

109 學年度大學部專題競賽



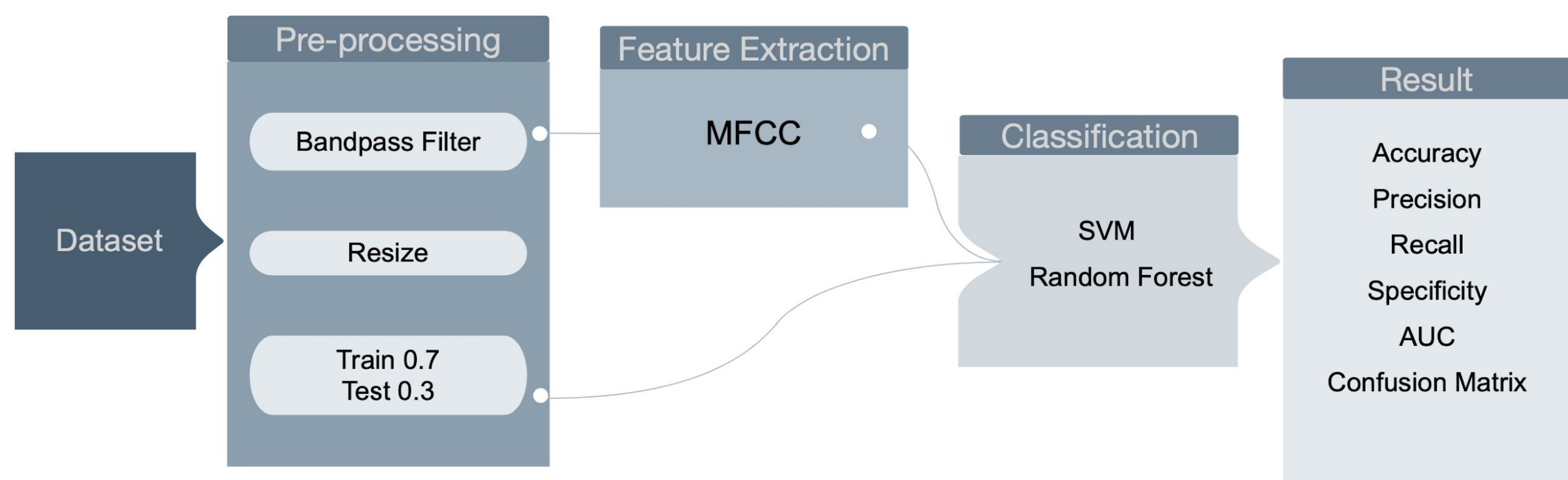
國立清華大學資訊工程學系

Department of Computer Science, National Tsing Hua University

Respiratory Sound Analysis of Chronic Obstructive Pulmonary Disease

慢性呼吸道疾病音訊分析

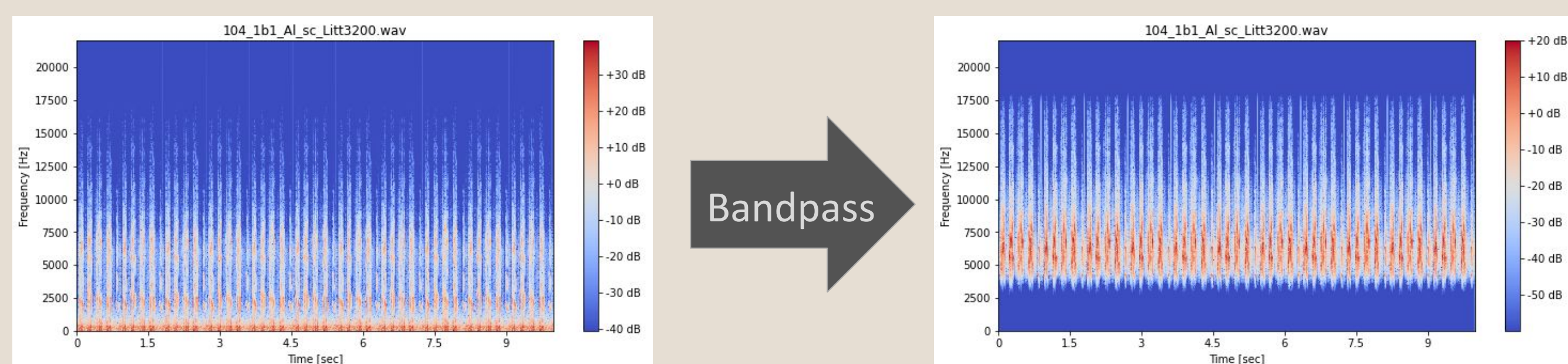
組員：張子宜、王士亞、李卓庭



此次專題目的即提出利用機器學習配合適當的資料預處理與特徵提取，辨認分析此呼吸音訊，以判斷患者是否罹患COPD，進而減少醫生聽取每筆音檔所花費的時間，提高診斷效率。

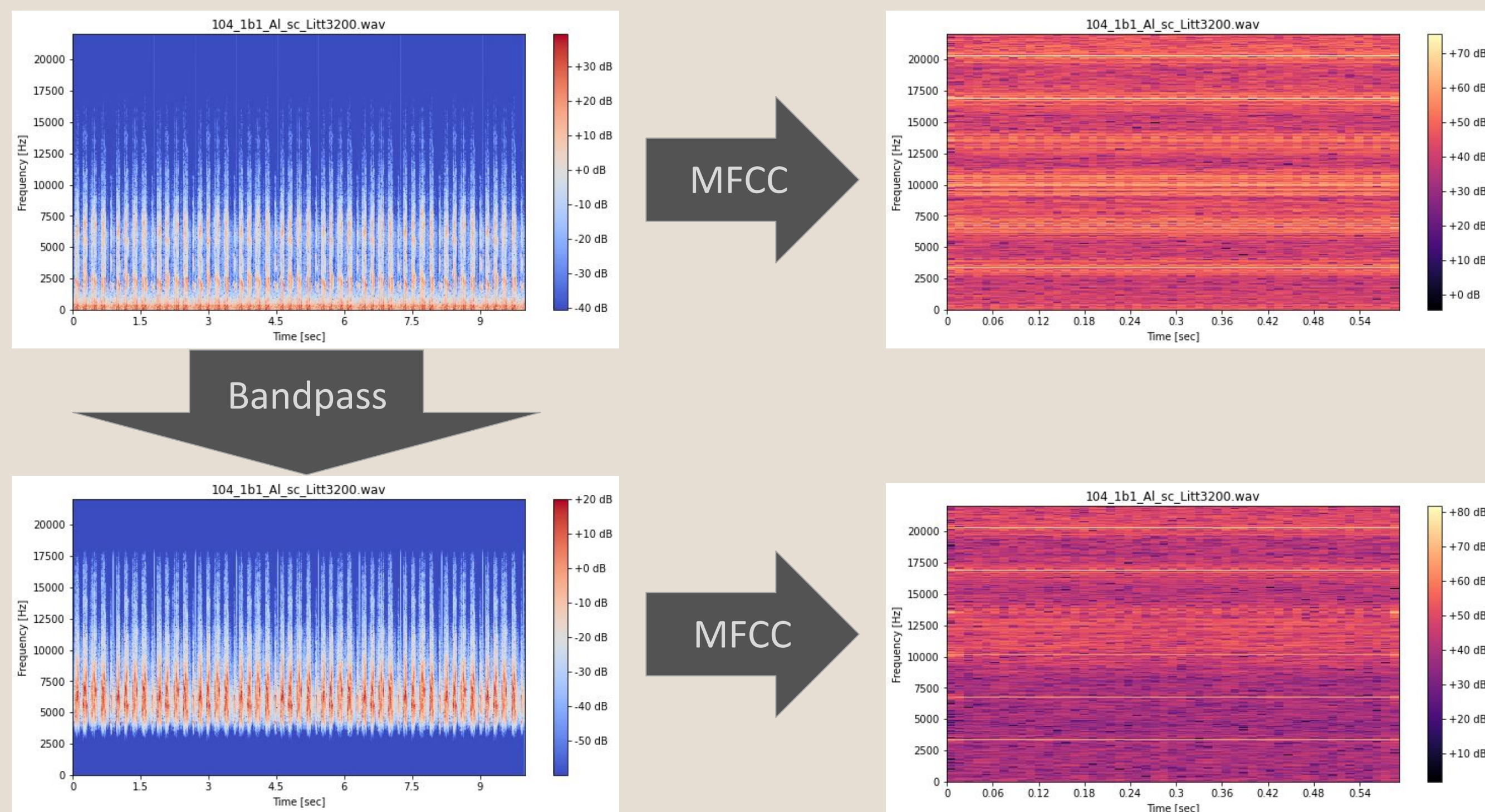
數據預處理 (Pre-processing)

我們截取每筆錄音檔的前20秒，使用帶通濾波器 (Bandpass Filter) 去除錄音檔的雜訊後，使用濾波器後的頻譜圖明顯去除了大量不必要的雜訊。



提取特徵 (Feature Extraction)

為了強化錄音檔中爆裂音與喘鳴聲，我們使用計算梅爾倒頻譜係數 (MFCC) 來提取特徵。



Confusion Matrix

真實情況與模型判斷下有四種可能：

- 一、情況為真且模型判斷為真 (True Positive)
- 二、情況為真且模型判斷為假 (False Negative)
- 三、情況為假且模型判斷為真 (False Positive)
- 四、情況為假且模型判斷為假 (True Negative)

四種可能性構成Confusion Matrix (如右圖)

從Confusion Matrix即可明瞭的看出分類器的成效。

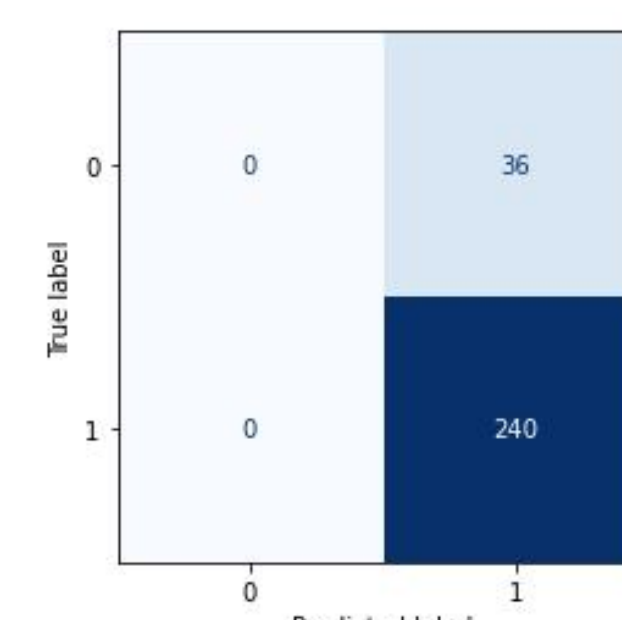
| | | Predicted class | |
|--------------|---|----------------------|----------------------|
| | | N | P |
| Actual Class | N | True Negatives (TN) | False Positives (FP) |
| | P | False Negatives (FN) | True Positives (TP) |

結果 (Result)

與將原始音訊放入分類器相比，透過MFCC提取特徵處理之後，的確有顯著進步，衡量指標都明顯提高；而原始數據，經過帶通濾波器去除雜訊後，再計算MFCC提取特徵音訊，其效能是所有實驗中最高的。

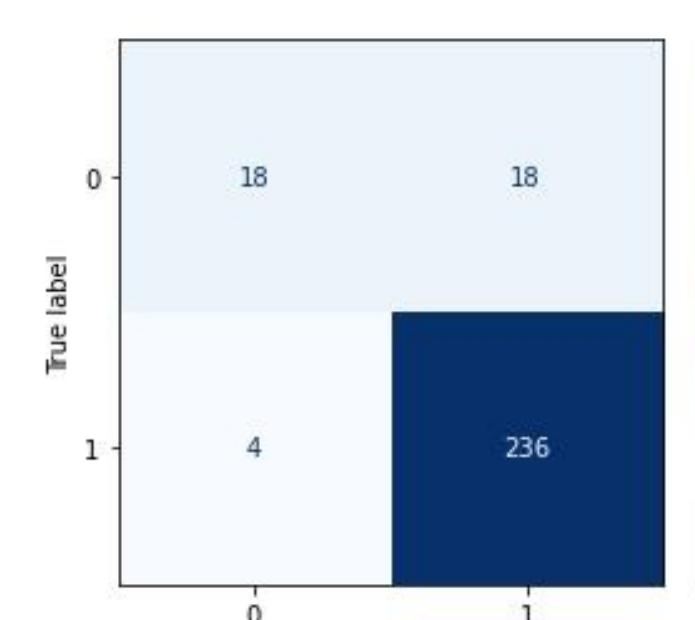
不做預處理

For Support Vector Machine:
Accuracy: 0.870
Precision: 0.870
Recall: 1.000
Sensitivity: 1.000
Specificity: 0.000
AUC: 0.500



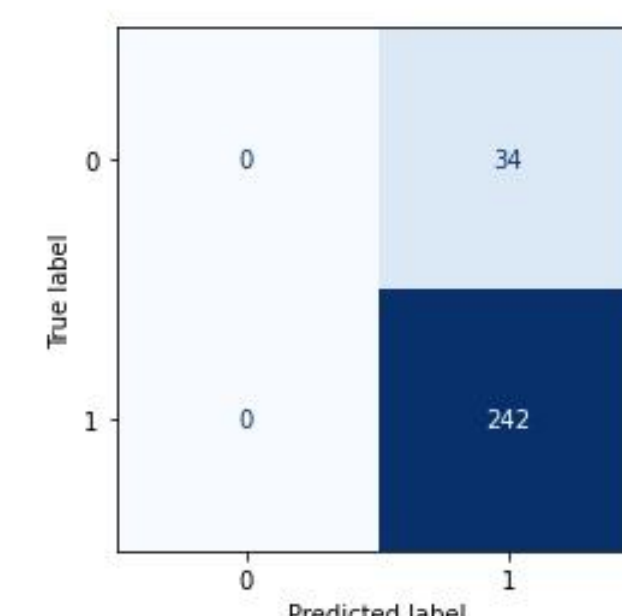
Random Forest

For Random Forest Classifier:
Accuracy: 0.920
Precision: 0.929
Recall: 0.983
Sensitivity: 0.983
Specificity: 0.500
AUC: 0.742

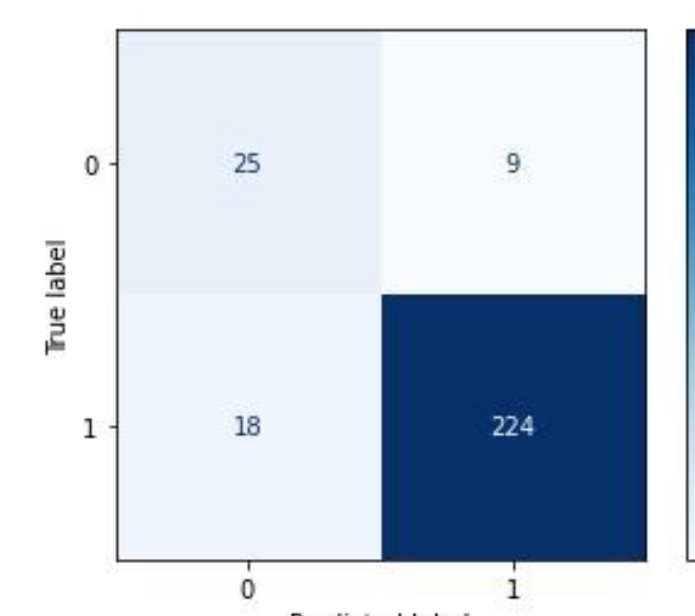


Bandpass濾波

For Support Vector Machine:
Accuracy: 0.877
Precision: 0.877
Recall: 1.000
Sensitivity: 1.000
Specificity: 0.000
AUC: 0.500

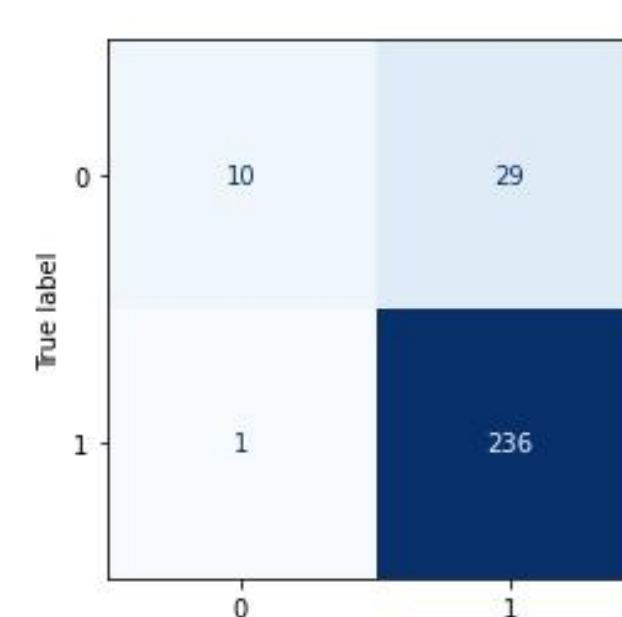


For Random Forest Classifier:
Accuracy: 0.902
Precision: 0.961
Recall: 0.926
Sensitivity: 0.926
Specificity: 0.735
AUC: 0.830

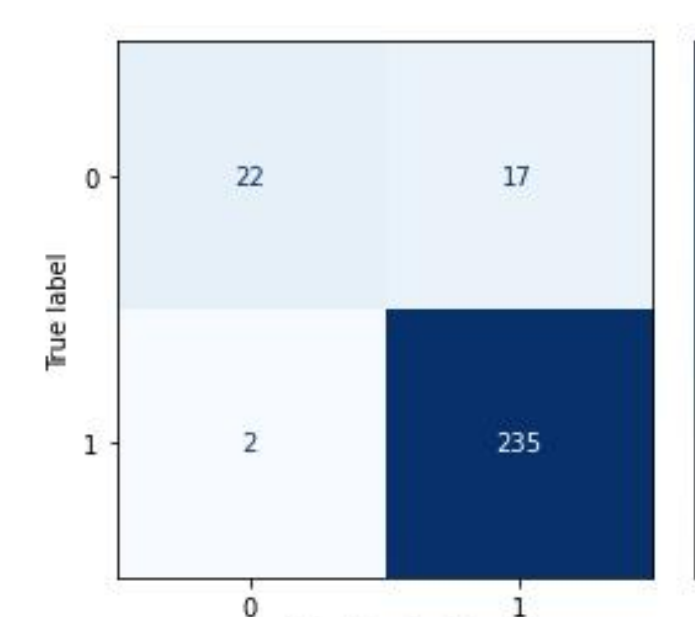


MFCC處理

For Support Vector Machine:
Accuracy: 0.891
Precision: 0.891
Recall: 0.996
Sensitivity: 0.996
Specificity: 0.256
AUC: 0.626

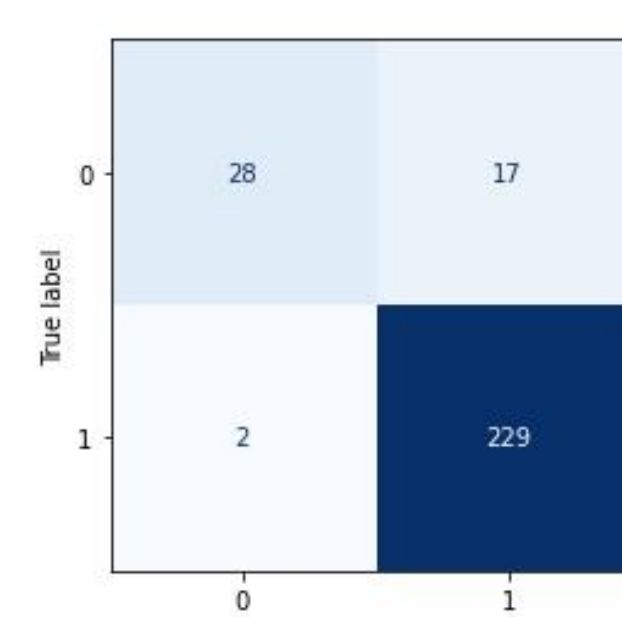


For Random Forest Classifier:
Accuracy: 0.931
Precision: 0.933
Recall: 0.992
Sensitivity: 0.992
Specificity: 0.564
AUC: 0.778

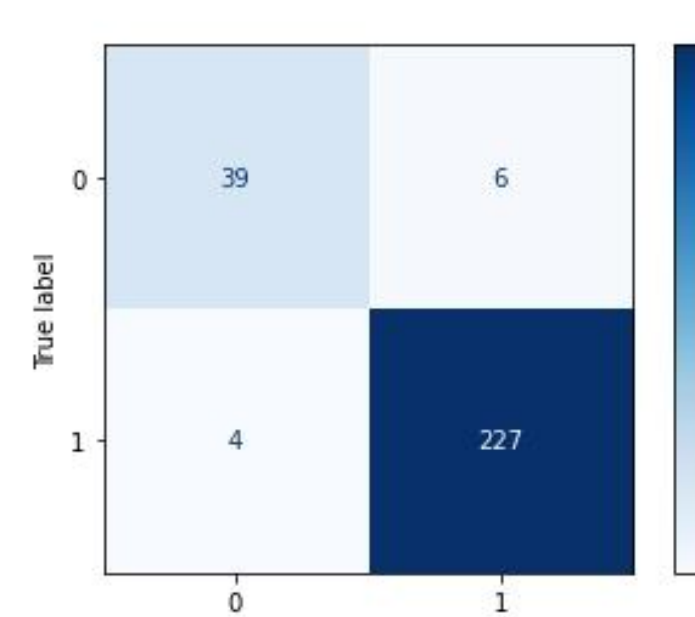


Bandpass濾波
+ MFCC處理

For Support Vector Machine:
Accuracy: 0.931
Precision: 0.931
Recall: 0.991
Sensitivity: 0.991
Specificity: 0.622
AUC: 0.807



For Random Forest Classifier:
Accuracy: 0.964
Precision: 0.974
Recall: 0.983
Sensitivity: 0.983
Specificity: 0.867
AUC: 0.925



從我們的實驗結果來看，在罹患COPD的患者中，模型可以將百分之95以上的患者篩檢出來，而過程中所花費的時間，相較醫生親自聽取每個病患的呼吸音再進行判斷要快得多。若機器學習應用在醫療界上，幫助醫生診斷病患，相信也能減輕負擔，提升醫療效率。