Junyu Yan, jy420, 01895990, Yibing Liu, yl11820, 01939400

1. Data Preparation

Task 1 As listed in the Appendix, Figure 1, Figure 2 and Figure 3 show the pressure, vibration, temperature (PVT) and electrode data of 6 objects detected by the F0 finger or F1 finger of BiTac[2] three trials.

From the result, we choose the time step 15 for further sampling as illustrated by the black line in figures. We observe that the pressure and electrode data of some objects fuse into other groups from time 100, and the vibration data begins to merge at about time step 25, therefore, the choice of time should be before the time step 25. Moreover, we omit the time steps at the very beginning to avoid possible influences caused by phase transition. At time instance 15, we observe that the separation of PVT and electrode data between different objects is large enough to differentiate the objects. Though some objects may mix at one variable, they can be distinguished at the other variables (e.g., as shown in Figure 1, pressures of acrylic, black foam and steel vase are similar, but their temperature and electrode data have wide differences). Also, from time step 15, the pressure and electrode data tend to saturation. In summary, 15 is a reasonable choice.

Task 2 We sample the F0 finger data at time step 15 to generate the new PVT and Electrodes datasets. In F0_PVT.mat, it contains 6 matrix of size 3×6 with different object names, where each row in the matrix represents the P, V, T data of the object at time instant 15 for 10 trials. Moreover,the F0_Electrode.mat file also contains 6 matrix with size 19×10, which are the 19 electrode data of different objects for 10 trials. The detailed structures of two .mat file are listed in the Figure 4.

Task 3 As shown in Figure 5, the PVT data of car sponge, flour sack and kitchen sponge are clustered. However, the black foam, acrylic and steel vase mix into other groups. The reasons are that both the steel vase and acrylic are smooth, hard, and have good thermal conductivity[1], therefore, their obtained PVT data will be too similar to be differentiated. From the perspective of the vibration and pressure plane, there is a clear distinction, since the car sponge, flour sack and kitchen sponge are all belong to the foam group with rough texture and springy structure, while steel vase and acrylic are of the opposite. Generally, the features of the black foam is similar to the foam group[3], however, it does not have good resilience and may be deformed if applied with large force. As a result, the black foam may not at its original shape for each trial so that at some trials it becomes hard and smooth with surface squeezed, leading to distributions on both sides. At the temperature axis, the black foam and car sponge are heat-insulating, therefore, their temperatures are slightly higher than other objects. Since all objects are placed at the room temperature, therefore, their temperatures do not significantly different.

2. Principle Component Analysis

Task 1 In this task, we determine the principal components of the PVT data following 5 steps. First, we normalize the data by subtracting the mean and rescaling the variance to one so that each variable contributes same to the analysis. Then, we obtain the covariance matrix, eigenvalues, and eigenvectors of the PVT data as listed in Table 1. The standardised data and principal components (PC) are shown in Figure 6. Thirdly, we project the data onto the top two priciple components to reduce the data into 2-dimension, and the result is shown in Figure 6. Finally, Figure 7 shows the distribution of data across 1-dimension lines by projecting the PVT data onto each principle component.

Covariance matrix	$\begin{bmatrix} 1.0000 & -0.3330 & -0.3008 \\ -0.3330 & 1.0000 & -0.0248 \\ -0.3008 & -0.0248 & 1.0000 \end{bmatrix}$
Eigenvalues	$\begin{bmatrix} 0.5387 & 0 & 0 \\ 0 & 1.0246 & 0 \\ 0 & 0 & 1.4366 \end{bmatrix}$
Eigenvectors	$\begin{bmatrix} -0.6973 & 0.0056 & 0.7167 \\ -0.5294 & 0.6701 & -0.5203 \\ -0.4832 & -0.7423 & -0.4643 \end{bmatrix}$

Table 1. T2.1 Covariance matrix, eigenvalues, and eigenvectors of PVT data

Figure 6 proves that after removing bias from all variables, the PVT data is centred at the origin $(0, 0, 0)$. Since the covariance matrix is negative except the diagonal, the pressure, vibration and temperature of objects are inversely correlated. Also, through dividing the eigenvalues of each principle components by the sum of eigenvalues, it can be concluded that the PC with the largest eigenvalue account for the 47.89% variance of PVT data, while the second one accounts for 34.15%, and the last one carries 17.96% variance. Therefore, the dimensions of data can be reduced to 2D with most information preserved. By comparing the original data and 2D projected data shown in Figure 6, we observe that the distribution of data on 2D principal components has similar characteristics as the original distribution as objects with similar features are close together. Therefore, reducing data onto 2D dimensions still contains most information of the data. Moreover, the variance of data has been increased so that the separation between data becomes larger, which is beneficial for classifying data into different groups. From the perspective of 1D reprojection, as shown in Figure 7, the patterns emerge. The first component has the largest variance with some data clustered. Therefore, there is some potential to discriminate and cluster the PVT data according to objects features. However, the second component and the third component has smaller variances, and do not seem to correlate cleanly.

Task 2 The variances of each principal component for the electrode data are shown in Figure 8. Since there is an "elbow" at the third principal component, it is proper to choose the number of PCs as three to contain the most information of electrode data while reducing the dimensions to decrease computation complexity. Figure 9 shows the projected electrode data onto the three principal components with largest variance. The result illustrates that in the 3D reprojected space, the data still preserve the characteristics of the original distribution with data of similar features clustered. However, it is much convenient to analyze the 3D data than the 19 dimensions. Also, PCA increase the data variance so that the separation between data of different objects can be enlarged. Consequently, PCA is a sufficient way to pre-process the data before classification.

3. Linear Discriminant Analysis

Firstly, we distinguish the black foam and car sponge using any two properties of pressure, vibration, and temperature. The results are shown in 10. Then, we apply LDA to the 3D PVT data, the results are shown in Figure 11.

Figure 10 and Figure 11, illustrate that the black foam and car sponge can be distinguished by using any two physical properties of PVT or the whole three properties. The 2D LDA illustrates that the black foam and car sponge can be better classified with pressure and vibration. That's because the car sponge is softer and more rough than black foam, but since they are all have heat preservation property and they are put in the room temperature, there is not huge differences in the temperature. In addition, Whether in 1D, 2D or 3D plot, car sponge data are more concentrated, which means the car sponges have better robustness.

We choose steel vase and kitchen sponge to repeat the LDA analysis. In the previous experiment, the black foam and car sponge have a certain similarity, therefore we prefer choosing two objects with totally different properties (hard vs soft, smooth vs rough, and good thermal conductivity vs heat insensitive). we want to test that using these two distinctly different objects, the classification after LDA analysis will be more obvious. The results are shown in Figure 12 and Figure 13. It is clear that, the data difference between the two objects is very obvious, especially when split the data in term of Pressure vs. Vibration, which also corresponds to the completely different thermal conductivity and texture of two objects. When projecting data to a two-dimensional plane or one-dimensional straight line, the data difference between the two objects is obviously larger than that of the previous group of objects. We determine that the classification results of the LDA depends on the differences between the objects' properties. The LDA analysis can always maximize the difference between different classes of data.

4. Clustering & Classification

Task 1 In this section, we first use the K-Means Clustering to cluster the PVT data with Euclidean distance metric. The result is shown in Figure 15, where different markers represent different clustered groups, and different color represent different object.

According to the result, we observe that the car sponge, kitchen sponge and flour sack are successfully clustered corresponding to the original groups, wheres the black foam, steel vase and acrylic are clustered wrongly into other groups. The reasons are that the steel vase and acrylic have significantly similar texture, thermal conductivity and stiffness, therefore, the trials under similar room temperature are clustered together for these two objects. For black foam, it is squishy, soft and compressible. Therefore, generally its characteristics are similar to foam. However, after one trial, it may not return to its original shape before a new trial, then its pressure will decrease and its surface will be smoothed due to squeezing holes on the surface. As a result, some trials data are clustered into the car sponge, while some data of it are clustered into metal and acrylic groups. For car sponge, kitchen sponge and flour sack, they are all rough and soft so that they are separated from steel and acrylic. However, all of them have individual prominent characteristics (e.g., car sponge is heat-insulating, flour sack is extremely rough) resulting in a clear distinction between these three groups.

We repeat the clustering with Manhattan distance to evaluate the effect of different distance metric. One result is shown in Figure 16. It can be seen that clustering with Euclidean always has the same result with the Manhattan, but sometimes Manhattan outperforms the Euclidean. That's because if two points are close on most variable but more discrepant on one of them, then the Manhattan will reduce its effect and pays more attention on the closeness of the other variables, while the Euclidean will exaggerate the difference. In our classification task, the data with two similar variables are always belong to one object group, therefore, the Manhattan is better than Euclidean at some times.

Though K-Means clustering we used achieved satisfactory results, it still has some disadvantages. First, the choice of centriod is quite important. Since we randomly choose the centroid in the work space, if the positions of centroids are not close to original groups, then the clustering will perform bad, which may even does not converge or cluster nothing into some groups. As a result, we repeat the experiment several times to achieve a good clustering result.

Task 2 In this section, we use bagging and decision trees to classify objects with the reduced 3D electrode data processed by PCA. For the number of the tree, we adopt 20 trees. Because as the Figure 17 shows, after the number reaches 20, continuing to increase the number does not reduce the Out-Of-Bag (OOB) error much. For overall 60 data, we split 36 for training and 24 for testing. To make sure we can get the data from every object, we randomly select 6 from each object groups to forming training set and the other is used for testing set. Using the function `TreeBagger` , we can get the decision tree, as shown in Figure 18. Running the trained model with test data, we can get the confusion matrix as shown in Figure 19.The class number 1 to 6 corresponds to the acrylic, black foam, car sponge, flour sack, kitchen sponge and steel vase, respectively. It can be seen that for 24 test data, there are only two misclassifications. The overall accuracy is 91.7%, which is acceptable. We also repeated experiment for several times, and the accuracy is always greater than 90%. The results for repeated experiments are shown in Figure 19.

As Figure 19 shows, the misclassification mainly happen between the acrylic and steel vase, since they are quite similar with hard structure, smooth surface and good thermal conductivity. As a result, one of the acrylic was incorrectly marked as 6 (steel vase) and one of the steel vase was marked as 1 (acrylic). In the data classification process, there may be some random error due to the small training datasets. As shown in Figure 20 (a) in the appendix, one of the acrylic is marked as car sponge incorrectly. which is very rare. When we repeat the experiment, misclassification between acrylic and steel vase is common. But they are rarely marked as others.

PCA step is useful, because it increase the variance of the data while reducing the dimensions of the data. Therefore, when we classify the data, using the data processed with PCA can be more concise and faster with accuracy preserved. Also, it can increase separation between data to facilitate the classification.

5. Conclusion

We have used the PCA, LDA, K-Means and decision trees to analyze the PVT and Electrode data of 6 different objects. The PCA method is helpful to increase the variance between data points, and reduce the dimensions of data to decrease computational complexity with accuracy preserved. Therefore, it is beneficial for separating data, and saving much cost especially for machine learning tasks. However, as PCA is an unsupervised algorithm, it does not consider the categories of data during the dimension reduction. LDA is another dimension reduction method, but it reprojects the data with good class-separability to avoid overfitting for classification task. Consequently, it is quite useful for multi-class classification tasks due to the enlarged separation between classes and simple data structures with less dimensions. The k-Means, bagging and decision trees are clustering method, which can help us to classify the data into different groups according to some features.

The presented results prove that when the features of objects are not very similar, it is possible to distinguish the objects only using touch. Nevertheless, if haptic properties of different objects are quite close, it is difficult to differentiate them with only physical interactions.

From figures listed in the Data preparation section, we find that most pressure and electrode data of different objects are separated from each other, wheres the temperature and vibration data are merged. Additionally, the LDA results suggest that the pressure and vibration features distinguish different objects best. Therefore, for sensors without electrode channel, the vibration and pressure are the most important properties for a cheaper tactile sensor to measure. However, the results of decision tree with only electrode data has clearly differentiate different objects with a high accuracy. As a result, the sensor with electrode channel can only measure the electrode data for object classification.

An alternative method to prepare the data is using different time steps for different sensors. As shown in Figure 1, Figure 2 and Figure 3, it is clear that the maximum separation occurs at different time steps for PVT and electrode data. More specifically, the temperatures of objects have significant differences in later time steps. However, the pressure and vibration are quite different at the beginning, and finally merge. In result, choosing different time steps for different variables can increase the separation between classes, which is in favor of object classification. However, since the PVT and electrode data are correlated, choosing time step for different properties may not correctly illustrate the features of an object, leading to a classification mistake. Moreover, this method may lead to overfitting.

References

[1] V. Chu, I. McMahon, L. Riano, C. G. McDonald, Q. He, J. M. Perez-Tejada, M. Arrigo, N. Fitter, J. C. Nappo, T. Darrell, et al. Using robotic exploratory procedures to learn the meaning of haptic adjectives. In *2013 IEEE International Conference on Robotics and Automation*, pages 3048–3055. IEEE, 2013.

[2] J. Fishel, G. Lin, and G. Loeb. Biotac® product manual. *SynTouch LLC, February*, 2013.

[3] Wikipedia contributors. Foam rubber — Wikipedia, the free encyclopedia, 2021. [Online; accessed 22-March-2021].

6. Appendix

6.1. Data Preparation

This section lists the figures related to original data preparation. Figure 1, Figure 2 and Figure 3 present the time series PVT or electrode data of 6 objects detected by the F0 finger or F1 finger of BioTac and with different trials. Figure 4 lists the formats of our generated F0_PVT.mat and F0_Electrode.PVT files. Figure 5 shows the 3D scatter plot of the full content of PVT.mat file.

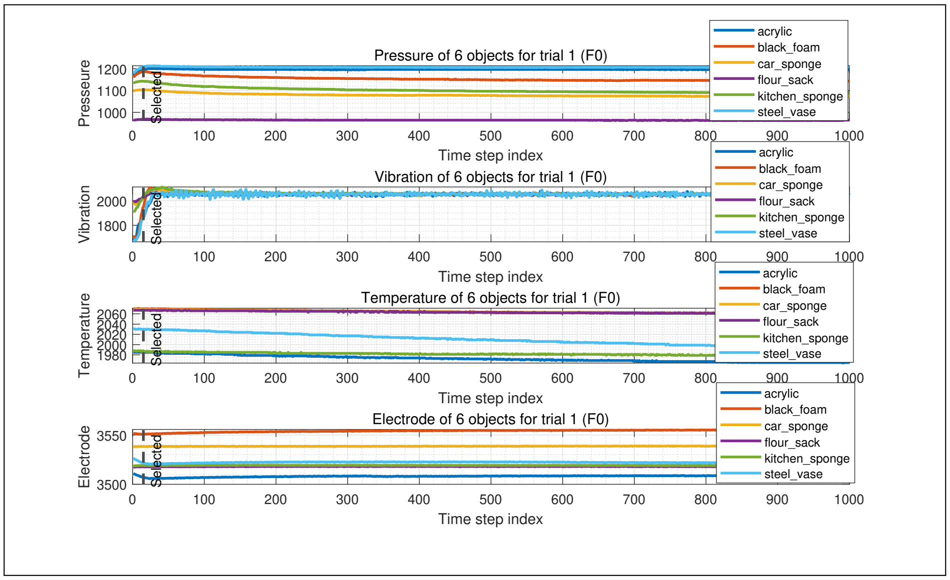

Figure 1. T1.1 The PVT and Electrode data taken by F0 finger for six objects in trial one

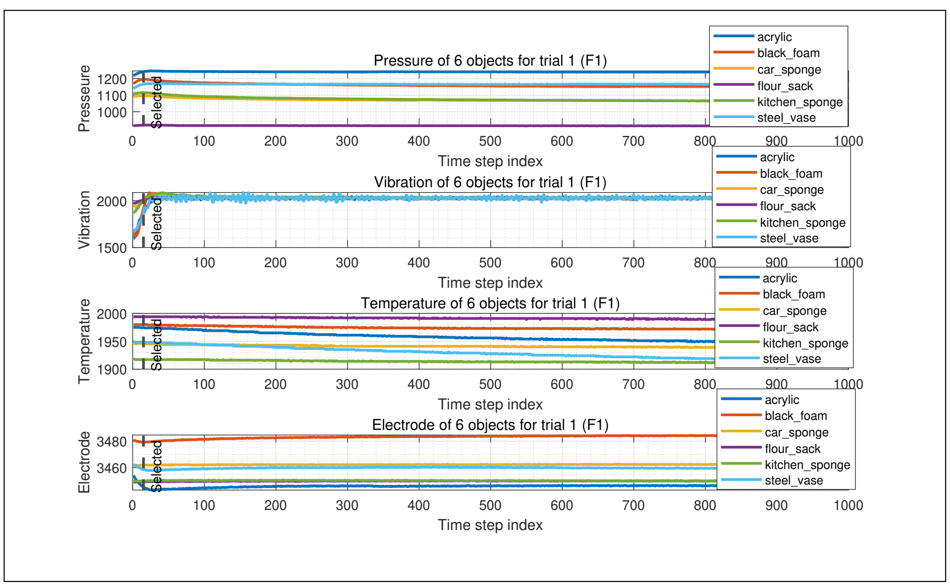

Figure 2. T1.2 The PVT and Electrode data taken by F1 finger for six objects in trial one

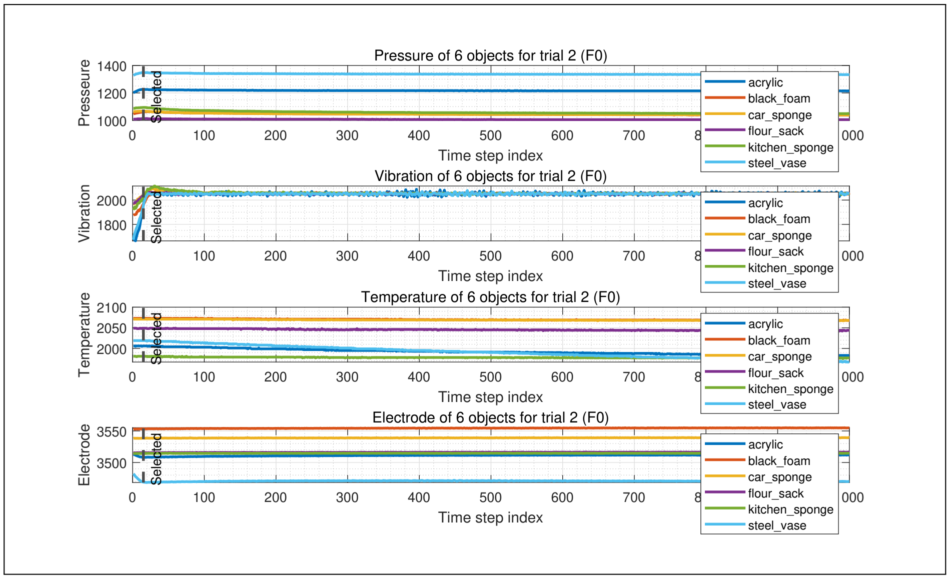

Figure 3. T1.3 The PVT and Electrode data taken by F0 finger for six objects in trial two

F0_PVT.mat (MAT 文件)

名称	值
acry_PVT	3x10 double
black_PVT	3x10 double
car_PVT	3x10 double
flour_PVT	3x10 double
kitchen_PVT	3x10 double
steel_PVT	3x10 double

(a)

F0_Electrode.mat (MAT 文件)

名称	值
acry_Electrode	19x10 double
black_Electrode	19x10 double
car_Electrode	19x10 double
flour_Electrode	19x10 double
kitchen_Electrode	19x10 double
steel_Electrode	19x10 double

(b)

Figure 4. Left: the format of F0_PVT.mat file; right: the format of F0_Electrode.mat file

6.2. Principle Component Analysis

Task 1 This section lists the PVT data processed by PCA. The left figure in Figure.6 shows the standardized data with principal components, while the right figure shows the projected PVT data onto two principal components with largest variance.

Task 2 This section lists the electrode data processed by PCA. Figure 8 shows the scree plot of variances of each principal component, while Figure 9 presents the reprojected electrode data with PCA.

6.3. Linear Discriminant Analysis

Task 1 This section lists the processed PVT data using LDA. The Figure 10 and Figure 12 show the LDA in 2D, the Figure 11 and Figure 13 show the LDA in 3D.

6.4. Clustering & Classification

Task 1 In this section, we listed the result for K-Means clustering. The original data before clustering is shown in Figure **??**. The K-Means clustering result is shown in Figure 15. The clustering results with Euclidean and Manhattan distance of different trials are listed in Figure 16.

Task 2 Figure 17 shows the relationship between the Out-of-Bag error with the number of trees.

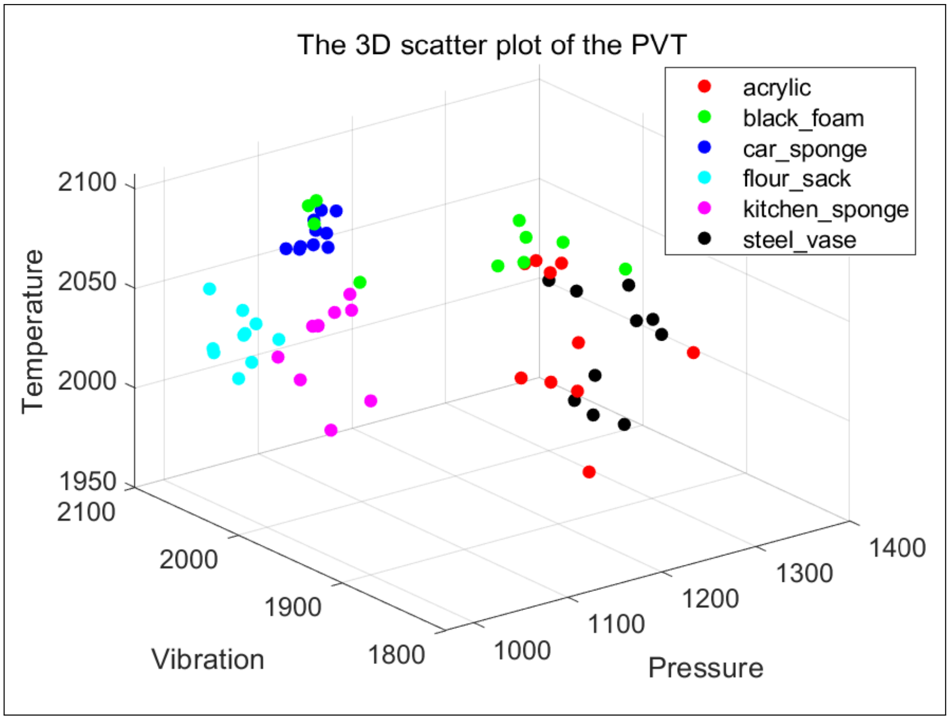

Figure 5. 3D scatter plot of PVT data with axes as pressure, vibration and temperature, and different colors represent different objects

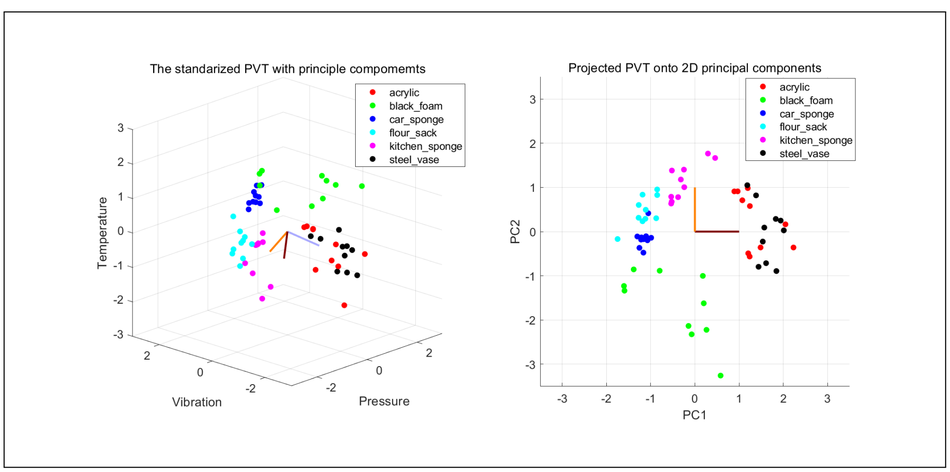

Figure 6. Left: The standardised PVT data with three principal components; Right: Projected PVT data on the 2D plane

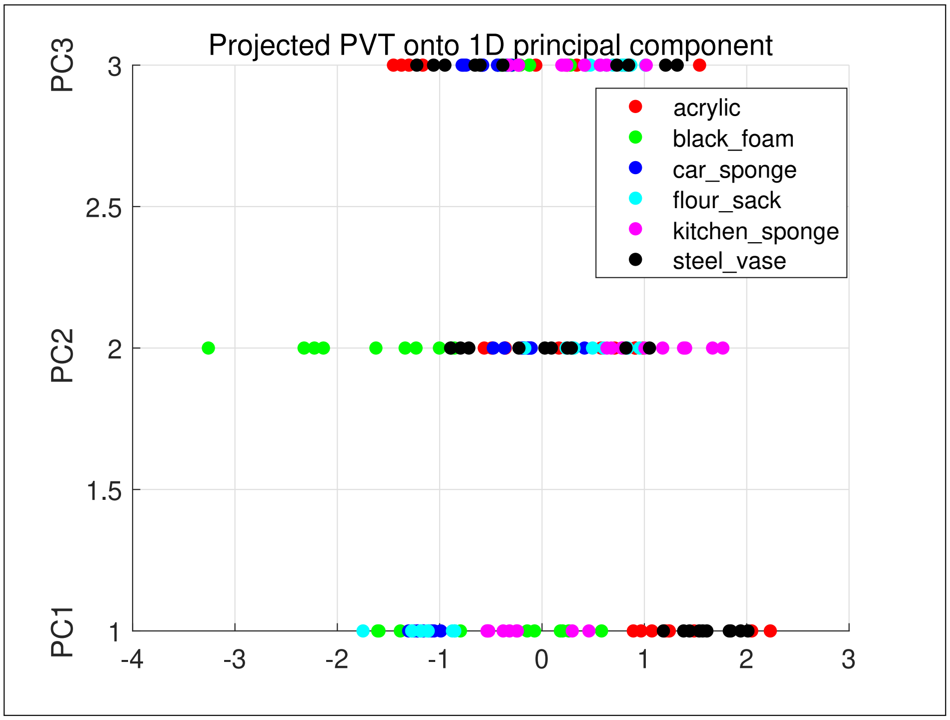

Figure 7. T2.3 Variance visualization

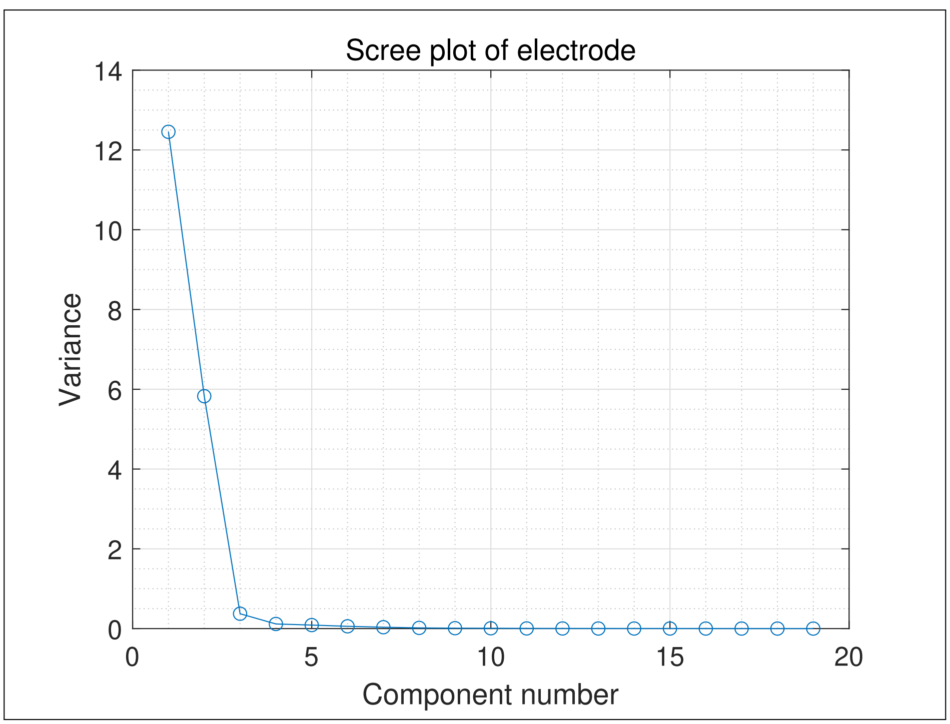

Figure 8. T2.2.a The variance of each principal component

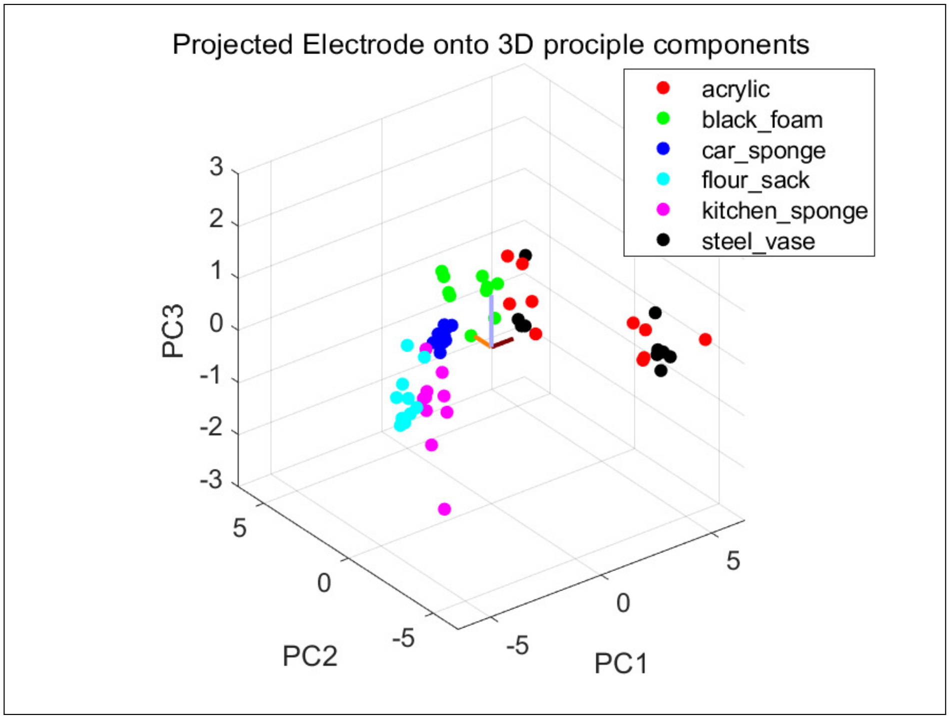

Figure 9. T2.2.b 3D scatter plot of the electrode data after projecting data onto three top principle components

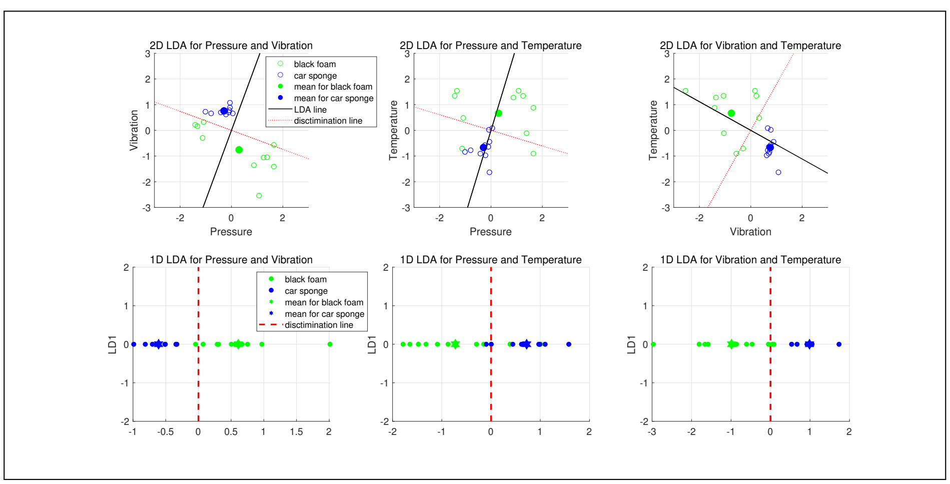

Figure 10. T3.1 Use LDA to split the the black foam and the car sponge

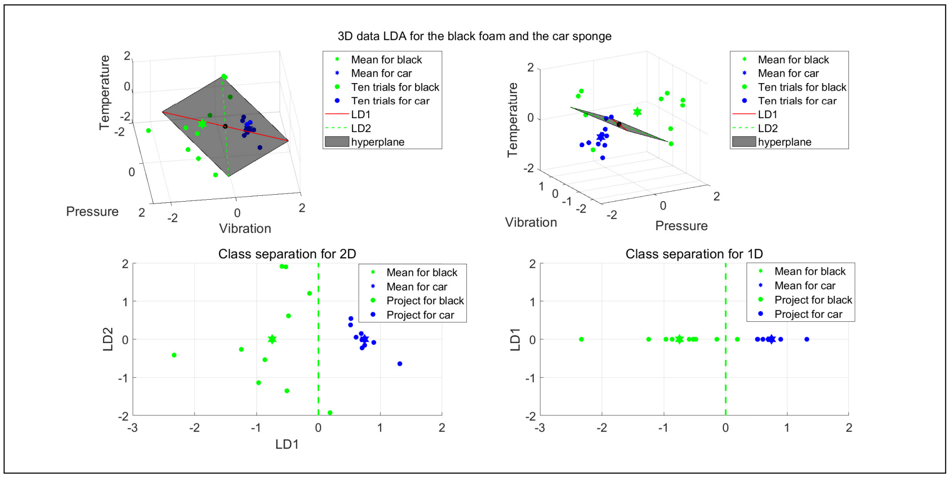

Figure 11. T3.2 Apply LDA to the three-dimensional PVT data. The two pictures above are hyperplane from different perspectives, The two figures below are LDA two-dimensional projection and one-dimensional projection

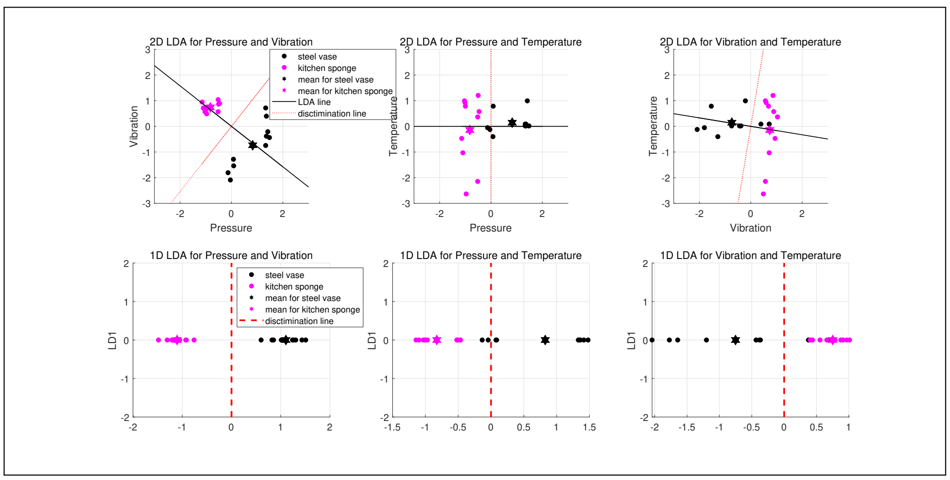

Figure 12. T3.4 Use LDA to split the the steel vase and kitchen sponge

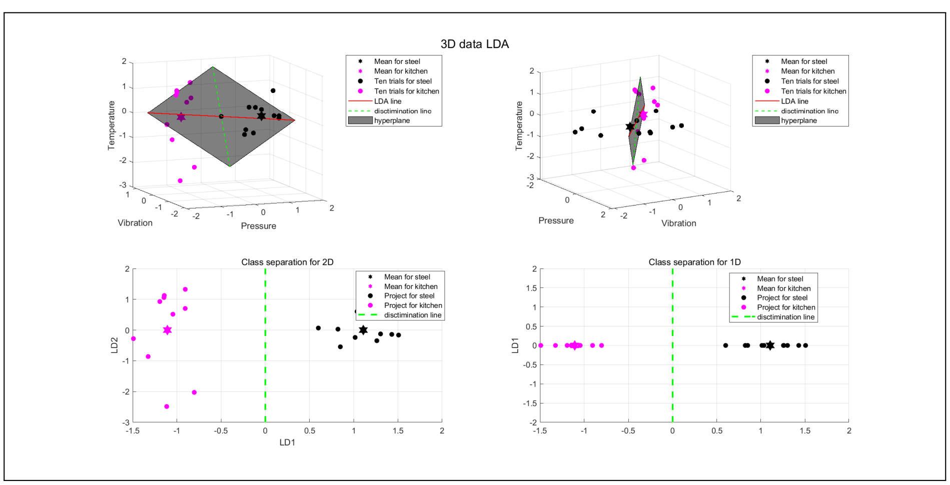

Figure 13. T3.4 Apply LDA to the three-dimensional PVT data for the steel vase and kitchen sponge. The two pictures above are hyperplane from different perspectives, The two figures below are LDA two-dimensional projection and one-dimensional projection

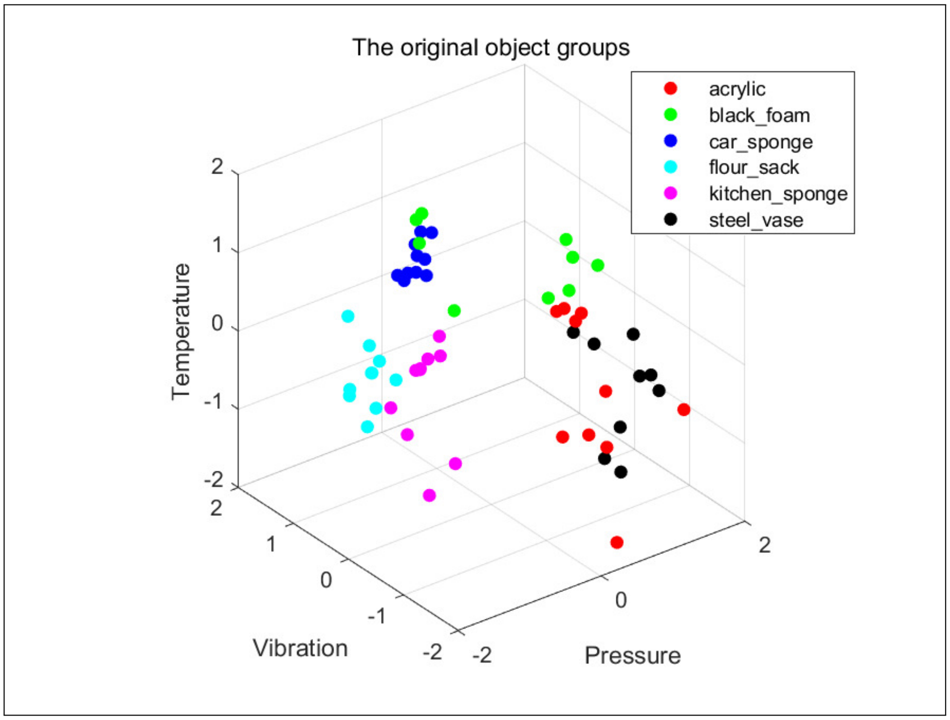

Figure 14. T4 Original data before clustering

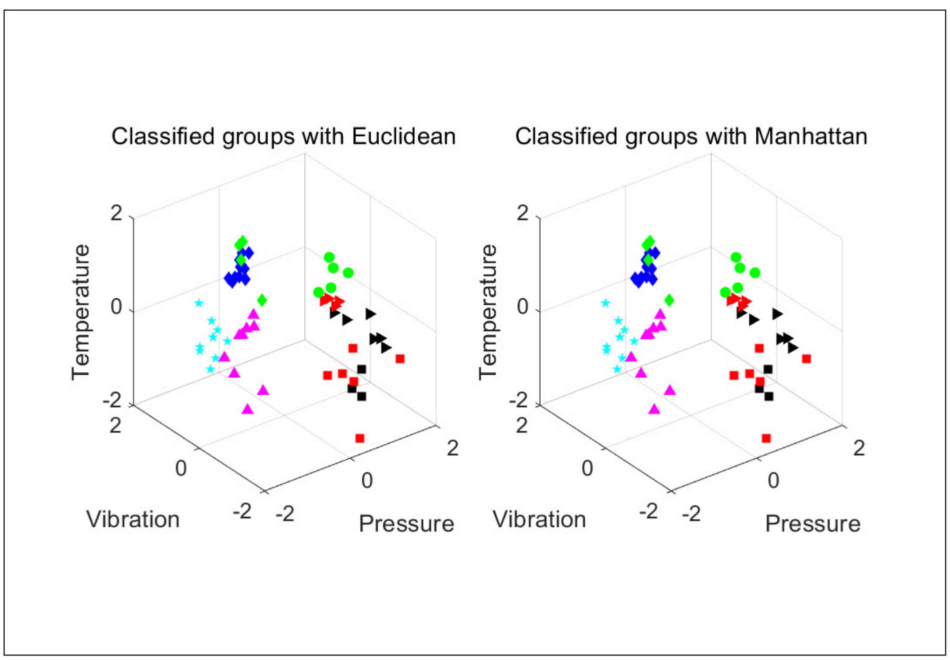

Figure 15. T4.1 Clustering results with Euclidean and Manhattan

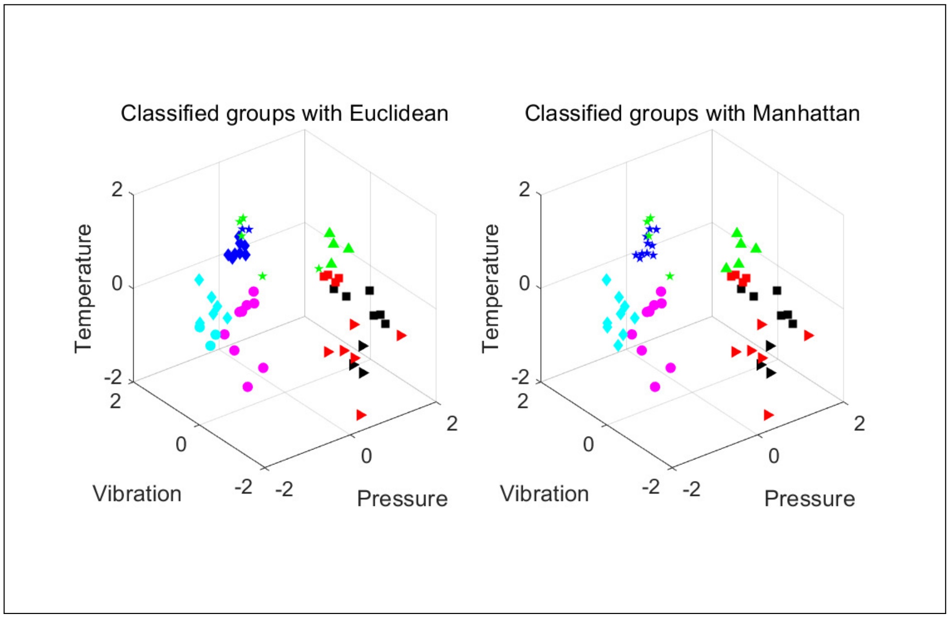

Figure 16. T4.2 K-Means clustering result with Euclidean and Manhattan distance of different trials

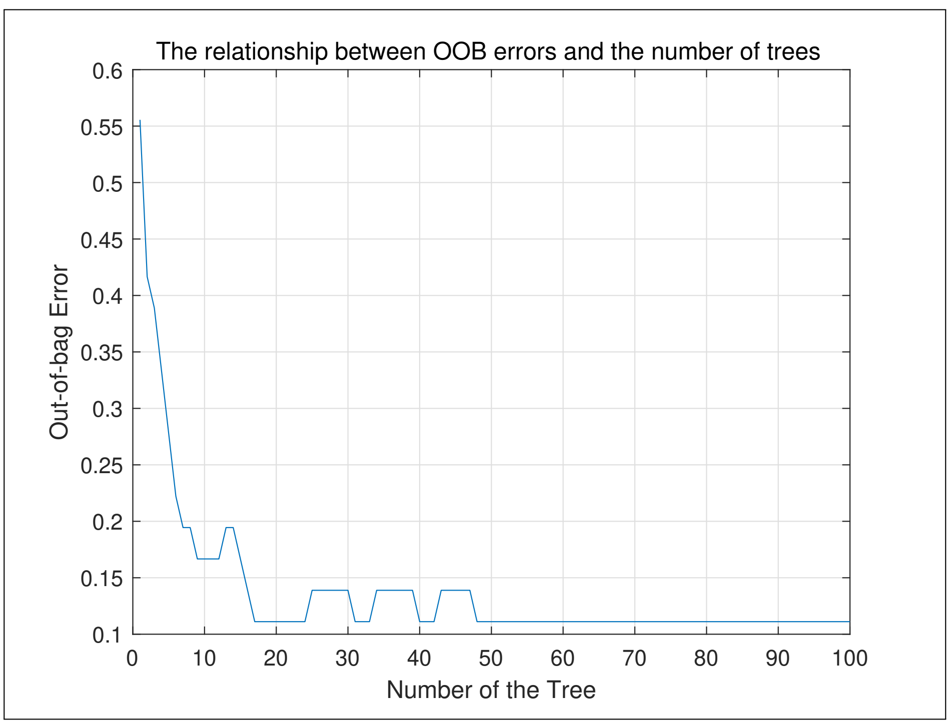

Figure 17. T4.3 Relationship between Out-of-Bag Error and the number of trees

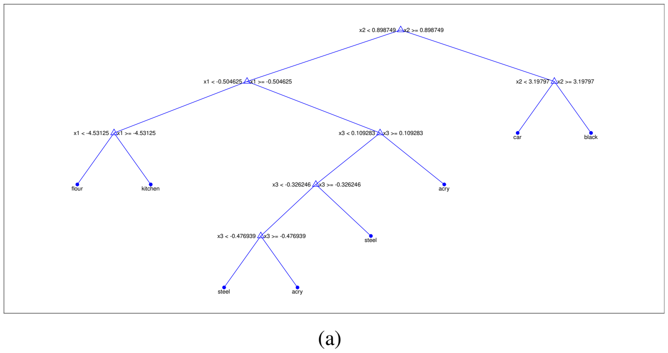

Figure 18. Two of the decision trees

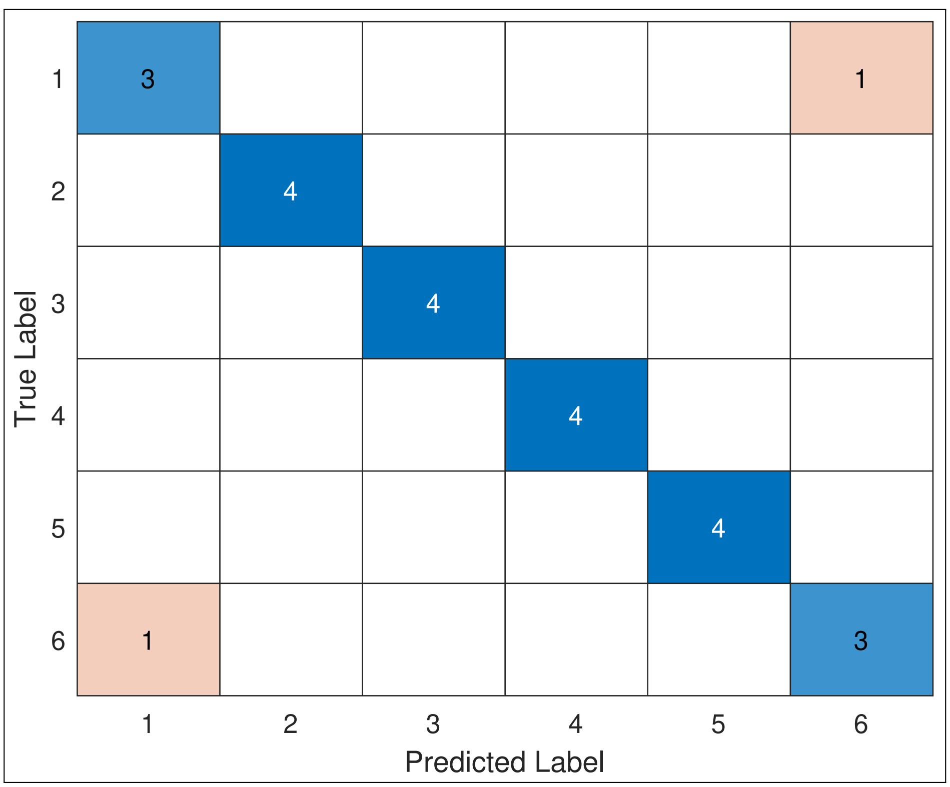

Figure 19. T4.3 Confusion Matrix

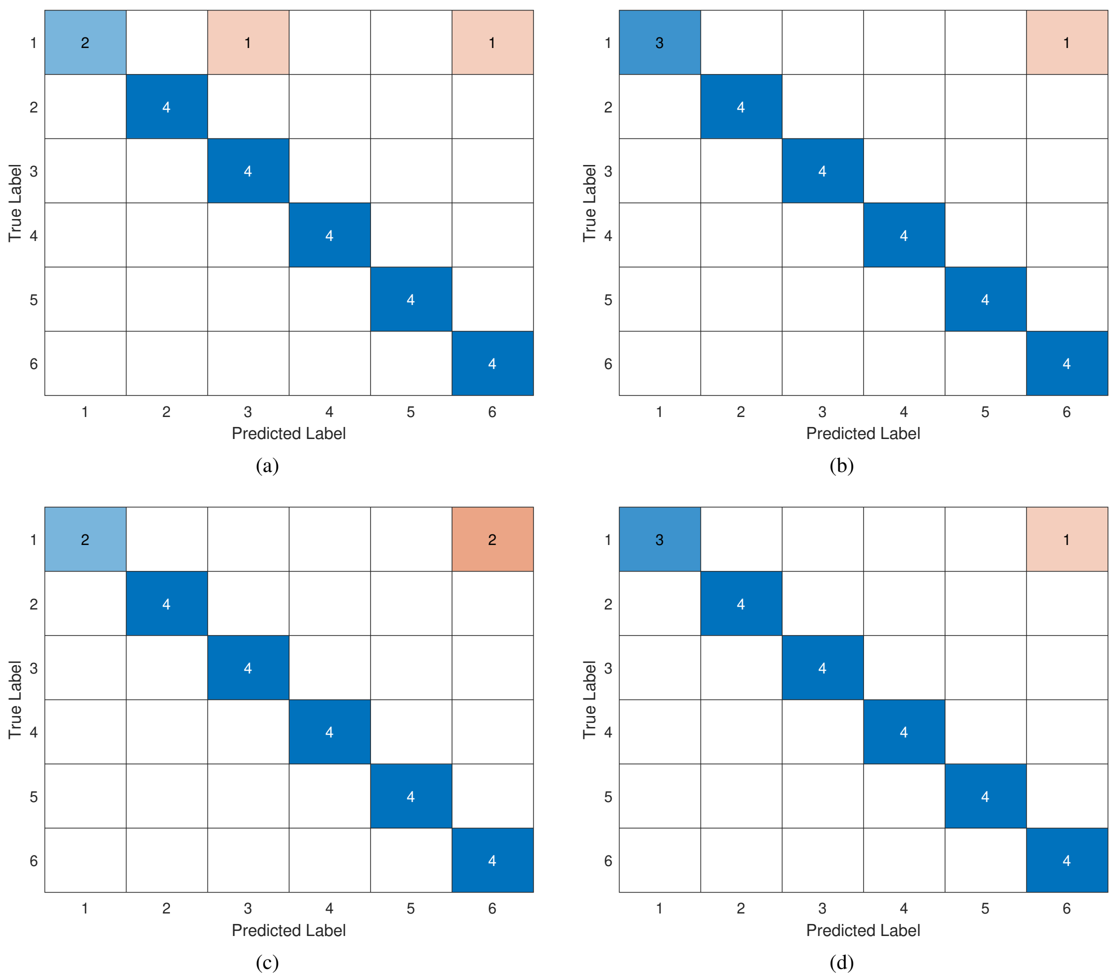

Figure 20. Some repeated experiments for Confusion matrix