DOI: 10.6717/JTMRT.202212_10(2).0009

運用深度學習針對牙科環景全口 X 光影像進行品質管控之研究

蘇心悅[1,5] 魏雅雯[2] 王友立[3] 黃詠暉[4] 陳泰賓[4]

國軍花蓮總醫院 放射科[1]
高雄長庚紀念醫院 牙科部[2]
國軍花蓮總醫院 門急診部[3]
義守大學 醫學影像暨放射科學系[4]
花蓮縣醫事放射師公會[5]

摘要

臨床上常利用環景全口 X 光(panoramic x-ray, PANO)對病人的齒槽狀況進行評估。一張擺位正確的 PANO 影像，可提供醫師完善的判讀資訊，故在影像產生後傳送至醫師前，應進行品質控制(quality control, QC)以確保影像符合診斷標準。然而，過去對影像進行 QC 的方式，通常是由放射師進行目視檢查，此種仰賴雙眼的品管方法可能存在缺陷，如：1.目視檢查取決於放射師對影像判讀的感知與經驗，影響一致性。2.工作負擔增加造成疲勞感增加，易使判斷能力下降。3.大量影像進行品管較費時。本研究自院內醫學影像儲存通訊系統(Picture archiving and communication system, PACS)選取 PANO 影像 553 張，使用 7 種卷積神經網路(convolutional neural network, CNN)模型：alexnet 、googlenet、mobilenet、resnet50、resnet101、vgg16 及 vgg19，並分別結合三種機器學習分類器：邏輯迴歸(logistic regression)、貝氏分類器(naïve Bayes classifier)、支持向量機(support vector machine, SVM)，共 21 種組合分類模型，利用人工智慧(artificial intelligence, AI)進行自動辨識影像中常見的各種擺位錯誤，並比較各模型分類結果找出最佳分類模型。模型效能評比項目包含準確度(accuracy)及一致性係數(kappa)。實驗結果證明，resnet101 結合 logistic regression 分類器模型之 Kappa 平均值為 0.211，為最佳分類模型。

關鍵字：PANO、CNN、QC

前言

臨床上經常利用環景全口 X 光(panoramic x-ray, PANO)對病人的齒槽狀況進行評估(圖 1)，包括：齒槽骨高度是否正常、牙齒數目是否正確、牙齒角度有無位移或傾倒、根尖病變、齲齒、假牙、阻生齒及牙周病等。為避免因影像品質不佳影響診斷或治療之正確性，在影像送達醫師工作站電腦前，應使影像通過品質管控(quality control, QC)，確保其符合鑑別診斷標準，亦可減少潛在醫療糾紛之風險[1-4]。

人工智慧(artificial intelligence, AI)於醫學影像的應用正在廣泛的被研究中，AI 在識別醫學影像異常方面，表現出高度的準確度和靈敏度，並可望強化基於組織的檢測和表徵[5]。運用 AI 進行 QC 在各領域中已有先例，如使用 Google Cloud 的視覺檢測 AI (visual inspection AI)於製造業檢查工作、焊接接縫檢查、行動電話 PCB 檢查和矽晶圓不良原因分析。過去對 X 光影像進行

通訊作者：陳泰賓
通訊地址：82445 高雄縣燕巢鄉角宿村義大路 8 號　　義守大學醫學影像暨放射科學系
聯絡電話：(07) 6151100
E-mail: chen.tb@gmail.com
2022 年 11 月受理，2022 年 12 月接受刊登

QC 的方式，通常是由放射師進行目視檢查。然而，目視檢查是一種高度人工作業的過程，非常耗時且容易出錯，此種仰賴雙眼的品管方法可能存在缺陷，如：1.目視檢查取決於放射師對影像判讀的感知與經驗，影響一致性。2.工作負擔增加造成疲勞感增加，易使判斷能力下降。3.大量影像進行品管較費時。而將 AI 應用於醫學影像 QC 有以下優點：1.對於影像評核標準一致，亦可透過反覆訓練增加或強化辨識效度。2.減少放射師工作負荷。3.幾秒鐘內檢測出影像品質是否達診斷標準。綜上所述，使用 AI 進行 QC 可以輔助臨床的放射師提高判斷影像品質的精確度，並加速判斷流程。

本研究針對環景全口 X 光影像，使用深度學習中的捲積類神經網路(convolutional neural network, CNN)演算法擷取影像特徵[6-8]，並透過監督式機器學習 SVM 分類模型建構影像自動辨識暨分類 AI 模型；並從中選取分類結果最佳之模型，藉此找出符合醫師診斷標準之影像及未達診斷標準之不良影像 (不良片)，降低人眼分辨錯誤率及減少臨床人力資源，提升檢查品質，提供醫師判讀時的完善資訊。

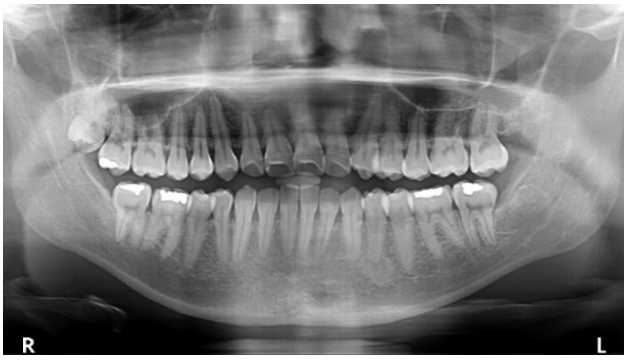

圖 1：PANO 影像

材料與方法

研究流程

本研究之研究流程如下：收集臨床 PANO 影像作為輸入來源，以監督式學習方式訓練神經網路，辨識影像中是否包含下列臨床常見的擺位錯誤：(i)患者前傾(slumped position)、(ii)下巴低傾(chin tipped low)、(iii)張開嘴唇(open lip)、(iv)頭部轉向一側(head turned to one side)、(v)頭部傾斜向一側(head tilted to one side)及(vi)舌頭抵住上顎(tongue against the palate)，共 6 項。

樣本收集

影像收集時間為 2017 年起至 2021 年止，自本院醫學影像儲存通訊系統(Picture archiving and communication system, PACS)選取 PANO 影像 553 張。本院使用之 PANO 造影儀廠牌為 Carestream，型號 CS9300，造影時間為 4-16 秒，影像像素為 64 × 1536 pixels，照野為 5 × 149mm。

影像前處理

本研究使用之 PANO 影像原始大小為 800 × 600之 JPG 檔，影像前處理共包含二步驟：(i)影像資料集各類別 80%作為訓練集、20%作為測試集(train ratio: 0.8)及 90%作為訓練集、10%作為測試集(train ratio: 0.9)(表 1)，準備進行影像深度學習之用。(ii)影像轉換後將影像邊緣不必要之資訊去除(如病人資訊)，再將影像大小縮小為180 × 120以降低運算時間。

表 1：影像樣本及劃分資料表。影像樣本共 553 張，以 train ratio 0.8 及 0.9 將影像資料集各類別分為訓練集和測試集

類別	樣本數	train ratio	訓練集	測試集
PANO	553	0.8	442	111
		0.9	498	55

CNN 模型及參數

本研究利用 MatlabR2018b 建立下列 7 種 CNN 模型，包含：alexnet、googlenet、mobilenet、resnet101、resnet50、vgg16、vgg19。使用 3 種分類器：logistic regression、naivebayes、svm 進行分類。樣本數大小(batch size)分別設定 3、6、9。

評估模型之方法

本研究使用準確度(accuracy, ACC)及一致性係數(Kappa 值)之平均值進行不同模型及不同 train ratio 間之評估。

準確度

準確度是診斷工具效能的一種參考指標，本

研究中準確度是指影像中真陽性加真陰性所佔的比例，計算公式如下：

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad (式 1)$$

真陽性(true positive, TP)：表示在包含膽囊的影像中，辨識結果也是陽性。
真陰性(true negative, TN)：表示在未包含膽囊的影像中，辨識結果也是陰性。
偽陽性(false positive, FP)：表示在未包含膽囊的影像中，辨識結果卻是陽性。
偽陰性(false negative, FN)：表示包含膽囊的影像中，辨識結果卻是陰性。

一致性係數(Kappa)

Kappa 檢驗是評分者實際評定一致的次數百分比與評分者理論上評定最大可能次數百分比的比率，經常用在下列兩種情況中：一是評價待評價的診斷實驗方法與金標準的一致性；另一種是評價兩種檢查方法對同一樣本的檢查結果的一致性，其公式如下：

$$K = \frac{P_o - P_c}{1 - P_c} \qquad (式 2)$$

P_o：觀測一致性(observed agreement)：前後(兩種)測量結果一致的百分比。
P_c：期望一致性(chance agreement)：前後(兩種)測量結果預期相同的機率。

Kappa 值可分為 5 組來表示不同等級的吻合度，如表 2。

結果

本研究透過不同訓練比例、不同 CNN 方法提取影像特徵、不同分類器進行影像辨識，探討下列問題：不同訓練比例對分類結果之影響、不同 CNN 方法結合 logistic regression 分類之結果、不同 CNN 方法結合 naïve Bayes 分類之結果、不同 CNN 方法結合 SVM 分類之結果以及最佳 CNN 結合分類器之模型，茲分列如後：

表 2：Kappa 值與吻合度之關聯表

Kappa 值	吻合度
0.0-0.20	極低的吻合度
0.21-0.40	一般的吻合度
0.41-0.60	中等的吻合度
0.61-0.80	高度的吻合度
0.81-1	幾乎完全吻合

不同訓練比例對分類結果之影響

針對不訓練比例對分類結果之影響如表 3.1。不出意外，當 90% (0.9)資料進行訓練時，其 Kappa 值和分類準確性(ACC = Accuracy)的平均分別為 0.152 及 0.168；二者比率所得之 Kappa 與 ACC 值之最小與最大差不多。

不同訓練比例的 Kappa 值直方圖如圖 3.1；當訓練比例為 0.9 時，其 Kappa 值超過 0.15 的次數高於訓練比例為 0.8。不同訓練比例的 ACC 值直方圖如圖 3.2，當訓練比例為 0.9 時，其 ACC 值超過 0.6 的次數高於訓練比例為 0.8。

不同 CNN 方法結合邏輯迴歸(logistic regression)分類之結果

不同 CNN 方法結合邏輯迴歸分類之 ACC 及 Kappa 值統計如表 3.1 及表 3.2，CNN 包括 alexnet、googlenet、mobilenet、resnet10、resnet50、vgg16、vgg19。在 80%比例所得訓練模型，以 vgg16 之平均 ACC 值為 0.175 最佳、resnet101 之平均 Kappa 值為 0.211 最佳；在 90%比例所得訓練模型，以 resnet101 之平均 ACC 值為 0.189 最佳、vgg16 之平均 Kappa 值為 0.202 最佳。

表 3. 1：不同訓練比例的分類準確性及 Kappa 值

Index	Split Ratio	Mean	Std	Minimum	Maximum
ACC	.8	.101	.213	.000	.975
	.9	.168	.259	.000	1.000
Kappa	.8	.150	.049	.066	.242
	.9	.152	.044	.056	.242

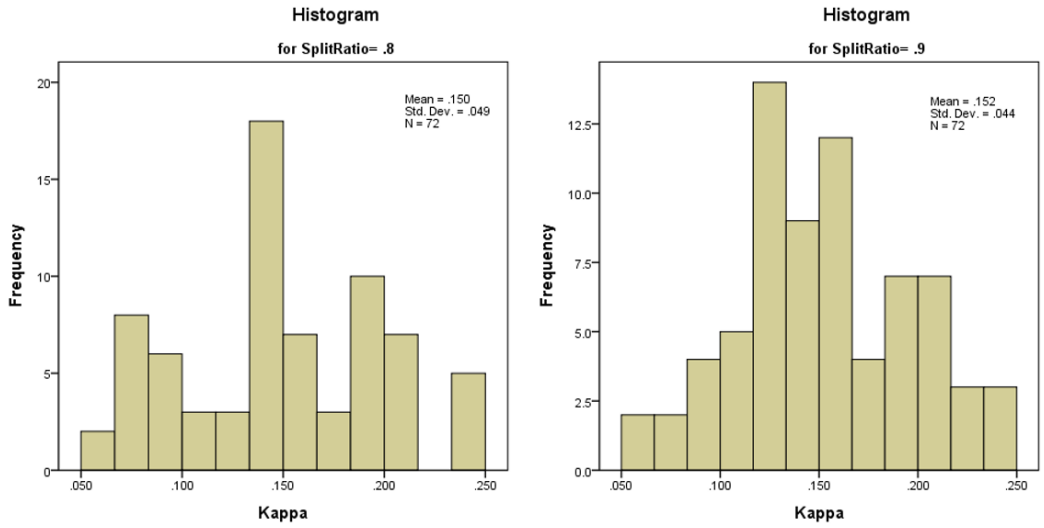

圖 3.1：不同訓練比例的 Kappa 值直方圖

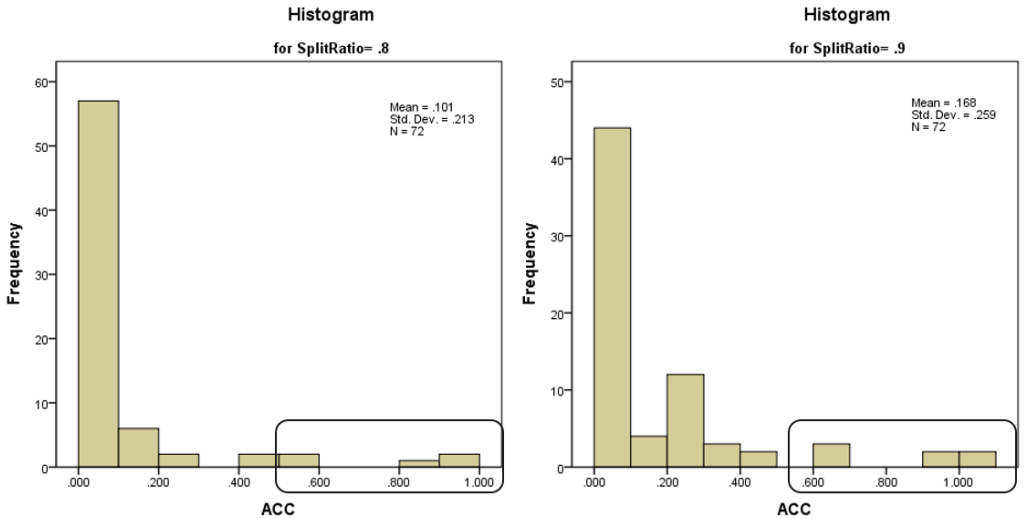

圖 3.2：不同訓練比例的 ACC 值直方圖

表 3.2：不同 CNN 方法結合邏輯迴歸(logistic regression)分類 ACC 值統計

Split Ratio	CNN	Mean	Std	Minimum	Maximum
	alexnet	.171	.007	.163	.177
	googlenet	.167	.015	.143	.187
	mobilenet	.134	.017	.120	.153
.8	resnet101	.170	.013	.157	.183
	resnet50	.162	.011	.150	.170
	vgg16	**.175**	.011	.167	.187
	vgg19	.166	.016	.150	.193
	Alexnet	.184	.027	.153	.200
	googlenet	.172	.024	.127	.200
	mobilenet	.158	.014	.147	.173
.9	**resnet101**	**.189**	.028	.167	.220
	resnet50	.165	.004	.160	.167
	vgg16	.162	.010	.153	.173
	vgg19	.169	.021	.153	.193

表 3.3：不同 CNN 方法結合邏輯迴歸(logistic regression)分類 Kappa 值統計

Split Ratio	CNN	Mean	Std	Minimum	Maximum
.8	alexnet	.205	.058	.139	.242
	googlenet	.179	.045	.116	.237
	mobilenet	.151	.058	.096	.212
	resnet101	**.211**	.028	.187	.242
	resnet50	.194	.039	.162	.237
	vgg16	.177	.059	.109	.215
	vgg19	.180	.016	.162	.194
.9	Alexnet	.138	.018	.121	.157
	googlenet	.184	.039	.146	.242
	mobilenet	.175	.025	.146	.192
	resnet101	.152	.018	.131	.162
	resnet50	.179	.042	.152	.227
	vgg16	**.202**	.066	.126	.242
	vgg19	.160	.050	.126	.217

不同 CNN 方法結合 naïve Bayes 分類之結果

不同 CNN 方法結合 naïve Bayes 分類之 ACC 及 Kappa 值統計如表 3.4 及表 3.5。在 80%比例所得訓練模型，以 alexnet 之平均 ACC 值為 0.192 最佳、vgg16 之平均 Kappa 值為 0.017 最佳；在 90%比例所得訓練模型，以 mobilenet 之平均 ACC 值為 0.196 最佳、mobilenet 之平均 Kappa 值為 0.035 最佳。

不同 CNN 方法結合 SVM 分類之結果

不同 CNN 方法結合 SVM 分類之 ACC 及 Kappa 值統計如表 3.6 及表 3.7。在 80%比例所得訓練模型，以 alexnet 之平均 ACC 值為 0.130、平均 Kappa 值為-0.44 皆為最佳；在 90%比例所得訓練模型，以 alexnet 之平均 ACC 值為 0.153 最佳、Kappa 值為-0.16 皆為最佳。

最佳 CNN 結合分類器之模型

綜上所述，當訓練比例為 0.8 時，以 resnet101 結合 logistic regression 之分類模型，其平均 Kappa 值 0.211 為所有組合中表現最佳。

表 3.4：不同 CNN 方法結合 naïve Bayes 分類 ACC 值統計

Split Ratio	CNN	Mean	Std	Minimum	Maximum
.8	**alexnet**	**.192**	.022	.170	.213
	googlenet	.183	.012	.173	.207
	mobilenet	.160	.026	.130	.180
	resnet101	.173	.015	.160	.190
	resnet50	.170	.036	.130	.200
	vgg16	.181	.003	.177	.183
	vgg19	.164	.030	.133	.193
.9	alexnet	.189	.017	.173	.207
	googlenet	.190	.024	.153	.213
	mobilenet	**.196**	.027	.180	.227
	resnet101	.176	.034	.140	.207
	resnet50	.178	.027	.147	.193
	vgg16	.173	.000	.173	.173
	vgg19	.164	.008	.160	.173

表 3.5：不同 CNN 方法結合 naïve Bayes 分類 Kappa 值統計

Split Ratio	CNN	Mean	Std	Minimum	Maximum
.8	alexnet	.031	.026	.004	.056
	googlenet	.019	.015	.008	.048
	mobilenet	-.008	.032	-.044	.016
	resnet101	.008	.018	-.008	.028
	resnet50	.004	.043	-.044	.040
	vgg16	**.017**	.005	.012	.020
	vgg19	-.003	.036	-.040	.032
.9	alexnet	.027	.020	.008	.048
	googlenet	.028	.029	-.016	.056
	mobilenet	**.035**	.032	.016	.072
	resnet101	.011	.040	-.032	.048
	resnet50	.013	.032	-.024	.032
	vgg16	.008	.000	.008	.008
	vgg19	-.003	.009	-.008	.008

表 3.6：不同 CNN 方法結合 SVM 分類 Acc 值統計

Split Ratio	CNN	Mean	Std	Minimum	Maximum
.8	**alexnet**	**.130**	.018	.110	.143
	googlenet	.087	.040	.047	.163
	mobilenet	.057	.006	.050	.060
	resnet101	.064	.002	.063	.067
	resnet50	.066	.002	.063	.067
	vgg16	.071	.009	.063	.080
	vgg19	.061	.013	.047	.070
.9	**alexnet**	**.153**	.014	.140	.167
	googlenet	.122	.047	.060	.187
	mobilenet	.073	.024	.047	.093
	resnet101	.069	.003	.067	.073
	resnet50	.051	.023	.027	.073
	vgg16	.087	.024	.067	.113
	vgg19	.071	.010	.060	.080

表 3.7：不同 CNN 方法結合 SVM 分類 Kappa 值統計

Split Ratio	CNN	Mean	Std	Minimum	Maximum
.8	**alexnet**	**-.044**	.021	-.068	-.028
	googlenet	-.096	.048	-.144	-.004
	mobilenet	-.132	.007	-.140	-.128
	resnet101	-.123	.002	-.124	-.120
	resnet50	-.121	.002	-.124	-.120
	vgg16	-.115	.010	-.124	-.104
	vgg19	-.127	.015	-.144	-.116
.9	**alexnet**	**-.016**	.016	-.032	.000
	googlenet	-.053	.057	-.128	.024
	mobilenet	-.112	.029	-.144	-.088
	resnet101	-.117	.005	-.120	-.112
	resnet50	-.139	.028	-.168	-.112
	vgg16	-.096	.029	-.120	-.064
	vgg19	-.115	.012	-.128	-.104

討論

依本論文之研究結果，顯示以下三個發現：第一，在所有模型與分類器組合中，訓練比例為 0.9 時，使用 mobilenet 模型結合 naïve Bayes 分類器，可達本次研究之最高辨識精確度(19.6%)；訓練比例為 0.8 時，resnet101 結合 logistic regression 分類器可達本次研究之最高 Kappa 值(0.211)。第二，train ratio 為 0.9(資料 90%為訓練集、10%為測試集)時，模型之分類表現大部分皆優於 train ratio 為 0.8，說明可能因 train ratio 不同而得到不同的分類結果，應嘗試多種劃分方式對初始數據集進行分割，如進行多次隨機劃分、重複試驗或取平均值，作為評估結果。第三，本研究所得之眾多結果中，在各種不同模型與分類器結合皆有其長處，並無最優之唯一組合，說明單一模型或分類器的辨識效果有限，本研究未來可朝融合 AI (Fusion AI)或強 AI (Strong AI)方向進行，結合二種或以上 AI 模型，可望更進一步提升辨識精度。

結論

本研究發現 resnet101 模型結合 logistic regression 分類器，針對 PANO 影像擺位錯誤之分類效度最佳。

本研究之目的，為自動辨識 PANO 影像中常見的各種擺位錯誤，提供 QC 人員一個速度快、效果佳且易於使用的方法。經本研究結果表明，resnet101 模型結合 logistic regression 分類器能達到一般的吻合度。未來可朝融合 AI (Fusion AI)或強 AI (Strong AI)方向進行，結合二種或以上 AI 模型，可望更進一步提升辨識精度。

致謝

感謝國軍花蓮總醫院 111 年醫療事業基金醫學研究發展計畫(計畫編號：HAFGH-D-111007)支持。

參考文獻

1. Wenzel A, Matzen LH, Spin-Neto R, Schropp L. Effect of computer-assisted-learning and simulation clinics on dental students' cognitive and performance skills: panoramic image errors related to patient's head position. Dentomaxillofac Radiol. 2020; 49(7):20200154.
2. Khator AM, Motwani MB, Choudhary AB. A study for determination of various positioning errors in digital panoramic radiography for evaluation of diagnostic image quality. Indian J Dent Res. 2017;28(6):666-670.
3. Dhillon M, Raju SM, Verma S, Tomar D, Mohan RS, Lakhanpal M, Krishnamoorthy B. Positioning errors and quality assessment in panoramic radiography. Imaging Sci Dent. 2012; 42(4):207-12.
4. Kitai N, Murabayashi M, Sugimoto H, Fujiwara A, Tome W, Katsumata A. Accuracy and head positioning effects on measurements of anterior tooth length using 3-dimensional and conventional dental panoramic radiography. Am J Orthod Dentofacial Orthop. 2017; 151(3): 607-615.
5. Kim HE, Kim HH, Han BK, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. Lancet Digital Health. 2020; 2: e138-e148.
6. Kunihiko Fukushima. 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36: 193-202.
7. LeCun Yann, Bottou Leon, Bengio Yoshua, and Haffner Patrick. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86:2278-324.
8. LeCun Yann, Bengio Yoshua, and Hinton Geoffrey. 2015. Deep learning. Nature 521: 436-44.

Quality Control of PANO Images by Using Deep Learning Methods

Hsin-Yueh Su[1,5], Wei Ya Wen[2], Yu-Li Wang[3], Yung-Hui Huang[4], Tai-Been Chen[4*]

Department of Radiology, Hualien Armed Forces General Hospital[1]
Department of Dentistry, Kaoshiung Chang Gung Memorial Hospital[2]
Department of Outpatient and Emergency, Hualien Armed Forces General Hospital[3]
Department of Medical Imaging and Radiological Sciences, I-Shou University[4]
Hualien Hsien Association of Radiological Techologists[5]

Abstract

In clinical practice, panoramic x-ray (PANO) is often used to evaluate the alveolar condition of patients. A correctly positioned PANO image can provide physicians with complete interpretation information. Therefore, quality control (QC) should be performed to ensure that the images meet the diagnostic criteria before being transmitted to physicians. However, in the past, the way of QC of images was usually visual inspection by radiologists. This quality control method relying on both eyes may have defects, such as 1. Visual inspection depends on the radiologist's perception and experience of image interpretation, Affect consistency. 2. The increased workload results in increased fatigue, which efficiently reduces the ability to judge. 3. Quality control of a large number of images is time-consuming. In this study, 553 PANO images were selected from the in-hospital medical image storage and communication system (PACS), and seven convolutional neural networks (CNN) models were used: alexnet, googlenet, mobilenet, resnet50, resnet101, vgg16 and vgg19, and combine three machine learning classifiers respectively: logistic regression (logistic regression), Bayes classifier (naïve Bayes classifier), support vector machine (support vector machine, SVM), a total of 21 combined classification models, using Artificial intelligence (AI) automatically identifies various common placement errors in images, and compares the classification results of each model to find the best classification model. Model performance evaluation items include accuracy (accuracy) and consistency coefficient (Kappa). The experimental results show that the average Kappa of resnet101 combined with the logistic regression classifier model is 0.211, which is the best classification model.

Keywords: PANO, CNN, QC

Corresponding Author: Tai-Been Chen
Department of Medical Imaging and Radiological Sciences, I-Shou University
Address: No. 8, Yida Rd., Jiaosu Village Yanchao District,Kaohsiung City 82445,Taiwan, R.O.C.
Tel: +886-7- 6151100
E-mail: chen.tb@gmail.com