# 摘要

本研究以圖像辨識模型為基礎，訓練分辨深度偽造影像的機器學習模型，並探討不同圖像辨識模型之準確率。

本研究使用取幀程式將深度偽造影片轉換為然後對轉換的圖片進行預處理：人臉裁切及人臉縮放。

本研究預計開發一款可在Android運行的深度偽造檢測應用程式，程式基於Kotlin語言開發而成，目前程式正在開發中。

本研究的資料集為Celeb-DF(v2)。經過預處理後的資料集共有40000張。最後我們使用預處理過後的資料集，分別使用2個不同的圖像辨識演算法，分別得到了99.11%與98.63%的準確率。

# 研究動機

近年來深度偽造技術日漸成熟，偽造技術過的圖片及影片極難分辨。這使有心人士有可乘之機，用不實的影像惡意造謠，侵犯隱私、傷害名譽、操縱選舉甚至引發暴力。

為了抵抗惡意使用，有了深度偽造檢測技術的出現，不過目前大眾可使用的檢測工具甚少，有些已開源但無包裝成GUI，有些僅供大型企業使用，而在網路上或應用程式商店查詢，基本無可供大眾使用的深度偽造技術辨識工具。僅在移動裝置端有幾個應用程式可供使用，但其功能並不完整，有些軟體只能辨識圖片；有些軟體只能辨識影片，且有影片長度限制(不可過短)；有些則需要上傳至雲端進行辨識，可能造成資安問題。

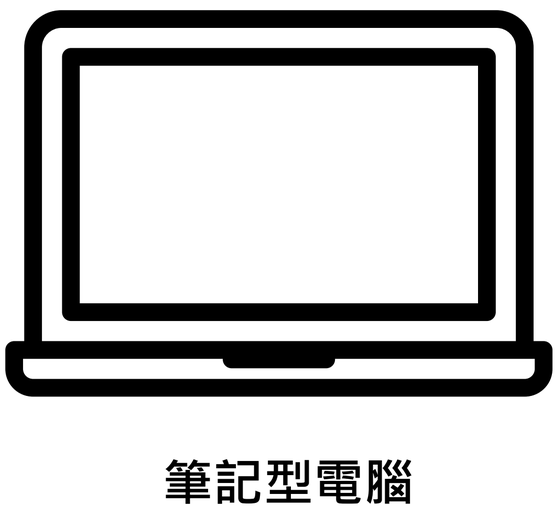
考慮手機的便攜性大於電腦，本研究期望開發出一款可在Android系統使用的深度偽造技術檢測應用程式，同時具有檢測深度偽造圖片及影片的功能，以彌補現在深度偽造辨識工具缺乏以及功能不全的缺點，並且考慮資安及隱私問題，檢測過程全程使用本地化的模型，避免將資料上傳至網路。

# 研究目的及研究問題

1. 使用人臉裁切器裁切人臉，以縮小偵測範圍，並建立、分割資料集。
2. 把這些人臉圖像輸入卷積神經網路訓練一個本地化的深度偽造技術檢測模型，基於深度偽造檢測的需求，開發出一套應用程式，讓使用者可檢測圖片或影片是否經過深度偽造。
3. 開發出美觀、現代，且易於使用者操作的圖形化使用者介面(GUI)，其中本研究欲使用Material Design設計語言。並且讓使用者可以查看該圖片/影片是否採用深度偽造技術，並讓使用者明確知道影片中偵測到深度偽造技術的時間區段。

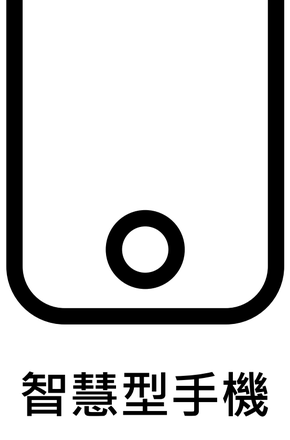
# 研究設備

## 硬 體



筆記型電腦

基本規格：
1. CPU：Intel i5-1135G7
2. GPU(Nvidia)：NVIDIA Geforce MX350



智慧型手機

基本規格：
1. 型號：Mi 9T pro
2. SoC：Qulacomm Snapdragon 855
3. 系統：Android 13 AOSP

## 軟 體

Kotlin

1. 用途：開發Android App

Python

1. 用途：資料預處理、機器學習
2. 使用的套件：

Android

1. 用途：測試Android App
2. 版本：Android 13 AOSP

Ubuntu

1. 用途：資料預處理
2. 版本：Ubuntu 22.10

Google Colab

1. 用途：機器學習

# 研究過程或方法

## 一、 研究方法

(一) 研究架構

本研究首先從網路上下載影片，利用程式將影片轉換成一張張圖片，使用人臉偵測器剪裁出人臉，最後將圖片調整成一樣的尺寸，輸入神經網路中訓練，並建立出深度學習模型，最後將神經網路模型轉換至 Tensorflow Lite 模型，將其輸入至開發好的 Android App。

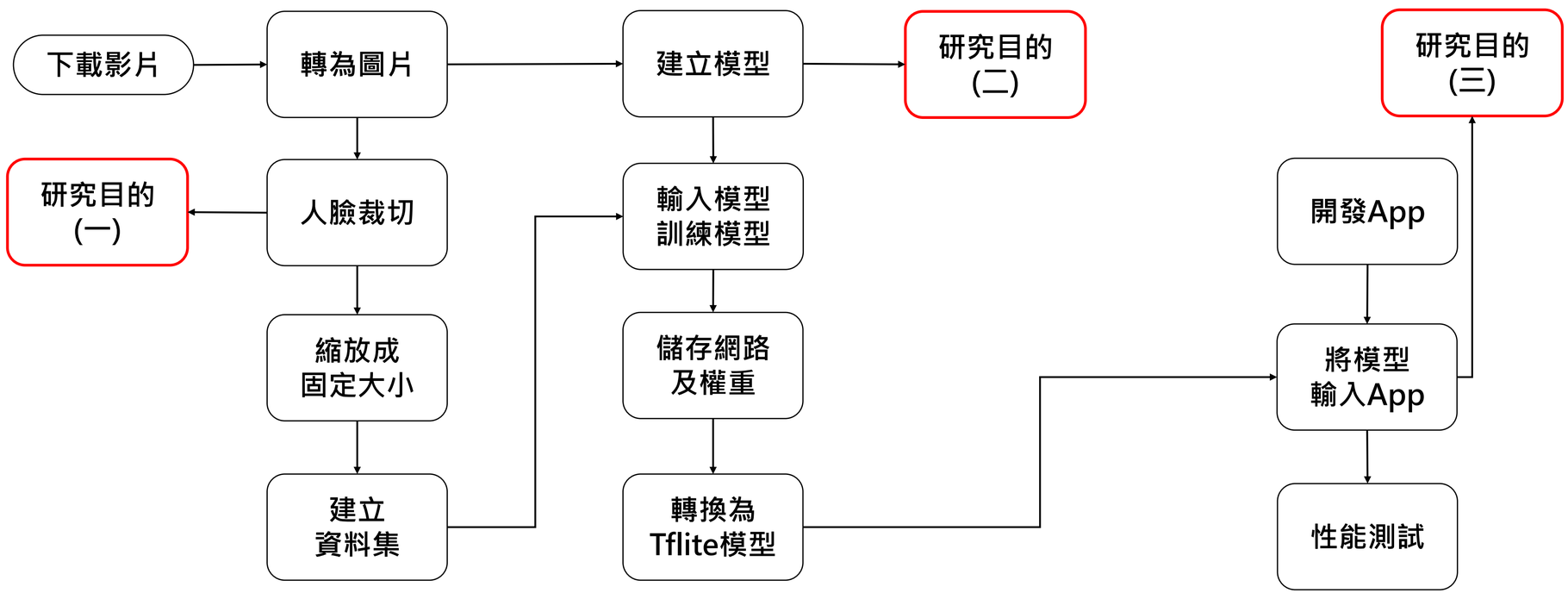



圖1、研究架構圖

## (二) 資料預處理

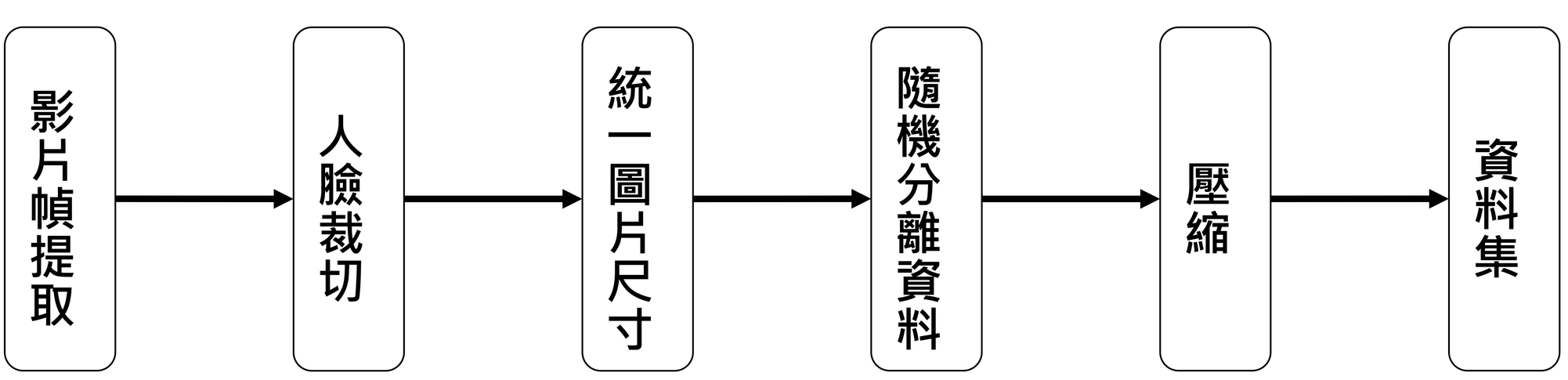



圖2、資料預處理流程圖

| Celeb-DF(v2) | Train | Test | Valid | 總計 |
|---|---|---|---|---|
| Real(真實) | 12000 | 4000 | 4000 | 20000(50%) |
| Fake(偽造) | 12000 | 4000 | 4000 | 20000(50%) |
| Total | 24000(60%) | 8000(20%) | 8000(20%) | 40000 |

## (三) 程式運行環境

機器學習之外的部分在 Ubuntu 22.10 中運行
以下為電腦的詳細配置：
CPU：Intel Core i5-1135G7
GPU：Nvidia MX350 2GB

機器學習部分在 Google Colaboratory 中運行
以下為 Google Colaboratory 的詳細配置：
CPU：Intel(R) Xeon(R) @ 2.20GHz
GPU：Nvidia Tesla T4 15GB

## (四) 機器學習

我們使用 VGG16與InceptionResNetv2 兩種神經網路，作為本研究的圖像辨識模型。

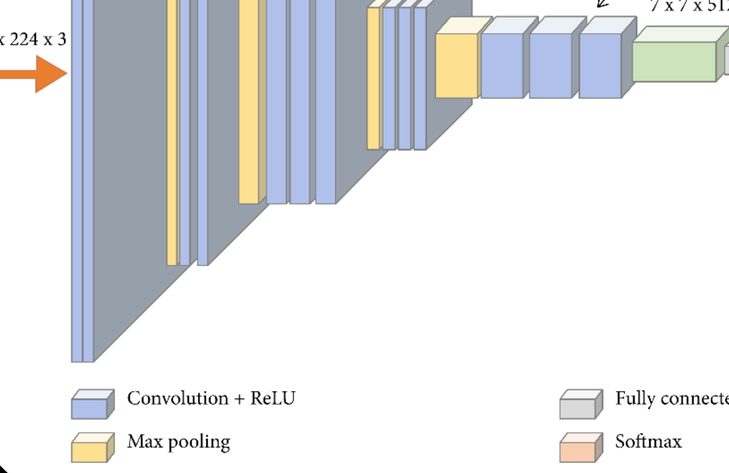

圖3、VGG16 架構圖

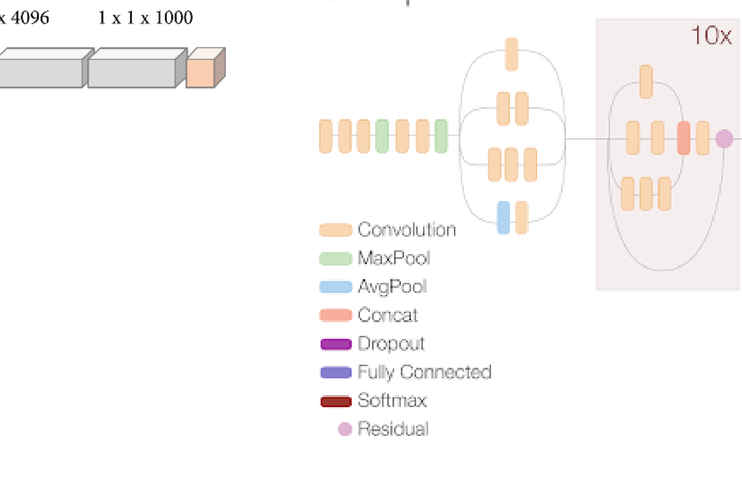

圖4、InceptionResNetV2 架構圖

# 研究結果

## 一、 機器學習成果

### (一) VGG16



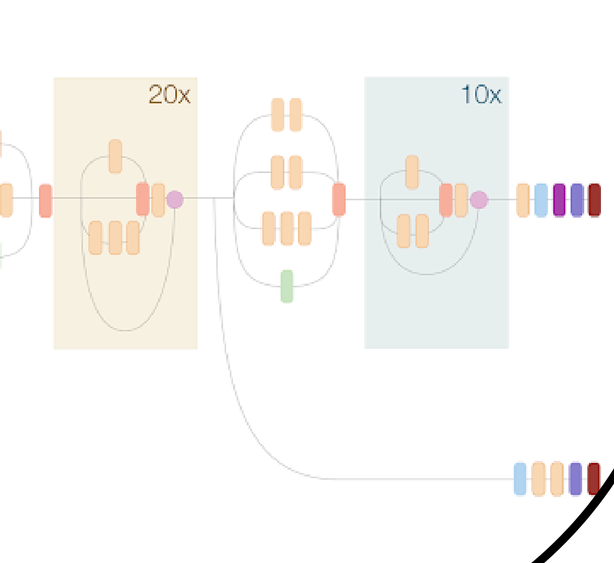
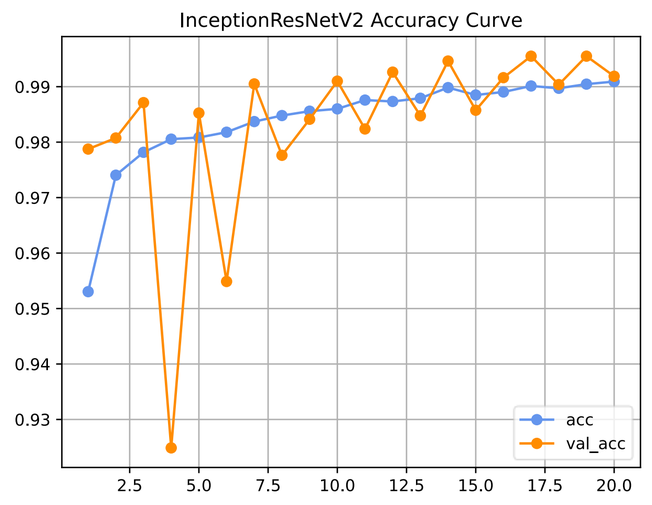
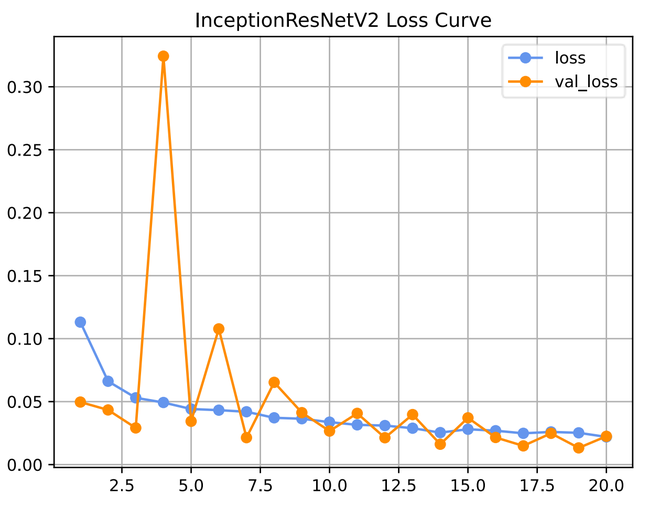




圖4、VGG16之準確率曲線(左) 、損失函數曲線(中)與混淆矩陣(右)

## (二) InceptionResNetV2



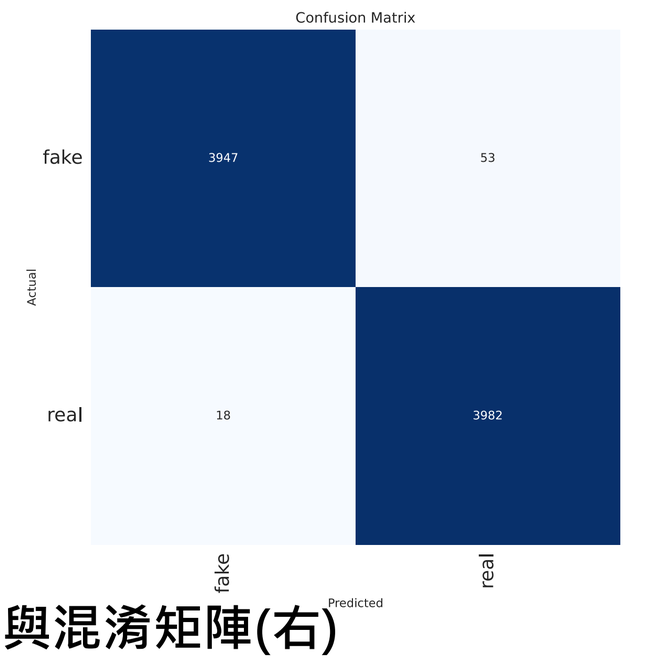




圖5、InceptionResNetV2之準確率曲線(左)、損失函數曲線(中)與混淆矩陣(右)

| 圖像辨識模型 | 正確張數 | 錯誤張數 | 準確率 |
|---|---|---|---|
| InceptionResNetV2 | 7929 | 71 | 99.11% |
| VGG16 | 7721 | 279 | 96.51% |

# 二、辨識資料集外影片

## (一)歐巴馬影片

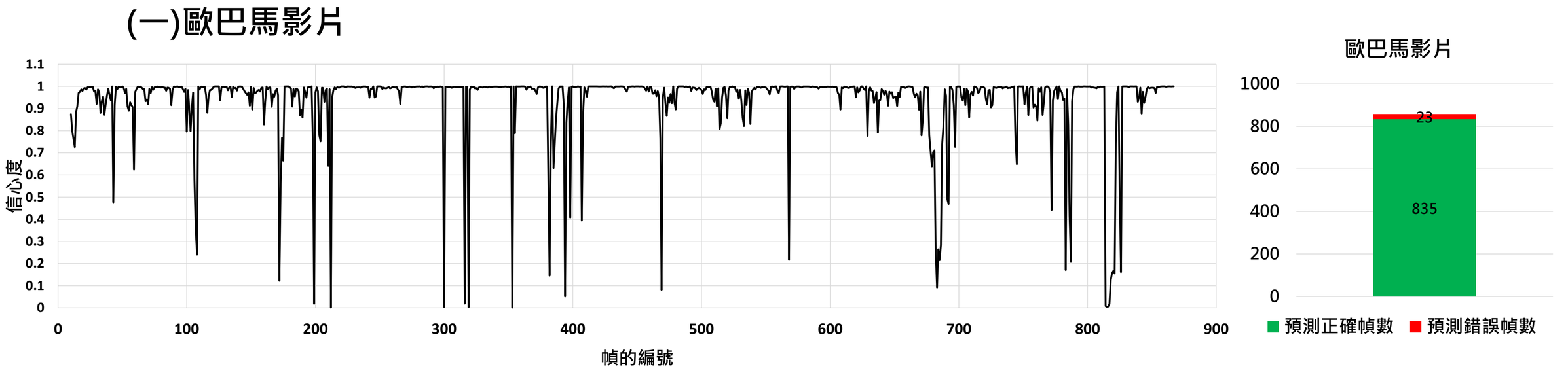
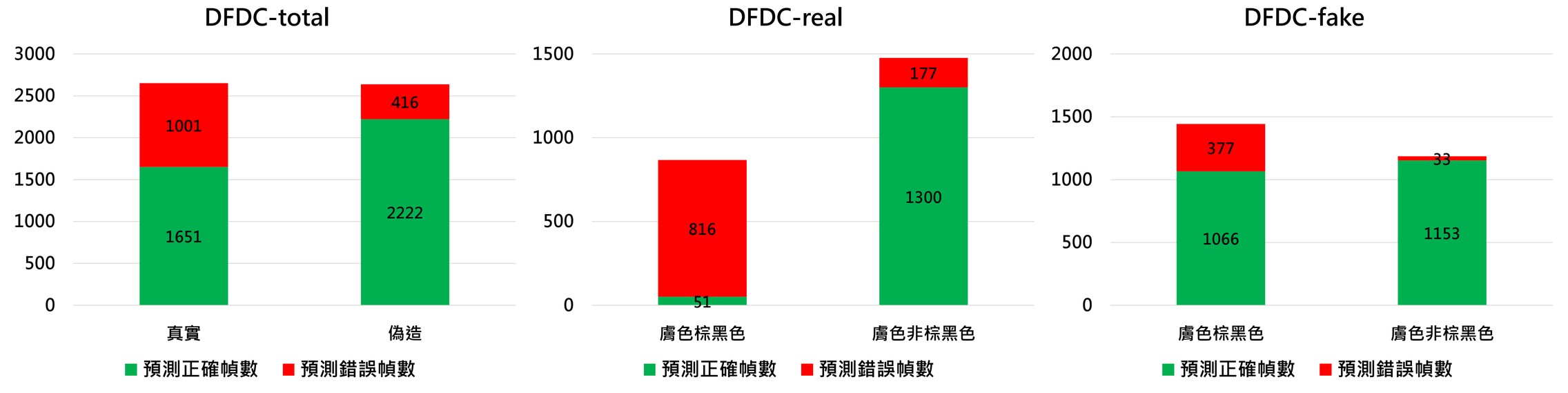



圖6、歐巴馬影片幀編號-信心度折線圖

## (二) DFDC Dataset(抽取20部)




## (三)應用程式


| 目標 (■ 完成 ■ Bug ■ 未完成) |
|---|
| UI - Material Design 設計 |
| UI - 顯示圖片 |
| UI - 顯示影片 |
| 資料預處理 - 提取影片的幀 |
| 資料預處理 - 人臉裁切 |
| 資料預處理 - 圖像縮放 |
| 圖像辨識 - 預測圖像 |

# 討論與結論

一、本研究在使用InceptionResNetV2作為圖像辨識模型，Celeb-DF(v2)測試集做測試時，可獲得最高為99.11%的準確率。
二、在測試DFDC資料集後，我們發現圖像辨識模型對深皮膚演員的辨識錯誤率較高，可能與Celeb-DF(v2)資料集中缺乏深皮膚演員有關。因此未來會獲取更多深色人種的影片。
三、圖像辨識模型在辨識美國前總統歐巴馬的深度偽造影片時，得到了97.31%的準確率。
四、應用程式目前有預測圖像真偽輸出錯誤的問題。

# 參考資料


羅濟威(民109)。基於人臉辨識加人臉對齊及數據增強強化後的深度造假檢測器。台北：國立台灣科技大學資訊工程學系。台北：國立台灣科技大學資訊工程學系。
Yuezum Li, Xin Yang, Pu Sun, Honggang Qi, & Siwei Lyu(2020). Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics. Retrieved February 1, 2023.
Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, Cristian Canton Ferrer(2020). The DeepFake Detection Challenge (DFDC) Dataset. Retrieved March 2, 2023.