N = \mu_\alpha^2 * (\frac{c \cdot v}{A})^2, \quad c \cdot v = \frac{\sigma}{\bar{X}} \tag{1}$$

N is the total sample size, μ_α is the horizontal probability; when the confidence level α is 0.05, μ_α=1.96; A is the relative guarantee error. In order to improve the accuracy of the research, A is set to 1.6%; $c \cdot v$ is the coefficient of variation , σ is the sample standard deviation, $\bar{X}$ is the sample mean [9].

Taking the clothing structure needs and the standard deviation of the characteristics as the standard, the weight, height, cervical vertebra point height, waist height, breast height, back length, upper thigh length, arm length, front and rear length of the upper crotch, total shoulder width, chest width, back width, and total Shoulder width, neck-root circumference, bust circumference, waist circumference, chest-waist difference, abdomen circumference, hip circumference, hip-waist difference, arm circumference, shoulder oblique angle, and side angle are defined as characteristics.

2.2.   Data Preprocessing

Python is used to perform singular value and vector normalization dimensionless processing on the original data. The dimensionless processing of the original data eliminates the difference in dimension and dimension unit, which provides a guarantee for the accuracy and scientificity of the follow-up research.

2.3.   Research Tools

Scikit-learn is an open source machine learning package based on the Python language. This research is done by using the "Scikit-learn" learning package of the Python programming software.

2.4.   Body Type Recognition

2.4.1.K-means clustering

K-means clustering is an unsupervised machine learning algorithm [10]. Dividing a limited data set into multiple clusters is concise and efficient, and is widely used in machine learning and other fields. The principle of clustering algorithm is shown in figure 1 [11].
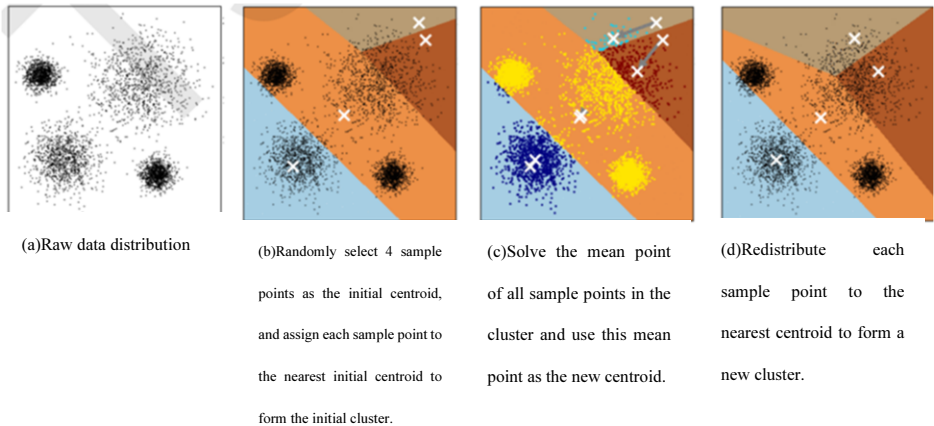
| (a)Raw data distribution | (b)Randomly select 4 sample points as the initial centroid, and assign each sample point to the nearest initial centroid to form the initial cluster. | (c)Solve the mean point of all sample points in the cluster and use this mean point as the new centroid. | (d)Redistribute each sample point to the nearest centroid to form a new cluster. |

Figure 1. K-Means clustering principle

The clustering algorithm aims to pursue "small differences within clusters and large differences outside clusters". In Sklearn, the Euclidean distance formula is used to measure this difference (see the following formula 2), which is the sum of the squares of the distances from all sampling points in the cluster to the centroid.

Add the sum of squares within the clusters of all clusters contained in the data set to get the sum of squares, as shown in the following formula 3 The smaller the total sum of squares, the more similar the samples in each cluster, the better the cluster effect [12]. When the total sum of squares (formula 4) . is the smallest, the center of mass remains unchanged.

The clustering effect depends entirely on the density of different categories and the degree of dispersion between different categories. In this study, the contour coefficient

(the following formula 5) was used as an evaluation index for the clustering effect of middle-aged women in Xinjiang. The range of the contour coefficient is [-1, 1]. The larger the contour coefficient, the closer the sample is to other samples in the cluster, and the farther the distance from other clusters is. The total profile coefficient is the average of the profile coefficients of all samples.

$$d(x,\mu) = \sqrt{\sum_{i=1}^{n}(x_i - \mu_i)^2} \qquad (2)$$

Where: x represents a sampling point in the cluster, μ represents the centroid in the cluster, n represents the number of elements in each sampling point, and i represents each feature of the constituent points.

$$\text{Cluster sum of square (CSS)} = \sum_{j=1}^{m}\sum_{i=1}^{n}(x_i - u_i)^2 \qquad (3)$$

Where: m is the number of samples in the cluster, and j is the number of each sample.

$$\text{Total cluster sum of square} = \sum_{i=1}^{k} CSS_i \qquad (4)$$

$$S = \frac{b-a}{max(a,b)} \qquad (5)$$

Where: a represents the similarity between a sample and other samples in the cluster, and b represents the similarity between the sample and the samples in other clusters.

$$\Delta\,\text{loss} = \frac{S_L^2}{n_L} + \frac{S_R^2}{n_R} - \frac{S_P^2}{n_P} \qquad (6)$$

Where: S_R is the gradient of the right node, SL is the gradient of the left node, S_P is the number of samples of the parent node, n_R is the number of samples of the right node, n_L is the number of samples of the left node, and n_P is the sample of the parent node Number.

2.4.2. LightGBM Algorithm

In the tree growth algorithm, LightGBM uses leaf growth (as shown in figure 2) as the tree growth strategy, which aims to reduce the loss of the model and maximize [13]. Find the leaf with the largest information gain from all the current leaves to split, and loop in turn to reduce the training error and improve the accuracy of the learner. In this study, the maximum depth of the tree is set to 3 to achieve the purpose of preventing over-fitting while improving the accuracy.

In the selection of segmentation nodes, the histogram algorithm is selected (figure 3). The basic idea is to discretize continuous eigenvalues into K integers, and at the same time construct a histogram with a width of K [14]. In order to improve the running speed, the histogram is used to make the difference (figure 4), and the histogram of the requested node is obtained by subtracting the histogram of the sibling node from the parent node [15], and the speed is doubled.

The steps to determine the optimal segmentation point are: first traverse all bins. Then, using the current bin as the dividing point, accumulate the gradient from the bins on the left to the current bin, S_L and the number of samples n_L, and use the histogram to make the difference, and obtain the gradient on the right by SR=SP-SL; n_R=n_P-n_L And the number of samples, the gain is calculated by the following formula6 [16]. Finally, the maximum gain is obtained in the traversal process, and the feature and bin feature value at this time are used as the feature and value of the split node.

The smaller the gradient of the sample, the smaller the training error of the sample. In LightGBM, the unilateral sampling method is used for adaptation. The sampling strategy retains all large gradient samples, and random sampling is performed on small gradient samples [20], which can be used without changing the original In the case of data distribution, the performance of the algorithm is improved, and the impact of sampling on the data distribution is reduced.

LightGBM uses the EFB (Exclusive Feature Bundling) algorithm to merge mutually exclusive features to reduce the feature dimension. The time complexity spent to construct histogram has also changed from O(#data*#feature) to O(#data*#bundle), where #feature≤#bundle[17].

LightGBM has the characteristics of supporting efficient parallelism, and currently supports feature parallelism and data parallelism. Feature parallelism is the process of finding the optimal partition point in the parallelized decision tree. When dividing the features, each worker finds the local optimal segmentation point, and uses point-to-point communication to find the global optimal segmentation point (figure 5). The goal of data parallelism is to parallelize the entire decision-making learning process. LightGBM uses the PV-Tree algorithm for data parallelism (figure 6). Each worker owns part of the data, builds a histogram independently and finds the top-k optimal partition feature. The central worker aggregates to obtain the optimal 2K global partition features. Then collect the histogram of top-2K features from each worker, merge them to obtain the optimal division, and broadcast it to all workers for local division.

Use cross-validation to adjust the parameters of the LightGBM model. The cross-validation method is to divide the sample data into 5 parts, use 4 parts of the sample data for model training, and use the other part of the data for model evaluation, and perform K times of verification in a loop to find the average of the K-th mean square error. Represents the performance of the model under the parameter state. Figure 7 shows that as the sample size increases, the scores of the training set and the test set are getting higher and higher, and the training score is closer to the prediction score.

In order to optimize the learning curve, the generalization error is used to measure the learning rate of the model on the unknown data set. The generalization error is jointly controlled by deviation, variance, and noise, as shown in the following formula 7. The deviation is determined by the fit of the training set, the variance is determined by the stability of the model, and the noise is uncontrollable. The deviation indicates whether R^2 is high enough or SME is low enough. Variance represents the stability of the model's performance in different data sets.

$$E(f; D) = bias^2 + var + \in^2 \qquad (7)$$

In the formula: $E(f; D)$ is the generalization error; $bias$ is the deviation; *var* is the variance; $\in$ is the noise.

In order to construct a body type recognition prediction model, a parameter curve diagram of 500 trees was established, and it was found that the R2 value was the highest when the number of trees was 51. See figure 8.

When performing parameter training, the LightGBM model needs to try multiple sets of parameter combinations, from which the parameter combination with the best model performance is selected as the final training completion state of the model. In this paper, the method of grid search is used to train the model parameters. The meaning of the model parameters and the parameter values are: Test-size=0.2; subsample=1; Max-depth=3; eta=0; gamma=0; Colsample-bytree=1; Colsample-bylevel=1; nfold=5. The parameter combination with the best model performance is used as the final parameter combination for model training.
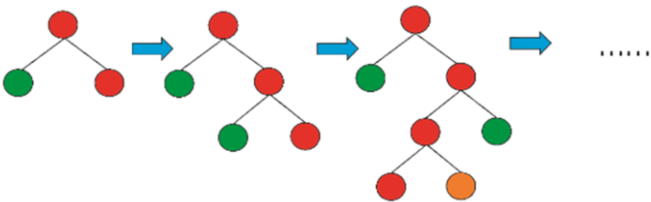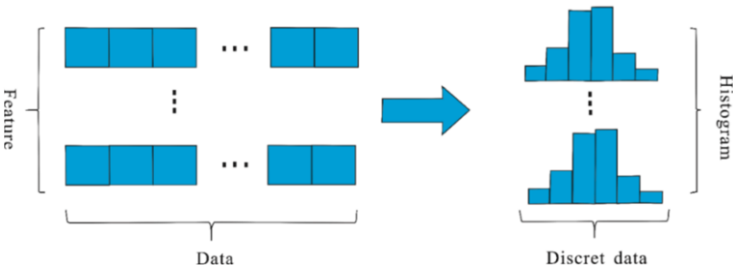
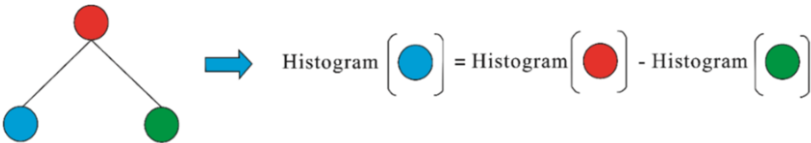**Figure 2.** Fitting diagram by leaf growth

Figure 3. Histogram algorithm
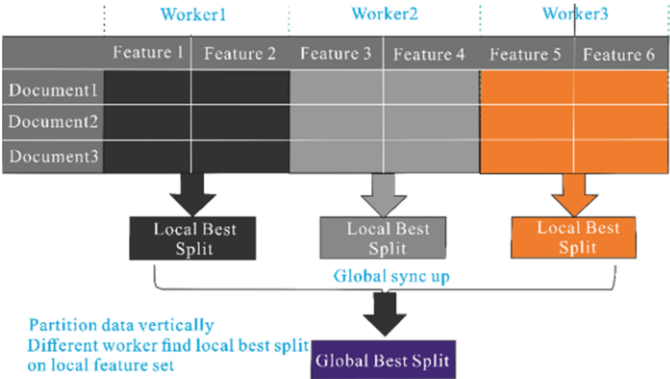

Figure 4. Histogram acceleration
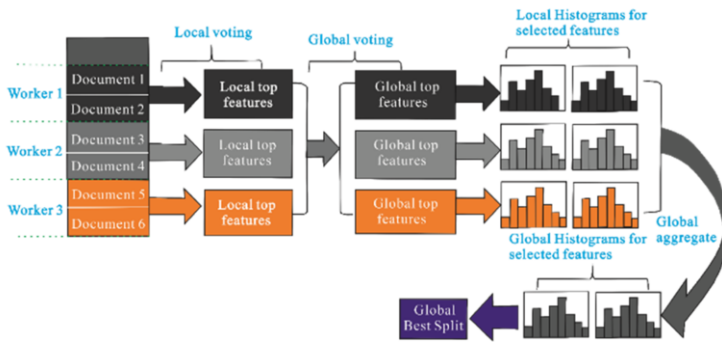

Figure 5. Feature parallel

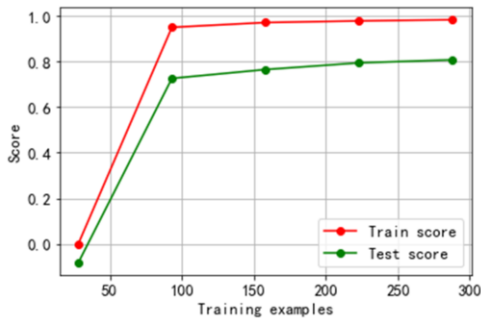**Figure 6.** Parallel algorithm voting of PV-Tree
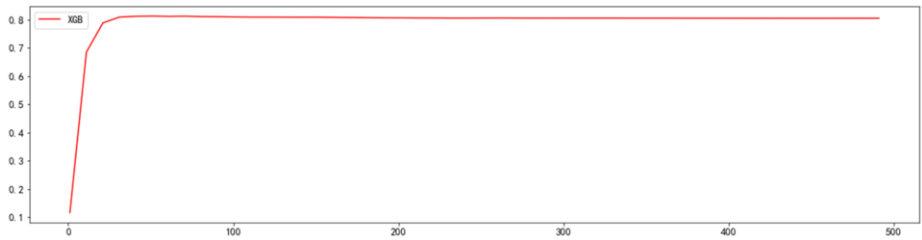

Figure 7. 5-fold cross-validation curve

**Figure 8.** N-estimate parameter selection curve

2.5. Proposed Prototype

It is necessary to analyze the correlation of characteristic variables before performing regression analysis. The correlation analysis of each important control part is shown in table 1 and table 2. The available length index is significantly correlated with the overall height, and the chest circumference is significantly correlated with the girth index and the width index.

To establish the linear regression equation of the prototype structure, it is necessary to analyze and verify the relationship of related characteristic variables. Taking the correlation between neck and root circumference and chest circumference of the body type 0 sample as an example, R^2 in table 3 shows that the impact of neck root circumference on chest circumference is 9.3%. Table 4 shows that the significance of the regression equation of neck-root circumference on chest circumference is less than 0.05, and the regression equation is meaningful. Table 4 shows that the chest circumference

coefficient is greater than 0, and the significance is 0, indicating that the chest circumference has a positive influence on the neck-root circumference.

New cultural prototypes, Donghua prototype and Liu Ruipu prototype are the three most commonly used in teaching and enterprises [18-19]. Table 5 is the revised prototype structure regression equation with reference to the new culture prototype, the Liu Ruipu prototype, and the Donghua prototype respectively. Figure 9 is a revised experimental prototype based on the body type 0 and body type 1 intermediates.
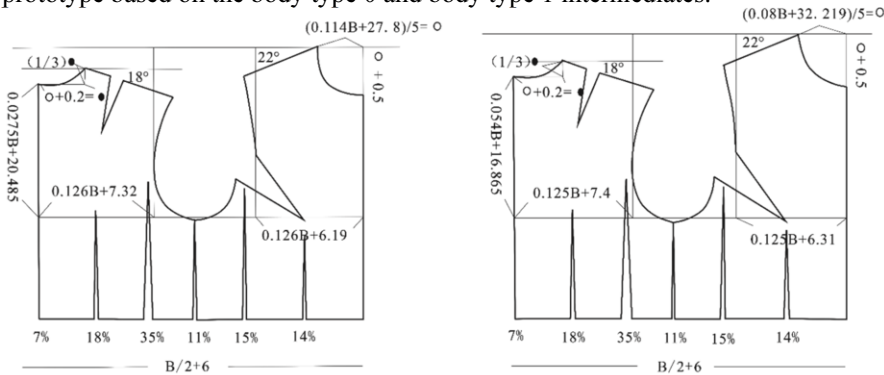
(a). Proposed version based on the new prototype (body type 0)

(b). Proposed version based on the new cultural cultural prototype (body type 1)
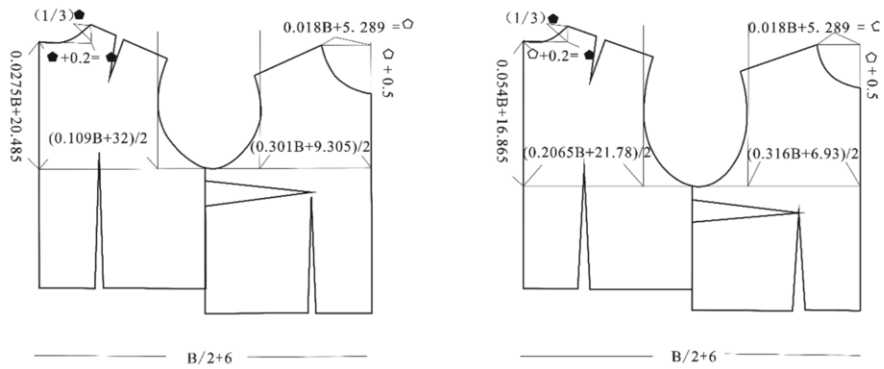
(a). Proposed version based on Liu Ruipu prototype (body size 0)

(b). Proposed version based on Liu Ruipu prototype (body size 1)
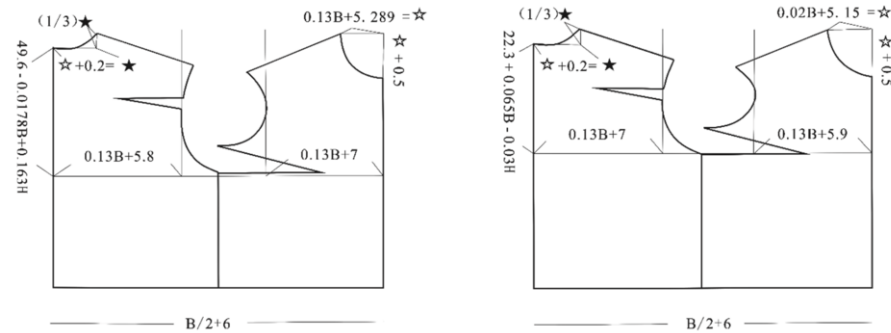
(a). Proposed version based on Donghua prototype (body size 0)

(b). Proposed version based on Donghua prototype (body size 1)

Figure 9. Proposed prototype

Table 1. Pearson correlation analysis table related to height

	height	Cervical point high	Breast height	Waist height	Back length	Arm length	Total shoulder width
height	1			.			
Cervical point high	.794**	1					
Breast height	.364**	.332**	1				
Waist height	.431**	.368**	.428**	1			
Back length	.077	.123**	.173**	-.008	1		
Strand length	-.017	-.130**	.235**	.228**	-.059		
Arm length	.469**	.406**	.152**	.155**	.096*	1	
Total shoulder width	.006	.019	.225**	.145**	.123**	-.027	1

Note: **. At the level of 0.01 (two-tailed), the correlation is significant; *. At the level of 0.05 (two-tailed), the correlation is significant.

Table 2. Pearson correlation analysis table related to bust circumference

	chest circumference	Neck-root circumference	waist	Chest width	Back width	Front waist length	Hips
chest circumference	1						
Neck-root circumference	.267**	1					
waist	.144**	.726**	1				
Chest width	.364**	0.162	.218**	1			
Back width	.532**	.194**	.354**	.426**	1		
Front waist length	.501**	.657**	.164**	0.217	0.107	1	
Hips	.264**	.754**	.657**	.164**	.326**	.345**	1

Note: **. At the level of 0.01 (two-tailed), the correlation is significant; *. At the level of 0.05 (two-tailed), the correlation is significant.

Table 3. Model summary

Model	R	R^2	After adjustment R^2	Standard estimate error
1	.305[a]	.093	.088	2.47845

Note: a. Predictor variable: (constant), bust

Table 4. Analysis of variance[a]

Model		Sum of square	Degree of freedom	Mean square	F	Significance
1	Regress	122.312	1	122.312	19.912	.000[b]
	Residual	1191.688	194	6.143		
	Total	1314.000	195			

a. Dependent variable: neck-root circumference

Table 5. Key points of prototype structure design

(a). The main points of the prototype structure design based on the Japanese new cultural prototype

Location	Body Type 0	Body Type 1
Base line width	40	40
Chest width	0.126B+6.19	0.125B+6.31
Back width	0.126B+7.32	0.125B+7.4
Back sleeve deep	0.0275B+20.485	0.054B+16.865
Front collar width	(0.114B+27.8) /5	(0.08B+32.219) /5
Front collar opening deep	(0.114B+27.8) /5+0.5	(0.08B+32.219) /5+0.5
Back collar width	(0.114B+27.8) /5+0.2	(0.08B+32.219) /5+0.2

Deep back collar	1/3[(0.114B+27.8) /5+0.2]	1/3[(0.08B+32.219) /5+0.2]
Front sleeve cage deep	15.81+0.14B- Front collar opening deep	0.158B+15.32- Front collar opening deep

(b). Key points of prototype structure design based on Liu Ruipu prototype

Location	Body Type 0	Body Type 1
Base line width	40	40
Chest width	(0.301B+9.305)/2	(0.316B+6.93)/2
Back width	(0.109B+32)/2	(0.2065B+21.78)/2
Back sleeve deep	0.0275B+20.485	0.054B+16.865
Front collar width	(0.018B+5.289)	(0.019B+5.377)
Front collar opening deep	(0.018B+5.289) +0.5	(0.019B+5.377) +0.5
Back collar width	(0.018B+5.289) +0.2	(0.019B+5.377) +0.2
Deep back collar	1/3 [(0.018B+5.289) +0.2]	1/3[(0.019B+5.377) +0.2]
Front sleeve cage deep	0.027B+20.485	0.037B+16.865

(c). Key points of prototype structure design based on Donghua prototype

Location	Body Type 0	Body Type 1
Base line width	0.034h+35.14	0.034h+35.135
Chest width	0.13B+5.8	0.13B+5.9
Back width	0.13B+7	0.13B+7
Back sleeve deep	49.6-0.0178B-0.163h	22.3-0.03h+0.065B
Front collar width	0.03B+5.289	0.02B+5.15
Front collar opening deep	0.03B+5.289+0.5	0.02B+5.15+0.5
Back collar width	0.03B+5.289+0.2	0.02B+5.15+0.2
Deep back collar	1/3[(0.03B+5.289+0.2)]	1/3[(0.02B+5.15+0.2)]
Front sleeve cage deep	0.13h+7.05- Front collar opening deep	0.13h+6.74- Front collar opening deep

3.　　Method

3.1. Body Type Classification

3.1.1 Classification Cluster

Set the number of sample clusters to 6, 5, 4, 3, and 2, respectively. In figure 10, the horizontal axis represents the profile coefficient, and the vertical axis represents the number of people included in each cluster. The red dotted line in figure 10 is used to evaluate the performance of the clustering model, and the larger the value of the contour coefficient, the better. When the contour coefficient is less than 0, it means that the sample is divided into misclusters. Theoretically, the closer the contour coefficient is to 1, the better, but there is currently no clear definition of how good a cluster is. It can be clearly seen from table 6 that the contour coefficient is the largest when the cluster is 2 and the samples with the contour coefficient less than 0 are also the least. Therefore, the effect is the best when the cluster is classified as 2.

Table 6. Cluster contour coefficients of middle-aged women in Xinjiang

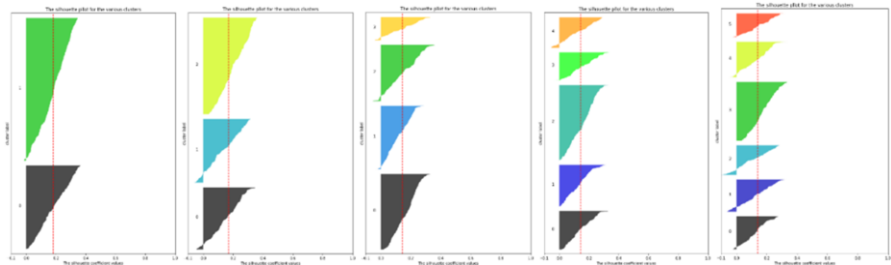| Clusters	2	3	4	5	6
Contour factor	0.1835	0.1705	0.1478	0.1424	0.1399

Figure 10. Contour coefficient graph of 2 clusters, 3 clusters, 4 clusters, 5 clusters, and 6 clusters

3.1.2 Andrews Curve Test.

Andrews curve is a method of visualizing high-dimensional data [20]. The attribute value of each sample is converted into Fourier sequence coefficients to create a curve. Curves belonging to the same category are usually similar. The curve is created by converting the characteristic value of each sample into Fourier sequence coefficients [20]. Different types of curves are marked with different colors, and an Andrews curve is drawn for each sample to achieve a visual clustering effect. The X axis in the Andrews graph represents t (-π<t<π), and the Y axis represents the value obtained by bringing all the characteristic attribute values into the Fourier sequence. The calculation formula of Andrews curve is shown in the following formula 8. It can be seen from figure 11 that the curves of the same category are basically the same, and the difference between the two types of curves is not obvious. This result also provides a basis for the smaller profile coefficient in this industry.

$$F(t) = \frac{x_1}{sqrt(2)} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + \cdots\cdots \qquad (8)$$

Where: The coefficient corresponds to the value of each dimension, -Π<t<Π.

According to the contour coefficient and Andrews curve, when the K-Means cluster is 2, the classification is the most reasonable. The average value is used as the intermediate of each body type, and the parameters of each intermediate are shown in table 7. From the waffle chart in figure 12, the proportion of each type of body type is obtained: "0" type is 37.02%, and "1" type is 62.98%. The virtual human body constructed according to the characteristic values of various body types is shown in figure 13-14. Based on the data analysis and the virtual human body diagram, the "0" body type is named obese body, and the "1" body type is named as belly fat convexity. body.
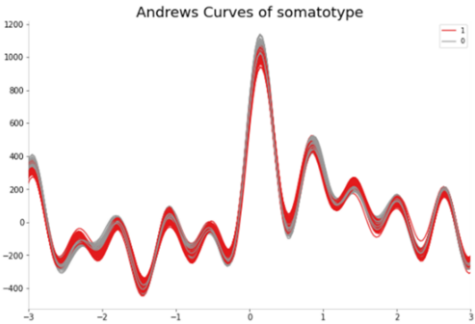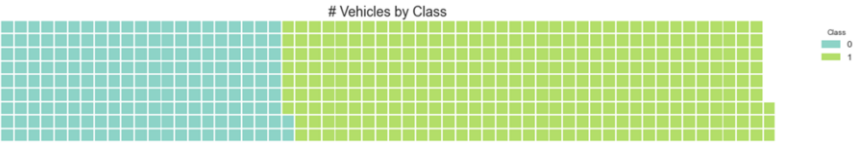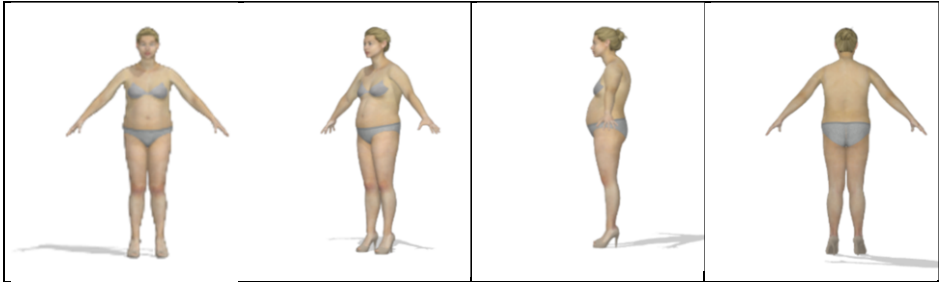

Figure 11. Andrews curve

Figure 12. Waffle chart
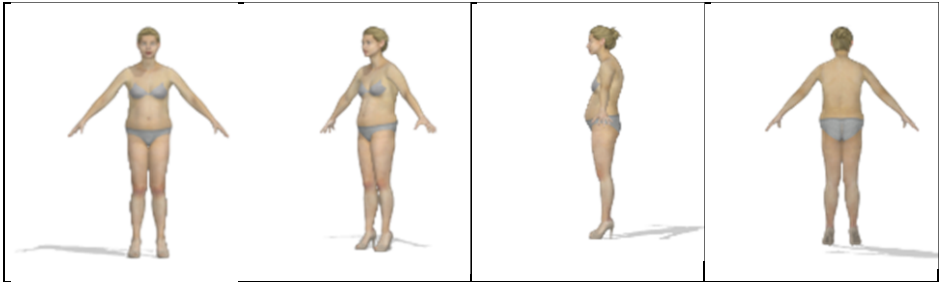

Figure 13. "0" obese body

Figure 14. "1" category are all called abdominal bulging body

Table 7. Intermediate size

cluster	weight	height	Cervical point high	Waist height	Breast height	Back length
0	67.57	161.92	137.51	97.91	115.58	42.34
1	62.63	162.72	138.29	99.31	114.69	41.56

Strand length	Arm length	Length of upper crotch front and rear	Total shoulder width	Chest width	Back width
26.88	52.80	76.35	41.64	39.14	39.72
27.08	51.87	74.46	40.82	38.36	38.25

Front waist length	Neck-root circumference	chest circumference	waist	Chest and waist difference	Abdominal circumference
46.72	40.44	105.31	98.01	7.30	104.03
45.87	39.35	94.10	86.84	7.25	92.22

Hips	Hip-waist difference	Arm circumference	Shoulder angle	Lateral angle
108.15	10.14	35.18	22.55	165.59
98.31	11.47	32.71	21.88	167.56

3.2. Evaluation of the Effect of Body Type Recognition

The mean square error MSE (formula 9) is used to measure the difference between the predicted value and the true value [21]. In this paper, R^2(formula 10) is used to measure the "law" of the data captured by the model. The numerator in R^2 is the total amount of information not captured by the model, and the denominator measures the proportion of the amount of information not captured by the model to the amount of information contained in the real label. The closer R^2 is 1 is better.

$$MSE = \frac{1}{m}\sum_{i=0}^{m}(y_i - \hat{y}_I)^2 \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=0}^{m}(y_i - \hat{y}_I)^2}{\sum_{i=0}^{m}(y_i - \bar{y})^2} = 1 - \frac{RSS}{\sum_{i=0}^{m}(y_i - \bar{y})^2} \tag{10}$$

In the formula, y is the true label, $\hat{y}$ is the predicted result, and $\bar{y}$ is the mean value.

In order to compare the effects of algorithms, LightGBM, random forest, and linear regression body shape recognition models were established respectively. The evaluation results of R2 and MSE are shown in table 8. The LightGBM algorithm is completely better than the random forest algorithm and linear regression algorithm in predicting the comfort of clothing structure.

If you want to predict the body shape of middle-aged women in a certain Xinjiang region, you only need to enter the characteristic measurement value in the built LightGBM body shape recognition model code (figure 15).

```
Mrs_W =[[120,158,101,107,105,113,115]]
#使用模型做出预测
pre = rfc.predict(Mrs_W)
print(pre)
```

Figure 15. LightGBM body type recognition model prediction code

Table 8. Comparative analysis of model effects

	LightGBM algorithm	Random forest	Linear regression
R^2	0.8558	0.7773	0.60275
MSE	0.03552	0.04859	0.08847

3.3. Feature Weight

The feature weight describes the degree of influence of each pressure point of each position on the effect of the LightGBM prediction model [22]. Figure 16 shows that waist circumference, chest circumference, hip circumference, and abdomen circumference are important features that have a greater impact on the body shape recognition of middle-aged women in Xinjiang.
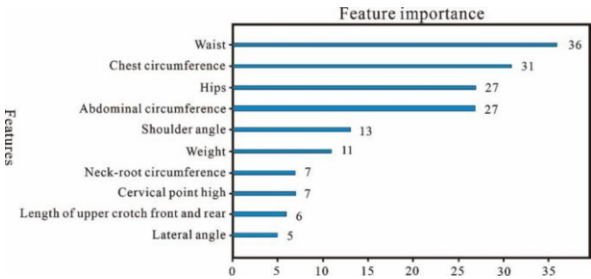
Figure 16. Histogram of feature weights

3.4. Binning Value

In order to facilitate the garment plate making to improve the coverage of the garment size, the formula 11 is used to calculate the grading value of each control part. According to the binary linear regression analysis method, the grading value of each control part is obtained from the two basic parts (height and bust circumference) in the clothing size series, as shown in table 9. The graded value of each model control component can be directly used in the model push plate, which is convenient for industrial production. According to the classification value of each part, push the files from each body shape intermediate to both sides to establish the clothing size standard for middle-aged women in Xinjiang, which is helpful to improve the regional clothing size system.

$$y = b(H) * \xi + b(B) * \tau \tag{11}$$

In the formula: y represents the grading value of each control part, $b(H)$ represents the partial regression coefficient of each part for the height, $b(B)$ represents the partial regression coefficient of each part for the chest circumference, ξ represents the number of height gradings, and τ represents the number of chest circumference gradings [24].

Table 9. Determination of the division value of each part

Location	0				1			
	Intermediate		Bin value		Intermediate		Bin value	
	Calculated	Adopted value	Calculated	Adopted value	Calculated	Adopted value	Calculated	Adopted value
height	161.92	162	5	5	162.72	163	5	5
Waist height	97.91	98	1.855	2	99.31	99.5	1.679	1.5
Back length	40.85	41	1.393	1.5	40.59	40.5	0.392	0.5
Arm length	52.80	53	1.858	2	51.87	52	1.779	1.5
Total shoulder width	41.64	41.5	-0.025	0	40.82	41	1.023	1
Collar	40.44	40.5	-0.105	0	39.35	39.5	0.227	0.2
Chest circumference	105.31	105.5	4	4	94.1	94	4	4
Waist	98.01	98	2.999	3	86.84	87	1.77	2
Hips	108.15	108	3.407	3.5	98.31	98.5	2.196	2

3.5. Comparison of the Body Shape of Middle-Aged Women in Xinjiang And That of the Country.

Table 10 shows the T-scores comparing the mean and standard deviation of the main measurement indicators of body shape between middle-aged women in Xinjiang and national adult women. It can be seen that the height, chest circumference, waist

circumference, and hip circumference of middle-aged women in Xinjiang are larger than the national average for adult women, and the carcass is about two standard deviations higher than the national level, indicating that the body shape of middle-aged women in Xinjiang is unique.

Table 10. Comparative analysis of body shape data of middle-aged women in Xinjiang and national adult women
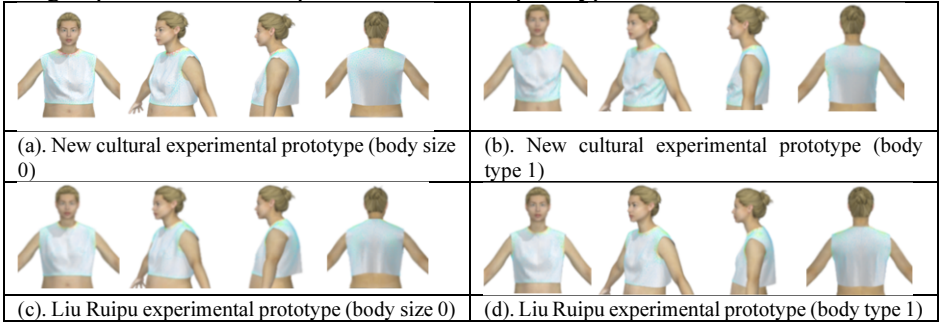
Location	Middle-aged women in Xinjiang		National adult women		T=50+10Z
	Mean X/cm	Standard deviation S	Mean $\bar{X}$/cm	Standard deviation $\bar{S}$	
Height	162.43	4.63	156.58	5.46	60.71
Chest circumference	98.24	7.44	84.06	6.64	71.36
Waist	90.97	8.39	69.76	8.26	75.68
Hips	101.95	7.01	90.37	5.58	70.75

Note: $Z=(x-\bar{x})/\bar{s}$, the average number of T scores is 50, and each standard deviation is 10, T greater than 50 indicates that the observation point is higher than the national average level of adult females, otherwise it means that the observation point is lower than the national adult female Average. The greater the T distance 50, the greater the difference between the body shape of the area and the national standard.

3.6 Prototype Effect

3.6.1 Virtual Try-On Evaluation

Use the CLO 3D virtual try-on software to obtain the virtual try-on pressure graphs of body type 0 and body type 1, respectively, showing the front, 3/4 side, left, and back try-on effects of each model's virtual try-on effect. The experimental prototypes established in accordance with the prototype of the new Japanese culture all showed high pressure at the front neckline, and it was easy to tighten the neck when wearing it, which caused people to breathe hard. In addition, the experimental prototype established based on the Japanese new cultural prototype showed that the front body was inverted, with many folds, and poor aesthetics. The experimental prototype built according to Liu Ruipu's prototype is generally more submissive, with only a slight tension in the collar. Based on the Donghua prototype, the prototype is loose and comfortable without tightness, and the body is flat and beautiful. In short, the experimental prototype proposed based on the Donghua prototype can best meet the needs of middle-aged women for comfort and aesthetics. The use of the Liu Ruipu prototype can also meet the body shape needs of this group better than the Japanese new cultural prototype.
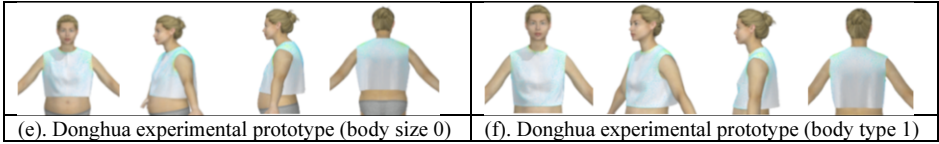

(a). New cultural experimental prototype (body size 0)

(b). New cultural experimental prototype (body type 1)

(c). Liu Ruipu experimental prototype (body size 0)

(d). Liu Ruipu experimental prototype (body type 1)

| (e). Donghua experimental prototype (body size 0) | (f). Donghua experimental prototype (body type 1) |

Figure 17. Virtual try-on pressure graph

3.6.2 Actual Try-On Inspection

Middle-aged women in Xinjiang with intermediate sizes of body type 0 and body type 1 were invited to have actual fitting experience. Each model performed standing, walking, sitting, running, squatting, and climbing stairs and scoring the wearing comfort. Three people are invited to try on each body type. In order to obtain more objective and true scoring results, each model tried on three times in the morning, noon, and evening, and performed the same action each time, and finally took the average value as the subjective evaluation result. The actual try-on found that the score is based on 0-9 points, the higher the score, the better the comfort and aesthetics. The result of the comprehensive evaluation is that the experimental prototype based on the new cultural prototype has an average score of 5.6 points; the experimental prototype based on the Liu Ruipu prototype has an average score of 7.8; the experimental prototype based on the Donghua prototype has an average score of 8.4 points. Similar to the virtual fitting evaluation results, the experimental prototype based on the Donghua prototype is most suitable for the body shape of middle-aged women in the region, and the experimental prototype based on the Liu Ruipu prototype meets the body shape needs of this group.

4.   Summary

By calling the "Scikit-learn" learning package in the Python programming software, the K-Means cluster analysis of dimensionless data, using the contour coefficient as the clustering effect evaluation index, divides the shape of middle-aged women in Xinjiang into two categories. Appropriate, they are obese (37.02%), and both are called abdominal bulging body (62.98%), and the reliability of the result is tested by Andrews curve.

The body shape recognition model of middle-aged women in Xinjiang based on LightGBM algorithm has better effect than random forest and linear regression.

Through the comparative analysis of the body shape of middle-aged women in Xinjiang and the national standard body shape, it is found that the chest circumference, waist circumference and hip circumference of middle-aged women in Xinjiang are larger than the national average level of adult women, and the carcass is two standard deviations higher than the national average level, reflecting The regional characteristics of middle-aged women in Xinjiang are analyzed.

4Through virtual try-on evaluation and actual try-on inspection, it is found that the experimental prototype established based on the Donghua prototype and Liu Ruipu's third-generation women's clothing prototype can well meet the needs of middle-aged women for wearing comfort and aesthetics. The Hua prototype has the best effect.

References

[1] J J Cai et al 2021. An assembly-level neutronic calculation method based on LightGBM algorithm. Annals of Nuclear Energy. vol 150. pp 107871-107881

[2] A Elen and E Avuçlu 2021. Standardized Variable Distances: A distance-based machine learning method. Applied Soft Computing Journal. vol 98. pp 106855-106869

[3] S Naaz 2021. Detection of Phishing in Internet of Things Using Machine Learning Approach. International Journal of Digital Crime and Forensics (IJDCF). vol 13 no 2. pp 1-15

[4] B K Ranjan et al 2021. Structural Mining for Link Prediction Using Various Machine Learning Algorithms. International Journal of Social Ecology and Sustainable Development. vol 12 no 3. pp 66-78

[5] V K Minul and G Joby 2020. Classification of Hot Spots using XGBoost and LightGBM Algorithms. International Journal of Engineering and Advanced Technology. vol 9 no 5. pp 722-724

[6] W Z Liang et al 2020. Predicting Hard Rock Pillar Stability Using GBDT,   XGBoost, and LightGBM Algorithms. Mathematics. vol 8 no 5. pp 765-782

[7] X Zhao 2018. The classification of female college students in Tianjin and the construction of experimental prototype. Tianjin Polytechnic University.

[8] Q Xie et al 2008. Parts and methods of anthropometric measurements for clothing. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, National Standardization Administration of China.

[9] Y M Liu and Dai Hong 2010. Research on the body shape of middle-aged and elderly women in Chengdu. Journal of Textile Research. vol 31 no 10. pp 110-115

[10] S Darpe S P and Singh 2017. Bearing damage assessment using Jensen-Rényi Divergence based on EEMD, Mech. Syst. Signal Process. vol 87. pp 307-337

[11] J P Mouton, M Ferreira and A S J Helberg 2020. A comparison of clustering algorithms for automatic modulation classification. Expert Systems With Applications. vol 151. pp 153-163

[12] M Sivaguru and M Punniyamoorthy 2021. Performance-enhanced rough k -means clustering algorithm. Soft Computing. vol 25 no 2. pp 1596-1616

[13] N Wang et al 2021. Vision and sound fusion-based material removal rate monitoring for abrasive belt grinding using improved LightGBM algorithm. Journal of Manufacturing Processes. vol 66. pp 281-292

[14] J P Han, D M Li and S Wang 2020. Based on PSO-RF two-way feature selection and LightGBM equipment fault detection. Computer Systems & Applications. vol 29 no 7. pp 228-232

[15] S Q Guo et al 2020. Classification for Glass Bottles Based on Improved SelectiveSearch Algorithm. Computers, Materials & Continua. vol 64 no 1. pp 233-251

[16] Ye B J et al 2020. A Methodology for Predicting Aggregate Flight Departure Delays in Airports Based on Supervised Learning. Sustainability. vol 12 no 7. pp 2749-2762

[17] H C Yuan et al 2020. Fish abnormal behavior detection based on LightGBM model. Fisheries Modernization. vol 47 no 1. pp 47-55

[18] Y L Lu 2005. Three-dimensional clothing design in computer virtual reality environment. Shanghai: Donghua University Textile.

[19] Z C Huang 2020. Research on the overall balance design of women's body relaxation. Textile Report. vol 39 no 11. pp 62-71

[20] C Tian and H Zhen 2019. DuanModel checking open systems with alternating projection temporal logic. Theoretical Computer Science. vol 774. pp 65-81

[21] Y T Qian and X X Song 2013. The structure design of hip-lift jeans. Shanghai Textile Science & Technology. vol 41 no 1. pp 89-91

[22] Y R Li et al 2020. A Model Combining Seq2Seq Network and LightGBM Algorithm for Industrial Soft Sensor. IFAC-PapersOnLine. vol 52 no 2. pp 12068-12073