

Automatic identification of dental implant brands with deep learning algorithms

Hasret Yuce, DDS^{1,*}, Gozde Acikgoz, DDS¹, Saif Al-Jumaili, PhD², Osman Nuri Ucan, PhD³,
Gulfem Ozlu Ucan, DDS⁴, Kadriye Ayca Dere, DDS⁵

¹Department of Oral, Dental and Maxillofacial Radiology, Faculty of Dentistry, Pamukkale University, Denizli 20160, Turkey

²Independent Researcher

³Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Altinbas University, Istanbul 34217, Turkey

⁴Department of Oral, Dental and Maxillofacial Radiology, Faculty of Dentistry, Istanbul Gelisim University, Istanbul 34310, Turkey

⁵Department of Oral, Dental and Maxillofacial Surgery, Faculty of Dentistry, Suleyman Demirel University, Isparta 32260, Turkey

*Corresponding author: Hasret Yuce, DDS, Department of Oral, Dental and Maxillofacial Radiology, Faculty of Dentistry, Pamukkale University, Kinikli, Suleyman Demirel Street, Denizli 20160, Turkey (kosarhasret@gmail.com)

Abstract

Objectives: To reduce the problems arising from the inability to identify dental implant brands, this study aims to classify various dental implant brands using deep learning algorithms on panoramic radiographs.

Methods: Images of 4 different dental implant systems (NucleOSS, Medentika, Nobel, and Impliance) were used from a total of 5375 cropped panoramic radiographs. To enhance image clarity and reduce blurriness, the contrast limited adaptive histogram equalization filter was applied. GoogleNet, ResNet-18, VGG16, and ShuffleNet deep learning algorithms were utilized to classify the 4 different dental implant systems. To evaluate the classification performance of the algorithms, Receiver Operating Characteristic (ROC) curves and confusion matrices were generated. Based on these confusion matrices, accuracy, precision, sensitivity, and F1 score were calculated. The Z-test was used to compare the performance metrics across different algorithms.

Results: The accuracy rates of the deep learning algorithms were obtained as 96.00% for GoogleNet, 84.40% for ResNet-18, 98.90% for VGG16, and 84.80% for ShuffleNet. A statistically significant difference was found between the accuracy rate of the VGG16 algorithm and those of GoogleNet, ShuffleNet, and ResNet-18 ($P < .001$, $P < .001$, and $P < .001$, respectively).

Conclusions: With the achievement of high accuracy rates, deep learning algorithms are considered a valuable and powerful method for identifying dental implant brands.

Keywords: automatic identification; deep learning; dental implant.

Introduction

Dental implants are a widely used treatment method for the oral rehabilitation of patients with partial or complete edentulism.^{1,2} Today, they are considered an ideal treatment option, particularly for individuals with good overall oral hygiene, as they provide both functional and aesthetic support. In recent years, the increasing demand for dental implants can be attributed to several key factors, including the rise in tooth loss associated with longer life expectancy, growing aesthetic concerns, increased social awareness, a higher incidence of partial or complete edentulism, and complications related to conventional prostheses.³ However, alongside the widespread use of implants, mechanical issues such as fractures of implant screws or bodies, as well as biological complications like peri-implantitis, may arise, negatively affecting the long-term success of dental implants.^{4,5}

In this context, complications associated with dental implants have become an increasing concern for the global dental community and represent a significant public health issue, imposing a substantial socioeconomic burden.⁶ The early detection and effective management of these complications require the accurate identification and classification of dental implant systems. However, in clinical practice, determining

the specific implant system used after implant surgery can be challenging due to factors such as the closure of dental clinics or the loss of patient records. The presence of more than 300 dental implant manufacturers worldwide and the variations in macro- and micro-designs among implant brands further complicate the identification process. According to the literature, approximately 3% of implants must be removed because the implant system used cannot be determined, preventing any prosthetic treatment or repair procedures.^{7,8}

One of the most commonly used methods for identifying and classifying implant systems is 2-dimensional dental radiography, including panoramic and periapical radiographs. However, factors such as similar designs, low resolution, and angular variations limit the reliability of manual identification methods. Consequently, there is an increasing need for more objective and automated approaches.⁹

In this regard, artificial intelligence-based systems have demonstrated significant potential. Deep learning, a subfield of artificial intelligence, offers a wide range of applications in the medical domain. Deep learning algorithms model biological neural networks within a computational framework. With the development of artificial neural networks, the concept of deep learning has emerged, leading to extensive

Received: 4 April 2025; Revised: 13 May 2025; Accepted: 31 May 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of the British Institute of Radiology and the International Association of Dentomaxillofacial Radiology. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

research in this field.¹⁰ These algorithms are widely employed in image recognition, classification, and object detection.¹¹ This innovative technology, based on deep neural networks with multiple layers and hidden units, has been associated with high accuracy rates in medical image analysis, particularly in edge detection, classification, and segmentation tasks.¹²

In the field of dentistry, deep and convolutional neural networks are increasingly preferred for analyzing 2- and 3-dimensional dental images.¹³⁻¹⁵ Studies in the literature have demonstrated significant advancements in the identification and classification of dental implant systems using deep learning algorithms based on 2-dimensional radiographic images.^{14,16-18} These studies indicate that, compared to conventional methods, deep learning-based models not only reduce processing time but also achieve higher accuracy rates than dental specialists in the field of implantology.^{13,18} Even when working with images containing limited information, such as panoramic or periapical radiographs, these algorithms are capable of distinguishing implant systems with high success rates by recognizing subtle differences in implant design, thereby enhancing classification performance. These findings suggest that deep learning models offer an objective, rapid, and reliable alternative for the identification of dental implants.^{19,20}

The aim of this study is to classify various dental implant brands using deep learning algorithms on panoramic radiographs as a means to mitigate the problems arising from the inability to identify dental implant brands due to various clinical and environmental factors.

Methods and materials

This study was approved by the Pamukkale University Non-Interventional Clinical Research Ethics Committee (Approval number: E-60116787-020-507049). This committee waived the need for individual informed consent, and thus, a written/verbal informed consent was not obtained from any participant, as this study featured a non-interventional retrospective design and all the data were analyzed anonymously.

Collection of images

In this study, images of 275 patients who underwent dental implant surgery at the Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, Pamukkale University between 2019 and 2024 and who had panoramic radiographs obtained with a Digital Orthopantomography (OP200D; Instrumentarium Corp., Imaging Department, Tuusula, Finland, 60 kVp, 6.3 mA, and 14.1 s scan time) at the Department of Oral and Maxillofacial Radiology were used.

Creating datasets

Panoramic radiographs of 275 patients were classified as NucleOSS (NucleOSS Dental Implant System, Izdent Medikal, Izmir, Turkey), Implance (Implance Dental Implant System, FMK Dental Medikal, Denizli, Turkey), Medentika (A Straumann Group Brand, Batigroup Dental Dental Products Trade Inc., Hügelsheim, Germany), and Nobel (Nobel Biocare NobelActive and Brånemark system, Gothenburg, Sweden) according to the dental implant archive records of the Department of Oral, Dental and Maxillofacial Surgery. Although there are different implant types within

each implant brand, no further sub-classification was made, and they were all included in the study under the same main brand name. Classified panoramic radiographs were cropped using Adobe Photoshop Elements (Adobe Systems, Inc., San Jose, CA, United States) to cover the entirety of a single dental implant for the preparation of data analysis. As a result of the cropping process, a total of 1075 radiographic images of dental implants were obtained, corresponding to 4 implant systems: NucleOSS (530 images), Medentika (193 images), Implance (184 images), and Nobel (168 images). A total of 1075 cropped images were saved in Portable Network Graphics (PNG) format.

A dataset consisting of cropped radiographic images classified according to dental implant brands was prepared. The prepared dataset was subjected to digital preprocessing using MATLAB Software (The MathWorks Inc., Natick, Massachusetts, United States). The images, originally in PNG format, were reduced to 224 × 224 pixels.

The dataset was expanded using various data augmentation techniques (reflection, rotation, scale, shear, and translation). The total number of images obtained after the data augmentation process was determined as 5375.

After this process, a contrast limited adaptive histogram equalization filter in MATLAB was used to sharpen the images and reduce blur. The contrast enhancement limit was set to 0.005 as the clip limit parameter.

The general schema of the dataset preparation process used in this study is presented in Figure 1. 80% of the collected datasets were used to train the artificial intelligence, and 20% were used to test the trained artificial intelligence.

Classification algorithms

In this study, GoogleNet, ResNet-18, VGG16, and ShuffleNet algorithms were utilized to classify dental implant brands radiographically.

The GoogleNet architecture is a deep neural network model consisting of a total of 27 layers, including pooling layers. It includes a 1 × 1 convolutional layer with 128 filters and ReLU activation function for dimensionality reduction. This is followed by a fully connected layer with 1024 units and a dropout layer with a 70% dropout rate. Finally, a linear layer with softmax loss, which is removed during inference, is included in the model.²¹

ResNet-18 consists of 22 layers in total, starting with a convolutional layer, followed by 8 residual blocks, and concluding with a fully connected layer.²²

VGG16 presents a deep network architecture with 13 convolutional layers and 3 fully connected layers. This model utilizes small 3 × 3 filters in each convolutional layer, enhancing depth and allowing for detailed feature extraction.²³

ShuffleNet is primarily based on grouped convolution and channel shuffle operations. Grouped convolution reduces computational cost by dividing the weights into smaller groups and processing each group separately. Channel shuffle addresses potential disconnections in feature maps caused by independent group processing, ensuring more efficient information flow.²⁴

These models were pre-trained using the ImageNet8 dataset, which comprises over 1 million images across 1000 classes.²³ However, since the dataset used in this study consisted of 4 classes, the final layer of each model was replaced with a

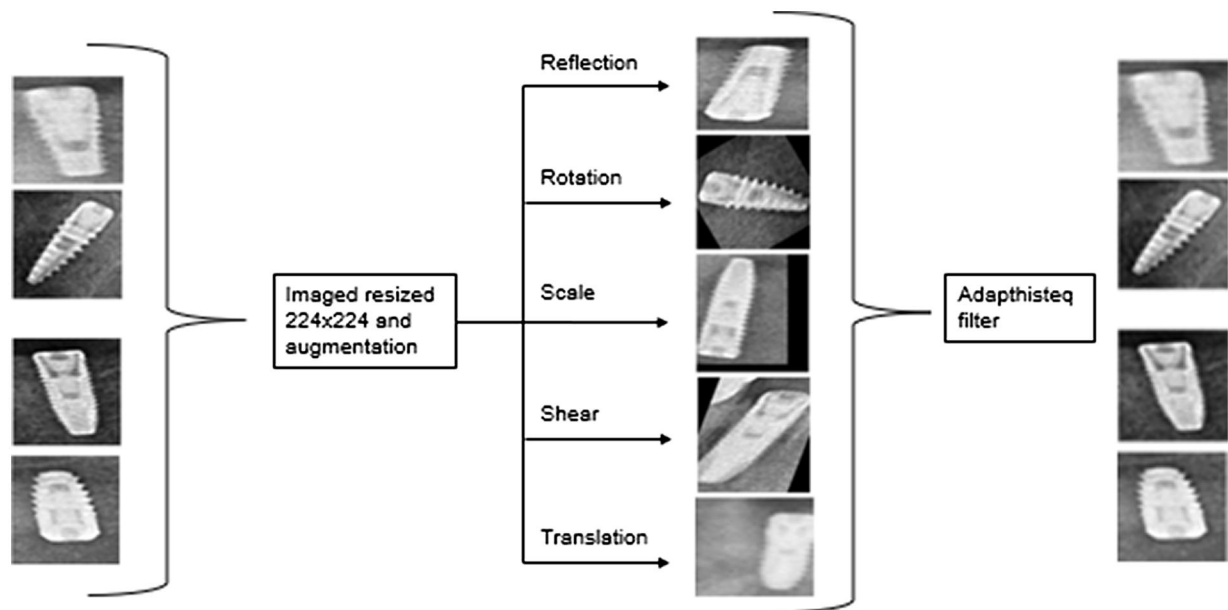


Figure 1. The general schema of the dataset preparation process.

fully connected layer with 4 outputs to classify dental implant brands.

Evaluation criteria for algorithm performance

In order to evaluate the classification success of the algorithms, a Receiver Operating Characteristic (ROC) curve was created using the detection rate-false positive rate values. In addition, true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values were determined, and confusion matrices were created. Accuracy, precision, recall, and F1 score were calculated based on this confusion matrix.

According to eqn (1), accuracy is defined as the ratio of correct predictions to the total predictions made by the model on the entire dataset. Recall indicates how many of the actual positive examples of a class were correctly predicted, and as shown in eqn (2), it is calculated as the ratio of true positive predictions to all actual positive examples. Precision, defined in eqn (3), shows how many of the positive predictions made by the model are truly correct. The F1-score, which balances precision and recall, is expressed as the harmonic mean of these 2 metrics in eqn (4). The use of the F1-score provides a reliable method for comprehensively evaluating the model's performance.

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

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 \text{ score} = \frac{2 \times (Recall \times Precision)}{Recall + Precision} \quad (4)$$

Model visualization

Gradient-weighted Class Activation Mapping (Grad-CAM) is a technique used to visualize the most influential regions in an image that contribute to a model's classification decision. By utilizing gradient information from the final convolutional

layer, Grad-CAM identifies the key regions and represents them using heatmaps, where more significant areas are highlighted with warmer colors (e.g., red). This approach enhances the interpretability of the model's decision-making process.²⁵ In this study, images were visualized and classified by Grad-CAM for each CNN model. The heat map generated using Grad-CAM was reconstructed with the last convolution layer.

Statistical method

Data were analyzed using Microsoft Excel (Microsoft Corporation, Redmond, WA, United States). Z test was used to compare the results of performance metrics according to algorithms. To reduce the risk of Type I error and to control p-value inflation in multiple Z-test analyses, the Bonferroni correction was applied to the p-values for all multiple comparisons. Significance level was taken as $P < .05$.

Results

Following the data augmentation process, a total of 5375 images were obtained. The dental implant brands used in the study are NucleOSS, Medentika, Nobel, and Implants, with the number of images corresponding to each brand being 2650, 965, 920, and 840, respectively.

The detection rate and false positive rate of GoogleNet, ResNet-18, VGG16, and ShuffleNet algorithms and the obtained ROC curves are presented in Figure 2. All ROC curves are close to the (0,1) point, and the models show good performance.

Confusion matrices of GoogleNet, ResNet-18, VGG16, and ShuffleNet algorithms are shown in Figure 3.

Accuracy, precision, recall values, F1 score, and overall accuracy of all 4 algorithms in dental implant classification are presented in Table 1.

Sample images of heat maps of 4 different dental implant brands are presented in Figure 4. The areas coded with warm colors in the visualization are generally focused on the body

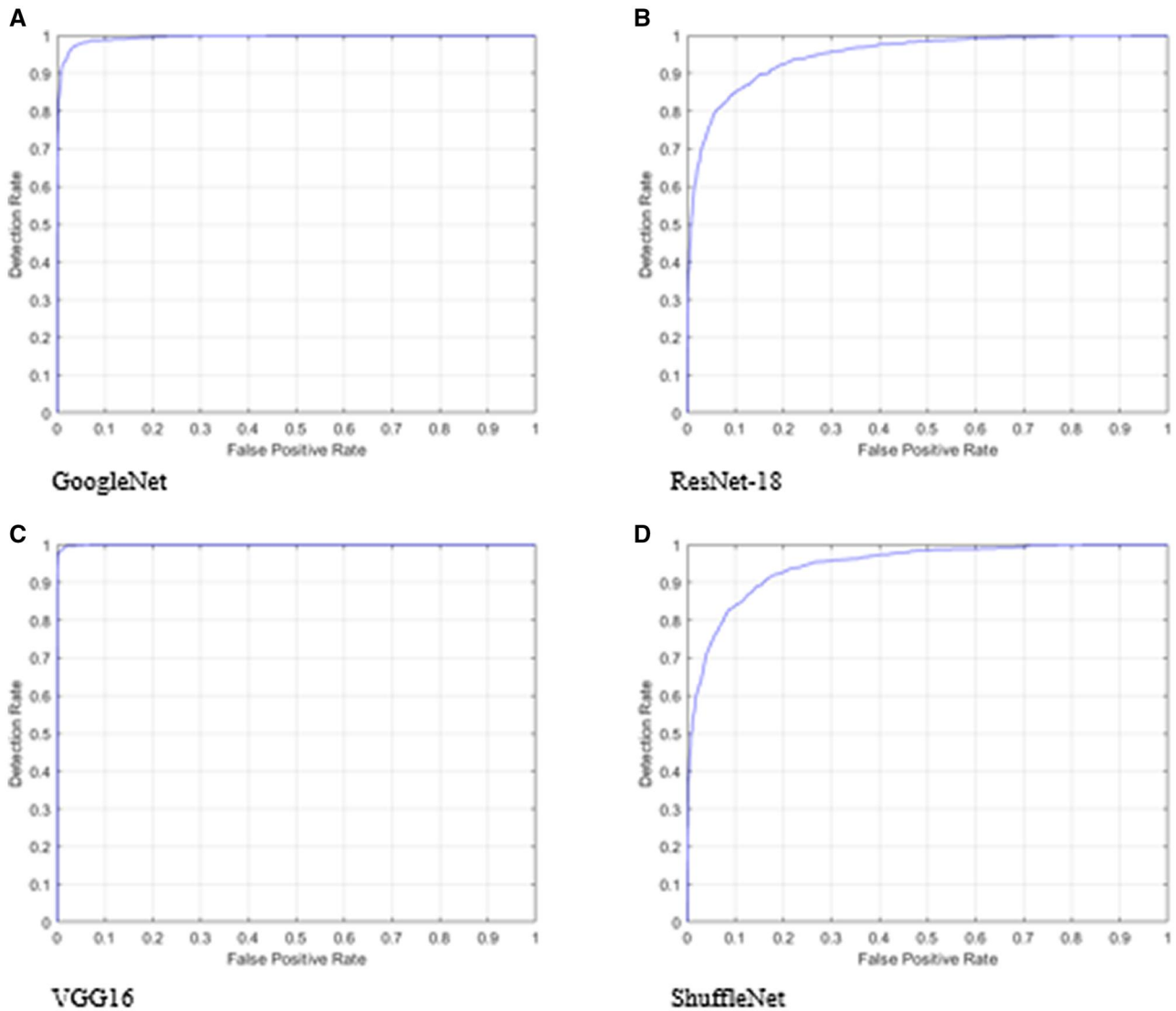


Figure 2. ROC curves of GoogleNet, ResNet18, VGG16, and ShuffleNet.

and threads of the dental implants. However, there are also areas coded with warm colors in the surrounding tissue.

The results of performance metrics were compared according to the algorithms. The comparison of accuracy rates is presented in Table 2. The comparison of precision values in Table 3, the comparison of recall values in Table 4, and the comparison of F1 scores in Table 5.

The comparison of overall accuracy values for GoogleNet, ResNet-18, VGG16, and ShuffleNet algorithms in dental implant classification is presented in Table 6.

Discussion

Deep learning algorithms are artificial intelligence-based systems with the potential to minimize human errors and enable faster, more accurate classification by analyzing large datasets.²⁶ Although a comprehensive dataset of 5375 cropped panoramic images in this study positively influences algorithm performance, factors such as data quality, diversity, overload, and data augmentation techniques also play a significant role, leading to variations in performance.²⁷ CNN accuracy rates were in the current study: GoogleNet 96.00%,

ResNet-18 84.40%, VGG16 98.90%, and ShuffleNet 84.80%. The classification of dental implants was successfully accomplished with high accuracy. An analysis of the ROC curves indicated that all 4 algorithms were close to the (0,1) point, supporting the finding that the models performed well. On the other side, it was observed that the models maximized true positives while minimizing false positives. These findings suggest that CNN is a valuable method for clinical applications.

Previous studies on dental implant classification show that Kim et al.²⁰ achieved an accuracy rate of 93.00% with GoogleNet, while Said et al.²⁸ reported 93.80%, which are relatively lower than the accuracy rate in this study. These differences may be attributed to the variation in the dental implant brands and the smaller datasets compared to this research. In contrast, Tiryaki et al.¹⁶ achieved the GoogleNet accuracy rate of 97.20%, which is slightly higher than that of this study, possibly due to the more extensive dataset used in their study. The high accuracy rates obtained with GoogleNet across various studies, including this one, can be attributed to its structural features, such as an Inception architecture-focused deep learning model.²¹

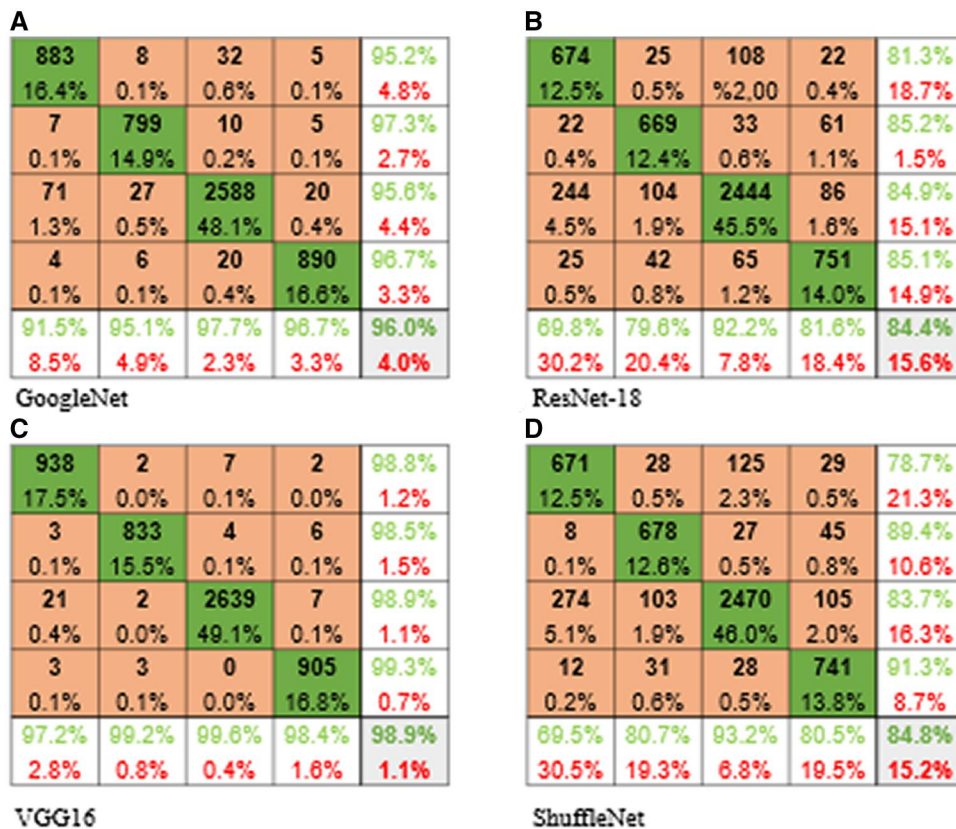


Figure 3. Confusion matrices of GoogleNet, ResNet18, VGG16, and ShuffleNet.

Table 1. Accuracy, precision, recall, F1 score, and overall accuracy of GoogleNet, ResNet-18, VGG16, and ShuffleNet algorithms in dental implant classification.

		Accuracy (%)	Precision	Recall	F1 score	Overall accuracy (%)
GoogleNet	Medentika	97.59	0.95	0.92	0.93	96.0
	Nobel	98.79	0.97	0.95	0.96	
	NucleOSS	96.62	0.96	0.98	0.97	
	Implance	98.85	0.97	0.97	0.97	
ResNet-18	Medentika	91.05	0.81	0.70	0.75	84.40
	Nobel	94.0	0.85	0.80	0.82	
	NucleOSS	87.64	0.85	0.92	0.88	
	Implance	93.77	0.85	0.82	0.83	
VGG16	Medentika	99.29	0.99	0.97	0.98	98.90
	Nobel	99.63	0.98	0.99	0.99	
	NucleOSS	99.24	0.99	1.0	0.99	
	Implance	99.61	0.99	0.98	0.99	
ShuffleNet	Medentika	90.54	0.79	0.70	0.74	84.80
	Nobel	94.96	0.89	0.81	0.85	
	NucleOSS	87.32	0.84	0.93	0.88	
	Implance	94.80	0.91	0.81	0.86	

Kim et al.²⁰ reported an accuracy rate of 98.00% with ResNet-18, which is higher than that of our study. Despite using a smaller dataset, the higher accuracy rate may be due to better optimization of hyperparameters or differences in image quality. Sukegawa et al.²⁹ achieved an accuracy rate of 97.87% in single-task learning and 98.03% in multi-task learning. These findings, along with ours, confirm that high accuracy rates can be obtained with ResNet-18 due to its “residual block” structures, which enable effective feature learning despite having fewer parameters than deeper architectures.²² However, it should be noted that the model’s

success depends on the dataset and the optimization of hyperparameters.^{22,30}

Sukegawa et al.¹⁴ reported an accuracy rate of 89.90% with standard VGG16, while fine-tuned VGG16 achieved 93.50%. This discrepancy was attributed to fine-tuned VGG16 benefits from more specific and optimized hyperparameters (1). Other studies reported accuracy rates of 97.60% (Tiryaki et al.¹⁶) and 92.61% (Leblebicioğlu Kurtuluş et al.¹⁷), which were relatively lower than that of our study. This high accuracy rate may be attributed to the effective use of data augmentation techniques and

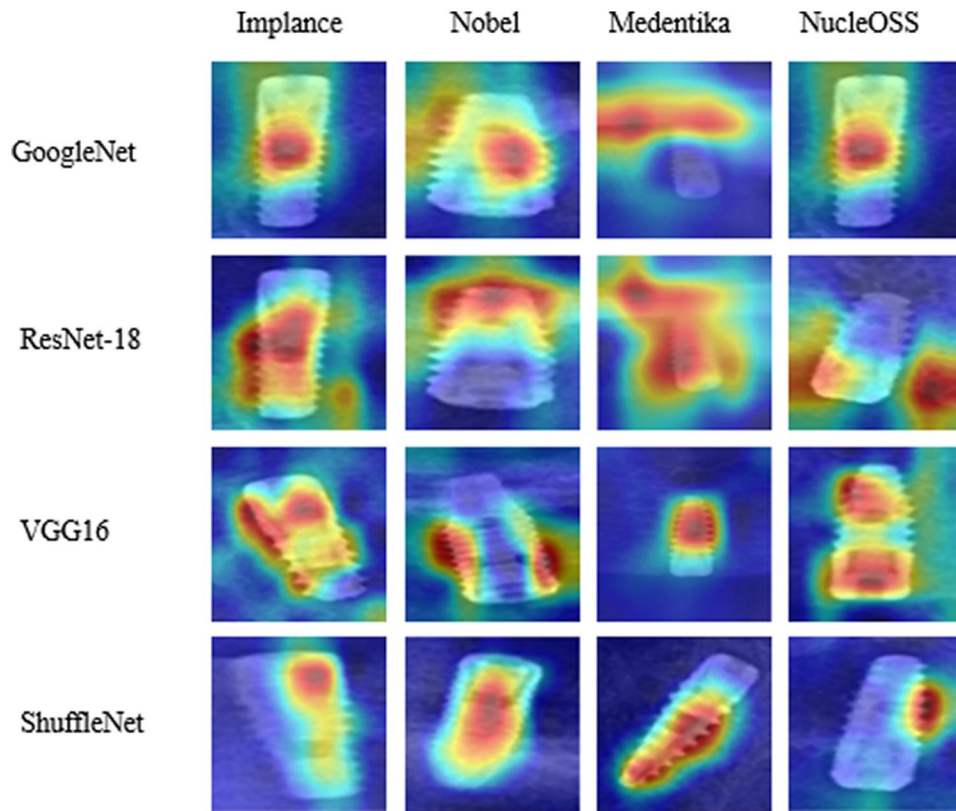


Figure 4. Example of heat activation maps of 4 different dental implants classified by Grad-CAM.

Table 2. Comparison of accuracy values between algorithms (Z test with a Bonferroni correction, P values).

Dental implant brand	Algorithm	GoogleNet	ResNet-18	VGG16
Medentika	GoogleNet	—		
	ResNet-18	<0.001*	—	
	VGG16	0.003**	<0.001*	—
	ShuffleNet	<0.001*	0.698	<0.001*
Nobel	GoogleNet	—		
	ResNet-18	<0.001*	—	
	VGG16	0.052	<0.001*	—
	ShuffleNet	<0.001*	0.413	<0.001*
NucleOSS	GoogleNet	—		
	ResNet-18	<0.001*	—	
	VGG16	<0.001*	<0.001*	—
	ShuffleNet	<0.001*	0.725	<0.001*
Implance	GoogleNet	—		
	ResNet-18	<0.001*	—	
	VGG16	0.062	<0.001*	—
	ShuffleNet	<0.001*	0.341	<0.001*

P-values for all multiple comparisons have been adjusted by applying the Bonferroni correction.

*P < .001.
**P < .05.

Table 3. Comparison of precision values between algorithms (Z test with a Bonferroni correction, P values).

Dental implant brand	Algorithm	GoogleNet	ResNet-18	VGG16
Medentika	GoogleNet	—		
	ResNet-18	<0.001*	—	
	VGG16	<0.001*	<0.001*	—
	ShuffleNet	<0.001*	0.272	<0.001*
Nobel	GoogleNet	—		
	ResNet-18	<0.001*	—	
	VGG16	0.189	<0.001*	—
	ShuffleNet	<0.001*	0.015**	<0.001*
NucleOSS	GoogleNet	—		
	ResNet-18	<0.001*	—	
	VGG16	<0.001*	<0.001*	—
	ShuffleNet	<0.001*	0.315	<0.001*
Implance	GoogleNet	—		
	ResNet-18	<0.001*	—	
	VGG16	0.002**	<0.001*	—
	ShuffleNet	<0.001*	<0.001*	<0.001*

P-values for all multiple comparisons have been adjusted by applying the Bonferroni correction.

*P < .001.
**P < .05.

hyperparameter optimization, which enhance model generalization and improve dental implant classification performance.

This study is the first, to our knowledge, to classify dental implants using ShuffleNet. The ShuffleNet architecture achieved high accuracy in this study, marking a notable contribution to the literature. This high accuracy rate is likely due to the effective implementation of group convolutions and channel shuffle mechanisms.²⁴ Despite its lightweight architecture

and low parameter count, ShuffleNet’s ability to maintain high accuracy demonstrates its potential as an effective solution for complex datasets such as dental radiographic images.²⁴ This suggests that models optimized for resource-limited systems or portable devices could be developed using ShuffleNet.

To our knowledge, this study represents the first statistical comparison of CNN models for dental implant classification. The classification accuracy rankings from highest to lowest

Table 4. Comparison of recall values between algorithms (Z test with a Bonferroni correction, *P* values).

Dental implant brand	Algorithm	GoogleNet	ResNet-18	VGG16
Medentika	GoogleNet	—	—	—
	ResNet-18	<0.001*	—	—
	VGG16	<0.001*	<0.001*	—
	ShuffleNet	<0.001*	1.000	<0.001*
Nobel	GoogleNet	—	—	—
	ResNet-18	<0.001*	—	—
	VGG16	<0.001*	<0.001*	—
	ShuffleNet	<0.001*	0.615	<0.001*
NucleOSS	GoogleNet	—	—	—
	ResNet-18	<0.001*	—	—
	VGG16	<0.001*	<0.001*	—
	ShuffleNet	<0.001*	0.167	<0.001*
Implance	GoogleNet	—	—	—
	ResNet-18	<0.001*	—	—
	VGG16	0.170	<0.001*	—
	ShuffleNet	<0.001*	0.581	<0.001*

P-values for all multiple comparisons have been adjusted by applying the Bonferroni correction.
**P* < .001.

Table 5. Comparison of F1 scores among algorithms (Z test with a Bonferroni correction, *P* values).

Dental implant brand	Algorithm	GoogleNet	ResNet-18	VGG16
Medentika	GoogleNet	—	—	—
	ResNet-18	<0.001*	—	—
	VGG16	<0.001*	<0.001*	—
	ShuffleNet	<0.001*	0.614	<0.001*
Nobel	GoogleNet	—	—	—
	ResNet-18	<0.001*	—	—
	VGG16	<0.001*	<0.001*	—
	ShuffleNet	<0.001*	0.098	<0.001*
NucleOSS	GoogleNet	—	—	—
	ResNet-18	<0.001*	—	—
	VGG16	<0.001*	<0.001*	—
	ShuffleNet	<0.001*	1.000	<0.001*
Implance	GoogleNet	—	—	—
	ResNet-18	<0.001*	—	—
	VGG16	0.002**	<0.001*	—
	ShuffleNet	<0.001*	0.075	<0.001*

P-values for all multiple comparisons have been adjusted by applying the Bonferroni correction.
**P* < .001.
***P* < .05.

Table 6. Comparison of overall accuracy values of algorithms (Z test with a Bonferroni correction, *P* values).

Algorithm	GoogleNet	ResNet-18	VGG16
GoogleNet	—	—	—
ResNet-18	<0.001*	—	—
VGG16	<0.001*	<0.001*	—
ShuffleNet	<0.001*	0.566	<0.001*

P-values for all multiple comparisons have been adjusted by applying the Bonferroni correction.
**P* < .001.

were VGG16, GoogleNet, ShuffleNet, and ResNet-18 in the current study. Statistical analysis revealed a significant difference between the accuracy rates of VGG16 and those of GoogleNet, ResNet-18, and ShuffleNet (*P* < .001 for all).

The deeper and broader architecture (more parameters and number of layers) of VGG16 has increased its ability to learn complex features, which may have resulted in higher accuracy. In contrast, the lower number of parameters and layers in GoogleNet, ResNet-18, and ShuffleNet may have resulted in limitations and disadvantages in their capacity to learn complex details when working with small datasets or complex data structures.

A significant difference was also observed between the accuracy rates of GoogleNet and both ResNet-18 and ShuffleNet (*P* < .001, *P* < .001), suggesting that GoogleNet's higher parameter count enhances learning capacity. However, no statistically significant difference was found between ResNet-18 and ShuffleNet (*P* = .566), indicating that both models may have similar learning capacities or that the dataset complexity was insufficient to highlight performance differences between them.

In this study, it was determined that the VGG16 algorithm achieved the highest accuracy rate of 99.63% for the Nobel brand, followed closely by the GoogleNet algorithm in 98.79%. However, the difference between them was not statistically significant (*P* = .052).

VGG16 algorithm achieved the highest accuracy rate of 0.99 for the Medentika, NucleOSS, and Implance brands, and the highest sensitivity value of 1.0 for the NucleOSS brand. Also, it achieved the highest F1 score of 0.99 for the Nobel, NucleOSS, and Implance brands. These results indicate that the VGG16 algorithm exhibits high performance, and the suitability of the dataset to distinguish subtle differences among dental implant brands. The deep architecture of VGG16, consisting of 16 layers, has proven particularly effective in achieving high sensitivity and accuracy values. The sensitivity value of 1.0 for the NucleOSS brand highlights the model's superior performance in detecting true positives. Similar findings have also been reported accuracy rates of 90% and above by Tiryaki et al.¹⁶, Leblecioğlu Kurtuluş et al.¹⁷, and Sukegawa et al.⁷.

When evaluating the general performance metrics of dental implant brands, it is observed that despite having the highest number of radiographic images, the NucleOSS brand exhibits the lowest accuracy rate across all algorithms. Similarly, Tiryaki et al.¹⁶ reported that the Implance brand, which had the largest number of images in their study, did not achieve the highest accuracy rates. This condition suggests that factors such as image angle and quality may affect classification accuracy.

While the Medentika brand demonstrated the lowest sensitivity and F1 scores across all algorithms, the Nobel brand recorded the lowest accuracy value only in the VGG16 algorithm. In contrast, for other algorithms, the lowest accuracy value is observed in the Medentika brand. These variations may be attributed to differences in the macroscopic characteristics of the implants.

The microscopic and macroscopic differences among dental implants complicate the classification.^{25,31} CNN studies increase the transparency of the classification process by identifying image regions that influence the model's decisions through model visualization, like the Grad-CAM technique.²⁵ In our study, Grad-CAM visualizations that the model primarily focuses on the implant body and threads while considering surrounding tissues. Similarly, Sukegawa et al.¹⁴ found that fine-tuned and transfer learning-based VGG16 and VGG19 algorithms improve classification

20. Kim JE, Nam NE, Shim JS, Jung YH, Cho BH, Hwang JJ. Transfer learning via deep neural networks for implant fixture system classification using periapical radiographs. *J Clin Med.* 2020;9: E1117-E1117. <https://doi.org/10.3390/JCM9041117>
21. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* 2015; 07-12-June-2015: 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
22. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* 2016; 2016-December: 770-778. <https://doi.org/10.1109/CVPR.2016.90>
23. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *International Conference on Learning Representations.* 2014.
24. Zhang X, Zhou X, Lin M, Sun J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* Published online December 14, 2018:6848-6856. <https://doi.org/10.1109/CVPR.2018.00716>
25. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: *Proceedings of the IEEE International Conference on Computer Vision.* 2017; 2017-October: 618-626. <https://doi.org/10.1109/ICCV.2017.74>
26. Saghiri MA, Freag P, Fakhrzadeh A, Saghiri AM, Eid J. Current technology for identifying dental implants: a narrative review. *Bull Natl Res Cent.* 2021;45:1-11. <https://doi.org/10.1186/S42269-020-00471-0>
27. Alan A, Karabatak M. Veri seti—sınıflandırma ilişkisinde performans etki eden faktörlerin değerlendirilmesi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi.* 2020;32:531-540. <https://doi.org/10.35234/FUMBD.738007>
28. Hadj Saïd M, Le Roux MK, Catherine JH, Lan R. Development of an artificial intelligence model to identify a dental implant from a radiograph. *Int J Oral Maxillofac Implants.* 2020;36:1077-1082. <https://doi.org/10.11607/JOMI.8060>
29. Sukegawa S, Yoshii K, Hara T, et al. Multi-task deep learning model for classification of dental implant brand and treatment stage using dental panoramic radiograph images. *Biomolecules.* 2021;11:815. <https://doi.org/10.3390/BIOM11060815/S1>
30. Firat H, Hanbay D. Comparison of 3D CNN based deep learning architectures using hyperspectral images. *J Fac Eng Archit Gazi Univ.* 2023;38:521-534. <https://doi.org/10.17341/gazimmfd.977688>
31. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell.* 2019;1:389-399. <https://doi.org/10.1038/s42256-019-0088-2>
32. Park WS, Huh JK, Lee JH. Automated deep learning for classification of dental implant radiographs using a large multi-center dataset. *Sci Rep.* 2023;13:4862-4868. <https://doi.org/10.1038/s41598-023-32118-1>