

Article

A Novel Pre-Processing Technique for Original Feature Matrix of Electronic Nose Based on Supervised Locality Preserving Projections

Pengfei Jia, Tailai Huang, Li Wang, Shukai Duan *, Jia Yan and Lidan Wang

College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China; jiapengfei200609@126.com (P.J.); 18580465830@163.com (T.H.); wangyu2007@email.swu.edu.cn (L.W.); yanjia119@163.com (J.Y.); ldwang@swu.edu.cn (L.W.)

* Correspondence: duansk@swu.edu.cn; Tel.: +86-139-8389-9976

Academic Editor: Carmen Horrillo Güemes

Received: 12 April 2016; Accepted: 20 June 2016; Published: 30 June 2016

Abstract: An electronic nose (E-nose) consisting of 14 metal oxide gas sensors and one electronic chemical gas sensor has been constructed to identify four different classes of wound infection. However, the classification results of the E-nose are not ideal if the original feature matrix containing the maximum steady-state response value of sensors is processed by the classifier directly, so a novel pre-processing technique based on supervised locality preserving projections (SLPP) is proposed in this paper to process the original feature matrix before it is put into the classifier to improve the performance of the E-nose. SLPP is good at finding and keeping the nonlinear structure of data; furthermore, it can provide an explicit mapping expression which is unreachable by the traditional manifold learning methods. Additionally, some effective optimization methods are found by us to optimize the parameters of SLPP and the classifier. Experimental results prove that the classification accuracy of support vector machine (SVM) combined with the data pre-processed by SLPP outperforms other considered methods. All results make it clear that SLPP has a better performance in processing the original feature matrix of the E-nose.

Keywords: sensor data; electronic nose; SLPP; data pre-processing; wound infection

1. Introduction

An electronic nose (E-nose) is an expert system which is composed of an array of gas sensors as well as a corresponding artificial intelligence algorithm. The E-nose is effective in dealing with problems of odor analysis [1,2], and has already been introduced into many fields such as disease diagnosis [3–5] and food engineering [6,7]. The sensor array of the E-nose has a characteristic of cross-sensitivity, namely different units in the array will make responses when facing the same smell, which can effectively avoid the decision-making risk brought by the single sensor. The original feature matrix which is extracted from the response of sensors is redundant, and much key information which plays a crucial role in helping the E-nose make the right judgment is submerged.

Previous work has confirmed that the E-nose can be used to detect wounds of patients in labs and it is feasible for the E-nose to detect bacteria including the in the investigation of bacterial volatile organic compounds (VOCs) from cultures and also from swabs taken from wound-infected patients [8–11]. In fact, there are many kinds of pathogenic bacteria which can lead to wound infection. Therefore, the same sampling experiment on one pathogenic bacterium must be repeated many times to make the E-nose learn about one kind of wound infection “deeply”; meanwhile, the sampling experiments based on different infection types are also necessary in order to let the E-nose distinguish more kinds of wound pathogens. Additionally, based on the original feature matrix extracted from the response

curves of sensors, the distance of points from different classes is larger than that of points from the same class, namely the data structure of the original feature matrix is nonlinear.

When this original feature matrix of wound infection data is put into the classifier directly, the classification accuracy of the E-nose is not ideal because of the redundancy of this matrix. So in order to capture more useful information which can be analyzed and to improve the classification accuracy of the E-nose as well, some methods must be applied to process the original feature matrix before it is put into the classifier.

Many seemingly complex systems described by high-dimensional data sets are in fact governed by a low number of parameters. The low-dimensional representation of such high-dimensional data sets not only leads to a more compact description of the data, but also enhances our understanding of the system [12]. Manifold learning is such a data processing method which can efficiently find the meaningful low-dimensional embedding from high-dimensional nonlinear data, and it processes the nonlinear data structure from the input matrix to a new matrix. In addition, the low-dimension output is very attractive to us, because this output will be processed by the classifier of the E-nose; if its dimension is low, then the computing complexity of the classifier is lower, and less computing and storage resources are needed. Many manifold learning methods have been proposed up to now [13,14], and a lot of work has been done to improve their performance [15–17]. Meanwhile, there are also many successful applications of manifold learning [18–22].

However, the traditional manifold learning cannot provide explicit mapping expression, which makes it unable to process the new sample points even if it can deal with the acquired data. Specifically, in our study, a trained E-nose will be used to predict the class label of unknown wound infection data, which means it is inevitable to encounter new data, and if the method employed by the E-nose cannot deal with these new points, it will make the E-nose incapable in practical application. Thus, supervised locality preserving projections (SLPP), a linear approximation of the nonlinear Laplacian eigenmaps, is applied. It not only shares many properties of the nonlinear techniques [23–27], but it also provides explicit mapping expression. As far as we know, SLPP has not been used in the field of the E-nose.

In this paper, we will use SLPP to process the original feature matrix and output a new matrix which can improve the classification accuracy of the E-nose. In Section 2, materials and experiments are demonstrated clearly. In addition, we will give the whole mathematical derivation of SLPP in Section 3. Then all considered pre-processing methods will be used to deal with the original feature matrix of wound infection data, and the classification results will be presented, analyzed and compared in Section 4. Finally, the conclusions of this paper are drawn in Section 5.

2. Materials and Experiments

2.1. Materials and Experimental Setup

In this study, sprague-dawley (SD) rats are chosen as the targets of wound pathogen infection. There are four kinds of rats, uninfected and those infected with *S. aureus*, *E. coli*, and *P. aeruginosa*, respectively. Each rat has a wound in its right hind leg and the pathogens are inoculated in the wound. The metabolites in the reproduction process of the three pathogens are shown in Table 1. According to the metabolites of pathogens and the response characteristics of gas sensors, 14 metal oxide sensors and one electronic chemical sensor are adopted to construct the sensor array of this paper (shown in Figure 1). In addition, the sensitive characteristics of the 15 sensors are shown in Table 2.

The practical E-nose system is shown in Figure 2. The gas sensor array is placed in a stainless steel test chamber coated by Teflon with the volume of 240 mL. A triple valve is used to change the gas circuit to make sure the desired gas can flow into the chamber. The flow velocity of gas is controlled by a flow meter and a mini-pump, and its value is set as 80 mL/min. The response curves of the sensor array obtained from the wound odor of rats are firstly conditioned through a conditioning circuit

and then sampled and saved in a computer via a 14-bit data acquisition system (DAS). A schematic diagram of the experimental system is shown in Figure 3.



Figure 1. Electronic nose sensor array.

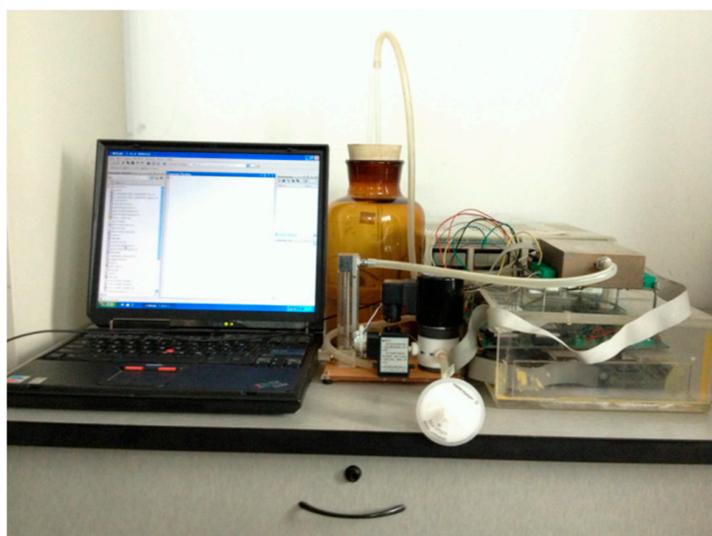


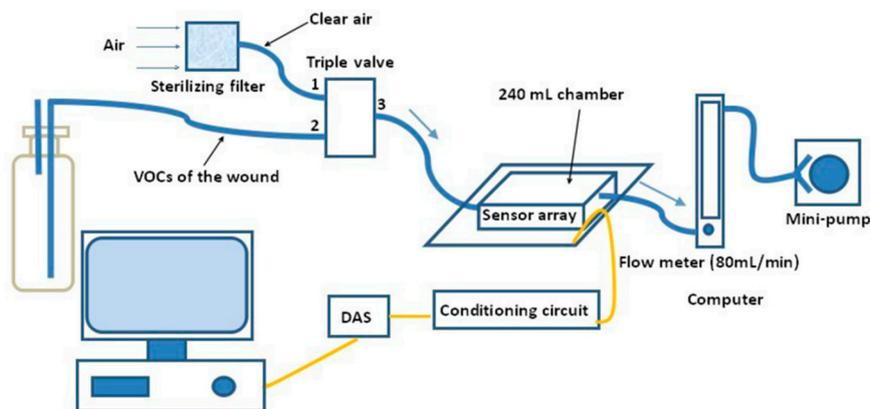
Figure 2. Practical E-nose system.

Table 1. Pathogens in wound infection and their metabolites.

Pathogens	Metabolites
<i>S. aureus</i>	Acetic acid, Aminoacetophenone, Ammonia, Ethanol, Formaldehyde, Isobutanol, Isopentyl acetate, Isopentanol, Methyl ketones, Trimethylamine, 1-Undecene, 2,5-Dimethylpyrazine isoamylamine, 2-Methylamine
<i>E. coli</i>	Acetaldehyde, Acetic acid, Aminoacetophenone, Butanediol, Decanol, Dimethyldisulfide, Dimethyltrisulfide, Dodecanol, Ethanol, Formaldehyde, Formic acid, Hydrogen sulfide, Indole, Lactic acid, Methanethiol, Methyl ketones, Octanol, Pentanols, Succinic acid, 1-Propanol
<i>P. aeruginosa</i>	Butanol, Dimethyldisulfide, Dimethyltrisulfide, Esters, Methyl ketones, Isobutanol, Isopentanol, Isopentyl acetate, Pyruvate, Sulphur compounds, Toluene, 1-Undecene, 2-Aminoacetophenone, 2-Butanone, 2-Heptanone, 2-Nonanone, 2-Undecanone

Table 2. Sensitive characteristic of gas sensors.

Sensors	Sensitive characteristic
TGS800	Methane, Carbon monoxide, Isobutane, Hydrogen, Ethanol
TGS813	Methane, Propane, Ethanol, Isobutane, Hydrogen, Carbon monoxide
TGS816	Combustible gases, Methane, Propane, Butane, Carbon monoxide, Hydrogen, Ethanol, Isobutane
TGS822	Organic solvent vapors, Methane, Carbon monoxide, Isobutane, n-Hexane, Benzene, Ethanol, Acetone
TGS825	Hydrogen sulfide
TGS826	Ammonia, Ethanol, Isobutane, Hydrogen
TGS2600	Gaseous air contaminants, Methane, Carbon monoxide, Isobutane, Ethanol, Hydrogen
TGS2602	VOCs, Odorous gases, Ammonia, Hydrogen sulfide, Toluene, Ethanol
TGS2620	Vapors of organic solvents, combustible gases, Methane, Carbon monoxide, Isobutane, Hydrogen, Ethanol
WSP2111	Benzene, Toluene, Ethanol, Hydrogen, Formaldehyde, Acetone
MQ135	Ammonia, Benzene series material, Acetone, Carbon monoxide, Ethanol, Smoke
MQ138	Alcohols, Aldehydes, Ketones, Aromatics
QS-01	VOCs, Hydrogen, Carbon monoxide, Methane, Isobutane, Ethanol, Ammonia
SP3S-AQ2	VOCs, Methane, Isobutane, Carbon monoxide, Hydrogen, Ethanol
AQ	Carbon monoxide, Methanol, Ethanol, Isopropanol, Formaldehyde, Acetaldehyde, Sulfur dioxide, Hydrogen, Hydrogen sulfide, Phenol, Dimethyl ether, Ethylene

**Figure 3.** Schematic diagram of the experimental system.

2.2. Data Collection

Each rat is placed in a jar of which the volume is 2.8 L with a rubber stopper. There are two holes in the rubber stopper with two thin glass tubes inserted. One glass tube is fixed above the wound of the infected rat as closely as possible. The output gas of the tube which contains VOCs of the rat wound flows out of the bottle and then flows into the chamber through a Teflon tube.

The dynamic headspace method is adopted in all the sampling experiments. Each sampling experiment contains the following three steps:

Step 1: Set the triple valve to make port 1 connect to port 3, so clean air is exposed to the chamber and lasts for 3 min to wash the sensor array;

Step 2: Set the triple valve to make port 2 connect to port 3, so the gas stream containing VOCs of the wound passes over the sensor array for 5 min;

Step 3: Set the triple valve to make port 1 connect to port 3, so the sensors are exposed to clean air again for another 15 min.

The clearance time between two experiments is 5 min. Twenty sampling experiments for each group of rats are made under the same conditions, and thus 80 sampling data sets can be collected. The response curves of 15 sensors on one wound infected with *P. aeruginosa* are shown in Figure 4. It can be seen that the obvious rise of each response curve is from the third min when the gas stream containing VOCs of the wound begins to pass over the sensor array, and the curves begin to drop from the eighth min when clean air is exposed to wash the sensor array.

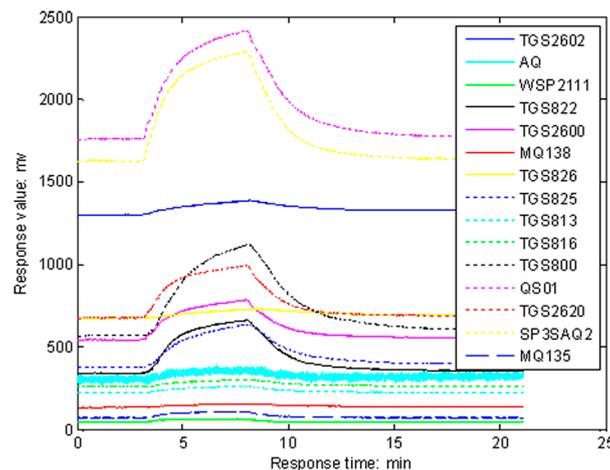


Figure 4. Response curves of 15 sensors on one wound infected with *P. aeruginosa*.

2.3. Original Feature Matrix

After all the sampling experiments have been finished, we succeed in extracting the maximum steady-state response value of each sensor during each sampling experiment, and then the original feature matrix of the wound infection data is constructed. The detailed information of this original feature matrix is shown in Figure 5. As it shows, there are 80 points in this matrix and the dimension of each point is 15. Each class of wound infection contains 20 points.

$$\begin{array}{cccc}
 \text{No-infection} & P. aeruginosa & E. coli & S. aureus \\
 \left(\begin{array}{cccc}
 \underbrace{x_{1,1} \cdots x_{1,20}} & \underbrace{x_{1,21} \cdots x_{1,40}} & \underbrace{x_{1,41} \cdots x_{1,60}} & \underbrace{x_{1,61} \cdots x_{1,80}} \\
 \vdots & \vdots & \vdots & \vdots \\
 x_{15,1} \cdots x_{15,20} & x_{15,21} \cdots x_{15,40} & x_{15,41} \cdots x_{15,60} & x_{15,61} \cdots x_{15,80}
 \end{array} \right)_{15 \times 80}
 \end{array}$$

Figure 5. Detailed information of this original feature matrix.

To study the data structure of the original feature matrix, we use Equation (1) to compute the average Euclidean distance of points in the original feature matrix.

$$\bar{d}_{ij} = \frac{1}{M_i} \frac{1}{M_j} \sum_{p=1}^{M_i} \sum_{q=1}^{M_j} dis(\mathbf{x}_p, \mathbf{x}_q), \quad i, j = 1, 2, \dots, 4 \quad (1)$$

where \bar{d}_{ij} is the average Euclidean distance between class i and j , M_i and M_j stand for the number of points in class i and j , and $dis(\cdot)$ is the Euclidean distance of points. The computed results are shown

in Table 3. For uninfected wounds and those infected with *S. aureus*, the distance of points from the same class is less than that of the points from different classes. However, the distance between *E. coli* and no infection is less than that of the points between *E. coli* and other different classes, and the same situation happens in *P. aeruginosa*. This proves that the data structure of the original feature matrix is nonlinear and complex.

Table 3. Average Euclidean distance of points in matrix X.

	No-Infection	<i>P. aeruginosa</i>	<i>E. coli</i>	<i>S. aureus</i>
No-infection	1155.5567	1372.7781	1325.8864	1344.9724
<i>P. aeruginosa</i>	1372.7781	1461.6700	1488.3676	1499.6072
<i>E. coli</i>	1325.8864	1488.3676	1416.4451	1523.1622
<i>S. aureus</i>	1344.9724	1499.6072	1523.1622	1100.3343

3. SLPP

Suppose there is a set of $\mathbf{x}_i, i = 1, 2, \dots, m$ in \mathbf{R}^N , find a transformation matrix \mathbf{A} mapping these points to a set of points $\mathbf{y}_i, i = 1, 2, \dots, m$ in \mathbf{R}^L , such that \mathbf{y}_i will represent \mathbf{x}_i , where $\mathbf{y}_i = \mathbf{A}^T \mathbf{x}_i$. The algorithmic procedure of SLPP can be formally stated below.

- (1) Constructing the neighborhood: \mathbf{x}_j becomes the neighbor of \mathbf{x}_i only if they are from the same class and are “close”, where both \mathbf{x}_i and \mathbf{x}_j are the points of \mathbf{X} and $i \neq j$. Additionally, two different ways can be employed to find the neighborhood of \mathbf{x}_i .
 - (a) ε -neighborhood: if $\|\mathbf{x}_i - \mathbf{x}_j\|^2 < \varepsilon, \varepsilon \in \mathbf{R}$, then \mathbf{x}_j can be taken as the neighbor of \mathbf{x}_i .
 - (b) k-nearest-neighbors: a judgment is made on whether \mathbf{x}_j is among the k-nearest neighbors of \mathbf{x}_i .
- (2) Describe the relationship between \mathbf{x}_i and \mathbf{x}_j : suppose that $w_{i,j}$ is a variable describing the relationship between these two points, and $w_{i,j}$ will be “larger” if \mathbf{x}_i and \mathbf{x}_j are “closer”. There are also two different methods available to realize it.
 - (a) simple-type: $w_{i,j} = 1$ if \mathbf{x}_j is the neighbor of \mathbf{x}_i ; otherwise, $w_{i,j} = 0$.
 - (b) heat-kernel:
$$w_{i,j} = \begin{cases} \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}) & \text{If } \mathbf{x}_j \text{ is the neighbor of } \mathbf{x}_i \\ 0 & \text{Otherwise} \end{cases}$$
- (3) Find the map: to make the relationship between \mathbf{y}_i and \mathbf{y}_j similar to that between \mathbf{x}_i and \mathbf{x}_j ; let \mathbf{Y} be a “good” map to minimize the following objective function [27].

$$F(\mathbf{y}_i, \mathbf{y}_j) = \sum_{i,j} (\mathbf{y}_i - \mathbf{y}_j)^2 w_{i,j} \quad (2)$$

under appropriate constraints, where \mathbf{y}_i and \mathbf{y}_j are the points of \mathbf{Y} and $i \neq j$. If \mathbf{x}_i and \mathbf{x}_j are “close” enough, then the value of $w_{i,j}$ will be much “larger”, and to make sure Equation (2) reaches its minimum, \mathbf{y}_i and \mathbf{y}_j must be “close” as well. In this way, Equation (2) transfers the local structure from matrix \mathbf{X} to \mathbf{Y} . Furthermore, because $\mathbf{y}_i = \mathbf{A}^T \mathbf{x}_i$, Equation (2) can be computed as

$$\begin{aligned} \frac{1}{2} \sum_{i,j} (\mathbf{y}_i - \mathbf{y}_j)^2 w_{i,j} &= \frac{1}{2} \sum_{i,j} (\mathbf{A}^T \mathbf{x}_i - \mathbf{A}^T \mathbf{x}_j)^2 w_{i,j} \\ &= \sum_{i,j} \mathbf{A}^T \mathbf{x}_i \mathbf{D}_{i,j} \mathbf{x}_i^T \mathbf{A} - \sum_{i,j} \mathbf{A}^T \mathbf{x}_i w_{i,j} \mathbf{x}_j^T \mathbf{A} = \mathbf{A}^T \mathbf{X} (\mathbf{D} - \mathbf{W}) \mathbf{X}^T \mathbf{A} = \mathbf{A}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{A} \end{aligned} \quad (3)$$

where $\mathbf{D}_{i,i} = \sum_{i,j} w_{i,j}$. \mathbf{A} constraint is imposed as follows [23]

$$\mathbf{Y}^T \mathbf{D} \mathbf{Y} = 1 \Rightarrow \mathbf{A}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{A} = 1 \quad (4)$$

Finally, the minimization problem is reduced to find

$$\arg \min_{\mathbf{A}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{A} = 1} \mathbf{A}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{A} \quad (5)$$

The transformation matrix \mathbf{A} minimizing Equation (5) can be given by the minimum eigenvalue solution to the generalized eigenvalue problem

$$\mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{A} = \lambda \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{A} \quad (6)$$

Let the vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_L$ be the solutions of Equation (6), and order them according to their eigenvalues, $\lambda_1 > \lambda_2 > \dots > \lambda_L$. Thus, the embedding is as follows

$$\mathbf{Y} = \mathbf{A}^T \mathbf{X}, \mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_L) \quad (7)$$

4. Results and Discussion

4.1. Experimental Results

To verify the effectiveness of SLPP, the original feature matrix of wound infection is processed by principal component analysis (PCA) [28], Fisher discriminant analysis (FDA) [29] and kernel FDA (KFDA) [30]. The original feature matrix (defined as matrix \mathbf{X}) which was introduced in Section 2.3 will be processed by these four different data processing methods to create a new feature matrix which has been denoted as \mathbf{Y} . Finally, matrix \mathbf{Y} is put into the classifier as its input. For SLPP, the k-nearest-neighbors method is employed to find the neighborhood, and if the size of the neighborhood is different, the local data structure will be changed which will finally influence the classification results of the E-nose. To solve this problem, the grid-searching method is adopted to set the number of the nearest neighbors in SLPP. The heat-kernel method is used to describe the relationship of points; the value of t will influence the performance of the heat-kernel, and so quantum-behaved particle swarm optimization (QPSO) [31] is used to find the optimal value of t .

In this paper, we employ support vector machine (SVM) [32] and k-nearest-neighbor (KNN) [33] as the classifiers, and the parameters of SVM are optimized by QPSO, and the size of the neighborhood in KNN is searched by the grid-searching method. The cross-validation method is adopted to train and test SVM, and the folds of cross-validation in this paper are 10, 40 and 80. The numbers of particles and iterations in QPSO are set as 60 and 500. Every single data processing method is evaluated by its corresponding classification results.

Tables 4–9 list the classification accuracy of SVM and KNN based on PCA, FDA, KFDA and SLPP when the folds of cross-validation are set as 10, 40 and 80, respectively. In addition, we also provide the classification results when matrix \mathbf{X} is put into SVM directly without being processed by any method (no-dealing). It is evident, for the classification accuracy of the total four classes, that the best performance is achieved by SLPP, and the worst one is achieved by FDA, no matter if the fold of the cross-validation is 10, 40 or 80.

Table 4. Classification results of 10-fold using different data processing methods (SVM).

Methods	L	Classification Accuracy (%)				
		No-Infection	<i>P. aeruginosa</i>	<i>E. coli</i>	<i>S. aureus</i>	Total
No-dealing	15	85	85	90	85	86.25
PCA	10	90	90	85	85	87.5
FDA	3	75	80	85	85	81.25
KFDA	3	90	95	95	95	93.75
SLPP	7	100	95	100	100	98.75

L is the dimension of matrix \mathbf{Y} , and for the no-dealing method, L is the dimensionality of matrix \mathbf{X} ; Total means the classification accuracy of the classifier in predicting the class label of the total 80 points; No-dealing means the original feature matrix is put into the classifier directly.

Table 5. Classification results of 40-fold using different data processing methods (SVM).

Methods	L	Classification Accuracy (%)				
		No-Infection	<i>P. aeruginosa</i>	<i>E. coli</i>	<i>S. aureus</i>	Total
No-dealing	15	85	90	90	75	85
PCA	10	90	80	90	85	86.25
FDA	3	75	80	70	95	80
KFDA	3	90	95	90	95	92.5
SLPP	7	100	95	90	100	96.25

Table 6. Classification results of 80-fold using different processing methods (SVM).

Methods	L	Classification Accuracy (%)				
		No-Infection	<i>P. aeruginosa</i>	<i>E. coli</i>	<i>S. aureus</i>	Total
No-dealing	15	80	80	95	75	82.5
PCA	10	85	85	90	75	83.75
FDA	3	75	80	70	95	80
KFDA	3	85	85	90	90	87.5
SLPP	7	100	85	90	100	93.75

Table 7. Classification results of 10-fold using different data processing methods (KNN).

Methods	L	Classification Accuracy (%)				
		No-Infection	<i>P. aeruginosa</i>	<i>E. coli</i>	<i>S. aureus</i>	Total
No-dealing	15	85	80	80	85	82.5
PCA	11	90	85	75	85	83.75
FDA	3	85	80	75	85	81.25
KFDA	3	95	90	90	90	91.25
SLPP	8	100	90	90	100	95

Table 8. Classification results of 40-fold using different data processing methods (KNN).

Methods	L	Classification Accuracy (%)				
		No-Infection	<i>P. aeruginosa</i>	<i>E. coli</i>	<i>S. aureus</i>	Total
No-dealing	15	80	80	75	80	81.25
PCA	11	85	80	75	85	81.25
FDA	3	75	75	70	80	77.5
KFDA	3	90	90	85	90	88.75
SLPP	8	100	90	85	100	93.75

Table 9. Classification results of 80-fold using different processing methods (KNN).

Methods	L	Classification Accuracy (%)				
		No-Infection	<i>P. aeruginosa</i>	<i>E. coli</i>	<i>S. aureus</i>	Total
No-dealing	15	75	75	75	85	77.5
PCA	11	80	80	75	85	80
FDA	3	75	75	70	85	76.25
KFDA	3	85	90	85	90	87.5
SLPP	8	100	80	85	100	91.25

4.2. Discussion

This paper focuses on the investigation of the performance of SLPP in pre-processing the original feature matrix of wound infection data. Three other methods are also used to deal with this original

feature matrix. When PCA is used to deal with the original feature matrix, its improvement is not obvious because PCA is good at finding and keeping the linear structure of data. FDA can make use of the class label information of the original feature matrix and find a linear transformation which can maximize the between-class scattering and minimize the within-class scattering to achieve a new feature matrix. However, the performance of FDA in predicting the class of wound infection is the worst among all the processing methods. As an enhanced technique of FDA, KFDA firstly maps the data to the high-dimension space, and then finds the same transformation as FDA, and its classification results of wound infection are better than that of FDA. However, SLPP manages to preserve the local structure of the data set through finding and keeping the neighbors of each point, and it can make use of the class label information during the course of finding the nearest neighbors. Further experimental results prove that the classification accuracy of the E-nose increases when SLPP is used to pre-process the original feature matrix of wound infection data, and meanwhile, it reduces the dimension of points from 15 to 7, which can greatly decrease the computational complexity of classifier.

5. Conclusions

Although the sensor array of the E-nose is good at cross-sensitivity, the original feature matrix extracted from the response curves of sensors is redundant; meanwhile, the data structure in this matrix is nonlinear. Traditional manifold learning methods are capable of solving the nonlinear problem, but they cannot provide an explicit mapping expression, which limits their application in the field of the E-nose. As a novel manifold learning technique, SLPP can efficiently find the meaningful low-dimensional embedding from high-dimensional nonlinear data, and it can process the nonlinear data structure from the input matrix to a new matrix; furthermore, the explicit mapping expression given by SLPP makes it possible for the E-nose to process the new sampling points.

The experimental results of this paper have proved that the classification accuracy of SVM combined with SLPP is much higher than that of other considered methods. All in all, SLPP is an ideal technique for the E-nose to pre-process its original feature matrix of wound infection data and improve its classification accuracy.

Acknowledgments: The work is supported by Program for New Century Excellent Talents in University (No. [2013] 47), National Natural Science Foundation of China (Nos. 61372139, 61101233, 60972155), Fundamental Research Funds for the Central Universities (No. XDJK2015C073), Science and Technology personnel training program Fund of Chongqing (No. Cstc2013kjrc-qnrc40011) and Fundamental Research Funds for the Central Universities (No. SWU115009).

Author Contributions: Pengfei Jia is the group leader and he was in charge of the project management and proposed the algorithm. Tailai Huang was responsible for data analysis and the discussion of the results. Shukai Duan provided valuable advice about the revised manuscript. Li Wang, Jia Yan and Lidan Wang were involved in discussions and the experimental analysis.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bicego, M.; Tessari, G.; Tecchiolli, G.; Bettinelli, M. A comparative analysis of basic pattern recognition techniques for the development of small size electronic nose. *Sens. Actuator B Chem.* **2002**, *85*, 137–144. [[CrossRef](#)]
2. Ciosek, P.; Wróblewski, W. The analysis of sensor array data with various pattern recognition techniques. *Sens. Actuator B Chem.* **2006**, *114*, 85–93. [[CrossRef](#)]
3. Yu, J.B.; Byun, H.G.; So, M.S.; Huh, G.S. Analysis of diabetic patient's breath with conducting polymer sensor array. *Sens. Actuator B Chem.* **2005**, *108*, 305–308. [[CrossRef](#)]
4. Pavlou, A.K.; Magan, N.; McNulty, C.; Jones, J.M.; Sharp, D.; Brown, J.; Turner, A.P. Use of an electronic nose system for diagnoses of urinary tract infections. *Biosens. Bioelectron.* **2002**, *17*, 893–899. [[CrossRef](#)]
5. Anh, D.T.V.; Olthuis, W.; Bergveld, P. A hydrogen peroxide sensor for exhaled breath measurement. *Sens. Actuator B Chem.* **2005**, *111–112*, 494–499. [[CrossRef](#)]

6. Trihaas, J.; Nielsen, P.V. Electronic nose technology in quality assessment: monitoring the ripening process of danish blue cheese. *J. Food Sci.* **2005**, *70*, 44–49.
7. Natale, C.D.; Macagnano, A.; Davide, F.; D’Amico, A.; Paolesse, R.; Boschi, T.; Faccio, M.; Ferri, G. An electronic nose for food analysis. *Sens. Actuator B Chem.* **1997**, *44*, 521–526. [[CrossRef](#)]
8. Dutta, R.; Dutta, R. “Maximum probability rule” based classification of MRSA infections in hospital environment: Using electronic nose. *Sens. Actuator B Chem.* **2006**, *120*, 156–165. [[CrossRef](#)]
9. Thomas, A.N.; Riazanskaia, S.; Cheung, W.; Xu, Y.; Goodacre, R.; Thomas, C.L.P.; Baguneid, M.S.; Bayat, A. Novel noninvasive identification of biomarkers by analytical profiling of chronic wounds using volatile organic compounds. *Wound Repair Regen.* **2010**, *18*, 391–400. [[CrossRef](#)] [[PubMed](#)]
10. Byun, H.; Persaud, K.C.; Pisanelli, A.M. Wound-state monitoring for burn patients using E-nose/SPME system. *ETRI J.* **2010**, *32*, 440–446. [[CrossRef](#)]
11. Jia, P.; Tian, F.; He, Q.; Fan, S.; Liu, J.; Yang, S.X. Feature extraction of wound infection data for electronic nose based on a novel weighted KPCA. *Sens. Actuator B Chem.* **2014**, *201*, 555–566. [[CrossRef](#)]
12. Goldberg, Y.; Zakai, A.; Kushnir, D.; Ritov, Y. Manifold learning: The price of normalization. *J. Mach. Learn. Res.* **2008**, *9*, 1909–1939.
13. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [[CrossRef](#)] [[PubMed](#)]
14. Tenenbaum, J.B.; Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323. [[CrossRef](#)] [[PubMed](#)]
15. Li, H.; Jiang, H.; Barrio, R.; Liao, X.; Cheng, L.; Su, F. Incremental manifold learning by spectral embedding methods. *Pattern Recognit. Lett.* **2011**, *32*, 1447–1455. [[CrossRef](#)]
16. Chatpatanasiri, R.; Kijssirikul, B. A unified semi-supervised dimensionality reduction framework for manifold learning. *Neurocomputing* **2010**, *73*, 1631–1640. [[CrossRef](#)]
17. Shen, J.; Bu, J.; Ju, B.; Jiang, T.; Wu, H.; Li, L. Refining gaussian mixture model based on enhanced manifold learning. *Neurocomputing* **2012**, *87*, 19–25. [[CrossRef](#)]
18. Han, Y.; Xu, Z.; Ma, Z.; Huang, Z. Image classification with manifold learning for out-of-sample data. *Signal Process.* **2012**, *93*, 2169–2177. [[CrossRef](#)]
19. Cheon, Y.; Kim, D. Natural facial expression recognition using differential-AAM and manifold learning. *Pattern Recognit.* **2009**, *42*, 1340–1350. [[CrossRef](#)]
20. Hadid, A.; Pietikäinen, M. Demographic classification from face videos using manifold learning. *Neurocomputing* **2012**, *100*, 197–205. [[CrossRef](#)]
21. Lu, X.; Wang, H.; Zhou, R.; Ge, B. Autonomic failure prediction based on manifold learning for large-scale distributed systems. *J. China Univ. Posts Telecommun.* **2010**, *17*, 116–124. [[CrossRef](#)]
22. Wachinger, C.; Yigitsoy, M.; Rijkhorst, E.G.; Navab, N. Manifold learning for image-based breathing gating in ultrasound and MRI. *Med. Image Anal.* **2012**, *16*, 806–818. [[CrossRef](#)] [[PubMed](#)]
23. He, X. Locality preserving projections. *Adv. Neural Inf. Process. Syst.* **2005**, *45*, 186–197.
24. Zhao, Z.; Zhang, L.; Zhao, M.; Hou, Z.; Zhang, C. Gabor face recognition by multi-channel classifier fusion of supervised kernel manifold learning. *Neurocomputing* **2012**, *97*, 398–404. [[CrossRef](#)]
25. Zheng, Z.; Yang, F.; Tan, W.; Jia, J.; Yang, J. Gabor feature-based face recognition using supervised locality preserving projection. *Signal Process.* **2007**, *87*, 2473–2483. [[CrossRef](#)]
26. Cheng, J.; Liu, Q.; Lu, H.; Chen, Y. Supervised kernel locality preserving projections for face recognition. *Neurocomputing* **2005**, *67*, 443–449. [[CrossRef](#)]
27. Belkin, M.; Niyogi, P. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Adv. Neural Inf. Process. Syst.* **2002**, *14*, 585–591.
28. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [[CrossRef](#)]
29. Zuo, W.; Zhang, H.; Zhang, D.; Wang, K. Post-processed LDA for face and palmprint recognition: What is the rationale. *Signal Process.* **2010**, *90*, 2344–2352. [[CrossRef](#)]
30. Baudat, G.; Anouar, F. Generalized discriminant analysis using a kernel approach. *Neural Comput.* **2000**, *12*, 2385–2404. [[CrossRef](#)] [[PubMed](#)]
31. Sun, J.; Feng, B.; Xu, W. Particle swarm optimization with particles having quantum behavior. In Proceedings of the Congress on Evolutionary Computation, Portland, OR, USA, 19–23 June 2004; pp. 325–331.

32. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
33. Dasarathy, B.V. Nearest neighbor classification techniques. *Stud. Fuzziness Soft Comput.* **1994**, *27*, 181–203.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).