# Contrastive Learning Dimensionality Reduction Method Based on Manifold Learning

Jinghao Situ

School of Mathematical Sciences, Beihang University, Beijing, 100191,China

**Abstract.** With the development of the times, more and more high-dimensional datasets come into people's view. In order to reduce the time complexity and space complexity of downstream tasks, data dimensionality reduction becomes the primary choice. Classical dimensionality reduction algorithms are mainly divided into linear dimensionality reduction algorithms and nonlinear dimensionality reduction algorithms. Some of the traditional dimensionality reduction methods have the problems of not considering the nonlinear structure of the original dataset and the existence of weak generalisation, which makes the dimensionality reduction effect not good or model need to be recalculated because of the addition of new samples. In order to solve these problems, the research in this paper is a comparative learning dimensionality reduction method based on manifold learning. The idea of manifold learning using geodesic distance can fully consider the nonlinear structure of the original dataset. In this paper, comparative learning is the main framework. When the neural network completes the training, it only need to take the new data as input to calculate, the result can be obtained, no need to reconstruct the model which means the generality is high. Starting from the related work, this paper briefly introduces manifold learning, comparative learning and neural network algorithms. Subsequently, an innovative model is proposed, including three modules, Isomap to extract nonlinear structure, expanding neighbourhood to make pseudo-labels, and comparative learning training. Detailed analyses are carried out through experiments, comparing with PCA and LLE algorithms with the geodetic distance retention rate as an indicator, which proves that the data dimensionality reduction method of this model is more effective and ubiquitous.

**Keywords:** dimensionality reduction; manifold learning;comparative learning; neural networks.

## 1. Introduction

### 1.1 Background and Significance of the Study

Existing classical dimensionality reduction algorithms are mainly divided into linear dimensionality reduction algorithms and nonlinear dimensionality reduction algorithms, of which the linear discrimination algorithm LDA and principal component analysis PCA algorithms are two famous linear dimensionality reduction algorithms, nonlinear dimensionality reduction algorithms are mainly manifold learning dimensionality reduction algorithms, tensor dimensionality reduction algorithms [1]. Generally speaking, nonlinear dimensionality reduction algorithms are better for extracting special data structures. But whether linear or nonlinear dimensionality reduction algorithms, traditional dimensionality reduction algorithms have the idea of calculating eigenvalues and eigenvectors to achieve the effect of dimensionality reduction. This means that when new data is added, the entire model needs to be reconstructed, which is less general and more time-consuming. In contrast, machine learning algorithms are a good choice to improve the efficiency of the algorithm and reduce the time and space complexity.

Machine learning algorithms can be divided into supervised learning and unsupervised learning, supervised learning refers to the training of samples with labels, while unsupervised learning in the training process samples without labels. Actually, most samples are unlabelled, so unsupervised learning is more widely used than supervised learning [2]. Machine learning dimensionality reduction is also a kind of unsupervised learning. Most of the existing machine learning dimensionality reduction algorithms are based on neural networks, and achieve better dimensionality reduction effect through a large number of training. Their advantage is that they are highly generalised and can be calculated directly without reconstructing the model when adding new data, and the disadvantage is that they do not consider the nonlinear structure of the dataset from a topological point of view.

As a machine learning algorithm, comparative learning inherits the above advantages and disadvantages. Comparative learning focuses on learning common features between similar instances and distinguishing differences between non-similar instances, and the basic idea is to map the original data into a feature space where positive pairs have the greatest similarity, and negative pairs have a smaller similarity [3]. Currently, most of the research focuses on applying comparative learning to recognition and classification [4], and very little research has been done on applying it to dimensionality reduction tasks.

In order to make up for the above drawbacks and shortcomings, this paper hopes to synthesise nonlinear dimensionality reduction algorithms and machine learning algorithms, focusing on applying comparative learning to the dimensionality reduction task, and researching and designing a data dimensionality reduction method that retains both nonlinear features and strong generalisability.

### 1.2 Main Research Content

As mentioned above, most of the traditional dimensionality reduction methods have some flaws. Most of the dimensionality reduction methods do not take into account the nonlinear structure of the original dataset, which makes the dimensionality reduction ineffective. The data of the Swiss Volume Surface presents a nonlinear structure. There are also cases where the Euclidean distance is small between data points of different categories. To compensate for the lack of using Euclidean distance, manifold learning algorithm tends to use geodesic distance. However, the manifold learning algorithm's dimensionality reduction method also has the problem of weak generalisability. Such as Isomap algorithm, when there is added to consider new data, need to recalculate the inner product matrix, etc., the calculation of the time complexity and space complexity is high.

The research in this paper is a comparative learning dimensionality reduction method based on manifold learning. The idea of using geodesic distance based on manifold learning can fully consider the nonlinear structure of the original dataset. In addition, with comparative learning as the main framework, after the neural network completes training, it only needs to take the new data as input to calculate to get the result, without reconstructing the model, and the generality is high.

### 1.3 Organisation

This paper is roughly organised as follows: the second part introduces the related work, detailing the algorithmic ideas of manifold learning and neural networks, while introducing a specific manifold learning algorithm Isomap algorithm as well as the comparative learning algorithm. The third part is a detailed description of the model, including three parts of Isomap algorithm to extract the nonlinear structure, expanding neighbourhood to make pseudo-labels, and comparative learning training. The fourth part is a comparative test to illustrate the superiority of the algorithm through quantitative data. The innovation of this model is that it combines the existing neural network dimensionality reduction method with manifold learning. As a dimensionality reduction method that can retain the nonlinear features of the original dataset, this model has good generality facilitating the continuation of downstream tasks.

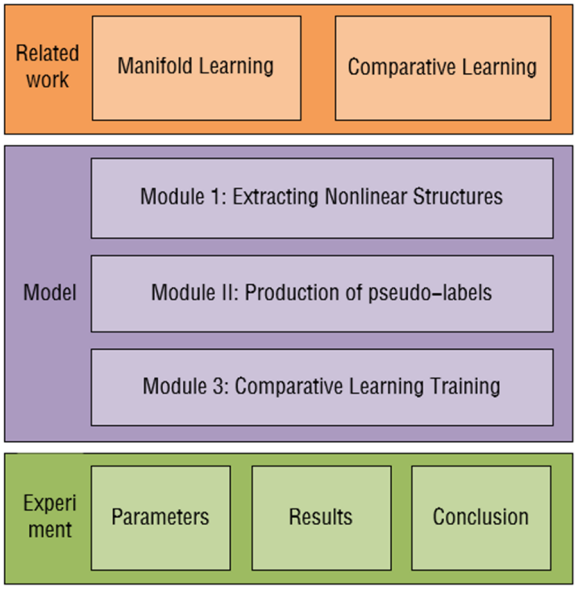


Figure 1: Architecture Diagram

## 2. Related Work

### 2.1 Manifold Learning

Manifold learning is a nonlinear dimensionality reduction method developed in the past 10 years or so, aiming at discovering the intrinsic geometric structure or intrinsic dimensionality of low-dimensional smooth manifolds embedded in a high-dimensional nonlinear data space, so as to facilitate deeper understanding and further processing of the data [5].

In 2000, TENENBAUM et al. proposed the isomap algorithm, which uses geodesic distance instead of Euclidean distance, and its dimensionality reduction effect is particularly good on the Swiss convoluted dataset [6]. Later, algorithms such as Local Linear Embedding (LLE) appeared, focusing on maintaining the local nature of the manifold, and the data before and after the dimensionality reduction are regarded as linear representations of other points in its neighbourhood [7]. The Laplace feature mapping algorithm of the same year is an algorithm based on spectral graph theory, the idea is to find a mapping that preserves the local properties of the data points in an average sense [8]. Incremental Alignment Method (IAM), the main idea is to gradually align the low dimensional coordinates of the input data, patch by patch, in order to iteratively generate a representation of the entire dataset with the advantage of high efficiency [9].

The Isomap algorithm has produced many optimisation schemes and variant algorithms after years of development. Wang Liguo et al. proposed an improved L-ISOMAP algorithm (KL-ISOMAP) based on the K-medoids clustering algorithm to obtain better visualization [10]. Xiangyuan Li proposed the ISOMAP algorithm based on the density scaling factor, which reduces the influence of noise on the dimensionality reduction and enhances the robustness of the algorithm [11]. The proposed L-Isomap-KELM prediction model solves the deficiencies of the traditional prediction methods, such as low prediction accuracy, slow prediction speed, etc., and is able to achieve an accuracy of 97.31%, and the calculation speed is fast, and has a very good ability to generalize [12]. Scholars from the University of Science and Technology of China proposed the FastIsomapVis method and experimented on social network datasets, and the model has outstanding results in reducing the time complexity [13].

### 2.2 Comparative Learning

Neural network model as an important carrier of deep learning provides a good condition for the rapid development of deep learning. Inspired by animal neural networks, neural network systems can learn large amounts of data. Through the introduction of activation function to enhance its nonlinear

expression ability, and then combined with the back propagation algorithm to make it has a strong learning ability for image, speech and other data [14].

In 1986, the back-propagation idea of BP neural network algorithm made the linear indivisibility problem effectively alleviated, which opened the fast track of neural network algorithm development [15]. Later, convolutional neural networks specialised in image processing appeared, using the idea of convolutional kernel to solve the problem of parameter explosion of fully connected neural networks [16]. Recurrent neural network is a special kind of recurrent neural network, whose main idea is to recursively follow the evolutionary direction of the sequence, and has a powerful ability in language processing [17]. Generative Adversarial Networks (GANs) want to use implicit adversarial instead of explicit mathematical representations, and "fool" the discriminator by training several times to get the desired result [18].

Comparative learning is based on neural network algorithms, the idea is to use twin neural networks for training, aiming to narrow the distance between the features of the same kind of images and push the distance between different kinds of images, in order to get a better feature extraction model [19]. Nowadays, the mainstream comparative learning generally generates positive and negative samples through different data augmentation methods, and then passes them through an encoder to get the encoding of the two, and calculates the similarity of the two to get the comparison loss [20]. However, the current stage of comparative learning does not pay attention to the problem of extracting the nonlinear structure of the dataset, and this aspect is also neglected when calculating the comparison loss. In addition, how to train the source and target data together, so that the model can get good results in the target domain is the core problem of unsupervised domain adaptation, so how to effectively combine the comparative learning method with the unsupervised domain adaptation method is also a problem worth studying [21].

At this stage, most of the research in the direction of comparative learning is the exploration of applying comparative learning to downstream tasks. Xie Zhuo et al. proposed a clustering method based on comparative learning for multi-relational attribute graphs (CCLMAG), which is superior to the current state-of-the-art baseline methods and has practical applications [22]. Cao Yibo et al. introduced unsupervised comparative learning into the field of trademark retrieval for the first time, and extracted semantic information through comparative learning to finally improve the average accuracy of trademark retrieval [23]. Joint Public Security of Tsinghua University proposed a loss function based on multi-instance relations to fully exploit the intra- and inter-modal associations between multiple positive and negative samples of masked faces and intact faces to improve the accuracy of masked face recognition [24]. The proposed method of introducing comparative learning for emotion category representation can better reflect the complex relationship between different emotion categories [25]. However, there is still no algorithm that superimposes the use of comparative learning and manifold learning, and this paper will be a good response to the problem of effectively combining the comparative learning method with the unsupervised domain adaptive method. Based on this idea, this paper proposes a data dimensionality reduction method based on comparative learning and manifold learning to achieve better dimensionality reduction effect for application in downstream tasks.

## 3. Model

The research in this paper is a comparative learning dimensionality reduction method based on manifold learning. The idea of using geodesic distance based on manifold learning can fully consider the nonlinear structure of the original dataset. In this paper, comparative learning is the main framework, after the neural network completes the training, it only needs to take the new data as the input to calculate to get the result, no need to reconstruct the model, and the generality is high. In this paper, we propose the innovative model, including three modules of Isomap to extract the nonlinear structure, expanding the neighbourhood to make pseudo-labels, and comparative learning training.

Table 1: Symbol Lable

| NOTATION | EXPLANATION |
|---|---|
| $\boldsymbol{x_i}$ | The $i$th item |
| $\boldsymbol{k_0}$ | Number of neighbours |
| $\boldsymbol{x_{i,t}}$ | The $t$th neighbor of the $i$th item |
| $\boldsymbol{s}$ | The batch size of the neural network |
| $\boldsymbol{m}$ | neighbourhood multiple |
| $\boldsymbol{t}$ | incremental factor |
| $\boldsymbol{n}$ | sample size |
| $\boldsymbol{r}$ | Geodetic distance neighbourhood retention rate |
| FUNCTION | — |
| $\boldsymbol{d(x_i, x_j)}$ | Calculate the Euclidean distance between two points |
| $\boldsymbol{d_2(x_i, x_j)}$ | Calculate the Geodesic distance between two points |
| $\boldsymbol{knn(x_i, x_j, k)}$ | K-nearest neighbour algorithm function, where 1 means $\boldsymbol{x_j}$ is one of the $k$ nearest neighbors of $\boldsymbol{x_i}$, 0 means $\boldsymbol{x_j}$ is not one of the $k$ nearest neighbors of $\boldsymbol{x_i}$ |
| $\boldsymbol{label(x_i, x_j)}$ | Label function of comparative learning, where 1 means mutually positive samples, 0 means mutually negative samples |
| $\boldsymbol{label_X(x_i, x_j)}$ | Label function before dimensionality reduction |
| $\boldsymbol{label_{X'}(x_i, x_j)}$ | Label function after dimensionality reduction |
| $\boldsymbol{\theta(x_i, x_j)}$ | Cosine function, calculates the magnitude of the cosine between two vectors. |
| $\boldsymbol{Loss}$ | Improved comparative learning loss function |

### 3.1 Isomap Extracting Nonlinear Structures

The Isomap algorithm needs to be considered first to set two parameters, one is the value of $K$ for the K-nearest neighbour algorithm, and the other is to set the dimensionality after dimensionality reduction. Only the points in the dataset that meet the K-nearest neighbour relationship are considered connected to each other and the others are considered disconnected, from which the collocation matrix is constructed by taking $K = k_0$. The distance matrix based on the geodetic distance is obtained by calculating the shortest path of the graph. Eigenvalue decomposition is performed after calculating the inner product matrix, then the first $t$ eigenvectors are the result of its dimensionality reduction. Dimensionality reduction of the dataset using Isomap algorithm has two advantages: 1) better extraction and preservation of the nonlinear structure of the original dataset 2) dimensionality reduction facilitates downstream tasks.

$$d(x_i, x_j) = \sqrt[2]{\sum_{t=1}^{n} (x_{i,t} - x_{j,t})^2} \tag{1}$$

$$d_2(x_i, x_j) = \begin{cases} 0 & knn(x_i, x_j, k_0) = 0 \\ d(x_i, x_j) & knn(x_i, x_j, k_0) = 1 \end{cases} \tag{2}$$

The calculation of geodetic distances is different from Euclidean distances. It is necessary to obtain proximity points based on the K-nearest-neighbour principle based on the Euclidean distances, and only mutually neighbouring points are considered to be connected. The shortest path algorithm is used for the distance matrix based on geodetic distances, the most commonly used shortest path algorithms are Floyd's algorithm and Dijkstra's algorithm. The reason for this is that it is desired to avoid the misclassification of the image classifier due to the short Euclidean distance between the two points for two data points with large differences.

The final result is obtained from the original dataset after extracting the nonlinear features. So far manifold learning algorithm to extract nonlinear structure module is completed, the following will be used as a basis to create pseudo labels.

### 3.2 Extended Neighbourhood Production Pseudo Labels

By extracting the nonlinear structure through Isomap, we get the dimensionality reduced data $Z$. In the following, we need to make pseudo-labels for comparative learning. Most of the traditional pseudo-labelling for comparative learning uses data enhancement, and the common method is to copy, rotate, crop the image to get positive samples, and all the rest are regarded as negative samples. This model here adopts another idea. The points with close geodesic distance can be regarded as the positive samples of the data points, and the opposite is set as the negative samples. Compared with the traditional method, the method of this model produces a larger number of positive samples, resulting in a larger value of the loss function, which is conducive to the training of the neural network model and accelerates neural network convergence.

The same K-nearest-neighbour principle is used to create pseudo-labels. This ensures that the number of points in each neighbourhood is constant, and there is no situation where the neighbourhood contains only that data point itself. It is worth noting here that at this point, $K = mk_0$ is taken, which means that the size of the neighbourhood of the K-nearest neighbour should be adjusted to $mk_0$, where $m > 1$. In this model, $m = 2$ is taken. It is not difficult to find out that, if $m \leq 1$ is taken, the resulting neighbourhood is exactly the same as that obtained by using the K-nearest-neighbour principle before extracting the non-linear structure. This is because the previous module also used the K-nearest-neighbour algorithm before calculating the geodetic distance, and its algorithm is based on the Euclidean distance between data points. So when $m \leq 1$, it is not possible to make good use of the results of the nonlinear structure extracted in the previous step.

This model takes $m = 2$ for two considerations. 1) The K neighbourhood of this data point is maintained from the perspective of Euclidean distance 2) The neighbourhood is expanded to $m$ times the original using manifold learning based geodetic distances, fully taking into account the nonlinear structure of the original dataset. This results in a pseudo-label matrix $L$, which treats the data points in the immediate neighbourhood of $mk_0$ as positive samples and the rest as negative samples. This method does not require manual labelling falls into the unsupervised category and is quite different from traditional data enhancement methods.

$$label(x_i, x_j) = \begin{cases} 0 & knn(x_i, x_j, mk_0) = 0 \\ 1 & knn(x_i, x_j, mk_0) = 1 \end{cases} \tag{3}$$

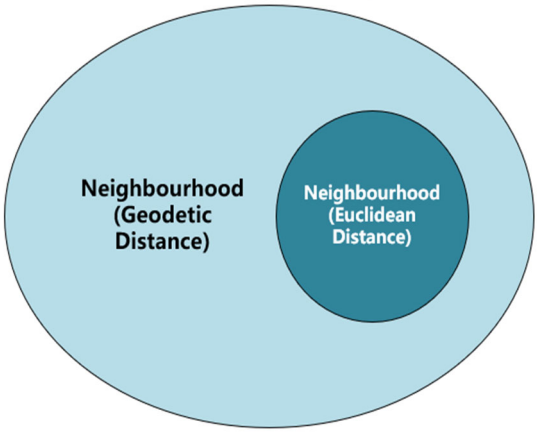


Figure 2: Neighbourhood Scope Map

### 3.3 Comparative Learning Training

The production of pseudo-labels as described above facilitates the training of comparative learning in the following ways: 1) the number of similar labels is more compared to the traditional comparative learning model, which avoids the function value of the LOSS function to be too small; and 2) the non-linear structure of the original data is taken into account.

The loss function of the comparative learning model is actually the same as the clustering idea, where the features of the data points are close to the positive samples and far from the negative

samples. When training, two encoders are used to extract features for sample A and its positive and negative samples respectively, and then distributed and mapped to a new representation space, where the model parameters are trained according to the comparative loss. The numerator of the loss function is the sum of the same kind of indices and the denominator is the sum of the whole indices, which is then divided by the batch size to find the evaluation loss. The exponential sum is the $\theta$ power of the natural logarithm $e$, and $\theta$ is the cosine of the angle of the data after dimensionality reduction. The closer the direction of the two data after dimensionality reduction, the larger the value of $\theta(x_i, x_j)$, and vice versa, the smaller the value of $\theta(x_i, x_j)$. As positive samples get closer in the direction after dimensionality reduction, the higher the power, the larger the denominator, and as negative samples get further away in the direction after dimensionality reduction, the denominator becomes smaller. Then the closer the fraction is to 1, the closer to 0 after taking $-ln$, the smaller the loss function is after taking the average. On the contrary, if the neighbourhood retention of geodetic distance becomes lower after dimensionality reduction, the loss function will increase. As the value of $\theta(x_i, x_j)$ is small, the absolute value is less than or equal to 1, here add the incremental coefficient $t$, $t$ take 0.1, amplify the value of the power exponent, so that the result of the Loss function is appropriately enlarged to accelerate the convergence.

$$\theta(x_i, x_j) = \frac{\langle x_i, x_j \rangle}{\|x_i\| \cdot \|x_j\|} \tag{4}$$

$$Loss = \frac{1}{s}\left[\sum_{i=1}^{n} -ln\left(\frac{\sum_{j,label(x_i,x_j)=1} e^{\theta(x_i,x_j)/t}}{\sum_j e^{\theta(x_i,x_j)/t}}\right)\right] \tag{5}$$

The model so obtained has the effect of data dimensionality reduction. Compared with the traditional data dimensionality reduction model, this model considers the nonlinear structure between data points, uses geodetic distances to make pseudo-labels, and is based on the idea of comparative learning, which downgrades the data that are not of the same kind after manifold learning into those that are relatively far away from each other, so that the data of the same kind are still close after downgraded. Using the above method to create pseudo-labels, the downscaled data still retains the nature of the original data geodesic distance, i.e., it still retains the non-linear structure of the original dataset. Applying these data to downstream tasks will allow for higher accuracy and better model robustness.

The next step is neural network training and testing. The training set data is used as an input to the neurons and activation functions of each layer to the output layer, where back propagation is started according to the newly defined loss function until the loss function decreases to a certain threshold and the training is completed. The final neural network obtained can be used as a feature extractor, which has the role of extracting the features of the dataset and reducing the dimensionality. The input of the feature extractor is the data to be processed and the output is the result of dimensionality reduction. The data before and after dimensionality reduction maintains the features of the distance (proximity of the geodesic distance), and can better retain the nonlinear structure of the original dataset.

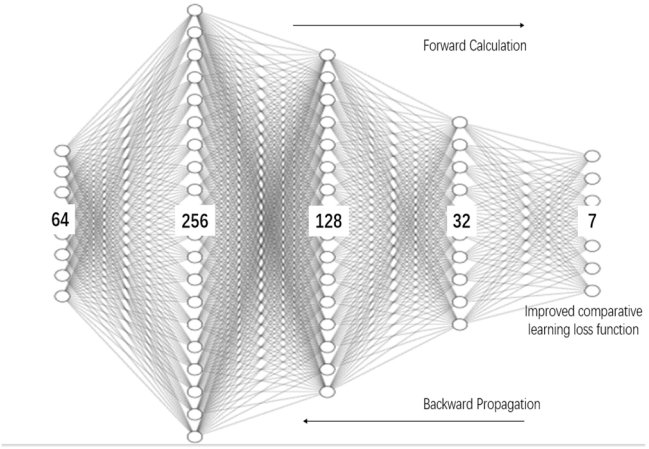

Figure 3. Neural Network Training Architecture

# 4. Experiments

## 4.1 Parameterisation

The experimental data is handwritten digit body dataset. The original handwritten digit body image is 8*8 size, and the vector is 64 dimensions after converted. For the selection of neural network, this model can use fully connected neural network and convolutional neural network. If the input data is in the form of pictures, convolutional neural network should be used, and if the input data is in the form of vectors, fully connected neural network should be used. This experiment falls into the latter category, and it is a solid choice to use a fully connected network. It should be noted that the loss function should be redefined as above. The size of the dataset is 1797, and 1500 data were used as the training set and the remaining as the test set.

In module 2, we choose the nearest neighbour of $k_0$ to be 10. Then we make the pseudo-labels and take $m$ to be 2. So the $mk_0$ points in the nearest neighbourhood belong to the neighbourhood in the sense of geodetic distance. Consider the points within this new domain as positive samples of the same kind and negative samples anyway. After that go to the comparative learning training module with the following specific parameters of the neural network.

Table 2: Neural Network Parameter List

| | |
|---|---|
| Learning rate | 0.01 |
| epochs | 15 |
| $\boldsymbol{t}$ | 0.1 |
| optimizer | Adam |
| Batch size | 32 |

The neural network has 5 layers, the number of neurons in the input layer is 64, the number of neurons in the hidden layer is 256, 128, and 32 in that order, and the output layer is 7. The activation function used is the ReLU function. In this experiment, the dimensionality is set to be reduced to 7 dimensions, and the reduced data is obtained after the training is completed.

## 4.2 Comparison Experiment

The original dataset is $X$ and the dimensionality reduced dataset is $X'$. Define the geodetic distance neighbourhood retention rate as

$$r = \frac{\sum_{i,j}(|label_X(i,j) - label_{X'}(i,j)| - 1)^2}{n^2} \tag{6}$$

The original intention of such a definition is to hope that before and after the dimensionality reduction, the point in the neighbourhood remains in the neighbourhood, not in the neighbourhood of the point remains not in the neighbourhood, that is, before and after the dimensionality reduction of the pseudo-labels to maintain consistency. When the pseudo-labelling is consistent, the absolute value is 0, minus 1 and then take the square to get 1; otherwise, the absolute value is 1 and the square gets 0.

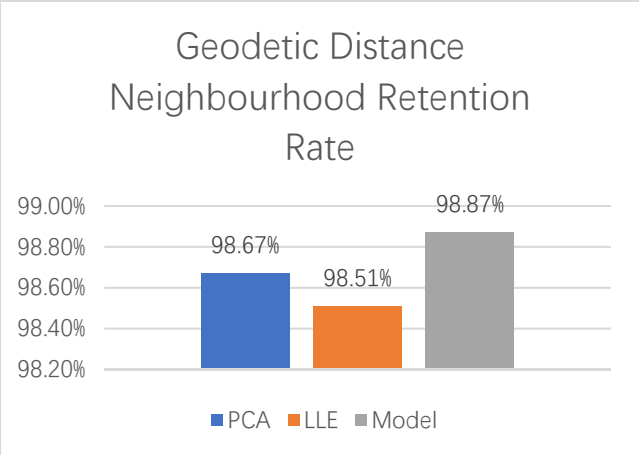


Figure 4. Geodetic Distance Neighbourhood Retention Rate

The neighbourhood retention rate of this model after training is slightly higher than that of the two dimensionality reduction methods of PCA and LLE, which indicates that the model is more practical. And when there is new data, it can be directly input into the neural network for calculation, and the neighbourhood retention rate before and after dimensionality reduction remains high. There are two reasons why the experimental model can achieve such an effect: 1) The first module extracts the nonlinear structure of the original dataset and retains its nonlinear features, i.e., the relational features of the geodesic distance. It is worth noting that different algorithms need to be used for different datasets in extracting their nonlinear structure 2) The loss function uses the idea of comparative learning, the positive samples still remain in the neighbourhood after dimensionality reduction, and the negative samples still remain outside the neighbourhood after dimensionality reduction, which keeps the neighbourhood features of the original data very well. In summary, the experimental model can well maintain the neighbourhood in the sense of geodetic distance of the original data set. In addition, the main body of the experimental model is a neural network, which is more generalisable. Such a data dimensionality reduction method with both feature extraction and generalisability is more effective.

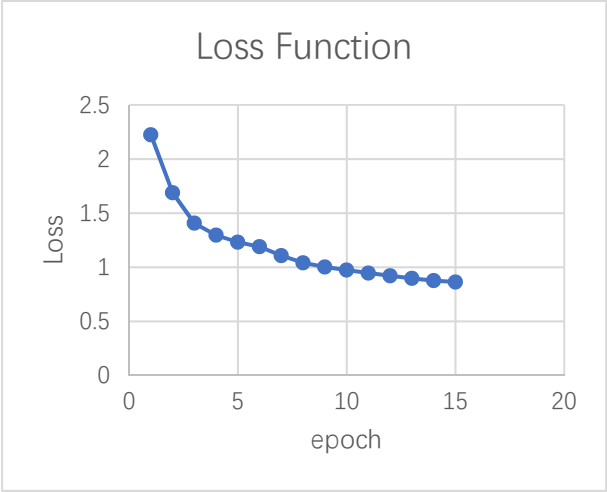


Figure 5..Loss Function

From the point of view of the Loss function of the neural network, the model has a rapid decline in the loss function in the initial stage of training, the subsequent decline is slow and fluctuations occur, and basically reaches the level where the loss value is less than 1 after 10 times of training.

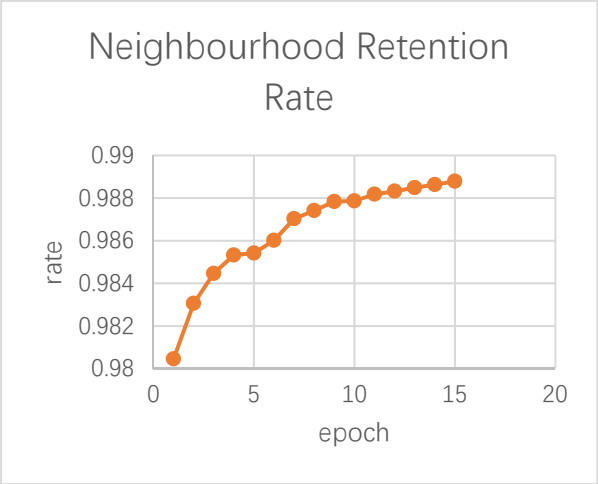


Figure 6. Neighbourhood Retention Rate of Training Group

Looking at the accuracy of the training set, the values have remained above 98%. It has reached over 98.5% at 5 epochs, surpassing the LLE algorithm. It fluctuates up in 5-15 epochs and finally reaches around 99%, indicating good results on the training set.

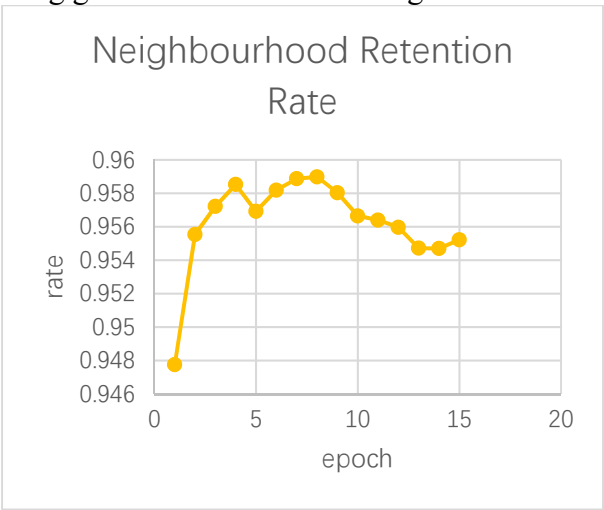


Figure 7.Neighbourhood Retention Rate of Testing Group

The accuracy of the test set also shows a trend of rapid increase and then levelling off, and finally fluctuates at a position close to 96%. The data dimensionality reduction method of this model achieves the expected effect, effectively extracts the nonlinear structure of the original dataset, and the geodetic distance neighbourhood is maintained at a high rate and has a strong generalisability.

## 5. Conclusion

In this paper, manifold learning and comparative learning algorithms are briefly introduced from the research literature. Subsequently, an innovative model is proposed, which includes three modules, Isomap to extract the nonlinear structure, expanding the neighbourhood to make pseudo-labels, and comparative learning training. Detailed analysis is carried out through experiments, which proves that the data dimensionality reduction method of this model is more effective and generalisable.

Looking ahead, data dimensionality reduction algorithms will play an increasingly important role in data analysis and machine learning. High-dimensional and large-volume data are also becoming more and more common, which puts higher requirements on data dimensionality reduction algorithms. For the processing of large-scale datasets, this model has unique advantages. It can be put into large-scale use after completing the training of neural network in a moderate volume training set, giving full play to its advantage of strong generality. In the future, more advantageous data dimensionality reduction methods will likely be applied in high-dimensional data processing, augmented learning and reinforcement learning, deep learning and artificial intelligence, and enhanced algorithm

interpretability. With the development of the machine learning, more data dimensionality reduction algorithms will be applied to these fields to optimise the learning process and improve the performance of intelligent systems, and play a more important role in information technology systems, 6G networks, and the new ecology of the Internet.

## References

[1] Chen SJ. Research on dimensionality reduction algorithm based on popular learning.Wireless Internet Technology,2022,19(09):147-149.

[2] Yin Ruigang,Wei Shuai,Li Han et al. A review of unsupervised learning methods in deep learning. Computer System Applications,2016,25(08):1-7.DOI:10.15888/j.cnki.csa.005283.

[3] Yang B,Luo Jiachen,Song Yantao et al. A time series clustering method based on comparative learning. Computer Science:1-13[2023-12-01].http://kns.cnki.net/kcms/detail/50.1075.TP.20231009.1736.065.html.

[4] YANG Lei,XIE Minghong,ZHANG Yafei et al. Unsupervised pedestrian re-identification based on hybridcomparative learning[J/OL]. Journal of Kunming University of Science and Technology(Natural Science Edition):1-15[2023-12-01].https://doi.org/10.16112/j.cnki.53-1223/n.2023.06.332.

[5] TAN Chao,GUAN Ji Hong,ZHOU Shuigeng. A review of incremental and evolutionary manifold learning.Journal of Intelligent Systems,2012,7(05):377-388.

[6] TENENBAUM J, DE SILVA V, LANGFORD J. A global geometric framework for nonlinear dimensionality reduction. Science, 2000, 290( 5500) : 2319-2323.

[7] ROWEI S, SAUL L. Nonlinear dimensionality reduction by locally linear embedding [J] . Science, 2000, 290( 5500) : 2323-2326.

[8] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduc- tion and data representation..Neural Computation, 2003, 15(6): 1373-1396.

[9] HAN Zhi, MENG Deyu, XU Zongben, et al. Incremental alignment manifold learning . Journal of Computer Science and Technology, 2011, 26( 1) : 153-1

[10] WANG Liguo,WU Fei. Colour visualization of hyperspectral images based on KL-ISOMAP. Journal of Nanjing University of Information Engineering (Natural Science Edition),2018,10(01):63-71.DOI:10.13878/j.cnki.jnuist.2018.01.006.

[11] Li Xiangyuan. ISOMAP dimensionality reduction algorithm based on density scaling factor and its application. Northwest Agriculture and Forestry University,2019.

[12] Xie GN,Huang RL,Liu M et al. L-Isomap-KELM model for coal and gas herniation. Control Engineering,2020,27(10):1802-1806.DOI:10.14107/j.cnki.kzgc.20180494.

[13] Yousaf M . Research on nonlinear Isomap dimensionality reduction method for high dimensional datasets. University of Science and Technology of China,2021.DOI:10.27517/d.cnki.gzkju.2021.000537.

[14] HUANG Dongrui,MAO Kebiao,GUO Zhonghua et al. A review of several classical models of neural networks. High Technology Letters,2023,33(08):860-871.

[15] RUMELHART D E, HINTON G E, WILLIAMS R. Learning representations by back-propagating errors. Nature,1986,323:533-536.

[16] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradientbased learning applied to document recognition. Proceedings of the IEEE, 1998,86(11):2278-2324.

[17] ELMAN J L. Finding structure in time. Cognitive Science, 1990,14(2):179-211.

[18] GOODFELLOW I,POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets,Proceedings of the 28th Conference on Neural Information Processing Systems. Montreal: Association for Computational Linguistics, 2014:2672-2680.

[19] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA: IEEE, 2006. 1735-1742

[20] Comparative learning in natural language processing.Network Security Technology and Application,2023(08):55-56.

[21] ZHANG Shuangxiang,CHEN Jie,LI Qilong et al. A review of deep comparative learning.Journal of Automation,2023,49(01):15-39.DOI:10.16383/j.aas.c220421.

[22] Xie Zhuo,Kang Le,Zhou Lijuan et al. A multi-relational attribute graph clustering method based on comparative learning. Computer Science,2023,50(11):62-70.

[23] CAO Yibo,YAN Binbin,YU Songsen et al. Unsupervised trademark retrieval based on comparative learning. Microelectronics and Computers,2023(10):74-82[2023-11-13].https://doi.org/10.19304/J.ISSN1000-7180.2022.0816.

[24] CHEN Xianming,LIN Qunxiong,LIU Weiqiang. Multi-feature fusion for mask-wearing face recognition based on comparative learning.Computer Application Research:1-7[2023-11-13].https://doi.org/10.19734/j.issn.1001-3695.2023.06.0266.

[25] Wang Xiangyu,Zong Chengqing. Text sentiment category representation based on supervised comparative learning. Journal of Software:1-12[2023-11-13].https://doi.org