採用網路音樂串流平台建構音樂風格分類機制：以音樂素養教育應用為例

A Music Genre Classification Mechanism for Enhancing Music Literacy Through Online Music Streaming Platform

陳昱安[1] 胡志堅[2]
CHEN, YU AN[1] HU, CHIH CHIEN[2]
[1] 大同大學 資訊經營學系研究所 研究生
[1] Tatung University Department of Information Management Student
E-mail：g11012005@o365.ttu.edu.tw
[2] 大同大學 資訊經營學系研究所 助理教授
[2] Tatung University Department of Information Management Assistant Professor
E-mail：holdenhu@gm.ttu.edu.tw

摘要

音樂美感儼然成為教育核心素養，透過音樂教學實踐引導學生認識各種音樂風格，習得音樂知識是音樂教學的重要途徑。為了協助學生運用音樂串流平台的資源，體驗並融入音樂教學活動。本研究採用卷積神經網路(CNN)模型架構進行音樂風格分類，藉此讓學生於聆聽音樂的過程中能夠習得音樂流派的特徵與知識。實驗結果顯示，所提出的模型在網路音樂串流平台的分類實證上，具有 87%的準確率。因此，若結合此方法於音樂串流平台，將可應用於音樂素養教學的輔助使用。未來研究，建議改善並增進音樂資訊檢索的準確率與效率。

關鍵詞：音樂風格分類、卷積神經網路、音樂資訊檢索、音樂教育

Abstract

Music aesthetics has become the core competence in education. Through music teaching practice, guiding students to understand various music genres and acquiring music knowledge is an important way of music education. This study uses a convolutional neural network (CNN) model to classify music genres, then students can acquire the characteristics and knowledge of music genres. Experimental results show that the proposed model has an accuracy rate of 87% in the classification of online music streaming platforms. Therefore, the proposed method can be integrated to a music streaming platform for music literacy education. For future research, it is suggested to improve the accuracy and efficiency of music information retrieval.

Keywords: music genre classification, convolutional neural network (CNN), music information retrieval (MIR), music literacy education

壹、 前言

資通訊技術的發展，促使音樂串流平台成為現今音樂通路的主要管道。截至 2022 年止，Apple Music 已逾 1 億首歌曲、Amazon Music 超過 9000 萬首、第三順位的 Spotify 亦有 8200 萬首作品(易起宇, 民 111 年)。音樂素養課程常藉由音樂欣賞(music appreciation)的學習活動來激發學生的學習動機(Asmus, 2021)，建立學生認知音樂與生活之間的關係。所以，音樂教育工作者常鼓勵學生藉由學校的音樂教育，或日常的音樂體驗來培養音樂素養(Green, 2006)。網路音樂串流平台大量風格多元的音樂作品，切乎音樂素養教育之需求。現今音樂素養導向教學中，以「表現」、「鑑賞」、「實踐」為課程架構，鼓勵學生參與體驗音樂活動、聆聽多種音樂風格作品，以理解音樂要素與型態(陳育恬 & 鄭勝耀, 2022; 鄭英傑, 2022)。企盼培養學生的音樂「審美感知」與「審美理解」(陳育恬 & 鄭勝耀, 2022)。

然而，特定音樂作品會流行、或受到人們的喜愛，往往與其音樂特徵有關，每首音樂作品皆有獨特的音樂特徵。每一種音樂類型的音樂特徵具有一定程度的相似性，聆聽者對於相似的音樂特徵之評價通常會趨於一致。因此，若能精準的評估音樂作品的音樂特徵，將學習者所喜愛的音樂作品以相似的特徵進行分類，則能協助教師有效地選擇合適的音樂作品融入音樂教學活動。同樣地，學習者自學的過程中，可以快速取得偏好的音樂風格進行聆聽體驗與學習。由於音樂串流平台擁有大量音樂資源，具有多樣風格與文化特色的音樂資源(陳育恬 & 鄭勝耀, 2022)，為了將其融入音樂教學活動，需要具備一音樂風格分類篩選機制。因此，本研究採用卷積神經網路（Convolutional neural network, CNN）模型架構(Ashraf et al., 2023)設計一音樂風格分類機制，藉此讓學習者可透過既有的網路音樂串流平台選擇不同類別的音樂作品，藉由聆聽串流音樂的過程夠體驗學習不同音樂流派的特徵與知識。研究過程為了尋求有效的分類模型，將音樂風格類別分為九類（即 blues、classical、country、disco、hip-hop、metal、pop、reggae、rock）。利用調節多種組合的模型訓練參數來改善音樂風格分類模型的準確率（accuracy），試圖強化音樂風格分類模型之泛化能力（generalization ability），擴展模型在測試用音樂庫（test set）的準確率。

實驗採用機械視覺技術，將音頻信號轉換成較相似於人耳聽覺感受的梅爾頻譜（Mel spectrogram）(Stevens, Volkmann, & Newman, 1937)，然後運用捲積神經網路將轉換後的梅爾頻譜圖資料進行處理，調整超參數（hyperparameter），並使用音頻資料增強模式（data augmentation）以提昇模型效能(Aguiar, Costa, & Silla, 2018)，以避免模型過度擬合（overfitting）問題，同時強化模型泛化能力。實驗過程發現，所採用的訓練用音樂庫（training set）透過超參數調整後，可以有效改善模型效能，具有 89%的準確率。

為了進一步評估所提出分類模型的效益，於是採用網路音樂串流平台「Apple Music」(易起宇, 民 111 年)資料庫的分類機制，並透過該資料庫取得測試資料（test set），共有九種類別，每一類別各隨機選擇 10 首作品，共 90 首，進行音樂風格分類模型的驗證分析。實驗結果得知，音樂風格分類具有 87%的準確率。因此，若將所建構的方法整合於網路音樂串流平台，方能輔助教師、與學習者，應用於音樂素養教學實踐中，並對日益遽增的大量音樂資料進行準確的自動化風格分類，省去以人工方式替資料進行標記作業，並可減少大量人力成本。

貳、 文獻探討

(一) 梅爾頻譜之轉換

研究指出(Wegel & Lane, 1924)，由於人耳的聽覺掩蔽效應（auditory masking）會導致人耳對於低頻頻率感受強度會高於相同能量的高頻頻率，因此(Stevens et al., 1937)提出了梅爾刻度（Mel Scale）。該方法可以將頻率（Hz）透過三角帶通濾波器（triangular bandpass filters）轉換為與人耳聽覺相似的梅爾刻度，轉換概念如式 1：

$$m = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \quad \text{式 1}$$

其中 m 代表梅爾單位，f 代表頻率單位，其逆向變換公式如式 2：

$$f = 700 \cdot \left(10^{\frac{m}{2595}} - 1\right) \quad \text{式 2}$$

繪製梅爾刻度的頻譜圖之前，須先將音頻數位訊號透過短時傅立葉轉換（short-time Fourier transform, STFT），將經過傅立葉變換（Fourier transform）的訊號，分析出時間、與頻域資料。隨即，使用 STFT 的複數運算結果$X(m,\omega)$，將其取絕對值並平方得到二維的實數矩陣如式 3：

$$S = |X(m,\omega)|^2 \quad \text{式 3}$$

進一步，使用代表人耳聽覺感受的梅爾刻度進行內積，內積相乘後的二維矩陣即可繪製梅爾頻譜（Mel spectrogram）。

(二) 分類方法與深度學習

音樂曲風分類具有多種方法(Bahuleyan, 2018)，早期研究常使用特徵提取（feature extraction）、及深度學習模型（deep learning model）的機器學習方法進行音樂曲風的分類。基於特徵提取的方法，首先需要以人工方式定義音頻特徵，然後使用機器學習演算法，如邏輯迴歸分析（logistic regression, LR）、隨機森林（random forest, RF）、以及支持向量機（support vector machine, SVM）等方法進行分類。然而，基於深度學習模型的方法，通常採用 CNN 網路為深度學習模型架構(Poonia, Verma, & Malik, 2022; Ashraf et al., 2023)，或以 CNN 網路為基礎的 VGG（visual geometry group）深度學習模型架構(Bahuleyan, 2018)；VGG 網路運用更多的隱藏層、捲積層（convolution layer）、池化層（pooling layer），扁平層（flatten layer）、以及全連接層（fully connected layer）等，並使用更多的參數訓練，以提高準確率。其透過捲積核(kernel)的資料於扁平層(flatten)轉換成一維向量，再透過全連接層的 SoftMax 激活函數（activation function）將擷取到的特徵，透過預先定義的類別進行分類。此種圖像辨識技術已廣泛運用於圖像分類(Krizhevsky, Sutskever, & Hinton, 2017)，首先將音頻轉換為頻譜圖(Spectrogram)，再使用 CNN 網路對頻譜圖進行音頻風格的分析與分類。從許多分類方式的性能評估研究中(Bahuleyan, 2018)，發現使用深度學習進行音樂曲風辨識的模型效能其準確率大多優於使用基於特徵提取的分類演算法。然而，採用 VGG 網路雖提升模型準確率，但卻大量地耗費訓練時間成本、與資源。

有鑑於此，本研究試圖在 CNN 網路架構下使用較少的參數，降低時間成本，並維持相近的分類準確率水準。根據研究指出(Aguiar et al., 2018; Costa, Oliveira, Koerich, & Gouyon)，於音樂

資訊檢索(music information retrieval, MIR)領域中對音頻資料進行採樣時，通常不會一整段內容直接分析，而是對全樂曲進行較短長度，且不重疊的切片。如此，可以去除音樂片段中不具代表性的部分，這種切片採樣策略可使模型處理具較高紋理特徵的圖像。亦有研究發現採用 CNN 學習分類方法可得到較優異的效能(Bahuleyan, 2018)，因此，本研究提出基於 CNN 架構的音樂風格分類機制。

(三) 捲積神經網路分類模型

CNN 網路中的捲積核（kernel）具有平移、翻轉、縮放、光照情形的不變性，使得實驗影像圖形內容並不會因所處的環境或背景不同，而對 CNN 網路的辨識能力造成影響(Poonia, Verma, & Malik, 2022; Ashraf et al., 2023)。相對地，不具備捲積核的全連接神經網路（fully-connect neural network, FNN），當圖片內容受環境影響產生變動，其輸出內容將會完全不同，而必須對變動內容進行重新學習，且不易捕捉特徵，導致效率低下，甚至難以進行學習。因此 CNN 神經網路於機器視覺領域中，經常受到廣泛運用及改良。式 4 用於表達捲積層的平移不變性，其中，x 表示輸入圖像，F 為捲積層的運算（Output = Input · Kernel），T 為平移的轉換。

$$F\big(T(x)\big) = T\big(F(x)\big) \qquad \text{式 4}$$

本研究將網路音樂串流平台的音樂作品分為 9 個類別，屬於多類別分類問題（Multi-Class Classification Problems）。因此，模型經過扁平層轉換多維度張量（tensor）輸出（output）成為一維度張量後，使用 SoftMax 激活函數，針對某一類別，加總所有可能的類別表徵特性，確保較小的值也有較小的概率，不會被捨棄掉；並產生對應於每一個分類結果的機率，其機率總和為 1。以機率方式來計算神經網路的輸出結果，最大值即代表預測結果。式 5 表示 SoftMax 函數之計算方式。其中，y_i代表為預測機率，C 代表為類別總數，z 為神經網路的預測值，i 為神經網路第 i 個輸出。

$$y_i = \frac{e^z i}{\sum_{j=0}^{C} e^z j} \qquad \text{式 5}$$

多類別分類問題經常採用 Categorical Cross-Entropy（CEE）損失函數，其使用 SoftMax 函數，並結合 Cross-Entropy 損失函數。因此，可計算每個類別的預測機率、以及與真實值之間的誤差。式 6 表達 CCE 函數計算方式，其中，y 為預期輸出結果，$\hat{y}$為 SoftMax 函數輸出結果，f為 SoftMax 函數，C 代表為類別總數，N為批次數量。

$$CCE = -\frac{\sum_{i=1}^{N}\sum_{j=0}^{C} y_{i,j}\log\Big(f\big(\widehat{y_{i,j}}\big)\Big)}{N} \qquad \text{式 6}$$

參、 研究方法

(一) 實驗資料處理

本研究採用(Tzanetakis & Cook, 2002)提出之 GTZAN 資料集，進行實驗。其研究中使用音色紋理、節奏內容、音高內容等特徵來定義音樂特徵，將音樂分為 10 個流派類別（包括：Blues、Classical、Country、Disco、Hip-Hop、Jazz、Rock、Reggae、Pop、Metal），並驗證提出特徵對歌曲性能的重要性。每個流派類別中包含 100 個片段，長度均為 30 秒的節選，資料源於廣播、CD、

或 mp3 等音頻文件(取樣率 22050Hz、16 位元深度、單聲道儲存)。雖此資料集已廣為研究運用，(Sturm, 2012)指出 GTZAN 資料集中存在著標籤誤植情形，且於 Jazz 類別中第 54 個樣本有檔案壞軌狀況。因此，本研究捨去 Jazz 類別，以 9 種類別進行分類及訓練。本文將 GTZAN 資料集樣本分割為 3 秒一個片段(Aguiar et al., 2018; Costa et al.)，每首歌長度均等長為 30 秒，因此一首歌會被分割成 10 個片段，每一個類別 100 首歌，經過分割後每一個類別為 1000 個片段分割，共 9 個類別，且每類別共 1000 個切片樣本，總共 9000 個資料樣本作為訓練集。

(二) 音頻資料增強模式

為防範 CNN 學習效果過度擬合（overfitting），而造成模型不具備良好泛化（generalize）能力，應避免訓練資料不足、參數過多、資料樣式過於單一、抑或是模型設計過為複雜等。樣本數量過少的資料集，可透過資料增強來使樣本更加多樣化(Litjens et al., 2017)。同樣的，資料增強處理亦常運用於機器視覺辨識上(Shorten & Khoshgoftaar, 2019)，例如影像縮放、添加雜訊、旋轉影像、顏色調整、光源調整等。然而，資料增強方式須注意應用限制。例如，資料形式為時間序列的資料，應避免時間序列順序遭破壞，必須根據資料特性調節使用。

本研究使用的音樂資料屬於時間序列，時間的先後順序尤其重要，若發生改變則會使資料本身遭到破壞，適得其反。本文參考(Aguiar et al., 2018)所提出的音訊資料增強模式：

a. 加入白噪音：隨機加入不同強度的白噪音，以 0～0.5 之間的亂數(1 為最大值)為參數值。
b. 時間拉伸：將音頻資料播放速度調慢一倍到加速一倍之間隨機選擇(以 0.5～1.5 之間的亂數為參數值。
c. 改變音高：目前國際音樂慣用音律為十二平均律，即將一個八度音間平均切分為 12 個半音階。因此，可採用 0～11 之間的整數型態亂數為參數值，使樂曲調性於 12 個不同的調之間隨機選擇(當第 12 個半音時為高八度，聽覺上會與原曲調一樣，但頻率上會變為原本的 1/2)。
d. 反轉相位極性：以 0 或是-1 隨機選擇整數亂數，若為-1 則將聲波的相位進行反轉，促使聽覺上感受到聲波失去立體感。
e. 隨機增益：以 0～5 之間的浮點型態亂數為參數值，將音樂隨機進行增益調整，創造出聽覺上音量變大之感受。
f. 保留原始狀態：不加以處理，保留原始狀態。

以上方式以隨機選擇方式，每次擇一種方法，對原始音頻資料進行處理。根據上述對音頻進行處理後，再轉換成梅爾頻譜圖模型訓練資料，藉此產生多樣化資料，改善過度擬合問題。

(三) 梅爾頻譜圖轉換機制

將音頻資料進行切分後，將其轉為梅爾頻譜（使用 librosa 套件的 feature.melspectrogram 函數），將 MP3 格式的音頻轉換為 PNG 格式的頻譜圖。

轉換梅爾頻譜前，須先經過 STFT 變換，STFT 的計算需要使用到以下參數：n_fft＝2048 為 FFT 採樣的樣本數， win_length= 2048 為代表一個音框（frame），通常一個音框的長度會等於採樣的樣本數，hop_length 為跳躍步長，表示當音框在進行移動時有多少部分與上一個音框是重疊的，計算方式為將 win_length÷4 取整數=512，因此 hop_length＝512，並且使用的窗函數（window）

類型為 Hann windows，計算方式如式 7：

$$w(n) = 0.5 \cdot \left(1 - cos\left(\frac{2\pi n}{N-1}\right)\right) \qquad 式\ 7$$

$$n = 1 \cdots N$$

其中 N 為 n_fft，接下來要將音框化（frame blocking）後的資料與窗函數$w(n)$進行相乘來得到 STFT 計算結果，計算方式如式 8：

$$S(m,k) = \sum_{n=0}^{N-1} x(n + mH) \cdot w(n) \cdot e^{-i2\pi n\frac{k}{N}} \qquad 式\ 8$$

$S(m,k)$為 STFT 計算結果。m為目前音框的起點，H為目前音框的終點，$w(n)$為窗函數，$e^{-i2\pi n\frac{k}{N}}$為離散傅立葉轉換。計算出結果為複數二維矩陣，並透過式 3 方法將結果取絕對值，進行平方運算，形成二維實數陣列，並與梅爾刻度進行內積。內積相乘後的二維矩陣，即可繪製梅爾頻譜，且最後再將梅爾頻譜值取對數，轉換為分貝單位(dB)，對梅爾頻譜圖進行可視化處理，如圖 1。

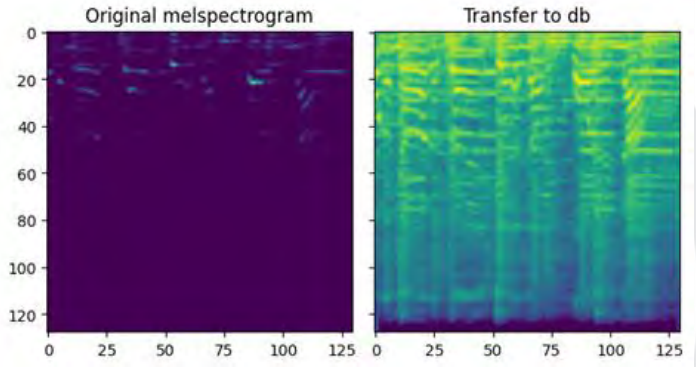

圖 1：經過可視化處理梅爾頻譜圖

(四) 模型訓練與測試

實驗過程將音樂資料訓練集、測試集、驗證集切割方式為 9:1:1，共有 8100 張頻譜圖作為訓練集，450 張作為測試集，450 張作為驗證集，圖片尺寸寬高為 288×432，色彩模式為 RGBA，因此通道數為 4。所以，訓練集、測試集、驗證集的圖形 shape 為（288,432,4），並設定每一個批次（batch）為 128，並不將測試集進行洗牌（shuffle），以利採用測試集繪製混淆矩陣，評估每類別辨識效能。為了比較不同參數設定下模型的效能，本研究使用以下超參數組合：(1) 影像增強資料的採用或不採用；(2) 網路輸出層標準化演算法（Batch Normalization） 的採用或不採用；(3) 不同的學習率（learning rate）之比較(0.00005、0.001、0.01、0.03）；(4)三種權重初始化之比較（RandomNormal = 0.01、glorot uniform、he normal）等。總共創造 48 種模型訓練組合（2 × 2 × 4 × 3 = 48），模型架構為 4 個捲積層，4 個捲積層皆使用 ReLU 激活函數，且分別都接上輸出大小為 2×2 的池化層，最後通過扁平層將資料壓縮為一維特徵張量，並傳入全連接層進行分類，全連接層輸出使用 SoftMax 激活函數，將輸出結果轉為機率形式；損失函數使用 Categorical Cross-Entropy，總迭代次數為 40 代，並且當正確率 Accuracy 超過總迭代的 1/3，即超過 12 次未上升時隨即停止訓練。最後機率數值最大則代表資料所在分類項目，輸出結果為 9 個類別，其模型中的使用資料來源、權重初始化參數，標準化方法、學習率參數皆依照不同的超參數組合變動。

肆、 實驗結果與討論

(一) 模型績效分析

由於模型訓練高達 48 種訓練組合，因此本研究僅保留準確率達 85%以上的模型(使用 Tensor Board)。觀察後發現，準確度較高的模型均有經過資料增強模式的處理，較低的學習率的參數可訓練出準確率較高的模型，但會降低模型收斂速度。避免此問題，應先以較高學習率執行訓練，並於準確率開始降低時再行調降學習率。

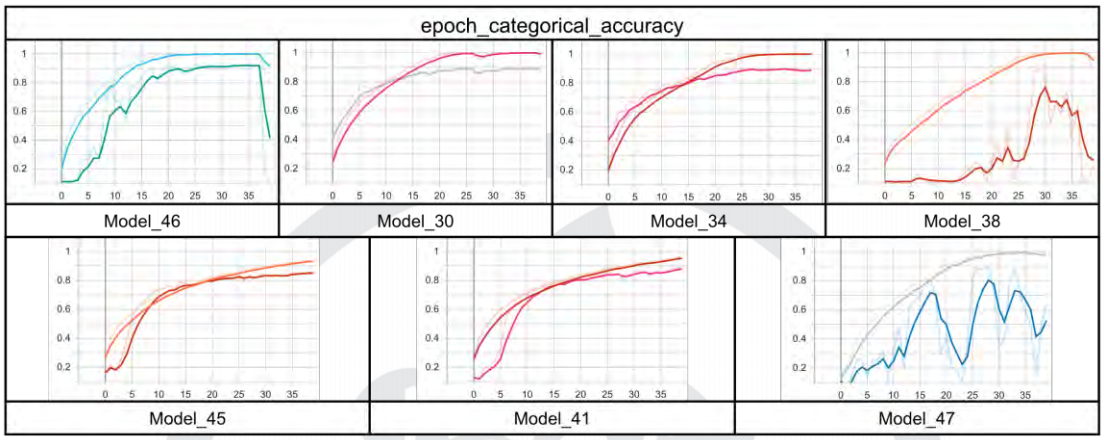

圖 2：模型分類效能圖表

此外，觀察超參數調整結果可發現，這些準確率 85%以上模型大多經過 Batch Normalization 處理，僅 30 及 34 號模型並未使用。觀察圖 2，發現大約於第 25~30 個迭代期間，模型準確率即停止上升；進一步與兩個準確率持續上升的模型（45 及 41 號）相比，訓練資料與驗證資料落差稍偏大。並繼續觀察超參數調整結果，資料增強模式、Batch Normalization、以及學習率等因素之調整，直接影響模型效能，然而權重初始化對模型的影響並不顯著。圖 2 中各模型之排列順序，由高準確率降冪至低準確率，然而準確率高的模型往往於學習階段後期的迭代產生了過度擬合現象，或因學習率較高而導致模型收斂速度加快使得學習波動率很大，起伏不定。另外，基於高準確率的模型效能總體指標，有可能驅使為達到高準確率而影響模型的訓練迭代，進而因為產生過度擬合現象，而導致準確率下降。因此，評估模型效能時，應該同時考量測試資料與實際場域驗證資料的差異性，選用較小損失率（loss rate）的模型，同時評估模型的泛化能力與準確率。綜合上述原因，最後挑選出 45 及 41 號模型作為候選模型。

(二) 混淆矩陣分析

針對所選模型 41 及 45 號，使用混淆矩陣（confusion matrix）進行分類成效的評估與分析，可藉此了解各個類別的分類品質。圖 3 呈現 41 號模型對測試集進行分類的表現，發現 41 號的 Country 風格音樂容易跟 Pop、以及 Rock 風格搞混，該類別之分類準確度為九個類別中最差。圖 4 則呈現 45 號模型對測試集進行分類的表現，從中發現混淆的項目分佈到了其它類別，且混淆的類別更多，與 41 號模型具有 Country、與 Rock 風格容易混淆的共通點。分析該原因，可追溯至 1960~1970 年代，發源於美國西部的鄉村搖滾音樂文化(维基百科编者, 2022)，大多由民間樂

手創作。因為，當時音樂經常會融合民族歷史、以及地方生活習慣等，於是當時的搖滾音樂亦經常結合鄉村音樂風格進行譜曲。綜合上述，得知模型 41 具有較佳的分類表現。

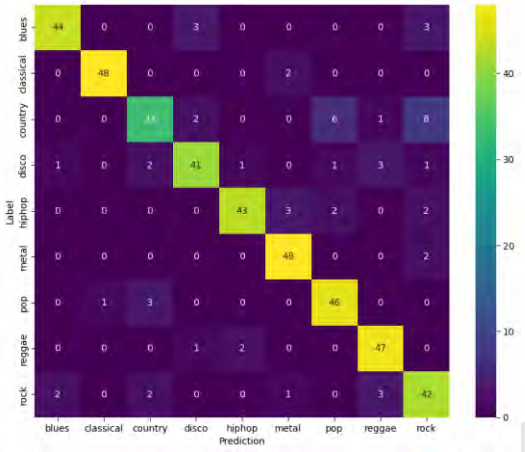

圖 3：模型 41 號混淆矩陣

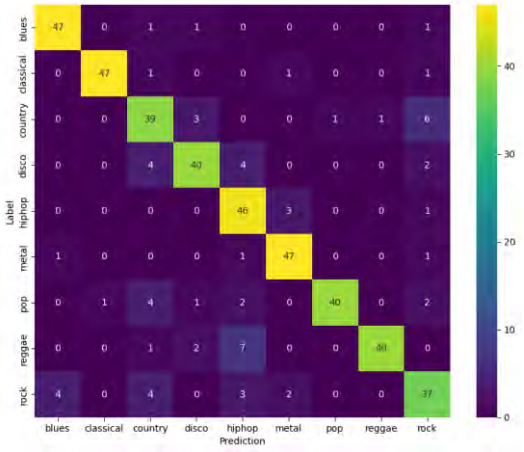

圖 4：模型 45 號混淆矩陣

(三) 網路音樂串流平台之音樂作品驗證分析

為了驗證模型泛化性、及強健性（robustness），本研究運用 Apple Music 音樂串流平台，將未於模型內訓練及定義標籤的新音樂，放入模型中進行分類；每類別共 10 首歌，共九類別，合計測試 90 首音樂，並分析其音樂風格以及分類錯誤的可能原因。資料前處理將歌曲截取 3 秒片段，並使用梅爾頻譜將頻譜圖片標準化，轉換成相同圖形尺寸的 PNG 檔，使用 41 號模型進行預測，並對預測結果進行統計分析。

如圖 5，可發現準確率最高的類別為 Classical，其次是 Metal，表現最差為 Rock 和 Disco 類別。其中，Rock 類別經常與 Country 以及 Blues 搞混，原因是 Rock 類別的音樂所使用的伴奏樂器通常為吉他，但音樂特性又並非如 Metal 重金屬般激動且吵雜；以音樂文化角度分析，Rock 音樂與 Country 音樂確實存在著互相影響的關係。另外，Disco 類別與 Hip-Hop 類別經常發生混淆；因為 Disco 音樂經常採用節奏鮮明的鼓點節奏作為前奏，Hip-Hop 同樣也屬於節奏鮮明的音樂，所以在前奏的預測上會經常與 Hip-Hop 類別發生混淆。

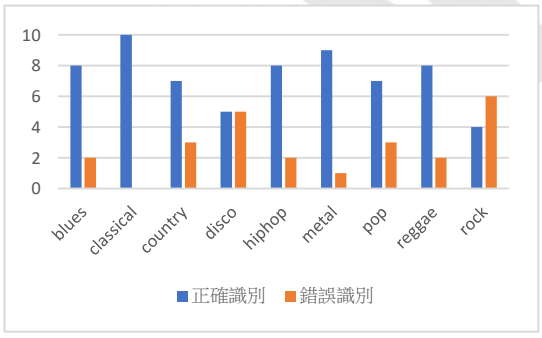

圖 5：模型預測結果統計分析

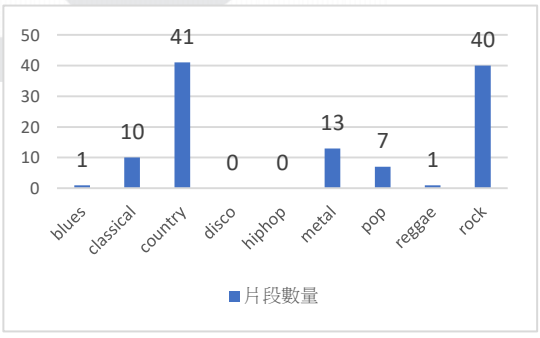

圖 6：全曲片段預測類別彙總

音樂創作靈感皆由心而發，音樂創作形式無須遵循公式，因此創作音樂亦經常融合了多種風格。作曲者個人的背景文化、偏好、與創意，將決定哪些風格的音樂會互相搭配(鄭英傑, 2022; 賴靈恩, 2015)，無固定的形式。由於音樂是一個注重整體感的藝術，當某個風格片段出現的次數明顯突出於其它風格時，便會帶給人就是此風格的感覺。因此，本研究進一步隨機取樣一音樂全曲，

進行 3 秒切片，並將每個切片轉換為頻譜圖，導入模型進行分析，試圖了解類別分布比率。於是，取樣 Country 類別中的第七首作品為例（原預測結果為 Rock 類別），使用此切片方法將原曲長 5 分 36 秒的作品切分為 113 個 3 秒片段，進行預測，結果如圖 6。觀察結果，Country 和 Rock 兩個類別的分布比率相當突出，充分解釋了為何 Country 類別的音樂會讓人們覺得有 Rock 類別的感覺。

總結上述分析，本研究所提出之方法與分類模型，其所驗證之分類結果與人類感受極為相似，顯示所提出之方法的強健性。因此教師或學習者對於實施數位媒體的教學使用，若導入本方法整合網路音樂串流平台，不僅可將大量音樂作品做為教學素材，還可以運用網路音樂串流平台進行創作展演與鑑賞之目的，實現學習者於日常生活中音樂參與的實踐效果。

伍、 結論

網路音樂串流平台的發展，成為現今音樂通路的主要管道，可運用於學校或日常的音樂體驗以培養學生音樂素養。網路音樂串流平台大量風格多元的音樂作品，切乎音樂素養教育之需求。因此，本研究採用卷積神經網路模型（CNN）架構一音樂風格分類機制。透過調節多種組合的模型訓練參數、音頻資料增強模式、以及梅爾頻譜轉換等策略，改善音樂風格分類模型的準確率，強化分類模型之泛化能力，擴展模型在教學實務應用上的音樂風格分類之準確率。實驗結果顯示，透過所提出之模型訓練策略，所建構的音樂風格分類模型，導入網路音樂串流平台之實證結果，具有 87%以上的分類準確率。因此，若將此功能結合於現有網路音樂串流平台，將能協助教師與學生有效地自動篩選合適的音樂作品，融入音樂教學活動。未來研究，將進一步提出增進音樂資訊檢索準確率與效率的改善機制。

參考文獻

Aguiar, R. L., Costa, Y. M. G., & Silla, C. N. (2018, 08-13 July). *Exploring Data Augmentation to Improve Music Genre Classification with ConvNets.* Paper presented at the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil.

Ashraf, M., Abid, F., Din, I. U., Rasheed, J., Yesiltepe, M., Yeo, S. F., & Ersoy, M. T. (2023). A Hybrid CNN and RNN Variant Model for Music Classification. Applied Sciences, 13(3), 1476.

Asmus, E. P. (2021). Motivation in music teaching and learning. *Visions of Research in Music Education, 16*(5), 31.

Bahuleyan, H. (2018). Music genre classification using machine learning techniques. *arXiv preprint arXiv:1804.01149*.

Costa, Y., Oliveira, L., Koerich, A., & Gouyon, F. (2013). *Music genre recognition based on visual features with dynamic ensemble of classifiers selection*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84-90.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., . . . Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60-88.

doi:10.1016/j.media.2017.07.005

poonia, s., verma, c., & Malik, N. (2022). Music Genre Classification using Machine Learning: A Comparative Study. *13*, 15-21.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data, 6*(1), 1-48.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The journal of the acoustical society of america, 8*(3), 185-190.

Sturm, B. L. (2012). *An analysis of the GTZAN music genre dataset.* Paper presented at the Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies.

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing, 10*(5), 293-302.

Wegel, R. L., & Lane, C. E. (1924). The Auditory Masking of One Pure Tone by Another and its Probable Relation to the Dynamics of the Inner Ear. *Physical Review, 23*(2), 266-285. doi:10.1103/PhysRev.23.266

易起宇. 上網日期：民 111 年, 10 月 9 日. Apple Music 歌曲量逾 1 億. *聯合新聞網*. 檢自 https://udn.com/news/story/6811/6672716

陳育恬, & 鄭勝耀. (2022). 運用 YouTube 進行音樂教學之展演實踐. **臺灣教育評論月刊, 11(9)**, 186-192.

维基百科编者. 上網日期：2022, December 13. 鄉村搖滾. *维基百科，自由的百科全書*. 檢自 https://zh.wikipedia.org/w/index.php?title=%E9%84%89%E6%9D%91%E6%90%96%E6%BB%BE&oldid=75036382

鄭英傑. (2022). 書評:[戰場] 轉移—評介《 流行樂, 媒體與青少年文化: 從 [節拍革命] 到 [位元世代]》. **教育研究集刊(68: 3)**, 113-124.

賴靈恩. (2015). 回歸部落之歌: 卑南族下賓朗部落的歌謠分類. **臺灣音樂研究(21)**, 97-135.