Heart Failure Prediction

《第八組》

資工三 4110056002 高韵晴

資管二 4111029032 蕭筠軒

指導教授:蔡孟勳教授



目錄



01 動機、目的、魚骨圖 05 結果與討論	01	動機、	目的,	魚骨圖	05	結果與討論
-----------------------	----	-----	-----	-----	----	-------

- 02 介紹資料集 06 結論
- 03 文獻探討 07 參考資料
- 04 研究方法

01

動機、目的、魚骨圖



動機



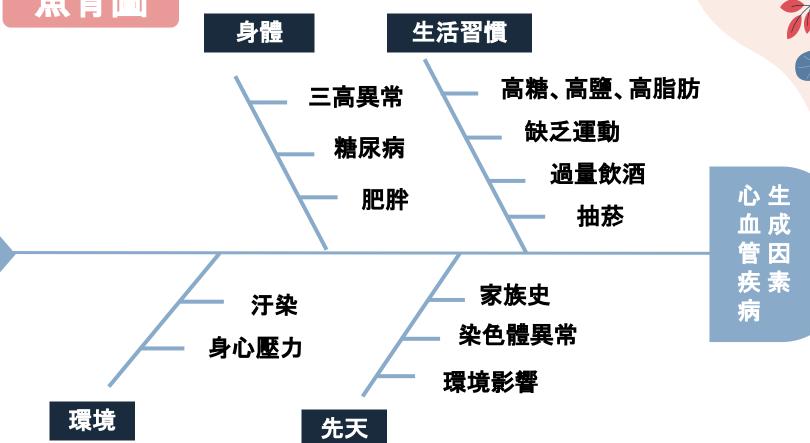
- 1、全球第一大死因, 佔所有死亡人數31%。
- 2、111年國人約2.3萬人死於心臟疾病,平均22分鐘1人。
- 3、誘發因子眾多,生活習慣的改變可對心臟血管健康產生重大影響。

目的



- 1、早期風險識別。
- 2、結合AI技術提升早期診斷的準確性。
- 3、提升大眾對心血管疾病意識,主動改善生活習慣。
- 4、了解發病機制與相關因子,促進科學研究進步。
- 5、醫療資源優化, 提高效率和質量。

魚骨圖



魚骨圖

生活習慣

治療

預 防 療 與 ──健康飲食──規律運動──戒菸戒酒

── 降膽固醇藥物 ─ 降血壓藥物 Ѣ尿病藥物 手術治療

血脂檢測 血壓檢測 — 血糖檢測 — 心電圖檢查 減輕壓力 - 心理輔導

積極生活

健康教育活動 社交媒體宣導 健康政策制定

監測健康數據

心理健康

宣導與教育

02

介紹資料集



介紹資料集 - SDGs

3.4

3.4 減少非傳染性疾病造成的死亡率 並促進心理健康

西元2030年前,透過預防治療,將非傳染性疾病的未成年死亡數減少三分之一,並促進心理健康。

3.d

3.d 完善全球健康風險預警系統

強化所有國家的早期預警、風險減少,以及國家與全球健康風險的管理能力,特別是開發中國家。



介紹資料集 - 資料內容



Heart Failure Prediction Datasets



918筆



12欄



https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction?rvi=1

介紹資料集 - 資料內容

欄位	說明	內容		
Age	年齢	數值型		
Sex	性別	M = Male F = Female		
ChestPainType	胸痛類型	TA = 典型心絞痛 ATA = 非典型心絞痛 NAP = 非心絞痛 ASY = 無症狀		
RestingBP	舒張血壓	mmHg		
Cholesterol	血清膽固醇	mg/dl		
FastingBS	空腹血糖	0 = 其他 1 = >120mg/dl		

欄位	說明	內容		
RestingECG	心電圖	Normal = 正常 ST = T段異常 LVH = 左心室肥大		
MaxHR	最大心率	每分幾次		
Exercise Angina	運動誘發心絞痛	N = 沒有 Y = 有		
Oldpeak	運動引起ST段抑制	運動時ST段低峰與休 息時之比值		
ST_SIpoe	運動高峰 ST段的斜率	UP = 向上傾斜 Flat = 平坦 Down = 向下傾斜		
Heart Disease	心血管疾病有無	0 = 沒有 1 = 有		

03

文獻探討



文獻探討

心血管疾病

常見原因:動脈粥狀硬化導致血管阻塞血流不順,引發各式病症

常見疾病: 冠心病、中風、心臟衰竭

診斷方式:

·抽血
·心臟超音波

· 運動心電圖 · 核醫心肌灌注掃描

·電腦斷層掃描

疾病治療:

·藥物治療:降三高藥、抗小板藥、血栓溶解劑

· 手術: 氣球擴張術、心臟支架、冠狀動脈繞道手術





文獻探討

心絞痛(Angina)

定義:動脈狹窄無法提供充分血量氧量,可能發生心肌梗塞,或造成胸骨下、心前區短暫陣發性疼痛,但心肌尚未有永久性損傷。

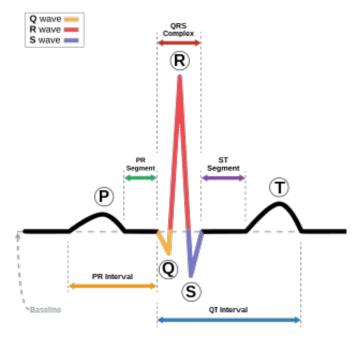
穩定型: 最常見, 運動引起的陣發性胸痛。

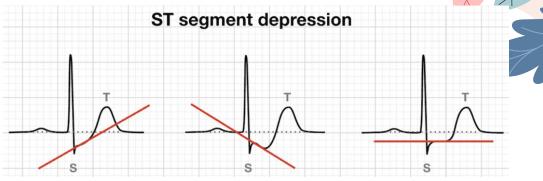
不穩定型:發作頻率、持續期間、疼痛強度隨時間增強,休息、睡眠、運動時都可能發生。

變異型:主要在休息。持續時間較長較痛,可能伴有嚴重心律不整,但找不到動脈硬化或阻塞點,發作時心電圖為ST段上升。有些個案可能只有局部尖銳疼痛、表淺性疼痛,亦可能沒任何疼痛的主訴,此稱為「靜默的心絞痛」。

文獻探討

心電圖





Upsloping

Downsloping

Horizontal

上升:表示心肌透壁性梗塞。此狀況, ST段較正常提升 >1mm, 且時間常過 80 毫秒。

下降:心內膜下梗死、低鉀血症。

較正常低:線段較平,可能代表冠狀動脈缺血缺氧。

04

研究方法



研究方法

01

資料預處理

03

特徵選取

02

資料視覺化

04

模型預測



缺失值

```
1 myData = pd.read_csv("/content/drive/MyDrive/資料探勘/heart.csv")
 2 myData.info()
<class 'pandas. core. frame. DataFrame'>
RangeIndex: 918 entries, 0 to 917
Data columns (total 12 columns):
    Column
                   Non-Null Count Dtype
                   918 non-nu11
                                  int64
    Age
                   918 non-nul1
                                  object
    Sex
    ChestPainType
                   918 non-nul1
                                  object
    RestingBP
                   918 non-nul1
                                  int64
    Cholesterol 918 non-null
                                  int64
    FastingBS 918 non-null
                                  int64
    RestingECG
                 918 non-nul1
                                  object
    MaxHR
                   918 non-nul1
                                  int64
    ExerciseAngina 918 non-null
                                  object
    01dpeak
               918 non-nu11
                                  float64
    ST Slope
             918 non-nu11
                                  object
    HeartDisease 918 non-null
                                  int64
dtypes: float64(1), int64(6), object(5)
memory usage: 86.2+ KB
```



2 myData.isnul1	().sum()
Age	0
Sex	0
ChestPainType	0
RestingBP	0
Cholesterol	0
FastingBS	0
RestingECG	0
MaxHR	0
ExerciseAngina	0
01dpeak	0
ST_Slope	0
HeartDisease	0
dtype: int64	

原本資料

	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	01dpeak	ST_Slope	HeartDisease
0	40	М	ATA	140	289	0	Normal	172	N	0.0	Up	0
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
2	37	М	ATA	130	283	0	ST	98	N	0.0	Up	0
3	48	F	ASY	138	214	0	Normal	108	Υ	1.5	Flat	1
4	54	М	NAP	150	195	0	Normal	122	N	0.0	Up	0
913	45	М	TA	110	264	0	Normal	132	N	1.2	Flat	1
914	68	М	ASY	144	193	1	Normal	141	N	3.4	Flat	1
915	57	М	ASY	130	131	0	Normal	115	Υ	1.2	Flat	1
916	57	F	ATA	130	236	0	LVH	174	N	0.0	Flat	1
917	38	М	NAP	138	175	0	Normal	173	N	0.0	Up	0
918 rd	ows x	12 col	umns									

轉換成數值型資料

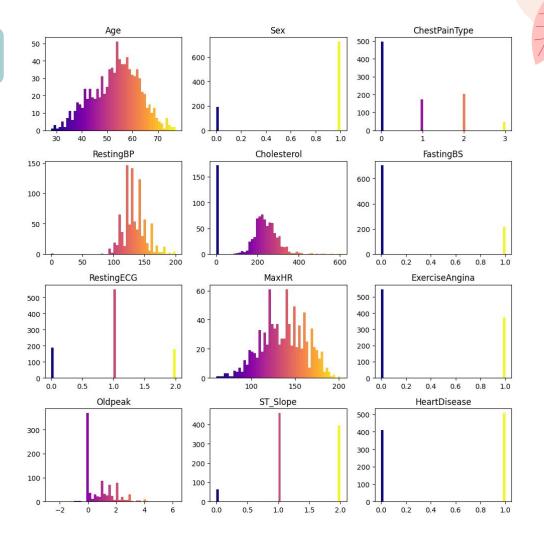
918 rows x 12 columns

```
1 # 從sklearn中導入Label Encoding
2 from sklearn.preprocessing import LabelEncoder
3 label_encoder = LabelEncoder()
4
5 myData['Sex'] = label_encoder.fit_transform(myData['Sex'])
6 myData['ChestPainType'] = label_encoder.fit_transform(myData['ChestPainType'])
7 myData['RestingECG'] = label_encoder.fit_transform(myData['RestingECG'])
8 myData['ExerciseAngina'] = label_encoder.fit_transform(myData['ExerciseAngina'])
9 myData['ST_Slope'] = label_encoder.fit_transform(myData['ST_Slope'])
10 myData
```

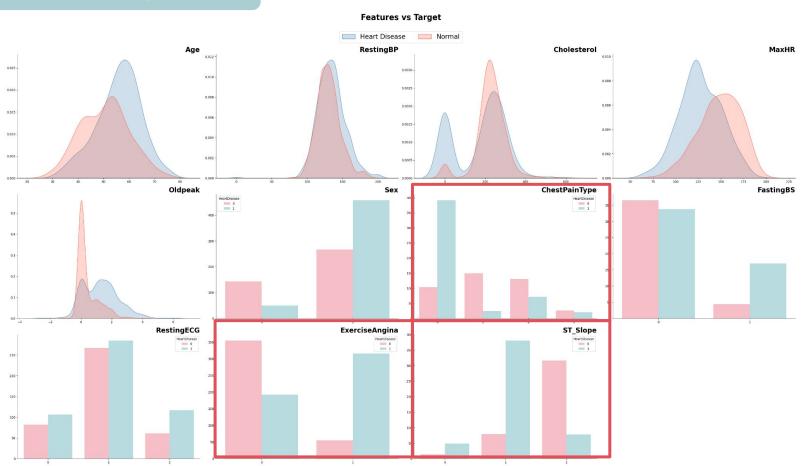
	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	01dpeak	ST_Slope	HeartDisease
0	40	1	1	140	289	0	1	172	0	0.0	2	0
1	49	0	2	160	180	0	1	156	0	1.0	1	1
2	37	1	1	130	283	0	2	98	0	0.0	2	0
3	48	0	0	138	214	0	1	108	1	1.5	1	1
4	54	1	2	150	195	0	1	122	0	0.0	2	0
913	45	1	3	110	264	0	1	132	0	1.2	1	
914	68	1	0	144	193	1	1	141	0	3.4	1	1
915	57	1	0	130	131	0	1	115	1	1.2	1	
916	57	0	1	130	236	0	0	174	0	0.0	1	1
917	38	1	2	138	175	0	1	173	0	0.0	2	0



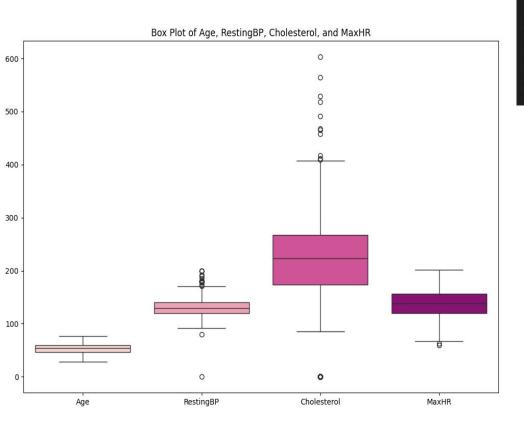
資料視覺化

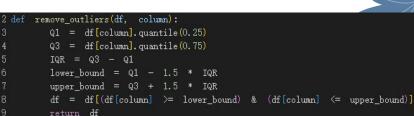


資料視覺化



刪除離群值





```
RestingBP
                                                        Cholesterol '
                         Sex ChestPainType
       735.000000 735.000000
                                  735.000000
                                             735.000000
                                                          735.000000
        52.903401
                                   0.843537 133.065306
                                                          241.038095
        9.495169
                    0.428693
                                   0.958972
                                            17.348365
                                                          51.164730
                                                          85,000000
        28.000000
        46.000000
                    1.000000
                                   0.000000 120.000000
        54.000000
                     1.000000
                                   1.000000 130.000000
                                                          236.000000
        59,000000
                     1.000000
                                            140.000000
                                                          274.000000
                                   3.000000 200.000000
        77.000000
                    1.000000
                                                          407.000000
        FastingBS RestingECG
                                   MaxHR ExerciseAngina
                                                            Oldpeak
       735.000000 735.000000 735.000000
                                              735.000000
                                                         735.000000
        0.164626
                    0.934694 140.277551
                                                0.386395
                                                           0.905170
mean
        0.371095
                              24.499218
                                                0.487254
                    0.000000 71.000000
                                                          -0.100000
        0.000000
                                                0.000000
                    1.000000 122.000000
        0.000000
                                                0.000000
                                                           0.000000
50%
                     1.000000 140.000000
                                                           0.500000
        0.000000
                     1.000000
                              160,000000
                                                1.000000
                                                           1.500000
         1.000000
                     2.000000 202.000000
                                                1.000000
                                                           6.200000
         ST Slope HeartDisease
       735.000000
                     735.000000
        1.413605
                      0.476190
mean
        0.597752
                      0.499773
        0.000000
                      0.000000
         1.000000
                      0.000000
50%
         1.000000
         2.000000
                      1.000000
```

特徵選取-P value

```
1 from sklearn import feature_selection as fs
 2 skb = fs. SelectKBest(fs.f_classif, k=5) #從此特徵執中選取K個最好的特徵值
 3 skb.fit(x,y)
 4 print (skb. pvalues )
 5 skb.fit_transform(x,y)
[4.31012494e-18 8.23636427e-16 5.59817836e-29 1.71945304e-06
 3.84881940e-03 7.45738033e-06 8.28844324e-01 4.13355298e-2
 5.54301629e-60 4.03560326e-48 2.43393737e-73]
array([[ 1. , 172. , 0. , 0. , 2. ],
      [ 2., 156., 0., 1., 1.],
      [ 1., 98., 0., 0., 2.],
      [ 0., 115., 1., 1.2, 1.],
      [ 1., 174., 0., 0., 1.],
      [ 2., 173., 0., 0., 2.]])
```

經由特徵選取可知胸痛類型、運動誘發心絞痛、Oldpeak運動引起ST段抑制、ST_SIpoe運動高峰ST段的斜率、MaxHR與是否罹心血管疾病的相關性較高,符合前面參考的文獻資料

	Age	Sex	ChestPainType	Resting	gBP Chol	lesterol	FastingBS	RestingECG	
0	40	1	1	1	140	289	0	1	
1	49	0	2	1	160	180	0	1	
2	37	1	1	1	130	283	0	2	
3	48	0	0]	138	214	0	1	
4	54	1	2	1	150	195	0	1	
913	45	1	3	1	110	264	0	1	
914	68	1	0	1	144	193	1	1	
915	57	1	0]	130	131	0	1	
916	57	0	1	1	130	236	0	0	
917	38	1	2		138	175	0	1	
					om o1				
_ \	MaxH		AND DESCRIPTION OF THE PARTY OF)ldpeak	ST_Slope				
0	17		0	0.0		?	0		
1	15		0	1.0		l	1		
2	9		0	0.0		2	0		
3	10		1	1.5		<u>.</u>	1		
4	12	Z	0	0.0		3	0		
019	10			1.0	• • • • • • • • • • • • • • • • • • • •				
913	13		0	1.2			1		
914	14		0	3.4		I.	1		
915	11		1	1.2			1		
916	17		0	0.0		E.	1		
917	17	3	0	0.0		2	0		
[735	rows	x 12	2 columns]						

熱力圖



Age -	1	-0.056	0.21	0.25	-0.095	0.2	0.21	-0.38	0.22	0.26	0.27	0.28
Sex -	-0.056	1	-0.18	-0.0051	0.2	-0.12	0.018	0.19	-0.19	-0.11	-0.15	-0.31
ChestPainType -	0.21	-0.18	1	0.075	-0.13	0.16	0.11	-0.27	0.32	0.3	0.35	0.46
RestingBP -	0.25	-0.0051	0.075	1	0.1	0.07	0.098	-0.11	0.16	0.16	0.075	0.11
Cholesterol -	-0.095	0.2	-0.13	0.1	1	-0.26	0.11	0.24	-0.034	0.05	-0.11	-0.23
FastingBS -	0.2	-0.12	0.16	0.07	-0.26	1	0.051	-0.13	0.06	0.053	0.18	0.27
RestingECG -	0.21	0.018	0.11	0.098	0.11	0.051	1	0.049	0.036	0.11	0.079	0.061
MaxHR -	-0.38	0.19	-0.27	-0.11	0.24	-0.13	0.049	1	-0.37	-0.16	-0.34	-0.4
ExerciseAngina -	0.22	-0.19	0.32	0.16	-0.034	0.06	0.036	-0.37	1	0.41	0.43	0.49
Oldpeak -	0.26	-0.11	0.3	0.16	0.05	0.053	0.11	-0.16	0.41	1	0.5	0.4
ST_Slope -	0.27	-0.15	0.35	0.075	-0.11	0.18	0.079	-0.34	0.43	0.5	1	0.56
HeartDisease -	0.28	-0.31	0.46	0.11	-0.23	0.27	0.061	-0.4	0.49	0.4	0.56	1
	Age -	Sex -	ChestPainType -	RestingBP -	Cholesterol -	FastingBS -	RestingECG -	MaxHR -	ExerciseAngina -	Oldpeak -	ST_Slope -	HeartDisease -

```
myData.corr()
    cor
   plt.figure(figsize = (10,10))
   sns. heatmap (cor, annot=True, fmt='.2f')
- 0.6
    根據熱圖可推得心血管疾病與
    胸痛類型、
- 0.2
    最大心率、
    運動誘發心絞痛、
    Oldpeak運動引起ST段抑制、
    ST_SIpoe運動高峰ST段的斜率
    相關性最高
```

模型預測



- Random Forest
- Gaussain Naïve Bayes
- DecisionTree
- ANN

模型預測

Random Forest

Gaussain Naïve Bayes

```
rfc = ensemble.RandomForestClassifier()
rfc.fit(X_train, y_train)
rfc_posi = rfc.predict_proba(X_test)[:, 1]
rfc_fp, rfc_tp, rfc_ts = metrics.roc_curve(y_test, rfc_posi)
rfc_pd = rfc.predict(X_test)
print("Random Forest confusion_matrix: ")
print(metrics.confusion_matrix(y_true = y_test, y_pred = rfc_pd))
print(metrics.classification_report(y_true = y_test, y_pred = rfc_pd))
```

```
gnb = naive_bayes.GaussianNB()
gnb.fit(X_train, y_train)
gnb_posi = gnb.predict_proba(X_test)[:, 1]
gnb_fp, gnb_tp, gnb_ts = metrics.roc_curve(y_test, gnb_posi)
gnb_pd = gnb.predict(X_test)
print("Bayes confusion_matrix: ")
print(metrics.confusion_matrix(y_true = y_test, y_pred = gnb_pd))
print(metrics.classification_report(y_true = y_test, y_pred = gnb_pd))
```

Confusion matrix

code

[[136 31] [24 140]] [[143 24] [24 140]]

	_	precision	recall	f1-score	support
Classification report	0 1	0. 85 0. 82	0. 81 0. 85	0. 83 0. 84	167 164
we	accuracy macro avg	0. 83	0.83	0. 83 0. 83	331 331 331

	precision	recall	f1-score	support
0	0. 86	0. 86	0. 86	167
1	0. 85	0. 85	0. 85	164
accuracy			0. 85	331
macro avg	0. 85	0.85	0. 85	331
reighted avg	0. 85	0. 85	0. 85	331

模型預測



DecisionTree

ANN

dtc = tree.DecisionTreeClassifier()
dtc.fit(X_train, y_train)
itc_pd = dtc.predict(X_test)
print(metrics.classification_report(y_true = y_test, y_pred = dtc_pd))
print(metrics.confusion_matrix(y_true = y_test, y_pred = dtc_pd))

mlp.fit(X_train, y_train)
mlp_pd = mlp.predict(X_test) #進行預測
print(confusion_matrix(y_true = y_test, y_pred = mlp_pd))
print(classification_report(y_true = y_test, y_pred = mlp_pd))
mlp.score(X_test, y_test) #準確度

Confusion matrix

code

[[140, 27], [40, 124]]

recall f1-score

support

[[144 23] [24 140]]

(Classification
	report
accurac	

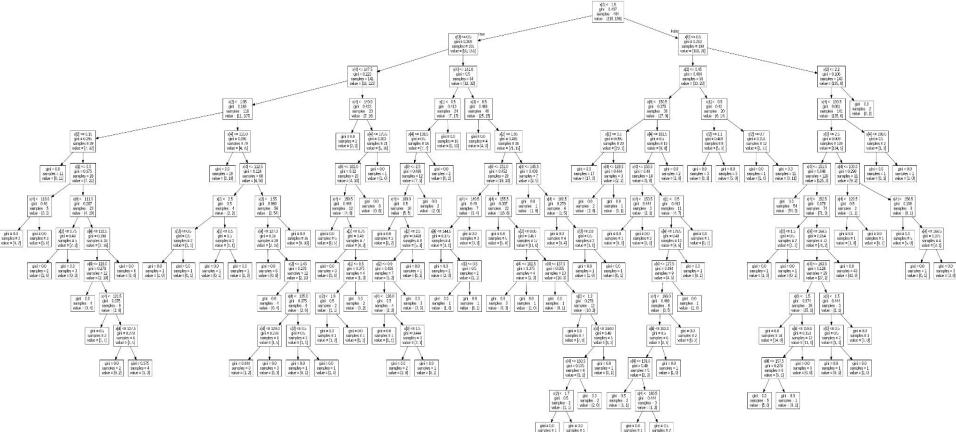
	-				
on	0	0.78	0.84	0.81	167
	1	0.82	0.76	0.79	164
	accuracy			0.80	331
	macro avg	0.80	0.80	0.80	331
wei	ighted avg	0.80	0.80	0.80	331
				<u> </u>	

precision

	precision	recall	f1-score	support
0 1	0.86 0.86	0.86 0.85	0.86 0.86	167 164
accuracy macro avg veighted avg	0.86 0.86	0.86 0.86	0.86 0.86 0.86	331 331 331

Decision Tree



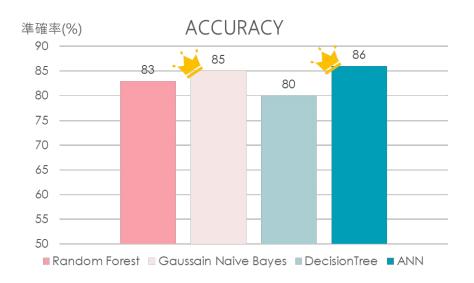


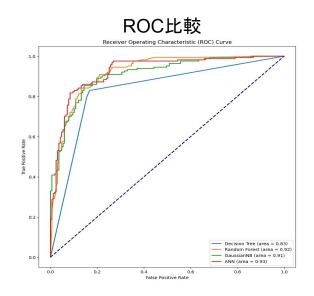
05

結果與討論



比較4種model





ANN的準確率最高
DecisionTree 準確率最低

ANN的ROC最高 DecisionTree的ROC最低

選擇ANN作為AI Model對我們的資料集最適合

06

結論



結論



心血管疾病與胸痛類型、運動誘發心絞痛、Oldpeak 運動引起 ST 段抑制、ST_Slpoe 運動高峰 ST 段的斜率相關度最高,與參考文獻相符

透過 ANN 的模型能達到 86% 的 F1-score,

心血管疾病近年來不再侷限於年長者,呼籲年輕族群注重飲食及健康習慣

07 參考資料

心血管疾病是什麼?一次了解心血管疾病症狀、治療以及如何預防 - 康健知識庫 (commonhealth.com.tw)

高雄榮民總醫院重症醫學部-心臟重症 (vghks.gov.tw)

https://health.udn.com/health/story/5964/3818713



Thanks for Listening!

