

**國立東華大學資訊工程系**

**National Dong Hwa University**

**112學年度大學部畢業專題研究報告**

112 CSIE Undergraduate Project Report

硬幣聲音辨識



| 指導教授Advisor： | 顏士淨教授 |
| --- | --- |
| 專題參與人員Team Member： | 羅禾筑 |
|  | 王天佑 |
|  | 王振權 |
|  | |
|  | |
|  | |

中 華 民 國　　113　 　年　　5　　月 　17 日

**國立東華大學資訊工程系**

**National Dong Hwa University**

**112學年度大學部畢業專題研究報告**

112 CSIE Undergraduate Project Report

硬幣聲音辨識



| 指導教授Advisor： | 顏士淨教授 |
| --- | --- |
| 專題參與人員Team Member： | 羅禾筑 |
|  | 王天佑 |
|  | 王振權 |
|  |  |
|  |  |
|  |  |

中 華 民 國　　113　 　年　　5　　月 　17 日

**國立東華大學資訊工程學系**

**專題報告原創性聲明**

**National Dong Hwa University**

**Department of Computer Science and Information Engineering**

**Statement of Originality**

本人鄭重聲明：

所呈交的專題報告是在指導老師指導下進行的研究工作及取得的研究成果。除文中已經註明引用的內容外，本報告不包含任何其他個人或集體已經發表或撰寫過的研究成果。對本文的研究做出重要貢獻的個人與集體，均已在文中以明確方式標明。若有違上述聲明，願依校規處分及承擔法律責任。

I hereby affirm that the submitted project report is the result of research under the supervision of my advisor. Except where due references are made, the report contains no material previously published or written by another person or group. All significant facilitators to the project have been mentioned explicitly. Should any part of the statement were breached, I am subject to the punishment enforced by the University and any legal responsibility incurred.

| 學號Student No. | 學生姓名Name | 親筆簽名Signature |
| --- | --- | --- |
| 410921234 | 羅禾筑 |  |
| 410912041 | 王天佑 |  |
| 410921211 | 王振權 |  |
|  |  |  |

日 期: Date **計畫摘要Abstract**

Keywords: 硬幣，聲音辨識，深度學習，卷積神經網路

最早的聲音辨識在電腦發明之前，人們用早期的聲碼器來視作語音辨識及合成的雛形，而1920年代生產的「Radio Rex」玩具狗是最早的語音辨識器，當這隻狗的名字被呼喚的時候，它能夠從底座上彈出來。最早的基於電腦的語音辨識系統是由AT&T貝爾實驗室開發的Audrey語音辨識系統，它能夠辨識10個英文數位。其辨識方法是跟蹤語音中的共振峰。而在這個項目中，我們使用了深度學習技術來識別不同硬幣的聲音。聲音文件被轉換成 MFCC（梅爾頻率倒譜係數）特徵，然後用於訓練和測試模型。項目涉及的主要步驟包括數據預處理、模型構建、訓練和評估。

目錄Contents

壹、前言--------------------------------------------------------------p.6

貳、研究動機與研究問題----------------------------------------------------p.7

參、研究方法與步驟

(一)前置觀念----------------------------------------------------------p.8~p.10

(二)實作步驟----------------------------------------------------------p.10~p.11

(三)實驗工具及安裝配置-------------------------------------------p.12

(四)擷取資料特徵方法----------------------------------------------p.13

(五)資料的色彩辨識-------------------------------------------------p.14

(六)模型訓練過程----------------------------------------------------p.15~p.16

肆、結果與討論

(一)實驗結果----------------------------------------------------------p.17

(二)問題討論及總結-------------------------------------------------p.18

(三)未來展望----------------------------------------------------------p.18

伍、參考文獻-------------------------------------------------------------------p.19

壹、前言Introduction

硬幣聲音辨識技術是一項旨在通過聲音特徵來識別不同硬幣類型的新興技術。隨著科技的不斷發展，聲音辨識技術逐漸應用於各個領域，其中包括硬幣的識別和辨識。本專題旨在利用深度學習方法，通過分析硬幣產生的聲音信號，實現對1元、10元和50元硬幣的自動識別，從而提高現代貨幣識別技術的效率和準確性。

聲音辨識技術的起源可以追溯到20世紀中葉。早期的聲音識別系統主要基於聲音的頻譜特徵和模式識別，用於語音識別和音樂分析。隨著計算機和人工智能技術的進步，聲音辨識應用逐漸擴展到更廣泛的領域，包括生物醫學、環境監測、安全監控和硬幣辨識等。在硬幣辨識方面，聲音辨識技術為硬幣識別提供了一種全新的解決方案。通過分析硬幣在落地或滾動時產生的聲音特徵，可以區分不同種類的硬幣，實現自動化的硬幣辨識和計數，為金融、商業和自動售貨等領域提供了便利和效率。

本文將探索硬幣聲音辨識技術的研究與應用，並通過構建深度學習模型，實現對不同硬幣的準確識別。

貳、研究動機與研究問題 Motivation and Research Problem

一、研究動機

研究的動機源於公車司機在日常工作中處理大量硬幣的需求。在公車服務中，司機需要收取乘客的車資並找零，這通常涉及大量的硬幣交易。然而，硬幣的辨識和計數是一項耗時且容易出錯的工作，特別是當硬幣種類繁多時。

我們希望利用現代技術的應用，尤其是聲音辨識技術，來解決這個問題。透過分析不同硬幣落地或滾動時產生的聲音特徵，我們可以建立一個智能系統，能夠自動識別並區分1元、10元和50元等不同種類的硬幣。這項研究不僅將提高公車司機在工作中處理零錢的效率，還能在其他領域中發揮重要作用。例如，在自動售貨機、自助服務行業和金融機構中，這種硬幣辨識技術也能夠大幅提升操作的便捷性和準確性，從而提供更好的用戶體驗和服務效率。

二、研究問題

我們針對硬幣聲音辨識的方法進行研究，歸納出以下問題：

1. 如何迅速又準確的判斷硬幣種類？
2. 機器學習的訓練模型要如何選擇？
3. 如何提高聲音識別的辨識率？

參、研究方法與步驟 Research Method

前置觀念

語音處理的概念:

人類講話的聲音可以以『示波器』(Oscilloscope)測量成一個隨時間變化的波形信號，這種信號對時間的關係，稱之為『時域』(Time Domain)，在電腦對信號進行分析時，通常信號會先被轉換成在不同頻率下對應的振幅及相位，稱之為『頻域』（frequency domain），轉換的公式稱為『傅立葉變換』（Fourier transform)，信號經過每隔一段時間取樣，就得到可以進行分析的數位音檔，附檔名是 wav。

根據 Nyquist 定理，如果想要從數位信號無損轉到類比信號，我們以最高信號頻率的2倍的採樣頻率進行採樣。我們硬幣的聲音的頻率大概在20kHz~1kHz ，因此語音辨識需使用40kHz的wav提取特徵。例如，40kHz取樣速率的音訊，經『傅立葉變換』後的頻率範圍為0-20KHz。

傅立葉變換在聲音辨識中的應用主要在於將聲音信號從時間域轉換到頻率域，從而能夠提取聲音的頻率特徵，進一步用於辨識和分類。以下是傅立葉變換在聲音辨識中的主要應用：

一.頻譜分析：

通過將聲音信號進行傅立葉變換，可以將其表示為在頻率上的成分，即頻譜。這使得我們可以分析聲音信號中各種頻率的分佈情況，從而了解聲音的結構和特徵。

二.特徵提取：

基於頻率域的表示，可以從頻譜中提取出各種特徵，例如梅爾頻率倒譜係數（MFCC）、調頻譜等。這些特徵能夠捕捉聲音的局部和全局特性，並且通常用於構建聲音辨識模型。

三.降噪和去除不需要的成分：

傅立葉變換可以將聲音信號分解為不同頻率的成分，這使得我們可以對信號進行降噪和去除不需要的成分。例如，可以將低能量的高頻噪音部分去除，以提高信號的品質。

四.分析聲音結構和語音特徵：

通過對頻譜進行分析，可以了解聲音的結構和語音特徵，例如音高、音量、語調等。這些信息對於語音辨識和語音分析具有重要意義。

所以，傅立葉變換在聲音辨識中起著關鍵作用，它通過將聲音信號從時間域轉換到頻率域，提取聲音的頻率特徵，進一步用於模型訓練和辨識。

為了方便作語音辨識，與影像一樣，我們會對語音作特徵抽取(Feature Extraction)，我們使用MFCC(Mel frequency cepstral coefficients)，特徵抽取前須先對聲音作前置處理：

一.幀(Frame)切割：

通常每幀是25ms，幀與幀之間重疊10ms，以避免邊界信號的遺漏。

二.信號加強：

針對高頻信號作加強，使信號更清楚。

三.加窗(Window)：

目的是消除各個幀兩端可能會造成的信號不連續性，常用的窗函數有方窗、漢明窗等。

去除雜訊圖. 聲音前置處理，圖片來源：<https://ithelp.ithome.com.tw/articles/10195763>

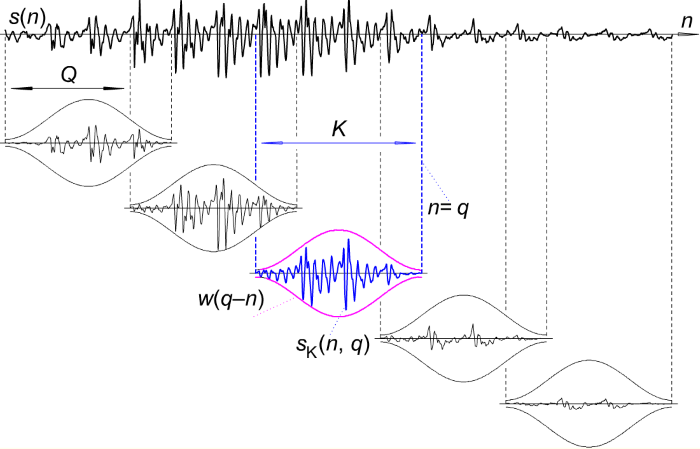


圖. Windowing and frame formation，圖片來源：<https://ithelp.ithome.com.tw/articles/10195763>

CNN卷積神經網路

CNN 指的是卷積神經網絡（Convolutional Neural Network），是一種深度學習模型，廣泛應用於圖像識別、影像處理、自然語言處理等領域。CNN在圖像識別方面取得了巨大成功，特別是在辨識具有層次結構的複雜圖像中表現出色。

CNN 的主要特點包括：

一.卷積層（Convolutional Layer）：

用於提取圖像的特徵，通過對圖像進行卷積操作，可以捕捉到圖像的局部特徵。

二.池化層（Pooling Layer）：

通常與卷積層交替使用，用於降低特徵圖的尺寸和計算量，同時保留重要的特徵。

三.激活函數（Activation Function）：

通常在卷積層之後使用，用於引入非線性，增加模型的表達能力，常見的激活函數包括 ReLU、Sigmoid、Tanh 等。

四.全連接層（Fully Connected Layer）：

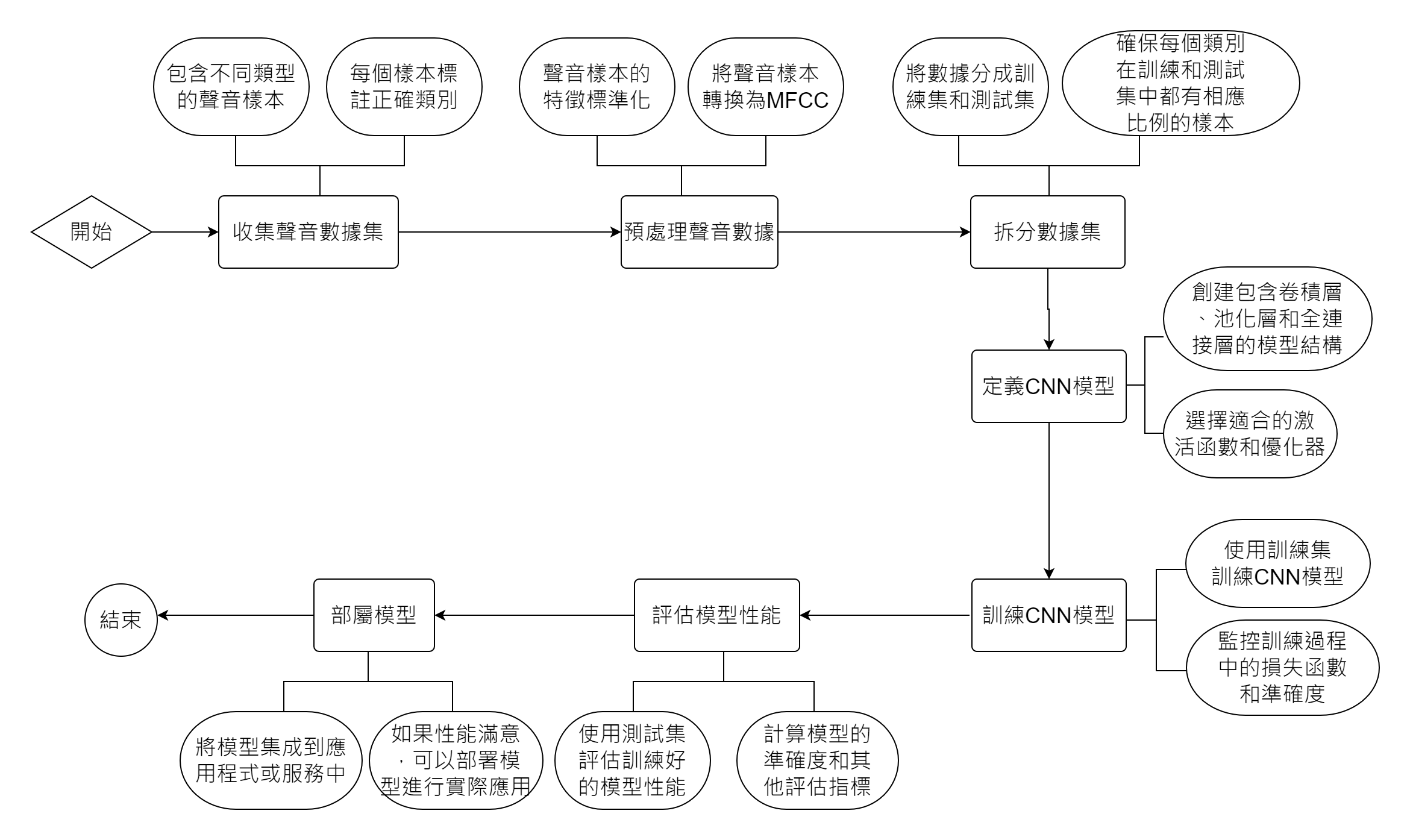
通常在 CNN 的最後幾層使用，用於將卷積層提取的特徵映射到最終的輸出空間，例如分類標籤。

CNN 的訓練過程通常使用反向傳播算法（Backpropagation）和梯度下降等優化算法進行模型參數的更新，以最小化訓練數據的損失函數。

在聲音識別中，CNN也被廣泛應用，特別是在聲音的特徵提取和模式識別方面。通過將聲音數據轉換為圖像形式（例如聲譜圖），CNN 可以有效地從聲音中提取特徵並進行分類。

實作:

1. 蒐集一元、十元、五十元的音檔各500筆，其中300筆作為訓練組，200筆做為測試組。
2. 利用 librosa 套件，對每一個音檔轉換成 MFCC 特徵向量。
3. 將 MFCC 特徵向量轉換成 CNN 的 input 格式。
4. 採取CNN模型訓練。
5. 輸出混淆矩陣來討論結果。
6. 任意指定一個硬幣音檔作測試，看能不能準確預測出來並做準確度分析



一.數據預處理：

save\_data\_to\_array(path, label, max\_pad\_len): 此函數用於將聲音文件轉換成 MFCC 特徵，並保存為 .npy 文件。它首先從指定路徑加載聲音文件，然後將其轉換為 MFCC 特徵向量。如果加載的文件存在問題（例如損壞的文件），則會發出警告並繼續處理下一個文件。

二.模型構建：

在 load\_dataset(path) 中，數據集準備階段涉及加載聲音文件並提取 MFCC 特徵。每個聲音文件對應一個 MFCC 特徵向量，然後用於訓練和測試模型。

三.模型訓練和評估：

建立了一個簡單的卷積神經網絡（CNN）模型，用於訓練和識別聲音文件的類別。該模型包括卷積層、池化層、Dropout 層和全連接層，並使用 softmax 激活函數輸出分類結果。

使用 keras 庫編譯模型，並使用訓練集進行模型訓練。訓練過程中的準確性和損失等指標被記錄下來，以便後續分析和優化模型。

最後，對模型進行評估並保存模型文件（.h5 格式），以備將來進行預測和部署。

實驗工具：

1. 錄音用的麥克風 (二) 投擲硬幣的容器



(三)一元硬幣 (四)十元硬幣 (五)五十元硬幣



安裝配置：  
librosa Version: 0.10.1

tqdm Version: 4.65.0

numpy Version: 1.24.3

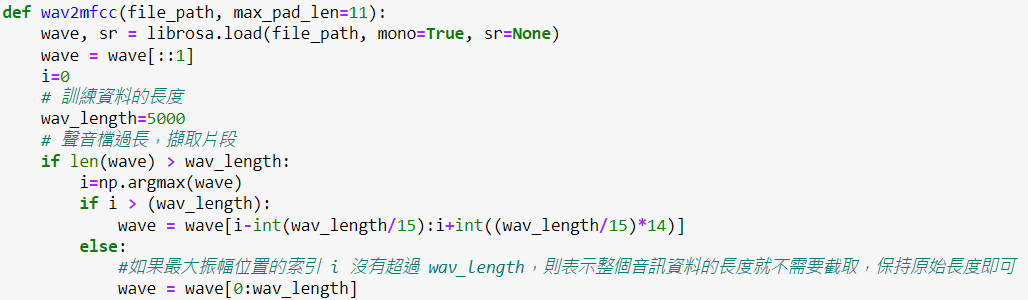
scikit-learn Version: 1.3.0

keras Version: 2.15.0

tensorflow Version: 2.15.0

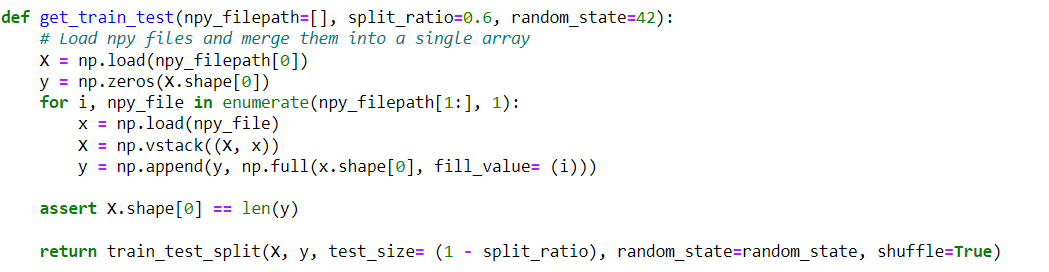
python 3.11.4

擷取資料特徵方法:



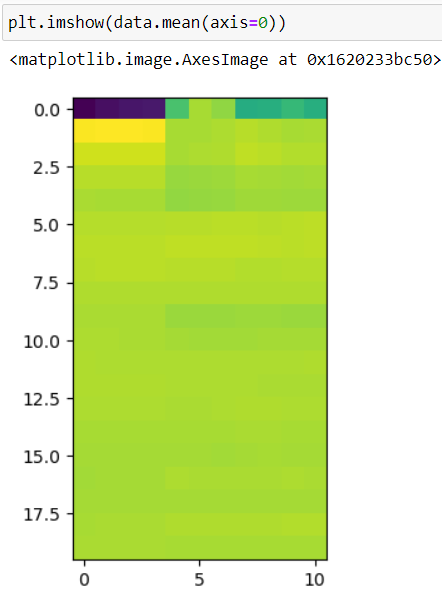
這段程式碼是為了從音訊檔案中提取 MFCC 特徵而設計的。在這段程式碼中，我們使用了 librosa 函式庫來載入音訊檔案，並利用其提供的功能來提取 MFCC 特徵。在實作中，我們發現當音訊檔案的最大振幅往往會在一開始硬幣掉下去第一聲的位置，所以為了盡可能採樣完整的片段，因此，我們將採樣範圍定在音訊中振幅最大的位置的前 1/15 到後 14/15 這段區間，這樣做能夠有效提升特徵提取的效果。並且能夠有效地保留有用的音訊資訊，同時減少了不必要的資料量，有助於提高後續模型訓練的效率和準確性。

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

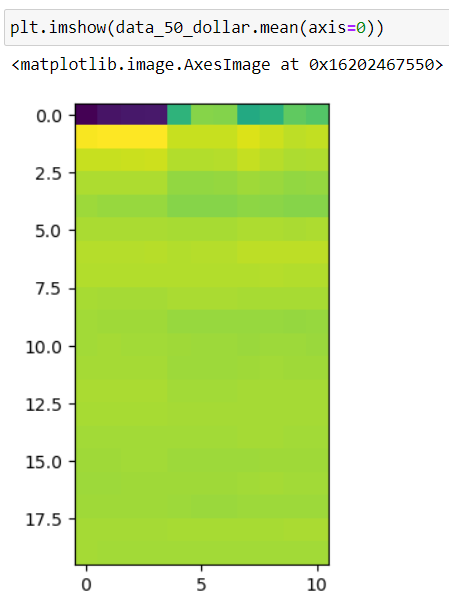
  
這段程式碼會將提供的 NumPy 數組檔案加載進記憶體中，然後根據指定的分割比例將資料分割為訓練組和測試組。訓練組和測試組包括特徵和對應的標籤，並且可以根據指定的隨機種子進行洗牌。最後，將分割後的資料返回給呼叫者。

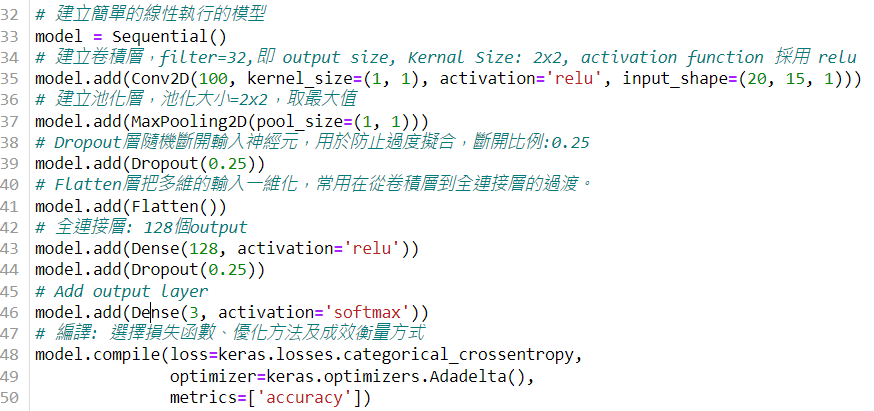
資料的色彩辨識:

1元 10元

50元

模型訓練過程:



這段程式碼是建立一個簡單的卷積神經網路模型，用於音訊檔案的聲音辨識，這個模型由一個卷積層、一個池化層、兩個 Dropout 層、一個 Flatten 層和兩個全連接層組成。它的輸入形狀為 (20, 15, 1)，對應於 MFCC 特徵的維度。該模型的最終輸出是一個大小為 3 的 softmax 分類器，對應於3個類別。模型使用的損失函數是 categorical\_crossentropy，優化方法是 Adadelta，並且使用 accuracy 作為衡量模型性能的指標。

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



這段程式碼執行了模型的訓練過程，並將訓練過程的歷史記錄存儲在名為 train\_history 的變數中。在這個 fit 函式中：

X\_train 是訓練組的特徵數據；

y\_train\_hot 是訓練組的標籤數據（經過 one-hot 編碼處理）；

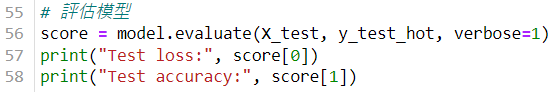
batch\_size=16 指定每次訓練時使用的樣本批次大小為 16；

epochs=200 指定訓練的迭代次數為 200 次；

verbose=1 指定訓練過程的輸出詳情，設置為 1 表示輸出進度條；

validation\_data=(X\_test, y\_test\_hot) 是用於驗證模型的測試組數據，包括特徵數據 X\_test 和對應的標籤數據 y\_test\_hot

在訓練過程中，模型將根據訓練數據進行梯度下降優化，並同時使用測試組數據進行驗證。訓練過程中的損失函數值和準確率等指標將被記錄下來，以便後續分析和可視化。

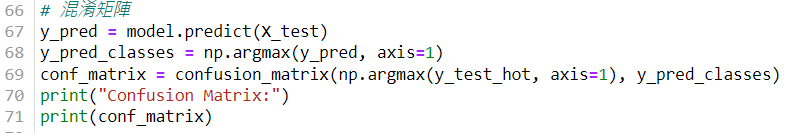


evaluate 函式將計算模型在測試組上的損失值和準確率。score 是一個包含兩個元素的列表，第一個元素是測試損失值，第二個元素是測試準確率。這兩個值將被打印出來，用於評估模型的性能表現。  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



wav2mfcc 函式將要測試的音檔的指定路徑轉換為 MFCC 特徵， 將獲得的 MFCC 特徵重新調整為模型所需的形狀；使用模型對新音訊檔案進行預測，並取得預測結果中概率最高的類別的索引，即該音訊檔案被分類為哪個類別。

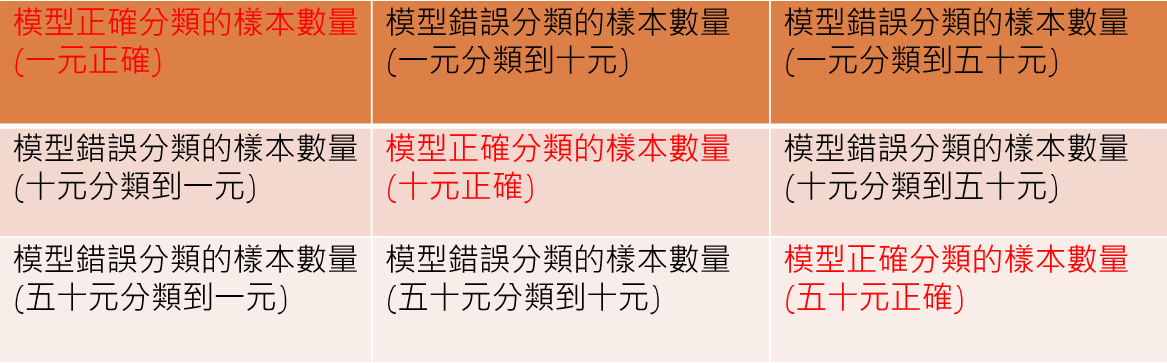
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



混淆矩陣:

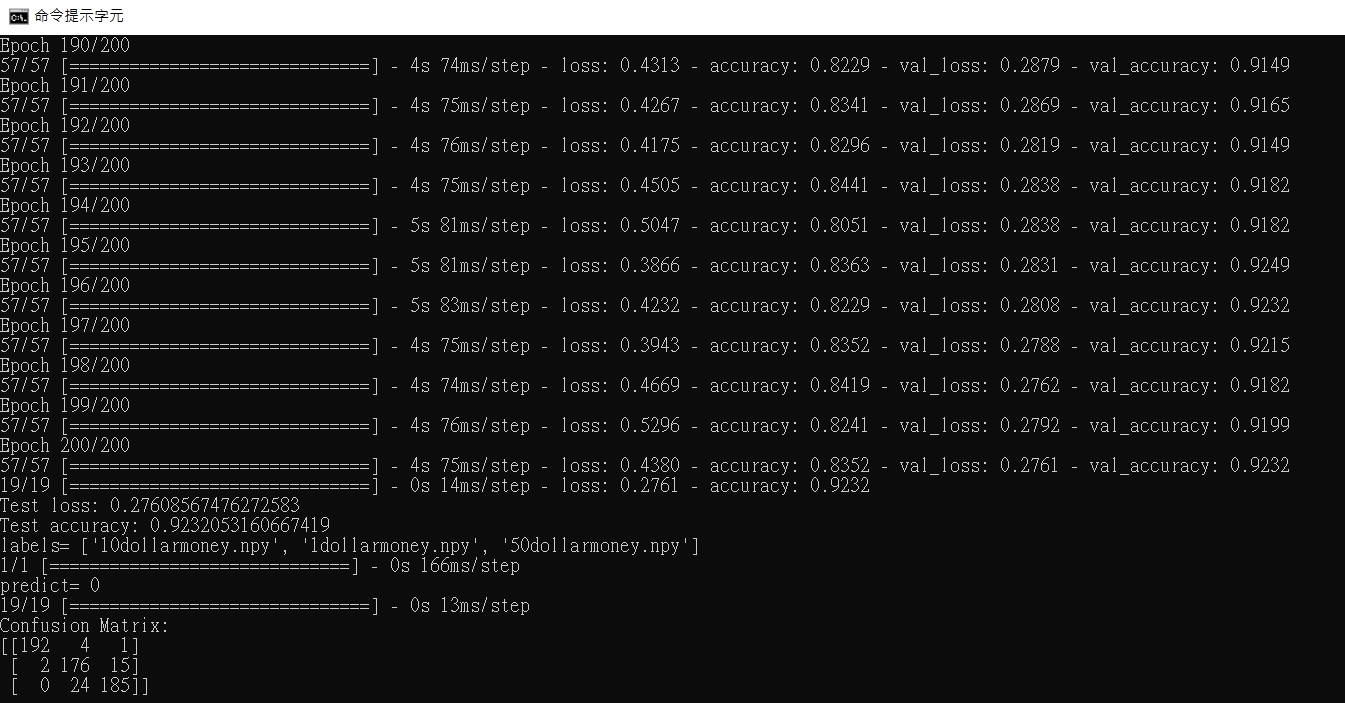
主對角線元素（True Positives）：主對角線上的元素表示模型正確分類的樣本數量，即實際類別與預測類別相同的樣本數量。

非主對角線元素：非主對角線上的元素表示模型錯誤分類的樣本數量，即模型將樣本錯誤分類為其他類別的情況

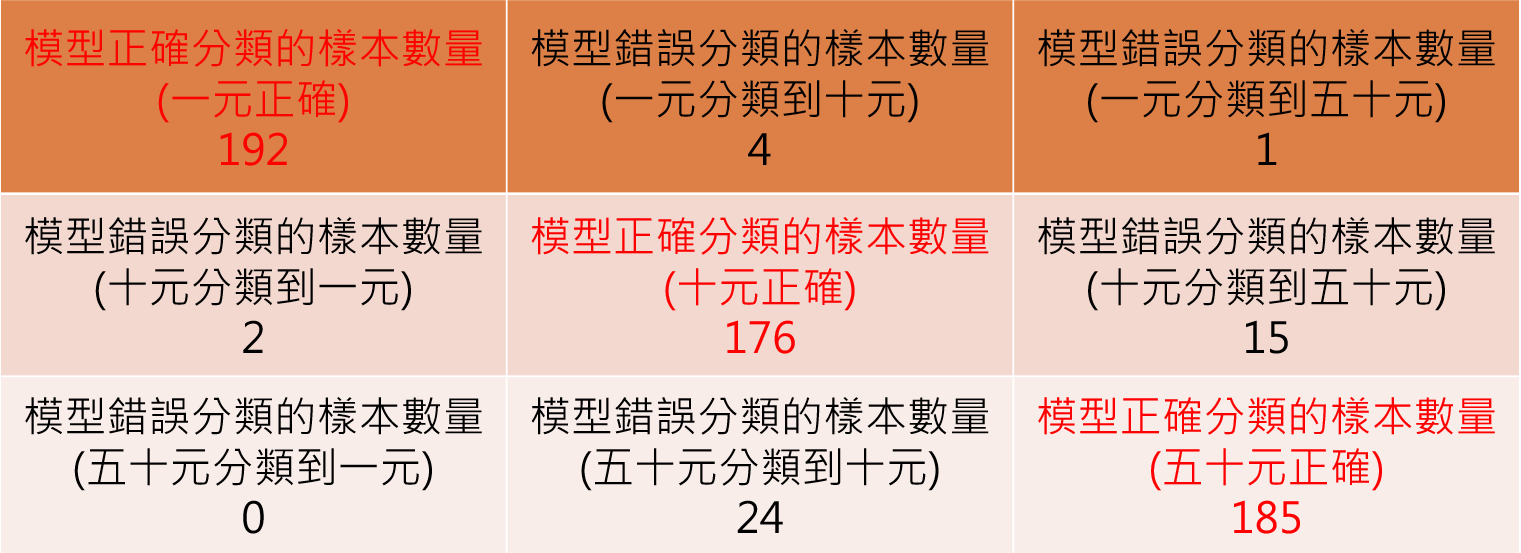


肆、結果及討論 Research Results an Conclusions：

(一)實驗結果：



我們的模型訓練後test set的準確率為92.32%，我們透過增加特徵抓取，雖然導致辨識一個音檔可能需要5分鐘，但換取來高準確率。



透過混淆矩陣可以發現辨實際屬於1元的資料的樣本總數為 197。其中，有 192 個樣本被正確預測為1元的資料（True Positives），有 4 個樣本被錯誤預測為10元的資料（False Negatives），有 1 個樣本被錯誤預測為50元的資料（False Negatives）。而實際屬於10元的資料的樣本總數為 193筆，其中，有 176 個樣本被正確預測為10元（True Positives），有 2 個樣本被錯誤預測為1元（False Positives），有 15 個樣本被錯誤預測為50元（False Negatives）。實際屬於50元的樣本總數為 209。其中，有 24 個樣本被錯誤預測為10元（False Positives），有 185 個樣本被正確預測為50元（True Positives），沒有樣本被錯誤預測為1元（False Negatives）。

(二)問題討論及總結

一、問題討論：

Q:如何迅速又準確的判斷硬幣種類？

A:根據我們的實驗結果，我們發現在擷取聲音片段也就是聲音振幅最大的部分，往往都在硬幣落下第一聲的位置。

Q:如何提高聲音識別的辨識率？

A:調整損失函數加大特徵抓取，有效的特徵數量越多辨識率越高但因為資料量變多所以會花更多時間

二、總結：

通過這個項目，我們使用深度學習技術處理音頻數據，並構建一個用於聲音分類的卷積神經網絡模型。通過合適的數據預處理和模型設計，並實現對硬幣聲音的自動識別，為聲音處理和分類任務提供了一個有趣的案例和示例。

(三)未來展望：

一、複數硬幣辨識: 透過完成更複雜的訓練模組來實現硬幣在多個硬幣的硬幣堆裡辨識出來。

二、不同國家貨幣辨識: 透過深度學習的模組來辨識出不同國家的貨幣。

伍、參考文獻 References：

(一)中文：

Day 25：自動語音識別(Automatic Speech Recognition) -- 觀念與實踐

<https://ithelp.ithome.com.tw/articles/10195763>

飛文彬(2022)。**深度學習之聲音辨識應用於金屬待測物。** ［碩士論文］，國立虎尾科技大學車輛工程系碩士班。  
<https://hdl.handle.net/11296/7q6zm5>

鄧氏陲殷(2017)。**基於深度學習之聲音辨識及偵測**。［碩士論文］，國立中央大學資訊工程學系。

<https://hdl.handle.net/11296/umdja4>

梅爾倒頻譜-維基百科

<https://zh.wikipedia.org/zh-tw/%E6%A2%85%E7%88%BE%E5%80%92%E9%A0%BB%E8%AD%9C>

語音辨識-維基百科

<https://zh.wikipedia.org/zh-tw/%E8%AF%AD%E9%9F%B3%E8%AF%86%E5%88%AB>

傅立葉變換

<https://hackmd.io/@jkrvivian/B1wHF21ib?type=view>

CNN卷積神經網路

<https://hackmd.io/@allen108108/rkn-oVGA4>

卷積神經網路的運作原理

<https://brohrer.mcknote.com/zh-Hant/how_machine_learning_works/how_convolutional_neural_networks_work.html>

(二)English：

Mel-Frequency Cepstral Coefficients Explained Easily

<https://www.youtube.com/watch?v=4_SH2nfbQZ8>