

結合影像與非影像資料之多模態架構效能分析-黑色素瘤分類

吳靜雯^{1,2}, 黃怡婷^{1,3}, 黃沁詠^{1,4}, 陳芷翎^{1,5}, 曾明性^{1,6,*}

¹ 中山醫學大學醫學資訊學系

E-mail: ² rita911130@gmail.com, ³ 414tinkerbell@gmail.com,

⁴ Joannaa192024277@gmail.com, ⁵ eileenchen98714@gmail.com, ⁶ mht@csmu.edu.tw

摘要

黑色素瘤是一種由黑色素細胞引發的惡性皮膚腫瘤。為了提升早期診斷的準確性與患者生存率，本研究集中於黑色素瘤的預測方法，結合皮膚鏡影像與非影像資料（如年齡、性別、病變部位及病變特徵如色素網絡、條紋等）進行分類，並評估影像單模態深度學習模型、四種結合影像及非影像的多模態深度學習架構、多模態深度集成學習架構，以及整合遷移學習與機器學習的多模態架構。本研究使用 ISIC2018 資料集，針對黑色素瘤的早期預測進行模型開發與評估[1,2]。本研究結果顯示，結合皮膚鏡影像與非影像資料之多模態深度集成學習架構在黑色素瘤二元疾病分類任務中擁有最佳的分類效能，可有效提升單純使用醫學影像進行疾病診斷的準確性，多模態資料處理架構在醫學資訊領域具有顯著的應用前景。

關鍵字：皮膚鏡影像分類、深度學習模型、機器學習模型、多模態架構、集成學習

Abstract

Malignant melanoma is a type of skin cancer originating from melanocytes. To enhance the accuracy of early diagnosis and improve patient survival rates, this study focuses on predictive methods for melanoma by integrating dermoscopic images and non-image data (such as age, sex, lesion location, and lesion features like pigment network and streaks). The research evaluates single-modality deep learning models for image data, four multimodal deep learning architectures that combine image and non-image data, multimodal deep ensemble learning architectures, and multimodal architectures integrating transfer learning with machine learning. The ISIC2018 dataset was used for model development and evaluation, aiming for early prediction of melanoma [1,2]. The results indicate that the multimodal deep ensemble learning architecture, which integrates

dermoscopic images and non-image data, achieves the best classification performance in the binary classification task of melanoma. This approach effectively improves the diagnostic accuracy compared to using medical images alone. Multimodal data processing architectures show significant application potential in the field of medical informatics.

Keywords : Dermoscopic Image Classification, Deep Learning Models, Machine Learning Models, Multimodal Architectures, Ensemble Learning

1. 研究背景

黑色素瘤是一種高度惡性的皮膚癌，早期發現對患者的生存和治療至關重要。隨著人工智慧、深度學習和機器學習技術的發展，基於影像的自動化診斷工具為早期檢測黑色素瘤提供了新可能性。多模態架構在皮膚鏡影像分類中展現顯著優勢，結合影像與非影像資料，提供更全面和準確的診斷結果。機器學習技術能處理和分析大量結構化數據，補充深度學習在非結構化數據處理方面的能力。通過比較不同深度學習、機器學習模型及多模態架構在皮膚鏡影像分類中的效能，本文希望找出最具效能的分類模型，提高黑色素瘤早期檢測準確性，促進臨床診斷和治療的改善。

2. 研究方法

本研究針對黑色素瘤分類任務，結合皮膚鏡影像與非影像資料進行深入分析，並採用 ISIC2018 所整理後的資料集來評估不同模型與架構的分類效能。經過篩選與整理後，最終資料集中包含 1,784 張標註影像，這些影像分為三類：脂漏性角化症、色素痣與黑色素瘤。其中，脂漏性角化症佔總影像數的 7.12%，色素痣佔 68.94%，而黑色素瘤則佔 24.06%。本研究特別關注黑色素瘤的早

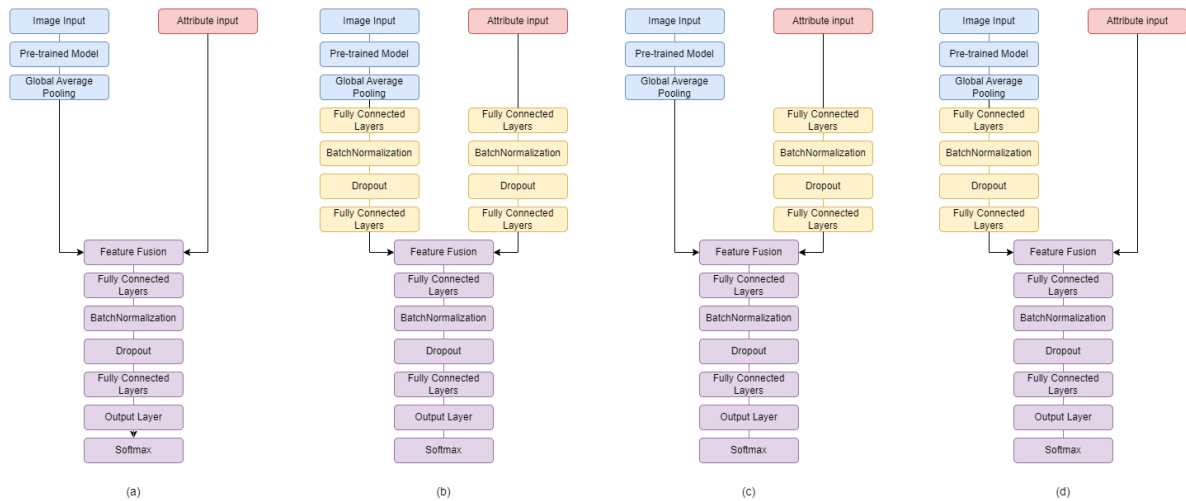


圖 1. 結合影像及非影像的多模深度學習架構

期分類與診斷，並透過比較單模態與多模態深度學習架構，探討結合影像與非影像資料對黑色素瘤分類效能的提升能力。

2.1 影像單模態深度學習模型

針對黑色素瘤皮膚鏡影像應用 Densnet201、EfficientnetV2B2、Resnet50 及 VGG16 等四種深度學習模型為主幹網路進行分類效能評估。

2.2 結合影像及非影像的多模深度學習架構

本階段同樣基於上述四種主幹網路，應用 2024 年 Tsai 學者等人提出四種結合影像及非影像的多模態深度學習架構[3]進行分類效能分析，如圖 1 所示。四種架構分別為：(a)將影像特徵與非影像特徵融合，經過全連接層、批次正規化及 Dropout 層，最後輸出層使用 Softmax 來進行分類。(b)將影像特徵與非影像特徵分別加上全連接層、批次正規化及 Dropout 層，再進行融合。(c)僅在非影像特徵加上全連接層、批次正規化及 Dropout 層。(d)僅在影像特徵加上全連接層、批次正規化及 Dropout 層。上述多模態架構旨在探討影像與非影像資料融合對黑色素瘤分類效能的提升作用，並評估不同特徵處理方式對分類準確性的影響。

2.3 增加性別、年齡及影像拍攝部位的元資料

基於上述四種深度多模態架構取得最佳結果的多模態模型，加入性別、年齡及影像拍攝部位等非影像元資料，再次評估分類效能。此步驟可望提升模型的準確性與泛化能力，使模型不僅依賴影像

特徵進行分類，還能全面考量與黑色素瘤相關的非影像因素，以提高臨床應用價值。

2.4 多模態深度集成學習架構

在多模態深度集成學習架構中，透過集成學習中的軟式投票機制，將四種多模態深度學習模型的預測機率進行加權平均，以提高分類準確度。並進一步調整各模型的權重值，以觀察不同權重組合對分類效能的影響，確定最佳集成分類效能。

2.5 整合遷移學習與機器學習的多模態架構

選擇影像分類表現最佳的深度學習模型，應用遷移學習對皮膚鏡影像進行特徵萃取，再加入非影像資料後，比較 HistGBClassifier、XGBClassifier、GBClassifier 等 12 種機器學習分類器的效能。

3. 研究結果與討論

根據表 1，以 Resnet50 和 VGG16 為主幹網路是單純皮膚鏡影像分類效能最高的單模態模型，測試準確率為 0.742，顯示出在單模態深度學習模型中此兩種模型具有最佳的分類能力。

在深度學習多模態架構中，僅對非影像特徵加全連接層、批次正規化及 Dropout 層（架構 c）取得了最佳結果，其中 VGG16 的測試準確率達到 0.798。進一步增加性別、年齡及影像拍攝部位的元資料後，VGG16 的準確率提高到 0.812，如表 2 所示，顯示非影像元資料在提升模型分類效能方面的重要性。

在多模態深度集成學習架構中，整合 EfficientnetV2B2 和 VGG16 兩種模型的結果最佳，測試準確率提高至 0.821，如表 3 所示。而表 4 顯示了使用 VGG16 整合遷移學習與機器學習的多模態架構的 HistGBClassifier 擁有較佳分類效能，測試準確率為 0.784，但仍低於多模態深度集成式學習模型的最佳結果 0.821。

研究結果顯示，結合影像與非影像資料的多模態深度集成學習架構在黑色素瘤的二元分類任務中表現最佳，顯示多模態資料在提升分類準確性上的潛力。針對資料不平衡問題，本研究未來將參考 Chang 等人（2022）的方法[4]，於 ML 架構中應用 SMOTE 進行過採樣，以平衡正負樣本數量；在 DL 架構中則採用資料擴增技術進行少量類別影像的擴增，以改善模型在少數類別上的學習效果。透過這些處理方法，期望能進一步提升模型的分類效能，使其在臨床應用中具備更強的泛化能力和穩定性。

表 1. 影像單模態深度學習模型分類效能比較

訓練模型	訓練準確率	測試準確率
Densnet201	1.000	0.697
EfficientnetV2B2	1.000	0.697
Resnet50	1.000	0.742
VGG16	1.000	0.742

表 2. 增加性別、年齡及影像拍攝部位的元資料

訓練模型	訓練準確率	測試準確率
Densnet201	1.000	0.784
EfficientnetV2B2	0.845	0.782
Resnet50	1.000	0.768
VGG16	1.000	0.812

表 3. 多模態深度集成學習架構效能評估

模型權重	訓練準確率	測試準確率
[1,1,1,1]	1.000	0.790
[1,1,1,0]	0.882	0.768
[1,1,0,1]	1.000	0.796
[1,1,0,0]	1.000	0.784
[1,0,1,1]	1.000	0.784
[1,0,1,0]	0.954	0.720
[1,0,0,1]	1.000	0.793
[0,1,1,1]	0.882	0.779
[0,1,1,0]	0.688	0.689
[0,1,0,1]	1.000	0.821
[0,0,1,1]	0.996	0.728

表 4.12 種機器學習分類器效能比較

機器學習分類器	訓練準確率	測試準確率
HistGBClassifier	1.000	0.784
XGBClassifier	1.000	0.762
GBClassifier	0.987	0.756
RBF SVM	0.877	0.754
Logistic Regression	0.884	0.751
Random Forest	1.000	0.739
Neural Net	1.000	0.725
DTBagging	0.978	0.720
Linear SVM	0.926	0.714
Nearest Neighbors	0.802	0.686
Decision Tree	1.000	0.639
Naive Bayes	0.629	0.569

4. 參考文獻

[1] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, Allan Halpern: "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)", 2018; <https://arxiv.org/abs/1902.03368>

[2] Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 doi:10.1038/sdata.2018.161 (2018).

[3] Tsai, H.-Y.; Tseng, M.-H. Evaluating Multimodal Techniques for Predicting Visibility in the Atmosphere Using Satellite Images and Environmental Data. Electronics 2024, 13, 2585.

[4] Chang, C.-C.; Li, Y.-Z.; Wu, H.-C.; Tseng, M.-H. Melanoma Detection Using XGB Classifier Combined with Feature Extraction and K-Means SMOTE Techniques. Diagnostics 2022, 12, 1747.