



# 研究感冒症狀對於聲音特徵值之影響用於發展感冒症狀辨別模型

Researching the effect of cold symptoms on the sound feature point use to develop cold symptoms to the discrimination models



指導老師：林瑞豐、林真如 教授 學生：簡郁昕、徐子婷

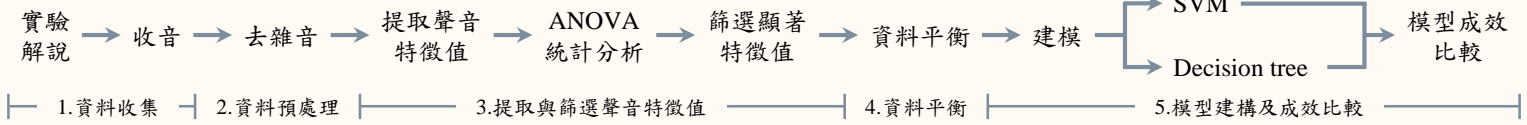
## 摘要

以現今的科技與研究，目前並未發現可以快速辨別或篩選患者具有何種感冒症狀的工具。有些患者在就診主述時，未能確切的說明自身的感冒症狀。因此如果有能辨別感冒症狀的模型作為醫師的輔助工具，將可以更好地了解患者自身的狀況。

為了驗證感冒症狀對於聲音特徵值是有影響的，我們利用單因子變異數分析和迴歸分析驗證感冒症狀確實會影響聲音特徵值，並了解兩者之間的相關性。

本研究為了發展能辨別感冒症狀的模型，藉由分析感冒症狀對於聲音特徵值的影響，進而篩選重要特徵值以**建構分類模型**，**預測單一感冒症狀**。

## 研究方法



- 於醫院、診所、網路上收錄音檔
- 本研究使用之資料共549筆
- 利用Audacity去除音檔雜音
- 使用ANOVA分析篩選顯著特徵值
- 利用openSMILE提取聲音特徵值
- 事後分析觀察特徵值與感冒症狀的相關性
- 以資料平衡解決資料不均問題
- 使用SVM、Decision tree建模
- 建模結果比較

## 研究結果

代號	編號	單一症狀	筆數
A	1	無症狀	357
B	2	流鼻水	43
C	3	喉嚨痛	36
D	4	有痰	31
E	5	鼻塞	28
F	6	頭痛	27
G	7	咳嗽	27
Total			549

**使用數據說明**

- 收案時間：2020/10/01~2020/12/25、2022/04/01~2022/11/15。
- 資料數據：由左方原資料筆數表格可知，無症狀的資料筆數為357筆，相對於其他單一感冒症狀的筆數而言多出了11倍左右的資料量。

共計 549筆

### 顯著特徵值表格

d	去個人差異化 (d)										ANOVA 以健康為準	ANOVA 以健康為準	
	Max Value	Min Value	Range	Max Position	Min Position	Amean	Lavgq1	Lavgq2	Lavgq3	Standard Deviation			Skewness
RMS													
MFCC 1													
MFCC 2													
MFCC 3													
MFCC 4													
MFCC 5													
MFCC 6													
MFCC 7													
MFCC 8													
MFCC 9													
MFCC 10													
MFCC 11													
MFCC 12													
ZCR													
voiceProb													
F0													
共61個													
共35個													

Legend: p-value ≤ 0.1 (green), p-value ≤ 0.05 (yellow), p-value ≤ 0.01 (orange), p-value ≤ 0.001 (red)

### 資料平衡 -> Undersampling

單一症狀	原資料筆數	只降無症狀	測試集 (70%)	訓練集 (30%)
無症狀	357	43	30	13
流鼻水	43	43	30	13
喉嚨痛	36	36	25	11
有痰	31	31	22	9
鼻塞	28	28	20	8
頭痛	27	27	19	8
咳嗽	27	27	19	8
Total	549	235	165	71

**Undersampling:**

由左方資料平衡表格可知，為避免使用原資料筆數建模時，因無症狀資料較多，而使其他六種單一感冒症狀皆被誤判為無症狀，又或者大多數無症狀皆被預測正確，使得模型準確率提高，讓人誤以為模型成效好。

因此透過Undersampling方法減少無症狀筆數至43筆，並平均抓取30%的單一症狀放入訓練集進行訓練。

## 研究目的

### 分析感冒症狀對聲音特徵值的影響

自變項：不同的感冒症狀(咳嗽、有痰、喉嚨痛等)  
 依變項：聲音特徵值(共384個)

### 發展能辨別感冒症狀之模型

自變項：顯著聲音特徵值  
 依變項：不同的感冒症狀(咳嗽、有痰、喉嚨痛等)

### 五次訓練綜合結果

精確率	健康	預測						實際症狀	五次正確預測總數	五次訓練皆未誤判症狀
		流鼻水	喉嚨痛	有痰	鼻塞	頭痛	咳嗽			
0.21	無症狀	25	11	13	7	0	7	2	25	鼻塞
實際	流鼻水	14	17	8	8	6	4	17	2	咳嗽
	喉嚨痛	8	18	8	10	3	6	8		
	有痰	9	17	3	5	4	7	0		
	鼻塞	10	11	8	2	7	5	2		
	頭痛	8	7	5	2	5	9	4		
	咳嗽	12	7	6	5	2	6	2		
	0.18	無症狀	24	11	7	7	9	3		
實際	流鼻水	17	15	10	8	10	4	15	2	咳嗽
	喉嚨痛	10	10	9	3	6	10	9		
	有痰	8	11	7	4	5	6	4		
	鼻塞	13	10	6	3	4	9	0		
	頭痛	9	11	1	2	8	6	3		
	咳嗽	8	9	6	6	5	4	2		
	0.15	無症狀	34	1	3	4	5	0		
實際	流鼻水	22	1	1	10	3	0	1	1	頭痛
	喉嚨痛	28	2	0	6	7	0	28	0	頭痛
	有痰	26	1	0	6	2	0	10	6	頭痛
	鼻塞	23	0	3	0	5	0	14	5	流鼻水、有痰、頭痛
	頭痛	19	0	0	2	4	0	15	0	流鼻水、喉嚨痛
	咳嗽	18	1	1	4	7	0	9	9	頭痛
	0.18	無症狀	42	8	0	1	11	1	2	42
實際	流鼻水	27	6	5	5	15	5	2	6	0
	喉嚨痛	30	4	1	4	8	3	5	1	
	有痰	10	9	3	0	12	4	7	0	
	鼻塞	19	8	4	3	5	5	1	5	
	頭痛	12	2	2	6	8	5	5	5	
	咳嗽	15	3	1	7	9	2	6	6	

由上方「決策樹、支援向量機」之建模結果表格可知：

- 決策樹在d+d'有顯著特徵之資料下的「五次模型訓練結果」，無症狀皆不會被誤判為鼻塞、有痰皆不會被誤判為咳嗽。
- 支援向量機在d+d'有顯著特徵之資料下的「五次模型訓練結果」，頭痛不會被無症狀、流鼻水、喉嚨痛、有痰、咳嗽誤判；鼻塞不會被誤判為流鼻水、有痰、頭痛；頭痛不會被誤判為流鼻水、喉嚨痛。

## 結論與未來研究方向

- 在「d+d'」有顯著特徵之資料下的模型訓練結果，可以發現五次訓練結果之平均精確率在決策樹時最佳。而在「d'」有顯著特徵之資料下的模型訓練結果，決策樹及支援向量機的五次訓練結果之平均精確率相同。以目前結果而言，決策樹相對於支援向量機，有較佳的結果。
- 未來可透過本研究結果，發展能辨別感冒症狀的模型作為醫師的輔助工具，可以更好地了解患者自身的狀況。未來亦可結合智能機器人，打造智慧醫療環境。