applied sciences

MDPI

Article

The Learning Curve of Artificial Intelligence for Dental Implant Treatment Planning: A Descriptive Study

Pathompong Roongruangsilp and Pathawee Khongkhunthian *

Center of Excellence for Dental Implantology, Faculty of Dentistry, Chiang Mai University, Chiang Mai 50200, Thailand; pathompong_r@cmu.ac.th
* Correspondence: pathawee.k@cmu.ac.th

Featured Application: Authors are encouraged to provide a concise description of the specific application or a potential application of the work. This section is not mandatory.

Abstract: Introduction: Cone-beam computed tomography (CBCT) has been applied to implant dentistry. The increasing use of this technology produces a critical number of images that can be used for training artificial intelligence (AI). Objectives: To investigate the learning curve of the developed AI for dental implant planning in the posterior maxillary region. Methods: A total of 184 CBCT image sets of patients receiving posterior maxillary implants were processed with software (DentiPlan Pro version 3.7; NECTEC, NSTDA, Thailand) to acquire 316 implant position images. The planning software image interfaces were anonymously captured with full-screen resolution. Three hundred images were randomly sorted to create six data sets, including 1–50, 1–100, 1–150, 1–200, 1–250, and 1–300. The data sets were used to develop AI for dental implant planning through the IBM PowerAI Vision platform (IBM Thailand Co., Ltd., Bangkok, Thailand) by using a faster R-CNN algorithm. Four data augmentation algorithms, including blur, sharpen, color, and noise, were also integrated to observe the improvement of the model. After the testing process with 16 images that were not included in the training set, the recorded data were analyzed for detection and accuracy to generate the learning curve of the model. Results: The learning curve revealed some similar patterns. The curve trend of the original and blurred augmented models was in a similar pattern in the panoramic image. In the last training set, the blurred augmented model improved the detection by 12.50%, but showed less accuracy than the original model by 18.34%, whereas the other three augmented models had different patterns. They were continuously increasing in both detection and accuracy. However, their detection dropped in the last training set. The colored augmented model demonstrated the best improvement with 40% for the panoramic image and 18.59% for the cross-sectional image. Conclusions: Within the limitation of the study, it may be concluded that the number of images used in AI development is positively related to the AI interpretation. The data augmentation techniques to improve the ability of AI are still questionable.

Keywords: artificial intelligence; deep learning; dental implant

check for updates

Citation: Roongruangsilp, P.; Khongkhunthian, P. The Learning Curve of Artificial Intelligence for Dental Implant Treatment Planning: A Descriptive Study. *Appl. Sci.* **2021**, *11*, 10159. https://doi.org/10.3390/app112110159

Academic Editor: Gabi Chaushu

Received: 11 October 2021
Accepted: 28 October 2021
Published: 29 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Bone quantity evaluation in dental implantology is crucial. It is an essential step for dental implant treatment planning. The result of faulty consideration during pre-surgical planning often leads to implant-related complications. The success of dental implant treatment is principally related to bone quality of patients, patient evaluation, and good treatment planning [1].

The precise measuring of bone architecture with conventional 2D radiography is limited by difficulty in assessing hard tissue morphology and both the quality and quantity of bone. There is little information about the buccolingual and cross-sectional dimension, resulting in inadequate identification of critical structures and restriction of spatial data for vital structures [2].

Cone-beam computed tomography (CBCT) provides a unique imaging analysis of proposed implant sites by reformatting the image data to create several imaging modalities. The specific software used for creating the images utilizes multiple tools, which can precisely mark the vital structures and provide a 1:1 image for accurate measurements. The utility of CBCT for dental implant treatment planning has been well investigated [2–4]. However, analyzing CBCT scans requires a specific level of training and expertise. This analysis is time-consuming, involving hundreds of images.

Learning is an essential part of the human brain, which creates intelligent behavior. Artificial intelligence (AI) is defined as the branch of computer science focused on a simulation of human intelligence adopted by machines based on the information they collected [5]. These principles relate to the algorithms used in knowledge representation and their implementation.

Computer vision is a field of AI that deals with the automation of the tasks mimicking the human visual system, which can be performed by enabling computers to derive significant information from digital visual inputs and take actions based on that information [6]. One of the most prominent applications in medical computer vision is the IBM Watson AI platform in the field of oncology. It has been deployed by cancer centers to help medical professionals quickly diagnose and determine the best treatments for patients [7].

Among AI technologies, the IBM PowerAI Vision platform offers a built-in deep learning algorithm that learns to analyze images. It includes a graphical user interface (GUI) to label objects in the images, which can be used to train and validate a model. The IBM PowerAI Vision platform provides a faster R-CNN algorithm as an image object detection model. The faster R-CNN algorithm has been optimized for accuracy [8–10]. It was first used as a proposal generator. Then, the feature vectors of generated proposals were encoded by deep convolutional neural networks, followed by making the object predictions. The developed model was then deployed for further application [11].

For limited data acquisition, the IBM PowerAI Vision platform also provides a data augmentation algorithm that processes the transformations of images, such as blur, sharpen, and rotate, to create new versions of existing images to increase the number of data samples for satisfactory AI performance.

With the continued adoption of both CBCT and AI, interpreting a large number of images is another area in which AI can promote benefaction. There are many developers focusing on AI that can automatically assess whole images of CBCT scans and propose the most potential diagnosis by reviewing similar images stored in databases.

These processes of interpretation can be performed by AI far faster than human technicians. Implant treatment planning in the posterior maxillary region is preferable for early AI development, because the implant's angulation is perpendicular to the occlusal plane and the vital structures, such as the maxillary sinus, are clearly visible.

Therefore, the purpose of this study was to investigate the learning curve of the developed AI for dental implant planning in the posterior maxillary region by using the images provided from CBCT.

2. Materials and Methods

This study was approved by the Human Experimentation Committee, Faculty of Dentistry, Chiang Mai University, Chiang Mai, Thailand, with the certification of ethical clearance No. 30/2019.

All of the 316 images used in this study were created from "Digital Imaging and Communications in Medicine" (DICOM) obtained from 184 CBCT scans of patients receiving treatment at the Center of Excellence for Dental Implantology, Faculty of Dentistry, Chiang Mai University, for the period from 2013 to 2018.

The selection criteria of CBCT scans were based on the missing posterior maxillary teeth, which included the teeth from the first premolar to the second molar. The DICOM considered must be displayed in the properly formed dental arches without discontinuities in the hard tissues and without artefacts.

One well-experienced dentist in digital dental implant treatment planning and AI was assigned to perform the study. DICOM reconstruction was performed in DentiPlan Pro version 3.7 software (DTP; NECTEC, NSTDA, Bangkok, Thailand) to create input images, which consisted of two types of images:

1. The panoramic images are created by the imaginary panoramic line, which is drawn in the axial view from the left condyle connecting through the center of each tooth until reaching the right condyle.

2. The cross-sectional images are then generated in the desired implant position. If there was more than one implant to be placed in a patient, every cross-sectional image in each position of implants was created. For example, if the patient needed three implants, three cross-sectional images in each position of implants were created.

A total of 316 images were then anonymously captured to ensure that no private information was revealed. The full-screenshot images were taken from DTP software with the joint photographic expert group (JPEG) format at 1920 × 1080 pixels with 24-bit depth (Figure 1). Three hundred images were used for model training, which were randomly sorted from 1–300 to create six data sets, including 1–50, 1–100, 1–150, 1–200, 1–250, and 1–300. These six data sets were called the training set. The remaining 16 images were selected for accuracy testing to assess the performance of the developed model; this set of 16 images was called the testing set.



Figure 1. The full screen image capture of DentiPlan Pro version 3.7.

The six original training sets were separately uploaded into the IBM PowerAI Vision platform (IBM Thailand Co., Ltd., Bangkok, Thailand), then the labelling process for model training was applied within the GUI of the platform. This user-defined data is significant for describing the desired output data.

The labelled area in each image was generated in a square shape by connecting the four lines from the farthest border of the alveolar bone available for implant placement. The mentioned four lines included the upper border, lower border, mesial border, and distal border in a panoramic image, together with the upper border, lower border, buccal border, and lingual border in a cross-sectional image. The demonstration of the labelling process is displayed in Figures 2 and 3a.



Figure 2. Image labelling in both cross-sectional and panoramic images.

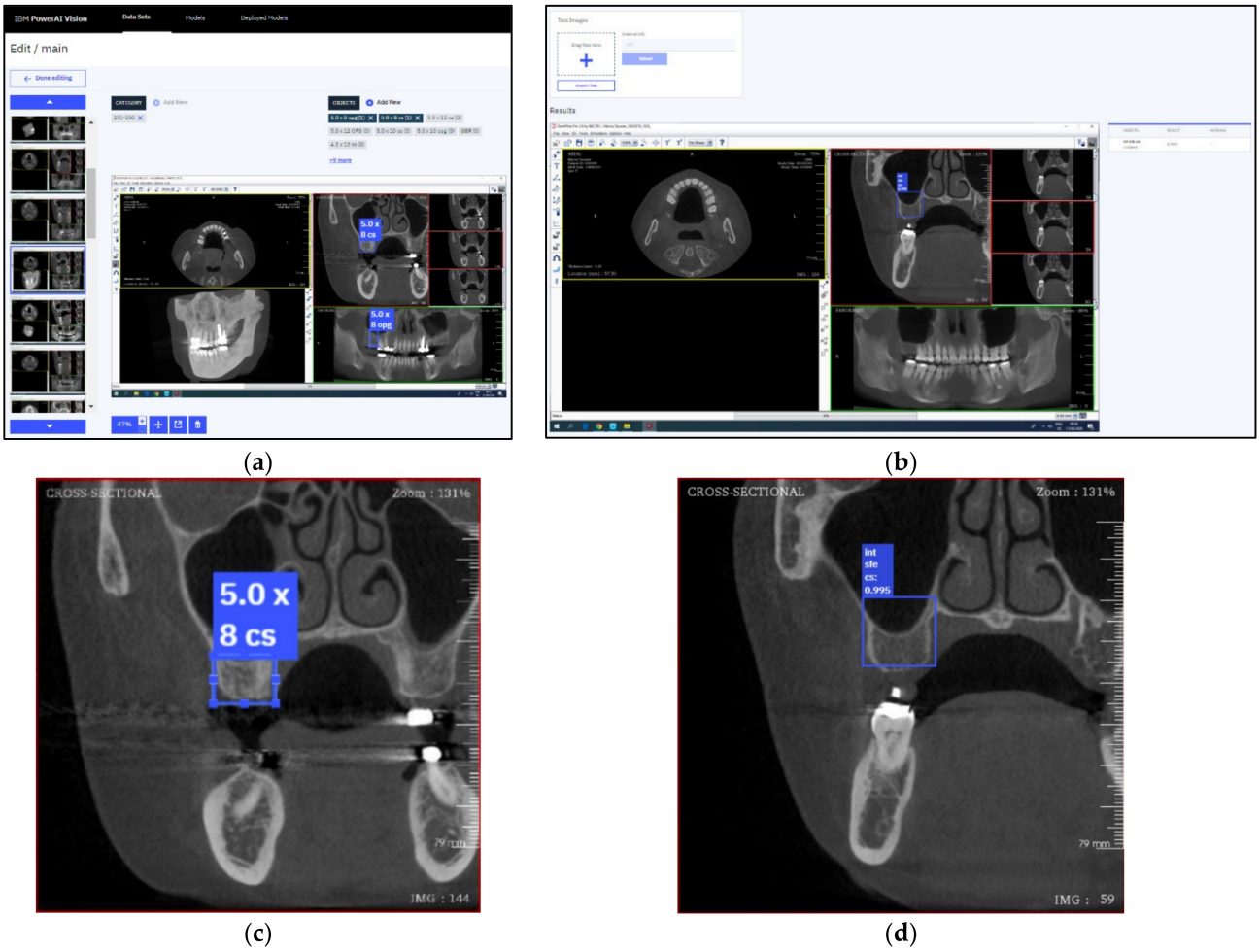


Figure 3. GUI of the IBM PowerAI Vision platform during the labelling process (**a**), during the testing process (**b**), the magnification of cross-sectional image during the labelling process (**c**), and the magnification of cross-sectional image during the testing process (**d**).

Since this was an early step of AI development for dental implant treatment planning. A dental implant system with its available 3-dimensional stereolithography file (.STL) was used in this study (NOVEM dental implant system, Novem Innovations Co., Ltd., Chiang Mai, Thailand; Figure 4). The available diameter and length, which will be used in the posterior maxilla are as follows:

- Implant diameter: 4.2 and 5.0 mm.
- Implant length: 8, 10 and 12 mm.

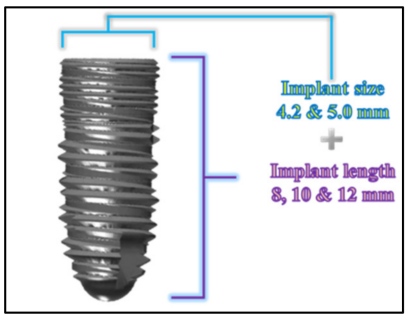


Figure 4. NOVEM dental implant system.

The implant selection was dependent upon the available space, which included bone width and bone height. Each labelled area was then annotated with the specific implant size and technique used. The criteria for annotation were described as follows:

Criteria for implant diameter selection in the mesiodistal dimension:

- In case of one implant placement: root space = 2x (T-I dist.) + Implant diameter.
- In case of two or more implant placements: root space = T-I dist. + implant diameter + 3 + implant diameter + 3 + . . . + implant diameter + T-I dist.

Note: T-I dist. refers to the distance between implant and tooth or the border of labelled space at marginal bone, which was 1.5 mm.

Criteria for implant diameter selection in the buccolingual dimension:

- Bone width = Buccal thickness + Implant diameter + Lingual thickness.

Note: The minimal buccal and lingual thicknesses needed are 1 mm.

Criteria for implant length selection in the coronoapical dimension:

- Bone height $\geq$ Implant length.

Note: If there is any exposure of the available shortest length implant out of the alveolar ridge, this condition is indicated as "Int SFE (internal sinus floor elevation)" or "Lat SFE (lateral sinus floor elevation)" depending on the residual bone height. If the residual bone height was $\geq$5 mm, the Int SFE is indicated, but if the residual bone height was <5 mm, the Lat SFE is indicated. When Int SFE or Lat SFE were indicated, it was unnecessary to indicate the implant diameter.

Subsequently, each labelled area was then annotated with eight different sets, including 4.2 × 8, 4.2 × 10, 4.2 × 12, 5.0 × 8, 5.0 × 10, 5.0 × 12, Int SFE, and Lat SFE for each panoramic and cross-sectional image. The annotations for both panoramic and cross-sectional images were independent because of their own characteristics, and they were indicated with the corresponding labelled box area (as displayed in Figure 3a).

When all the images of the six original training sets were annotated, they were then continuously supplemented by using a data augmentation algorithm within the IBM PowerAI Vision platform. These features were composed of blur, sharpen, color, crop, vertical flip, horizontal flip, rotate, and noise. Data augmentation created the new data set that contained all the existing images, plus the newly generated images. In this study context, which focused on the posterior maxillary teeth, four image processing features, including blur, sharpen, color, and noise, were used separately in each former data set to create new data sets. Each form of data augmentation was described as follows:

Blur image processing. This study selected the maximum amount of Gaussian and motion blur. Gaussian blur made the entire image appear out of focus by reducing detail and noise. Motion blur made the image appear as if it was in motion. Ten new images were generated in the range of each former data set, which included 1–50, 1–100, 1–150, 1–200, 1–250, and 1–300. The former data sets were supplemented to 1–550, 1–1100, 1–1650, 1–2200, 1–2750, and 1–3300, respectively.

Sharpen image processing. This study selected the maximum amount of sharpening. Five new images were generated in the range of each former data set. The former data sets were supplemented to 1–300, 1–600, 1–900, 1–1200, 1–1500, and 1–1800, respectively.

Color image processing. This study selected the maximum amount of change in the brightness, contrast, hue, and saturation of the images. Five new images were generated by using randomly selected values in the selected ranges by the algorithm itself. The former data sets were supplemented to 1–300, 1–600, 1–900, 1–1200, 1–1500, and 1–1800, respectively.

Noise image processing. This study selected the maximum amount of noise to add to the new images. The IBM PowerAI Vision platform determined the need for a reasonable amount of noise for the images to remain usable. If 100% is selected, none of the generated images will have 100% noise added. Instead, the output images may possibly have the maximum amount of noise added while still remaining usable. Five new images were generated with noise added in the range of each former data set. The former data sets were supplemented to 1–300, 1–600, 1–900, 1–1200, 1–1500, and 1–1800, respectively.

Thereafter, the six original training sets and the new 24 augmented training sets were then utilized to train the AI. The faster R-CNN algorithm was used to develop the AI. The model hyperparameters were configured with the default setting. After 30 training sets were completely trained, the trained model was then deployed and evaluated with the 16 prerequisite images in the testing set which included 2 images of every picture that must be indicated to 4.2 × 8, 4.2 × 10, 4.2 × 12, 5.0 × 8, 5.0 × 10, 5.0 × 12, Int SFE, and Lat SFE. Each testing image was annotated as the ground truth. The testing process required uploading each image one at a time, and the deployed model then displayed the detection with bounding boxes, annotation, and confidence percentage. Each testing image that was uploaded into the deployed model was not included in the development of the model. An example of the testing process is demonstrated in Figure 3b. If the labelled object in the images was indicated with more than an 80% confidence threshold, the annotation results were then recorded throughout the study. The recorded data were analyzed for detection and accuracy. Detection was defined as the event resulting in object detection, and accuracy was defined as the matching outcome between the annotation of the resulting object detection and the actual outcome that was interpreted by the human developer.

For better understanding, the equations of detection were written as "Detection = NoE/NoT", and the equations for accuracy were written as "Accuracy = NoM/NoE". The full terms of each acronym were defined as follows: NoE was used to represent the "number of events of object detection that occurred during the test process", NoT was used to represent "number of the total images used in the test process", and NoM was used to represent "number of matches between the annotation outcome from the model and the actual outcome from the human developer". These two numerations were performed in the event that they occurred in both the panoramic and cross-sectional images, and the final calculated data were then incorporated into the learning curve of the model.

The actual detections during the test were analyzed in the confusion matrix table. These tables included both rightness and mismatching of implant selection, as well as the technique used in various kinds of developed models.

3. Results

The distribution of 300 images used in the training process is displayed in Table 1. The performance of the deployed model is demonstrated in Tables 2–5. Figures 5 and 6 provide the confusion matrix, which demonstrates the actual detection throughout the test, both correct and incorrect detections.

Table 1. The distribution of each image's categories.

Image Categories	Number of Images
4.2 × 8	13
4.2 × 10	10
4.2 × 12	68
5.0 × 8	23
5.0 × 10	20
5.0 × 12	47
Int SFE	55
Lat SFE	64

Table 2. The panoramic detection of each model.

Number of Images	Original Model [a]	Blurred Augmented Model [a]	Sharpened Augmented Model [a]	Colored Augmented Model [a]	Noise Augmented Model [a]
50	12.50% (2/16)	12.50% (2/16)	18.75% (3/16)	12.50% (2/16)	12.50% (2/16)
100	37.50% (6/16)	37.50% (6/16)	37.50% (6/16)	43.75% (7/16)	37.50% (6/16)
150	62.50% (10/16)	68.75% (11/16)	68.75% (11/16)	43.75% (7/16)	50.00% (8/16)
200	56.25% (9/16)	62.50% (10/16)	68.75% (11/16)	62.50% (10/16)	43.75% (7/16)
250	50.00% (8/16)	62.50% (10/16)	81.25% (13/16)	56.25% (9/16)	68.75% (11/16)
300	62.50% (10/16)	75.00% (12/16)	43.75% (7/16)	31.25% (5/16)	50.00% (8/16)

[a] The number in parenthesis of the detection column explains the actual number from the detection equation, which was NoE/NoT.

Table 3. The cross-sectional detection of each model.

Number of Images	Original Model [a]	Blurred Augmented Model [a]	Sharpened Augmented Model [a]	Colored Augmented Model [a]	Noise Augmented Model [a]
50	37.50% (6/16)	56.25% (9/16)	50.00% (8/16)	31.25% (5/16)	56.25% (9/16)
100	50.00% (8/16)	56.25% (9/16)	68.75% (11/16)	87.50% (14/16)	68.75% (11/16)
150	81.25% (13/16)	81.25% (13/16)	75.00% (12/16)	56.25% (9/16)	68.75% (11/16)
200	62.50% (10/16)	62.50% (10/16)	87.50% (14/16)	75.00% (12/16)	75.00% (12/16)
250	75.00% (12/16)	75.00% (12/16)	68.75% (11/16)	87.50% (14/16)	68.75% (11/16)
300	62.50% (10/16)	75.00% (12/16)	62.50% (10/16)	56.25% (9/16)	50.00% (8/16)

[a] The number in parenthesis of the detection column explains the actual number from the detection equation, which was NoE/NoT.

Table 4. The panoramic accuracy of each model.

Number of Images	Original Model [b]	Blurred Augmented Model [b]	Sharpened Augmented Model [b]	Colored Augmented Model [b]	Noise Augmented Model [b]
50	100% (2/2)	100% (2/2)	100% (3/3)	100% (2/2)	50.00% (1/2)
100	83.33% (5/6)	83.33% (5/6)	66.67% (4/6)	85.71% (6/7)	83.33% (5/6)
150	70.00% (7/10)	81.81% (9/11)	72.72% (8/11)	71.42% (5/7)	75.00% (6/8)
200	66.67% (6/9)	80.00% (8/10)	81.81% (9/11)	80.00% (8/10)	57.14% (4/7)
250	62.50% (5/8)	70.00% (7/10)	61.53% (8/13)	77.78% (7/9)	54.54% (6/11)
300	60.00% (6/10)	41.66% (5/12)	71.42% (5/7)	100% (5/5)	62.50% (5/8)

[b] The number in parenthesis of the accuracy column explains the actual number from the accuracy equation, which was NoM/NoE.

Table 5. The cross-sectional accuracy of each model.

Number of Images	Original Model [b]	Blurred Augmented Model [b]	Sharpened Augmented Model [b]	Colored Augmented Model [b]	Noise Augmented Model [b]
50	50.00% (3/6)	44.45% (4/9)	25.00% (2/8)	40.00% (2/5)	44.45% (4/9)
100	62.50% (5/8)	55.56% (5/9)	54.54% (6/11)	64.28% (9/14)	63.63% (7/11)
150	69.23% (9/13)	76.92% (10/13)	75.00% (9/12)	88.89% (8/9)	81.81% (9/11)
200	70.00% (7/10)	80.00% (8/10)	64.28% (9/14)	83.33% (10/12)	75.00% (9/12)
250	58.33% (7/12)	83.33% (10/12)	81.81% (9/11)	78.57% (11/14)	81.81% (9/11)
300	70.00% (7/10)	50.00% (6/12)	80.00% (8/10)	88.89% (8/9)	87.50% (7/8)

[b] The number in parenthesis of the accuracy column explains the actual number from the accuracy equation, which was NoM/NoE.

		Ground truth								
		Lat SFE	Int SFE	4.2x8	5.0x8	4.2x10	5.0x10	4.2x12	5.0x12	Total
actual detection from original model	Lat SFE	4	2							6
	Int SFE	1	3							4
	4.2x8									0
	5.0x8				5					5
	4.2x10			2				5		7
	5.0x10				3		5			8
	4.2x12							7		7
	5.0x12							1	7	8
actual detection from blurred augmented	Lat SFE	5	2							7
	Int SFE	1	4							5
	4.2x8		1							1
	5.0x8				8		1			9
	4.2x10						1	4		5
	5.0x10				5		5			10
	4.2x12							5		5
	5.0x12								9	9
actual detection from sharpened augmented	Lat SFE	9	2							11
	Int SFE		3							3
	4.2x8					1				1
	5.0x8				7					7
	4.2x10			1			1	6		8
	5.0x10				3		5			8
	4.2x12							8		8
	5.0x12								5	5
actual detection from coloured augmented	Lat SFE	6	1							7
	Int SFE		2							2
	4.2x8		1			1				2
	5.0x8		1		3					4
	4.2x10							4		4
	5.0x10						7			7
	4.2x12							7		7
	5.0x12								8	8
actual detection from noise augmented model	Lat SFE	6								6
	Int SFE	4	1							5
	4.2x8									0
	5.0x8				4					4
	4.2x10							7		7
	5.0x10				2	1	4		1	8
	4.2x12							8		8
	5.0x12								4	4
	Total	36	23	3	40	3	29	62	34	230

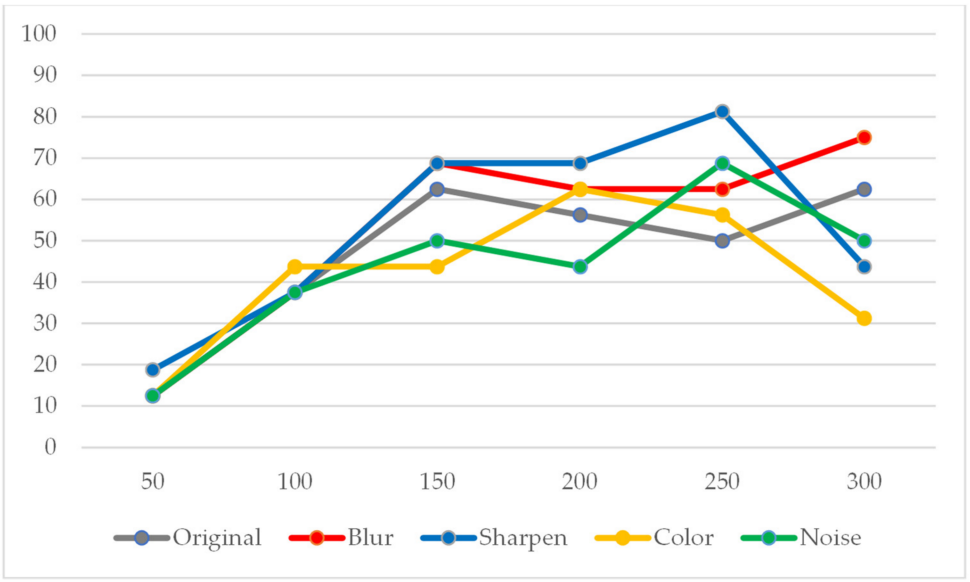
Figure 5. The confusion matrix table between each detection in each model for the panoramic images.

		Ground truth								
		Lat SFE	Int SFE	4.2x8	5.0x8	4.2x10	5.0x10	4.2x12	5.0x12	Total
actual detection from original model	Lat SFE	7	2							9
	Int SFE	2	7							9
	4.2x8		2		1	1				4
	5.0x8		2	1	3					6
	4.2x10			1			2	4		7
	5.0x10		1		3	1	4			9
	4.2x12							7	1	8
	5.0x12							1	9	10
actual detection from blurred augmented	Lat SFE	6	1							7
	Int SFE	1	10							11
	4.2x8		4							4
	5.0x8		1	1	5					7
	4.2x10			1			1	5		7
	5.0x10					2	4			6
	4.2x12							10	4	14
	5.0x12							1	8	9
actual detection from sharpened augmented	Lat SFE	7	4							11
	Int SFE		9							9
	4.2x8	4								4
	5.0x8	3			6					9
	4.2x10		1	1			1	5		8
	5.0x10				2	1	3			6
	4.2x12							9	2	11
	5.0x12							1	9	10
actual detection from coloured augmented	Lat SFE	9								9
	Int SFE	1	7							8
	4.2x8	3			1	1				5
	5.0x8		2		5					7
	4.2x10							4		4
	5.0x10				1		7			8
	4.2x12							10	1	11
	5.0x12							1	10	11
actual detection from noise augmented model	Lat SFE	8	1							9
	Int SFE		10							10
	4.2x8		5							5
	5.0x8		2		4					6
	4.2x10						2	4		6
	5.0x10				1		4			5
	4.2x12							10	2	12
	5.0x12							2	10	12
	Total	51	71	5	32	6	28	74	56	323

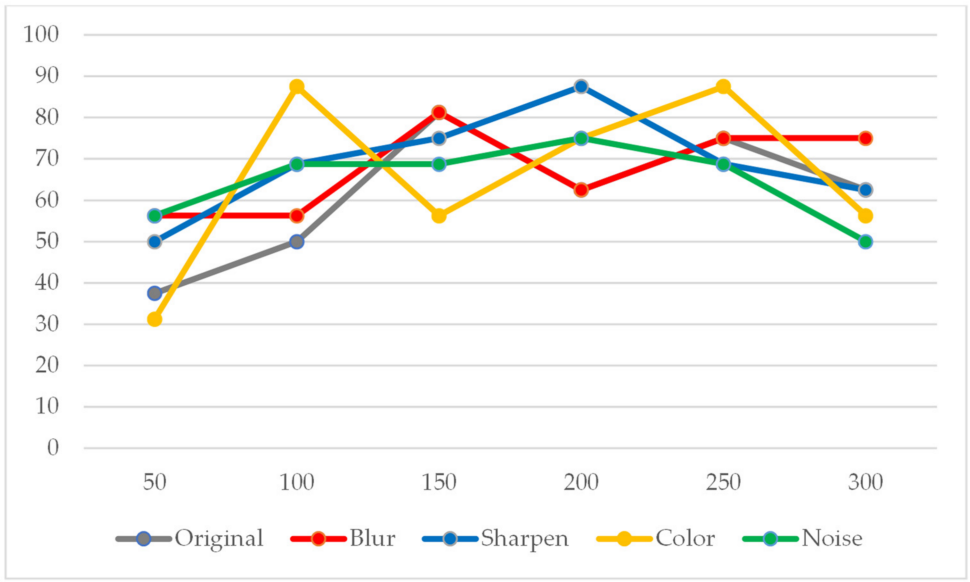
Figure 6. The confusion matrix table between each detection in each model for the cross-sectional images.

The performance of the detection in all models was incorporated into the learning curve, which are displayed in Figure 7a,b. The learning curve of the original model

demonstrated interesting events. The detection curves of the cross-sectional and panoramic images were continuously increasing until they reached the highest point at the 150 data set. At that point in learning, the detection of the panoramic image was 62.50% while that of the cross-sectional image was 81.25%. Afterwards, both detection curves swung around 12.50% for the panoramic image and 18.75% for the cross-sectional image. Both detection curves reached their final results at the value (62.50%). The learning curve of detection was varied in all the data augmented models.



a



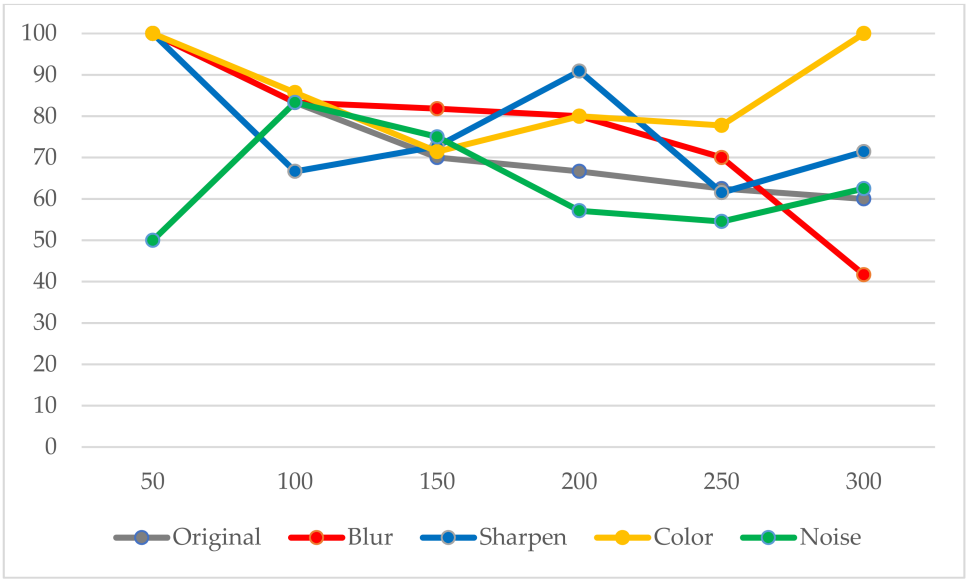
b

Figure 7. The learning curves of panoramic detection (**a**) and cross-sectional detection (**b**) of each trained model.

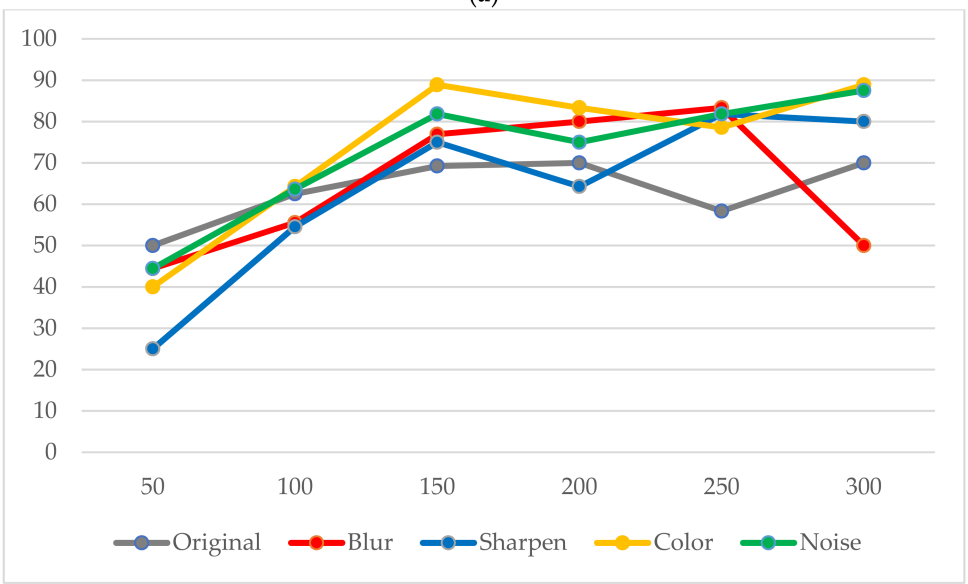
In the blurred, sharpened, colored, and noised augmented models, the detection curve of the panoramic images and cross-sectional images were also illustrated in Figure 7a,b. Overall, the learning curve of detection revealed some similar patterns. In panoramic detection, the curve trend of the original and blurred augmented models was in a similar pattern, but the blurred augmented model improved the performance by 12.50% in the last

data set. Moreover, the other three augmented models were continuously increasing and decreased at the last set of training. In the cross-sectional detection, the curve trends of the blurred and colored augmented models were in a similar pattern since they swung heavily.

On the other hand, the performances of the accuracy testing in all models were incorporated into the learning curve, which are displayed in Figure 8a,b. The learning curve of the accuracy curve in the original model also pointed out multiple informative events. The accuracy curve of the panoramic image started at 100% and continuously decreased throughout the training process. Conversely, the accuracy curve of the cross-sectional image initiated at 50.00% and continuously increased until it reached 69.23% at the 150 data set. Both accuracy curves exhibited an interesting swing after the 150 data set was trained, their swings were 10.00–11.67%. The learning curve of accuracy was improved significantly in all the data augmented models, which represented the different characters in each image processing used.



(**a**)



(**b**)

Figure 8. The learning curves of panoramic accuracy (**a**) and cross-sectional accuracy (**b**) of each trained model.

In the blurred, sharpened, colored, and noised augmented models, the accuracy curve of the panoramic images and cross-sectional images were also illustrated in Figure 8a,b. Eventually, the learning curve of accuracy revealed some similar patterns. In panoramic accuracy, the curve trends of the original and blurred augmented models were in a similar pattern, but the blurred model showed less performance than the original model by 18.34% in the last data set. Furthermore, the other three augmented models had different patterns. They were all trending upward in the final training set. In cross-sectional detection, the curve trends of the original, sharpen, color, and noise models were in a similar pattern, since they were trending upward, while the blurred augmented model was decreased greatly in the final data set.

4. Discussion

This study attempted to develop a novel approach for implant treatment planning by utilizing an available AI platform with a faster R-CNN algorithm. The annotations of the model outcomes were the main focus in this study, since the detected bounding boxes were still not precise.

The learning curve of detection rate and accuracy were almost similar in the original model. Since they were still swinging after the 150 data set was trained, their swings were around 10–20%, which were varied between each image. Based on the learning curve patterns of faster R-CNN, it could assume that the larger data sets are needed to see the further trend for the learning curve that can determine the minimum effective data set needed.

This study also integrated a data augmentation algorithm to overcome the limitation of data acquisition. The augmented models were improved differently. The detection curve in the blurred augmented model demonstrated the highest detection, as the model was improved by 12.50% for both the panoramic and cross-sectional images in the last training set, with an overall direction that was trending upward. On the other hand, the accuracy curves were totally different. The accuracy curves in the blurred augmented model were reduced around 20% in both the panoramic and cross-sectional images in the last training set. These results implied that the data augmentation algorithm with blurred image transformation made the outline of the interesting area seem widen, which made the images easier for detection, but with those blurry images, it was difficult to classify the corrected annotation. Moreover, the study of image classification in chest radiographs also reported that the Gaussian blurred augmented model showed 78.00–82.00% accuracy, which was not improved when compared to the non-augmented model (81.00–83.00%) [12]. Besides, the study comparing data augmentation strategies for image classification reported that Gaussian distortion was the worst form of augmentation tested, leading to changes in the accuracy of only +0.05% [13].

In contrast, the other three augmented models, including sharpen, color, and noise, were likely to improve, but the detection dropped on the last training set. The results were even worse than the original model. This phenomenon is still unclear, more data set for training may reveal the reason for this detection dropped. However, the accuracy showed some improvement, and their trends were also upward. The colored augmented model demonstrated the best improvement with 40% for the panoramic image and 18.59% for the cross-sectional image. From these results, color modification seemed to be the better alternative data augmentation algorithm. Furthermore, an interesting issue about color transformation outcomes was published in a study of data augmentation with skin lesion analysis. Scenario C (saturation, contrast, brightness, and hue) resulted in better accuracy than scenario B (saturation, contrast, and brightness), followed by scenario A (no augmentation) in the DenseNet-161 and Inception-v4 models [14].

Since accuracy is the most important issue in dental implantology, the error analysis was utilized by the confusion matrix to evaluate false positives and false negatives. When the models were tested with the images that should be detected with the Int SFE, the original model had 28% misdetection. On the other hand, the results of the blurred and

sharpened augmented models showed a slightly improved error of 20.00–27.27%. Likewise, the error that occurred in the colored augmented model was greatly improved at 13.33%. Conversely, the result in the noise augmented model appeared different, the error increased at 42.10%. These controversial outcomes could not summarize that data augmentation has minimized the error incidence.

As the cut-off point between Int SFE, Lat SFE, and simple implant placement with 4.2 × 8 or 5.0 × 8 is very close, these infer that the decision-making in the posterior maxillary region is critical. The misdetection of cross-sectional images in the original model from these images was around 60%. However, the total misdetection in all augmented models was 50.61%. This caused us to conclude that adding one type of image for training would not obviously improve object detection performance for that one type of image classification. The result may differ for each type of data augmentation technique used. Adding all kinds of images may improve the overall performance of the object detection model.

The authors speculated that the positive predictive value for the 4.2 × 12, 5.0 × 12, Int SFE, and Lat SFE images for the cross-sectional and panoramic images were approximately 77.78–90.00% and 66.67–100%, respectively. These images were loaded around 47–68 images for each image category, which were quite predictable when compared to the wide range of 0–100% in 4.2 × 8, 5.0 × 8, 4.2 × 10, and 5.0 × 10 images that had only 13–23 images. So, it could be predicted that, when the images are packed with at least 50 images, the trained model will respond to at least 70.00%, with regard to positive prediction. Moreover, the balance of data distribution had a great impact on the model's performance.

Faster R-CNNs are the preferred solutions for imaging analysis and have been employed in many fields [8–10,15,16]. Some successful applications in which this model has been applied in the field of dentistry are discussed in the paragraphs that follow.

A dental-related faster R-CNN model about automatic detection of periodontally compromised teeth in the panoramic image model reported 81.00% accuracy when 100 digital panoramic images were trained [17]. Likewise, this study reported almost the same 83.33% for 100 panoramic images trained in the original model, which was similar to our findings.

Moreover, the study of a faster R-CNN model associated with a rule-based module for tooth detection and numbering system also reported 91.00% accuracy with an enormously large number (800) of periapical radiographs were trained [18]. In our study, 300 images trained in the original model reported only 60.00–70.00% accuracy. However, the closest samples were observed in our data augmented model. The blurred augmented data of 100 images became 1100 images, which resulted in 83.33% accuracy in the panoramic image. Moreover, 150 images of the colored and noise augmented data became 900 images, and the accuracy of the cross-sectional image for both the colored and noise augmented data were 88.89% and 81.81%, respectively. Overall, their data set was about 800 images, which was greater than our 300 original images. The heterogeneity in their data was also higher than that in our data, which probably meant their model was developed from various kinds of images. It may be assumed that the development of the model should be focused on the great amount of original data rather than applied as a data augmentation technique to overcome the identical quantity of the data set. However, the development of the model may benefit from both data augmentation and addition of new images. Nevertheless, these different outcomes may denote that, with different types of images used in the training process, such as periapical radiographs, panoramic images, and cross-sectional images, the individual outcome may not be compared.

This study was unique from the input data used in the training process. The input images were obtained from the actual patients. Each DICOM resulted in one image, which made each image have its own characteristic. Besides, the results were also calculated from the testing images created by real humans as the ground truth, which did not even include the training process. Nevertheless, the limitation is the process of JPEG image extraction from the DICOM was still performed manually, which makes it difficult to obtain the same scale between each image. Although there is a digital ruler attached within the GUI, it

would be easier and faster for the AI to learn the scale-adjusted image. Further studies are required with more data acquisition from the automated DICOM reconstruction.

5. Conclusions

Within the limitations of the study, it may be concluded that the number of each image category used in AI development is positively related to the AI interpretation. Fifty images are the minimum image requirement for over 70% positive prediction. Data augmentation techniques to improve the ability of AI are still questionable.

Author Contributions: Conceptualization, Supervision, and Funding acquisition, P.K.; Methodology, Software, Validation, Investigation, Formal analysis, Visualization, Writing—original draft, Writing—review & editing, P.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was approved by the Human Experimentation Committee, Faculty of Dentistry, Chiang Mai University, Chiang Mai, Thailand, with the certification of ethical clearance No. 30/2019.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: We would like to gratefully thank INET—Internet Thailand Public Co., Ltd. and Thailand's National Electronics and Computer Technology Centre (NECTEC) for their support in terms of the IBM PowerAI Vision platform and DentiPlan Pro version 3.7 software used in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goodacre, C.J.; Kan, J.Y.; Rungcharassaeng, K. Clinical complications of osseointegrated implants. *J. Prosthet. Dent.* **1999**, *81*, 537–552. [CrossRef]
2. Nagarajan, A.; Perumalsamy, R.; Thyagarajan, R.; Namasivayam, A. Diagnostic imaging for dental implant therapy. *J. Clin. Imaging Sci.* **2014**, *4*, 4. [CrossRef]
3. Choi, Y.J.; Jun, S.H.; Song, Y.D.; Chang, M.W.; Kwon, J.J. CT scanning and dental implant. In *CT Scanning-Techniques and Applications*; IntechOpen: London, UK, 2011.
4. Venkatesh, E.; Elluru, S.V. Cone beam computed tomography: Basics and applications in dentistry. *J. Istanb. Univ. Fac. Dent.* **2017**, *51*, S102–S121. [CrossRef] [PubMed]
5. Luger, G.F. *Artificial Intelligence: Structures and Strategies for Complex. Problem Solving*, 5th ed.; Pearson Education Limited: Boston, MA, USA, 2005; ISBN-13 978-0-321-54589-3.
6. Forsyth, D.; Ponce, J. *Computer Vision: A Modern Approach*, 2nd ed.; Prentice Hall: Hoboken, NJ, USA, 2011; ISBN-13 978-0136085928.
7. Zhou, N.; Zhang, C.T.; Lv, H.Y.; Hao, C.X.; Li, T.J.; Zhu, J.J.; Zhu, H.; Jiang, M.; Liu, K.W.; Hou, H.L.; et al. Concordance study between IBM Watson for Oncology and clinical practice for patients with cancer in China. *Oncologist* **2019**, *24*, 812–819. [CrossRef] [PubMed]
8. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
9. Alganci, U.; Soydas, M.; Sertel, E. Comparative research on deep learning approaches for airplane detection from very high-resolution satellite images. *Remote Sens.* **2020**, *12*, 458. [CrossRef]
10. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1137–1149. [CrossRef]
11. Luo, S.; Kindratenko, V. Hands-on with IBM Visual Insights. *Comput. Sci. Eng.* **2020**, *22*, 108–112. [CrossRef]
12. Ogawa, R.; Kido, T.; Kido, T.; Mochizuki, T. Effect of augmented datasets on deep convolutional neural networks applied to chest radiographs. *Clin. Radiol.* **2019**, *74*, 697–701. [CrossRef] [PubMed]
13. O'Gara, S.; McGuinness, K. Comparing data augmentation strategies for deep image classification. In Proceedings of the Irish Machine Vision and Image Processing Conference (IMVIP), Dublin, Ireland, 28–30 August 2019.
14. Perez, F.; Vasconcelos, C.; Avila, S.; Valle, E. Data augmentation for skin lesion analysis. In *OR 20 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*; Stoyanov, D., Taylor, Z., Sarikaya, D., McLeod, J., González Ballester, M.A., Codella, N.C.F., Martel, A., Maier-Hein, L., Malpani, A., Zenati, M.A., et al., Eds.; Springer: New York, NY, USA, 2018; pp. 303–311, ISBN 978-3-030-01201-4.
15. Jiang, H.; Learned-Miller, E. Face detection with the faster R-CNN. In Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017.

16. Urbonas, A.; Raudonis, V.; Maskeliūnas, R.; Damaševičius, R. Automated identification of wood veneer surface defects using faster region-based convolutional neural network with data augmentation and transfer learning. *Appl. Sci.* **2019**, *9*, 4898. [CrossRef]
17. Thanathornwong, B.; Suebnukarn, S. Automatic detection of periodontal compromised teeth in digital panoramic radiographs using faster regional convolutional neural networks. *Imaging Sci. Dent.* **2020**, *50*, 169–174. [CrossRef] [PubMed]
18. Chen, H.; Zhang, K.; Lyu, P.; Li, H.; Zhang, L.; Wu, J.; Lee, C.H. A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films. *Sci. Rep.* **2019**, *9*, 3840. [CrossRef] [PubMed]