Xuanyu Lai, 02448599 , Shutuo Guo, 01857692

1. Section A: Data Preparation

Task 1 Initially, all files are processed to extract and save data. Direct plotting, however, would result in 60 lines, complicating the visualization process. To streamline the analysis, the maximum and minimum values from these samples are computed, as demonstrated with F1 data in Figure 1. This method simplifies the distinction between variables.

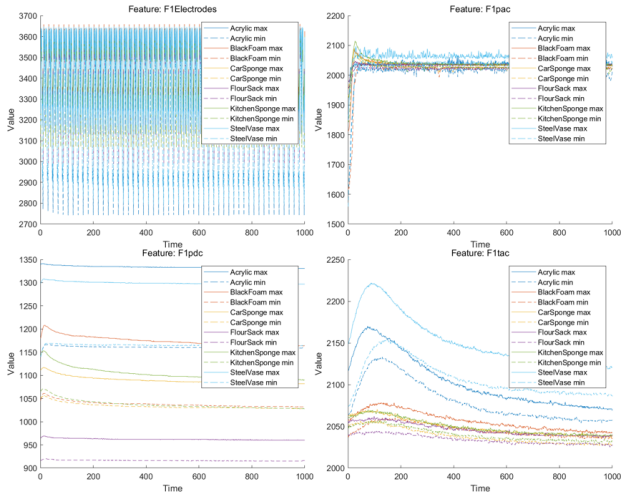

Figure 1. F1 Max-Min line for all objects

In Figure 2, **Black Foam** and **Car Sponge** are selected to illustrate their temperature variables, while **Kitchen Sponge** and **Car Sponge** are chosen to display their pressure variables.

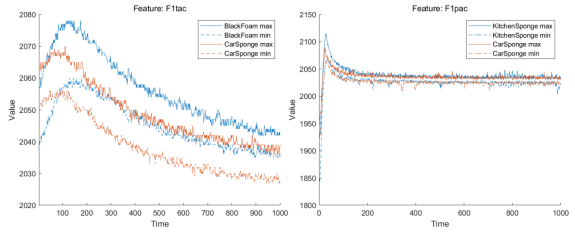

Figure 2. Black Foam, Car Sponge temperature Max-Min line & Car Sponge, Kitchen Sponge vibration Max-Min line

In the study, tactile data initially indicated an increasing overlap between **Black Foam** and **Car Sponge** from 0 to 100, with minimal overlap observed near 200. However, an exclusive focus on these materials proved insufficient. A notable overlap in data between **Kitchen Sponge** and **Car Sponge** was observed, necessitating a compromise on the accuracy of certain objects to enhance differentiation between these two sponges. Analysis of pressure data revealed distinct trends in response, leading to the selection of 30 as the optimal time instance after thorough examination.

Task 2 In this experiment, Finger F1 data was chose to analysis. After this, extraction and storage functions were developed to retrieve the corresponding time series data for storage in three files, notably **F1_PVT.mat**,**F1_PVT.mat**, **F1_another.mat**.

Task 3 Figure 3 shows a 3D scatter for **F1_PVT.mat** file, with different colours used for different objects.

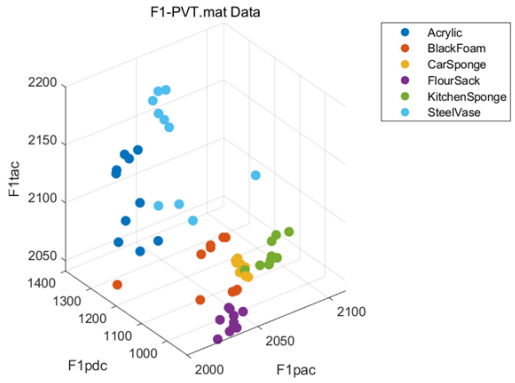

Figure 3. 3D scatter for F1 PVT

2. Section B: Principal Component Analysis

Task 1.a As matrix show below are covariance matrix, eigenvalues, and eigenvectors for the data.

Covariance matrix:

$$\begin{bmatrix} 1.0000 & -0.1287 & -0.2819 \\ -0.1287 & 1.0000 & 0.8239 \\ -0.2819 & 0.8239 & 1.0000 \end{bmatrix}$$

Eigenvalues:

$$\begin{bmatrix} 1.9165 & 0.9222 & 0.1612 \end{bmatrix}$$

Eigenvectors:

$$\begin{bmatrix} -0.3040 & 0.9429 & -0.1362 \\ 0.6601 & 0.3115 & 0.6835 \\ 0.6869 & 0.1179 & -0.7171 \end{bmatrix}$$

Task 1.b&c Figure 4 illustrates the visualization of standardized F1 PVT data with principal components and reduction to two dimensions.

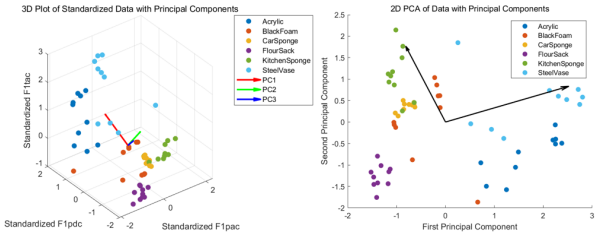

Figure 4. 3D scatter and dimension reduction for F1 PVT

Task 1.d Figure 5 illustrates the distribution of F1 PVT data across all principal components.

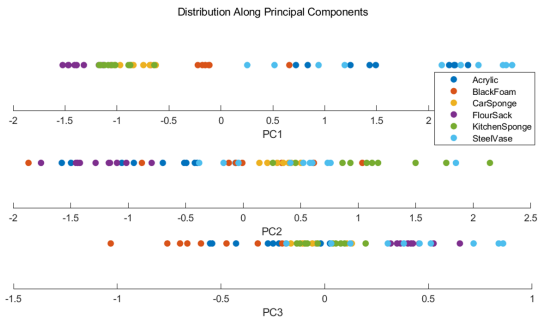

Figure 5. PVT distributions along all PC directions

Task 1.e Some categories such as Black Foam, Car Sponge, Flour Sack, and Kitchen Sponge appear to be quite tightly grouped across all principal components. However, categories like Acrylic and Steel Vase show a wider spread, especially along PC1. While objects like Flour Sack can be distinguished easily, there is a huge overlap for some other objects like Black Foam and Car Sponge. This would suggest the classification will not be straightforward and may need help with some sophisticated classification algorithm.

Task 2.a&b Figure 6 shows the Scree plot of F1 electrode data and 3D scatter of three largest principal components.

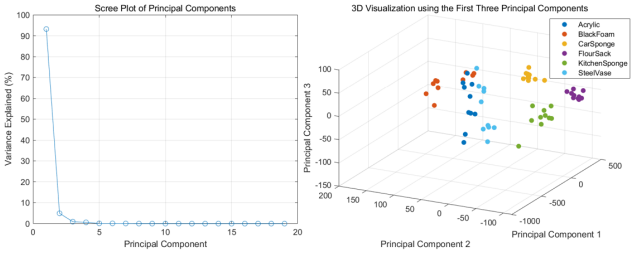

Figure 6. Scree Plot and 3D Scatter

Task 2.c The Scree Plot from PCA on nineteen similar electrodes demonstrates a steep decline, indicating that the initial principal components significantly capture the dataset's main differences. This suggests they are effective in simplifying the data through dimension reduction. The trend highlights that a few components contain most of information. Furthermore, 3D scatter plot in figure 6 reveals that three principal components can represent the data from nineteen electrodes effectively.

3. Section C: Linear Discriminant Analysis

Task 1.a Figure 7 compares the Pressure,Vibration and Temperature by using LDA. In the figure, the black dotted line indicates the decision boundary, green line represent the best separation of the different classes.

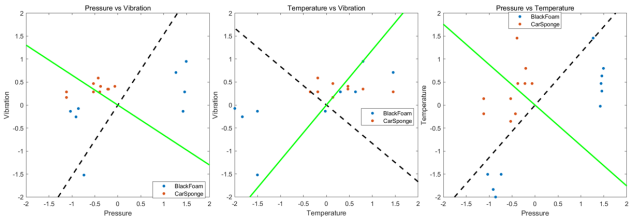

Figure 7. CarSponge and BlackFoam LDA 2D Scatter

Task 1.b Figure 8 shows the 3D scatter plot by applying LDA to the three-dimensional PVT data.

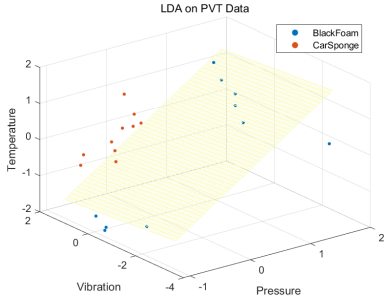

Figure 8. CarSponge and BlackFoam LDA 3D Scatter

Task 1.c Graph analysis reveals that LDA effectively distinguishes between materials using pressure and temperature, but struggles with vibration data, leading to errors. Vibration significantly impacts experimental outcomes. Tensile modulus, indicating material elasticity, shows that Black Foam and Car Sponge, being low-density and lightweight, likely have similar physical property, causing data overlap.

Regarding temperature , Car Sponge and Black Foam have porous structures but differ in insulation. Foam's closed cells hinder air flow, reducing heat transfer. In contrast, sponge's open cells, relying on air for insulation, are more prone to convective heat loss, as suggested by their higher heat exchange coefficient shown in figure 7.

For compression pressure, foam's air bubbles offer more push back, whereas sponge's interconnected pores allow easier air movement under pressure, leading to more compression and distinct pressure responses, also illustrated in figure 7.

Task 1.d **Acrylic** and **Steel Vase** were selected for Linear Discriminant Analysis (LDA) due to their distinct physical properties compared to other materials. Both acrylic and steel possess smooth, hard surfaces, offering unique feedback when subjected to external forces. This is in stark contrast to soft, porous materials such as **black foam** or **sponge**, which tend to absorb pressure and vibration.

The analysis also aimed to explore their thermal properties, with an expectation that steel vases would exhibit efficient heat transfer characteristics owing to their high thermal conductivity. In contrast, acrylic was anticipated to demonstrate thermal insulating properties, attributed to its low thermal conductivity. This examination is intended to facilitate a comprehensive understanding of the physical properties of these materials.

Figure 9 hows the 3D scatter plot result by applying LDA on **Acrylic** and **Steel Vase** PVT data

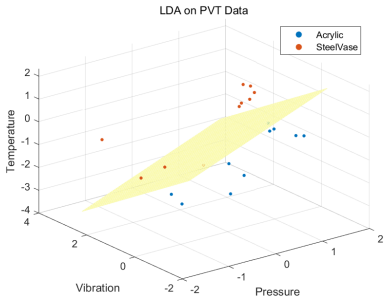

Figure 9. Acrylic and Steel Vase LDA 3D Scatter

4. Section D: Clustering & Classification

Task 1.a Figure 10 shows the result of clustering algorithm to the PVT data by using euclidean distance metric.

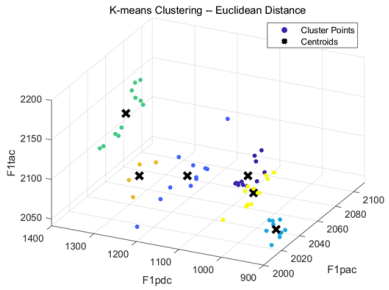

Figure 10. Euclidean Distance 3D scatter

Task 1.b Figure 10 shows the clusters formed by K-means clustering with Euclidean distance. This method has successfully partitioned the data into distinct clusters, each represented by a distinct colour, and the black 'X' marks represent the centroids of these clusters. The method has successfully clustered the Black Foam and Flour Sack. However, for the rest of the points, the method fails to classify them. This could be due to the problem that K-means assumes that clusters are spherical and of similar size, which is not the case for real-life data.

Task 1.c The distance metric is switched into Manhattan distance as shown by the second graph of figure 11. The clusters shown here have a different composition than those created with Euclidean distance due to the nature of the Manhattan metric, which can result in clusters that are more rectangular in shape as opposed to the spherical shape with Euclidean clusters.

Centroids are placed on a better spot that further clusters Acrylic and Steel Vase. However, since Car Sponge and Kitchen Sponge are overlaping so much, this method still fails to identify them and creates an extra centroid.

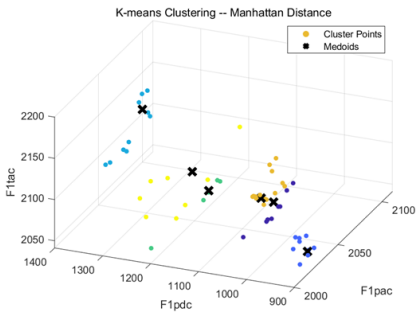

Figure 11. Manhattan Distance 3D Scatter

Task 2.a According to Figure 14, as model complexity incrementally rises, there is an initial increase in accuracy, followed by fluctuations. Yet, upon surpassing the threshold of 25 trees, the model demonstrates a trend towards enhanced stability, maintaining accuracy rates consistently

above 90% across various scenarios. Taking into account computational efficiency alongside the derived benefits, it is concluded that selecting **26** trees strikes an optimal balance between high accuracy and robust generalization performance.

Task 2.b Figure 12 shows the result of decision tree.

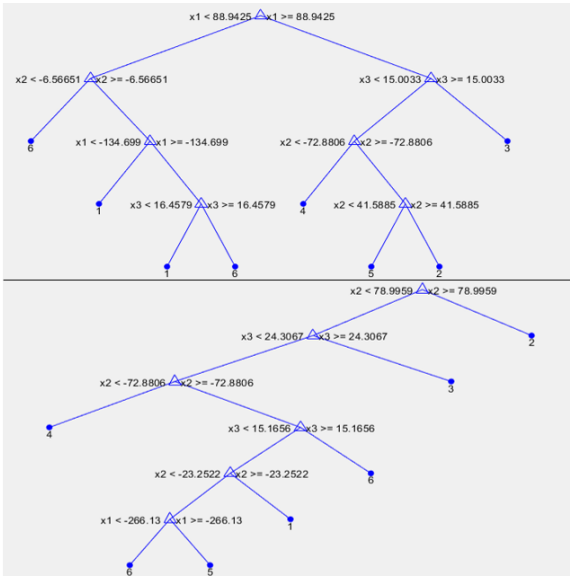

Figure 12. Decision Tree. Different objects are represented as numbers in order. 1. Acrylic, 2. Black Foam, 3. Car Sponge, 4. Flour Sack, 5. Kitchen Sponge, 6. Steel Vase

Task 2.c Figure 13 shows the confusion matrix Accuracy of the bagged ensemble on the test set: 95.83%. This is a significantly high accuracy based on the limited amount of data and the 16.67% accuracy if you choose one category randomly.

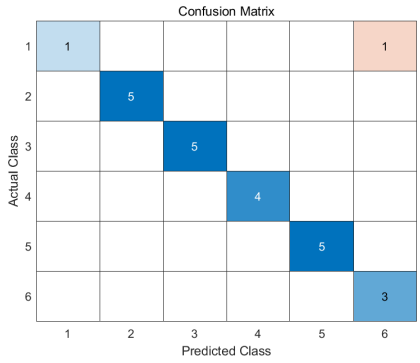

Figure 13. Confusion matrix of classification. Different objects are represented as numbers in order. 1. Acrylic, 2. Black Foam, 3. Car Sponge, 4. Flour Sack, 5. Kitchen Sponge, 6. Steel Vase

Task 2.d Misclassification in the task should represent a similarity between materials. However, the only misclassification happens between Acrylic and Steel Vase. One possible reason could be the noise from the electrode when capturing data, as, for example, Car Sponge and Kitchen Sponge should have a more similar characteristic and should have contributed to more misclassification. The PCA step helps this task in terms of dimensionality reduction. It helps to eliminate noise and less informative features; therefore improving model performance by focusing on the most relevant information.

5. Section E: Conclusion

Task 1.a Pattern recognition techniques play a crucial role in analyzing complex datasets. When the dataset is highly dimensional and abundant, dimensionality reduction techniques (like PCA, LDA) and clustering techniques (like K-means) aid the analysis by transforming raw data into meaningful insights, facilitating decision-making, and therefore finding underlying characteristics.

Task 1.b Yes, BioTac tactile sensors can tell the difference between objects by using touch only. In LDA part, from figure 8 and figure 9 can find that objects can be classified correctly. In bagging part, accuracy from the experiment has been a profound proof that the objects are distinguishable. The only error with the Acrylic and Steel Vase could be improved by training the model with more data or a possibly more complex pattern recognition method.

Task 1.c In the analysis, the first three Principal Components, illustrated in Figure 4, predominantly align with F1tac, indicating temperature fluctuations. Thus, the temperature sensor shows the best potential when constrained to a single sensor choice. However, as Figure 6's scree plot shows, a minimal number of electrodes is requisite to achieve a satisfactory level of precision in object detection. This could flavor the case when more electrodes are allowed.

Task 1.d Time series analysis can be used instead. This method considers the sequence of data points ordered in time. This method analyses the changes in sensor readings over time to recognize patterns.

Pros: This approach can capture temporal dependencies and patterns that unfold over time. Also, by leveraging historical data points, time series analysis can improve prediction accuracy for future states.

Cons: However, analysing time series data is generally more complex than static data analysis. More computational resource is needed and large amounts of data to capture the full range of temporal patterns.

6. Appendix

Appx.1 Figure 14 compares **Acrylic** and **Steel Vase** Pressure,Vibration and Temperature data by using LDA. In the figure, the black dotted line indicates the decision boundary, green line represent the best separation of the different classes.

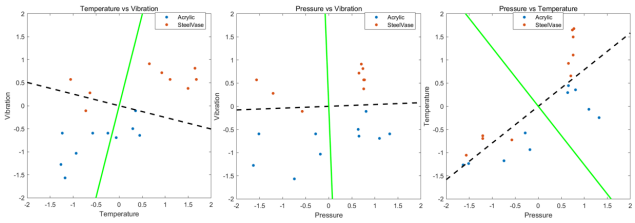

Figure 14. Acrylic and Steel Vase LDA 2D Scatter

Appx.2 The figure 15 depicts the ensemble learning model's accuracy improving and then stable as the number of trees increases, with a notable stabilization and high accuracy achieved beyond 25 trees.

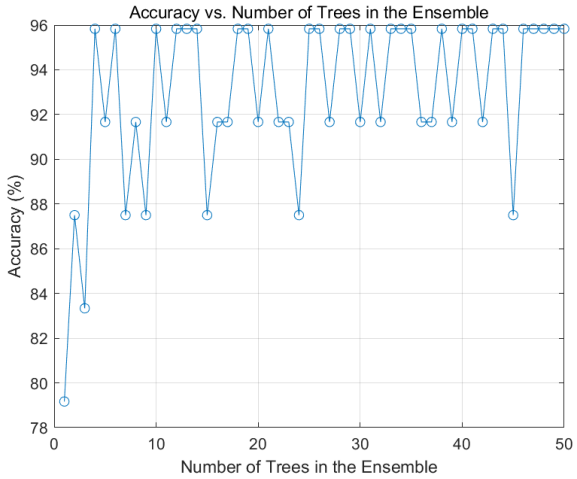

Figure 15. Accuracy/complexity trade-off of random forest