



Advanced Dental Implant System Classification with Pre-trained CNN Models and Multi-branch Spectral Channel Attention Networks



Srinivasa Rao Vemula¹, Maruthi Vemula², Ramesh Vatambeti^{3*}

¹ FIS Management Services, 27705 Durham, North Carolina, USA

² North Carolina School of Science and Mathematics, 27705 Durham, North Carolina, USA

³ School of Computer Science and Engineering, VIT-AP University, 522237 Vijayawada, India

* Correspondence: Ramesh Vatambeti (ramesh.v@vitap.ac.in)

Received: 02-22-2024

Revised: 04-25-2024

Accepted: 05-03-2024

Citation: S. R. Vemula, M. Vemula, and R. Vatambeti, “Advanced dental implant system classification with pre-trained CNN models and multi-branch spectral channel attention networks,” *Acadlore Trans. Mach. Learn.*, vol. 3, no. 2, pp. 94–105, 2024. <https://doi.org/10.56578/ataiml030203>.



© 2024 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: Dental implants (DIs) are prone to failure due to uncommon mechanical complications and fractures. Precise identification of implant fixture systems from periapical radiographs is imperative for accurate diagnosis and treatment, particularly in the absence of comprehensive medical records. Existing methods predominantly leverage spatial features derived from implant images using convolutional neural networks (CNNs). However, texture images exhibit distinctive patterns detectable as strong energy at specific frequencies in the frequency domain, a characteristic that motivates this study to employ frequency-domain analysis through a novel multi-branch spectral channel attention network (MBSCAN). High-frequency data obtained via a two-dimensional (2D) discrete cosine transform (DCT) are exploited to retain phase information and broaden the application of frequency-domain attention mechanisms. Fine-tuning of the multi-branch spectral channel attention (MBSCA) parameters is achieved through the modified aquila optimizer (MAO) algorithm, optimizing classification accuracy. Furthermore, pre-trained CNN architectures such as Visual Geometry Group (VGG) 16 and VGG19 are harnessed to extract features for classifying intact and fractured DIs from panoramic and periapical radiographs. The dataset comprises 251 radiographic images of intact DIs and 194 images of fractured DIs, meticulously selected from a pool of 21,398 DIs examined across two dental facilities. The proposed model has exhibited robust accuracy in detecting and classifying fractured DIs, particularly when relying exclusively on periapical images. The MBSCA-MAO scheme has demonstrated exceptional performance, achieving a classification accuracy of 95.7% with precision, recall, and F1-score values of 95.2%, 94.3%, and 95.6%, respectively. Comparative analysis indicates that the proposed model significantly surpasses existing methods, showcasing its superior efficacy in DI classification.

Keywords: Dental implant; Modified aquila optimizer algorithm; Multi-branch spectral channel attention network; Texture complication; Convolutional neural network

1 Introduction

The ability to place DIs has revolutionized dentistry and significantly benefited patients worldwide [1]. Technological advancements reduce the likelihood of adverse alveolar bone conditions and improve the long-term prognosis, leading to increased usage of DIs [2]. Before implant placement, surgeons address alveolar bone atrophy by designing implants with various shapes and textures, such as threading and platforms, to enhance alveolar ridge augmentation and sinus lift surgery. The demand for DIs has attracted numerous manufacturers, resulting in over 220 different brands of implants on the market since 2000, with this number continuing to rise [3, 4].

Categorizing implant brands is challenging due to the wide variety of styles, structures, and associated tools, including fasteners, abutments, and superstructures. Implant maintenance, such as retightening to prevent loosening, is directly influenced by the manufacturer’s unique screws used to fix prostheses [5, 6]. Identifying the brand of implant is crucial, especially when different dentists may use different implants and screws for the same patient, and the types of implants used may vary over time [7]. Panoramic radiography can provide comprehensive data on the jawbone and teeth in a single image, offering the potential to identify the brand of a patient’s implant. However, manual identification requires significant human skill and effort [8].

Recent advancements in deep learning (DL) and neural network technologies, particularly deep CNNs (DCNNs), have shown promise in detecting and classifying various medical conditions, including bone fractures. While studies on DI fractures are lacking, efforts have been made to increase accuracy in detecting other dental fractures using radiographs [9–11].

Recent advances in computer vision and machine learning (ML) have opened up new possibilities for enhancing the diagnostic capabilities of dental practitioners [12]. In particular, CNNs have shown remarkable success in various image classification tasks, including medical imaging. By leveraging the hierarchical representation learning capabilities of CNNs, it is possible to automatically extract relevant features from radiographic images of DIs, enabling more accurate and efficient classification of implant failures [13].

However, traditional CNN-based approaches typically focus on spatial domain characteristics, which may not fully capture the underlying structural and textural information present in DI images. Digital signal processing theory suggests that images contain valuable frequency domain information, which can provide additional insights into the underlying patterns and structures. By analyzing the frequency content of DI images, it may be possible to uncover hidden features that are not readily apparent in the spatial domain [14].

Therefore, this study aims to evaluate the validity and reliability of DI fracture identification and classification using panoramic and periapical radiography images with two distinct architectures. Features of the input images were extracted using VGG16 and VGG19 and classified using MBSCA. The MAO model was employed to properly adjust the MBSCA parameters. The remaining sections of the study include a review of relevant literature in Section 2, details of the data sources in Section 3, and a presentation of the proposed model in Section 4. Finally, Section 6 concludes the research, while Section 5 discusses the analysis of results.

2 Related Works

Yang et al. [15] proposed a two-stream implant position regression network (MSPENet) to address the challenge of accurately classifying implant fixture systems using periapical radiographs. By augmenting initial annotations with supervisory data for implant region detection (IRD) training, richer characteristics were incorporated without additional labeling expenses. A multi-scale patch embedding module was developed within the MSPENet to adaptively extract features from images with varying tooth spacing. The MSPENet encoder, integrating transformer and convolution for enhanced feature representation, utilized a global-local feature interaction block. Additionally, the region of interest (RoI) mask obtained from the IRD was employed to enhance prediction outcomes. Experimental trials on a DI dataset using five-fold cross-validation demonstrated that the proposed TSIPR model outperformed state-of-the-art techniques.

Ramachandran et al. [16] introduced a state-of-the-art prediction method utilizing ML models to classify DI materials and predict potential mechanical deterioration. Important parameters examined included corrosion potential and acoustic emission (AE) weight-loss estimations, with a particular focus on pure alloys. With ML prototype models achieving over 90% accuracy, the proposed approach validated its viability for predicting tribocorrosion, demonstrating its potential as a reliable predictive modeling method for DI monitoring.

Rekawek et al. [17] developed a ML model to maximize the success rate of DIs by predicting implant failure and peri-implantitis development. Utilizing ensemble methods and logistic classifiers on retrospective data from 398 patients and 942 DIs, the random forest model outperformed others in predictive performance. Significant factors associated with implant failure included local anesthetic dose, implant length and diameter, antibiotic usage before surgery, and hygiene visit frequency. Similarly, factors correlated with peri-implantitis included diabetes mellitus, hygiene visit frequency, implant characteristics, and antibiotic usage.

Park et al. [18] evaluated the effectiveness of an automated DL algorithm in classifying different DI systems (DIS) using a large-scale multicenter dataset. After analyzing panoramic and periapical radiographs from various facilities, the DL algorithm achieved high classification accuracy, demonstrating reliable performance in DIS classification across extensive datasets without noticeable variations between periapical and panoramic images.

Park et al. [19] compared two artificial intelligence (AI) algorithms for DI length categorization using periapical radiographs, employing DL and clustering analysis. Both AI models demonstrated reliable classification performance, with statistically significant improvements observed after fine-tuning. The study highlights the potential clinical utility of AI models in DI length categorization validated across multiple centers.

Chen et al. [20] proposed a novel method for assessing peri-implantitis damage utilizing periapical films (PA) and CNN models. With its high accuracy in implant localization and peri-implantitis damage assessment, the CNN-based method offers potential for precise evaluation of peri-implantitis damage, aiding in implant dentistry and patient care.

Park et al. [21] assessed the effectiveness of DL in identifying and classifying different DI schemes using a large dataset of panoramic and periapical radiographs. DL demonstrated reliable classification accuracy, outperforming both specialized and nonspecialized dental experts in categorizing DIS encountered in clinical practice. Additionally, DL showed efficiency in reading and categorization time compared to dental experts, suggesting its potential as an effective decision support tool in dental implantology.

3 Materials and Methods

The Institutional Review Boards reviewed and authorized the study design and protocol; informed or written consent was obtained [22]. All procedures in this study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations for the journalism and execution of experimental studies [23, 24] and the updated Declaration of Helsinki.

3.1 Dataset

Retrospective data was collected at the Veterans Health Service Medical Center (VHSMC) starting in January 2019. This research involved the analysis of clinical photographs and digital radiography images of dental structures by two board-certified clinical practitioners. The aim was to identify a total of 21,398 DIs distributed across 7,281 patients. Three dental practitioners referenced in the study were responsible for excluding radiographic images exhibiting excessive noise, haziness, or distortion, particularly those obtained through the conventional paralleling technique in periapical radiography. Subsequently, a periodontist named JHL categorized the remaining images based on their respective anatomical regions, which were of two main types: panoramic (1,402 images) and periapical. Ultimately, 251 DIs were classified as intact, while 198 were identified as fractured, forming the core dataset for this investigation. A prior study examining fracture patterns facilitated the classification of fractured DIs into three distinct types, with Type I denoting horizontal fractures confined within and around the crestal module, Type II representing vertical fractures, and Type III encompassing horizontal fractures. However, due to the limited occurrence of Type III fractured DIs (n=4) within the datasets, this study exclusively focused on Type I and Type II fractures. Detailed information regarding all panoramic and periapical images of each DI, regardless of their fracture status, along with associated characteristics and quantities, is presented in Table 1.

Table 1. Various panoramic and periapical images and DIs

	Dataset	
	Frequency	Percentage (%)
Intact DIs		
Panoramic imageries	110	43.9
Fractured DIs, type I		
Panoramic imageries	42	48.9
Periapical imageries	43	51.2
Fractured DIs, type II		
Periapical images	58	52.7

Source: Daejeon Dental Hospital at Wonkwang University and VHSMC.

Note: Type I fractured DIs are restricted to the area immediately surrounding the implant's crestal module, including horizontal and vertical cracks;

Type II fractured DIs are a vertical crack that extends beyond the implant's crestal module.

3.2 Pre-processing

The radiographic pictures used in the study were scaled to 224×224 pixels for the VGG19 construction and 299×299 pixels for the VGG16 architecture. For model construction and accuracy performance predictions, the dataset was randomly divided into three parts: training, validation, and test. Each part received 60% of the total. Pixel normalization is part of the pre-processing, and one-hot encoding was used to decrease dataset abnormalities. 100 instances of rotation (with a range of 30°) were randomly added, along with width and height shifting (with a variety of 0.2), zooming and flipping to the training dataset. The validation and test datasets did not undergo any sort of augmentation operation.

4 Proposed Methodology

4.1 Feature Extraction Using Pre-trained CNNs

The study employed an elementary CNN architecture featuring three instances of the VGG19 model. Leveraging transfer learning methodologies, both the VGG16 and VGG19 architectures underwent fine-tuning and training processes. Notably, the VGG16 model exhibits a profound layer depth, comprising 13 convolutional layers and three fully connected layers, as detailed in prior research [25]. Trained on an extensive dataset of over 1 million images spanning 1,000 classes, with iterative optimization exceeding 370,000 iterations, these models demonstrate a robust capacity for feature extraction and classification. VGG19, recognized for its prowess in image recognition tasks, notably secured first place in the classification and localization division of the 2014 Large Scale Visual Recognition Challenge. Comprising 19 weighted layers, 16 of which are convolutional layers, and organized into five distinct groups by max pooling layers [26], VGG19 encapsulates a deep hierarchical representation of image features. In the

experimentation phase, the learning rates were meticulously configured at 0.0001 for the fundamental CNN model, while the fine-tuning process for the VGG16/19 models also adhered to the same learning rate setting. Convolutional layers apply filters (also known as kernels) to the input data to extract features. Each filter slides across the input image, computing the dot product between the filter weights and the values in the receptive field, generating feature maps that highlight specific patterns such as edges, textures, or shapes.

The research used a total of five CNN study groups:

- A CNN model at its most fundamental level, consisting of six convolutional layers.
- A VGG16 model that learned from its pre-trained weights (VGG16 transfer).
- A VGG16 model that learned from its pre-trained weights was then fine-tuned (VGG16 fine-tuning).
- VGG19 transfer, a method of learning a model using pre-trained weights.
- Fine-tuning the VGG19 model using pre-trained weights and transfer learning.

Following each convolutional layer, a rectified linear unit (ReLU) activation function is commonly applied element-wise to introduce non-linearity into the network. ReLU sets all negative values in the feature maps to zero, enabling the network to learn complex relationships between features. Momentum stochastic gradient descent (SGD) was utilized for optimizing the four VGG models, whereas Adam was employed for the basic CNN. Transfer learning was used for dataset training, with the training dataset randomly divided into 128 batches for each epoch, and a maximum of 700 iterations (epochs) determined by the validation loss behavior. The present method's efficacy was assessed using fourfold cross-validation. This cross-validation procedure ensures generalization and helps avoid overfitting. Every architecture, including the basic CNN, VGG16/19 transfer, and tuning, underwent this procedure for both training and evaluation. The Keras library (<https://keras.io>) and the TensorFlow engine were used for building, training, and predicting deep-learning representations.

4.1.1 Model visualization

The most important elements utilized for categorization may be better understood with the aid of CNN model visualization. Using gradient-weighted class activation [27], the most important pixels for classification in the picture were determined, which allowed for both finding possible accurate classifications based on wrong characteristics and obtaining a better understanding of the classification process. Map representations are heatmaps of the gradients, where the locations of higher relevance for feature extraction are represented by the "hotter" hues. This work used the last convolutional layer to rebuild the gradient-weighted class activation mapping (Grad-CAM) heat map.

4.2 Classification Using MBSCAN

The MBSCAN refers to a sophisticated DL architecture designed for image analysis tasks. It incorporates multiple branches, each equipped with spectral channel attention mechanisms, which dynamically adapt the network's focus to relevant image features. Spectral channel attention enhances the network's ability to selectively attend to informative channels across different spectral frequencies, thereby facilitating robust feature extraction and representation learning. Through its multi-branch design, MBSCAN harnesses diverse pathways to capture intricate spatial and spectral dependencies within the input data, leading to enhanced performance in tasks such as image classification, object detection, and semantic segmentation. This advanced network architecture holds promise for various applications in computer vision and medical imaging, where precise feature discrimination and context-aware analysis are paramount.

For the dental image classification challenge, the study begins by presenting the overall structure of the proposed MBSCAN. Following this, a detailed illustration of the MBSCAN's main module of multi-branch spectral is provided.

4.2.1 General structure

MBSCAN, which can be trained effectively for rapid inference, is based on Residual Network (ResNet) 18, the fundamental network used in the study. The MBSCAN is composed of MBSCAs and ResNet18 fundamental blocks for each module. The topological structure of a basic block remains unchanged when the MBSCA is added into its end. Because there aren't enough dental photos to train a network from beginning, this approach lets us reuse ResNet18's weights. High- and low-frequency data interact more favorably with the network. The network can learn more important information and less useless information via the channel attention technique. It is suggested to utilize channel features of the input images, because texture images generally concentrate their primary information in the low-frequency domain. The attention module requires minimal additional parameters, offering the benefits of enhanced feature operations with a small increase in computational complexity. Furthermore, MBSCAN's modularity allows easy integration, simply by adjusting the number of output channels to incorporate the attention module while retaining channel structures consistent with preceding layers.

4.2.2 MBSCA unit

In reality, the feature extraction model ignores all data over a certain frequency and only uses data below that frequency. Despite the fact that the model delves into the many 2D DCT frequency components, each frequency component is only utilized to depict a portion of a channel in a feature map. There is no modeling of the more realistic

and deserving individual channels represented by numerous frequency components. Hence, to address this restriction, the multi-attention module is suggested.

In the MBSCA, distinct branches pay attention to distinct aspects of the input based on their frequency of occurrence. Generally speaking, such a branch can do cross-latitude interactive computation in a variety of ways, including channel attention, spatial attention, and others. This study aims to accomplish spectral channel attention by repeating the model's comparable computing procedure. Any two branches can record distinct frequency components since each branch uses its own unique component. The multi-branch structure allows for the exploration of many frequency components, which solves the problem of partial use of the image's frequency information. This allows for the realization of interaction between the frequency components. Both structures are redesigned by the MBSCA. The input $X \in R^{H \times W \times C}$ is denoted as $\{X_0, \dots, X_{K-1}\}$. In each subdivision, consistent 2D assignments for the input X_k are expressed as:

$$Freq_k = 2DDCT_{\Omega_k}(X_k) = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} X_k B_{h,w}^{\Omega_k} \quad (1)$$

s.t. $k \in \{0, 1, \dots, K-1\}$, where, $Freq_k \in R^c$ is the k -th branch, i.e., $Freq_k \in R^c = \text{compression}(X_k)$. The 2D DCT corresponds to X_k . Ω_k represents the 2D indices of the frequency components. Then the weights of every channel in were predicted using $Freq_k$ and were scaled subsequently.

$$attm_k = \text{sigmoid}(fc(Freq_k)) \quad (2)$$

$$\bar{X}_k = attm_k X_k \quad (3)$$

Eqs. (2) and (3) can be used for predicting weights $attm_k$ and scaling input X_k , respectively.

Applying Eqs. (1) to (3) across all K branches of the MBSCA extracts comprehensive feature representations $\{\bar{X}_0, \dots, \bar{X}_{K-1}\}$. These branches are aggregated to form the final output of the MBSCA module:

$$\Upsilon = AVG(\{\bar{X}_0, \bar{X}_1, \dots, \bar{X}_{K-1}\}) \quad (4)$$

Various components on each individual channel can interact, and AVG is the average pooling to fuse X_k . The MBSCA can be easily integrated into different basic system without modifying its topology, thanks to the fact of using the shape as the input X . This allows for reusing the pretrained weights. Ablation experiments were conducted to determine the relative relevance of various frequency branches (using MAO), and then the top-k frequency components were chosen using the best solution.

4.2.3 Finding K value using MAO algorithm

The MAO algorithm [28] was used to determine the recommended classifier's k-value. The ability of the aquila to swoop down and seize its prey is crucial to the aquila optimizer (AO). AO's worth was quickly proven in the domain of complicated and nonlinear optimization, which is a population-based approach. The work on the search control factor (SCF) from the improved aquila optimizer (IAO) motivated MAO to make further changes to the AO. Nevertheless, the precision of the epochs in IAO is hindered by the convergence characteristics of SCF. Some of the difficulties encountered in seeking the best outcome could be attributable to these characteristics. In order to address these issues, a revised IAO was implemented, which has a modified search control factor (MSCF) tailored to the second and third search phases. The next part describes the MAO method in depth, focusing on the changes that were made and how they affected the optimization method. The aquila's movement was reduced in terms of epochs when the search range was controlled using the MSCF. This means that there is a much smaller search space than there was with the previous SCF. Also, compared to the previous method, the optimal answer was found much faster. The updated MSCF takes the following form:

$$M - SCF(t) = 2 \times \exp\left(1 - \left(\frac{t \times (t \times 0.1)}{T}\right)\right) \times dir \quad (5)$$

$$dir = \begin{cases} 1 & \text{if } r < 0.5 \\ -1 & \text{else} \end{cases} \quad (6)$$

where, dir is the direction control factor, and the r parameter is a random number between zero and one. These characteristics are crucial in determining the aquila's fighting style. By limiting the aquila's mobility, the MSCF function aims to achieve rapid convergence. In addition, optimization delays were reduced. The updated method outperformed the original AO because it quickly identified the best set of solutions. With 250 and 250 epochs, respectively, both optimization methods were run. The proposed method has the following four search phases, which were incorporated into the MSCF function:

Step 1: Vertical dive attack (S_1)

Before diving into its hunt, the aquila swoops down to survey the area it intends to prey upon and choose the best spot to perch. These kinds of assaults are known as "vertical dive attacks," which can be expressed as follows:

$$S_1(t+1) = S_{best}(t) \times \left(1 - \frac{t}{T}\right) + (S_M(t) - S_{best}(t) \times r) \quad (7)$$

where, $S_1(t+1)$ is the key contender of $(t+1)$ epochs, r is the accidental integer in $[0, 1]$ the intermezzo, and $S_{best}(t)$ is the i -th generation of $\left(1 - \frac{t}{T}\right)$.

Step 2: Adapted full search with a short glide attack (MS)

This is known as a shorter glide assault, and it occurs before the aquila strikes its target. The aquila searches the key space using a variety of directions and speeds as follows:

$$MS_2(t+1) = S_R(t) + M - SCF(t) \times (S_{best}(t) - S(t)) \times r \times (y - x) \quad (8)$$

The above equation describes the outcomes using the point that forms the spiral during stage (x and y), a random number (r) between 0 and 1, and the MSCF (t). To avoid the issue of becoming stuck in a locally optimal solution, MSCF was used instead of LF distribution.

Step 3: Adapted search around prey and attack (MS)

The MS_2 search step is followed to correctly locate the prey's district. In what is known as an attack, the aquila carefully investigates the area surrounding the target and uses fake attacks to gauge the prey's reaction.

$$MS_3(i, j) = lb_j + r (ub_j - lb_j) + r \times (S_R(j) - S_{best}(j)) \times M - SCF(t) \times \left(1 - \frac{t}{T}\right) \quad (9)$$

where, $S_R(j)$ represents the accidental set of keys, and $MS_3(i, j)$ specifies the current key for t epochs.

Step 4: Walk and grab attack (S)

The last strategy for finding prey involves an aquila's aerial attack, which is triggered by the prey's motion. One way to describe this hunting technique is "walk and grab prey" as follows:

$$S_4(t+1) = Q_F \times S_{best}(t) - (G_1 \times S(t) \times r) - G_2 \times lev(D) \quad (10)$$

$$Q_F = t^{\frac{2 \times r - 1}{(1-r)^2}} \quad (11)$$

$$G_1 = 2 \times \text{random} - 1 \quad (12)$$

$$G_2 = 2 \times \left(1 - \frac{t}{T}\right) \quad (13)$$

where, $S_4(t+1) lev(D)$ displays the Levy distribution for the D variety, while G_1 represents the solution that has been achieved thus far. The quality is used to process and epitomize each type of movement an eagle makes during a hunt, and G_2 is the hunting combat slope.

$$Fitness = \max(P) \quad (14)$$

$$P = \frac{TP}{TP + FP} \quad (15)$$

TP and FP stand for the values of true positive and false positive, respectively. An essential part of the MAO approach is the fitness option. Encoder performance was used to find the best possible option. At this point, the most important criterion used to create a fitness function (FF) is its performance value.

On the other hand, while the suggested model did use many frequency components to enhance characteristics, each component of the channel features only used one frequency component. The insufficient depiction of a single channel led to inadequate channel modeling, as it failed to account for the interplay between these components. Out of all of them, the proposed MBSCAN showed the most promising trial outcomes.

The MBSCAN is an advanced DL architecture specifically designed for image analysis tasks. At its core, MBSCA consists of multiple branches, each incorporating spectral channel attention mechanisms. These attention mechanisms enable the network to dynamically adjust its focus on relevant image features by selectively attending to informative channels across different spectral frequencies. This capability is particularly valuable in scenarios where images contain complex spatial and spectral characteristics, such as medical imaging or remote sensing applications.

The rationale behind leveraging pre-trained CNN models like VGG16 and VGG19 lies in their established efficacy in feature extraction and representation learning. VGG16 and VGG19 are renowned for their deep architectures, comprising multiple layers of convolutional and pooling operations, followed by fully connected layers for classification. These models have been pre-trained on large-scale image datasets, such as ImageNet, which contain millions of labeled images across thousands of classes. As a result, the learned features in these models capture a broad range of visual patterns and semantics, making them well-suited for transfer learning. In MBSCA, the pre-trained VGG models serve as feature extractors within each branch of the network. By leveraging the hierarchical representations learned by VGG16 and VGG19, MBSCA can effectively capture and encode complex image features across different spectral channels. This not only enhances the network's discriminative power but also enables it to generalize well to unseen data, especially in tasks where labeled data is limited.

Furthermore, the choice between VGG16 and VGG19 may depend on factors such as the complexity of the image dataset and the computational resources available. With more convolutional layers and a deeper architecture, VGG19 may be able to capture more complex features, but it also needs more computing power for training and inference than VGG16. Therefore, the selection between these models should be based on a trade-off between model complexity and computational efficiency, tailored to the specific requirements of the image analysis task at hand.

5 Results and Discussion

Accuracy, precision, recall, and F1-score are performance metrics used to evaluate diagnostic performance at the picture level, in line with the study's recommendations for processing binary balanced data [29]. Network training hyperparameters include an initial learning rate of 0.001 and a decay rate that halves the current rate after five iterations. By utilizing the MAO optimizer with a value of 0.9, the loss function was controlled towards a global minimum and prevented from reaching suboptimal solutions. All models underwent training using label smoothing and cosine learning rate decay over 100 epochs. The PyTorch [30] framework was employed to execute all experiments on a server equipped with an RTX 3090 GPU. Figure 1 illustrates the suggested model's accuracy on both the training and validation datasets.

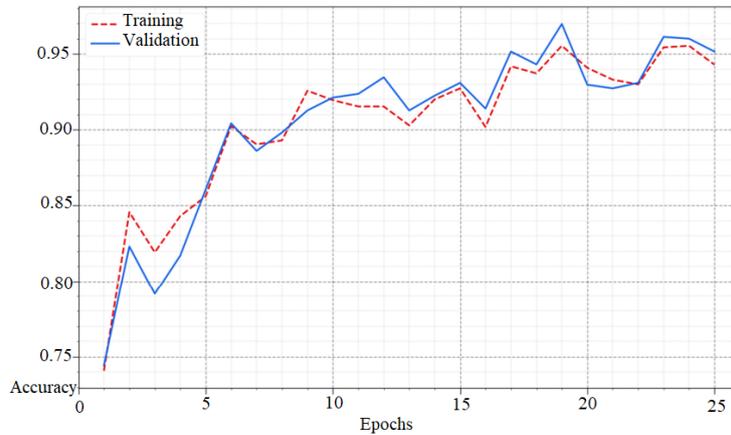


Figure 1. Accuracy of the proposed model on training and validation datasets

5.1 Validation Analysis of the Proposed Model

The accuracy of the proposed model was evaluated across various images, categorised into different classes.

Table 2. Investigation of the proposed model in terms of accuracy

Types	Normal	Intact DIs	Fractured DIs (Type I)	Fractured DIs (Type II)
Image_1	0.99756	0.98193	0.99514	0.96451
Image_2	0.99775	0.98975	0.99552	0.97972
Image_3	0.99758	0.98710	0.99518	0.97454
Image_4	0.99738	0.98974	0.99479	0.97969
Image_5	0.99777	0.97334	0.99556	0.94807
Image_6	0.99701	0.98348	0.99404	0.96751
Image_7	0.99761	0.97269	0.99523	0.94684
Image_8	0.99729	0.99024	0.99459	0.98068
Image_9	0.99895	0.98740	0.99790	0.97512
Image_10	0.99882	0.92817	0.99765	0.86597

Table 2 investigates the proposed model in terms of accuracy. In the analysis of Image_1, the proposed model demonstrated an accuracy of 0.99756 for the normal category, 0.98193 for intact DIs, and 0.99514 and 0.96451 for fractured DIs, Type I and Type II, respectively. Image_2 achieved accuracies of 0.99775 for the normal category, 0.98975 for intact DIs, and 0.99552 and 0.97972 for fractured DIs, Type I and Type II, respectively. In Image_3, the accuracy was 0.99758 for the normal category, 0.98710 for intact DIs, and 0.99518 and 0.97454 for fractured DIs, Type I and Type II, respectively. Image_4 displayed accuracies of 0.99738 for the normal category, 0.98974 for intact DIs, and 0.99479 and 0.97969 for fractured DIs, Type I and Type II, respectively. For Image_5, accuracies of 0.99777 were achieved for the normal category and 0.97334 for intact DIs, while for fractured DIs, Type II, the accuracy was 0.94807. Image_6 indicated accuracies of 0.99701 for the normal category, 0.98348 for intact DIs, and 0.99404 and 0.96751 for fractured DIs, Type I and Type II, respectively. In Image_7, the normal category had an accuracy of 0.99761, while intact DIs achieved 0.97269. Accuracies for fractured DIs, Type I and Type II were 0.99523 and 0.94684, respectively. For Image_8, the accuracy for the normal category was 0.99729, 0.99024 for intact DIs, and 0.98068 for fractured DIs, Type II. Image_9 showed an accuracy of 0.99895 for the normal category, 0.98740 for intact DIs, and 0.99790 and 0.97512 for fractured DIs, Type I and Type II, respectively. Finally, for Image_10, the accuracy of the normal category was 0.99882, while intact DIs had 0.92817, and fractured DIs, Type I and Type II showed accuracies of 0.99765 and 0.86597, respectively.

Figure 2 shows a visual analysis of the proposed model.

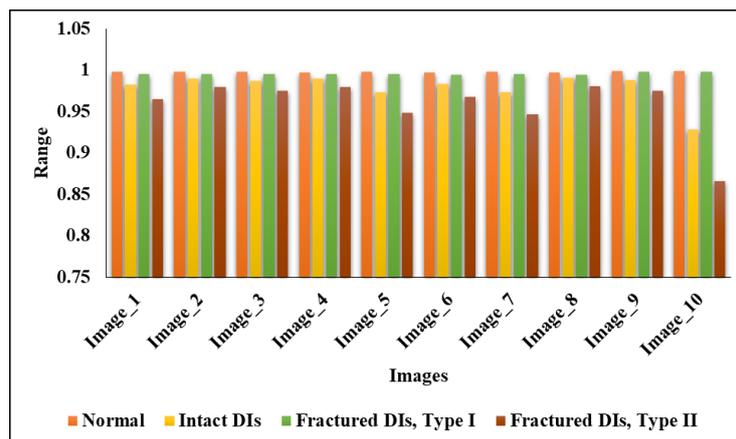


Figure 2. Visual analysis of the proposed model

5.2 Comparative Analysis of the Proposed Model with Existing Procedures

In terms of the effectiveness of the wished-for model, this study considered existing models from related works and tested them with a dataset. Researchers have not used the dataset extensively. Therefore, all models were implemented.

Table 3 shows a relative analysis of the proposed model with existing techniques. Firstly, the MSPENet scheme [15] achieved an accuracy of 89.7%, with precision, recall, and F1-score values of 89.1%, 88.0%, and 89.4%, respectively. Next, the random forest scheme [16, 17] exhibited an accuracy of 92.7%, with precision, recall, and F1-score values

of 92.4%, 91.5%, and 92.7%, respectively. The Support Vector Machine (SVM) scheme [17] demonstrated an accuracy of 94.3%, with precision, recall, and F1-score values of 94.1%, 93.2%, and 94.2%, respectively. The VGG16 scheme [19] yielded an accuracy of 91.6%, with recall and F1-score values of 90.5% and 91.4%, respectively. The CNN scheme [20, 21] displayed an accuracy of 87.2%, with precision, recall, and F1-score values of 87.3%, 86.7%, and 86.3%, respectively. Finally, the MBSCA-MAO scheme showcased an accuracy of 95.7%, with precision, recall, and F1-score values of 95.2%, 94.3%, and 95.6%, respectively.

Table 3. Comparative analysis of the proposed model with existing techniques

Architectures	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
MSPENet [15]	89.7	89.1	88.0	89.4
Random forest [16, 17]	92.7	92.4	91.5	92.7
SVM [17]	94.3	94.1	93.2	94.2
VGG16 [19]	91.5	91.6	90.5	91.4
CNN [20, 21]	87.2	87.3	86.7	86.3
MBSCA-MAO	95.7	95.2	94.3	95.6

Figure 3 shows a visual representation of the proposed model. Figure 4 shows a graphical description of different models for DIS.

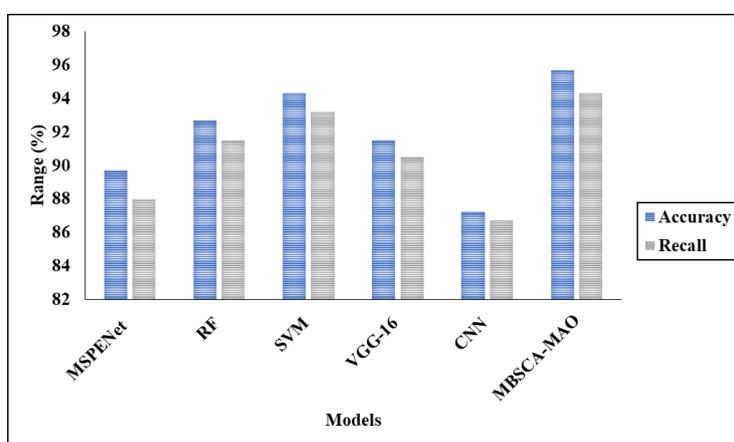


Figure 3. A visual representation of the proposed model

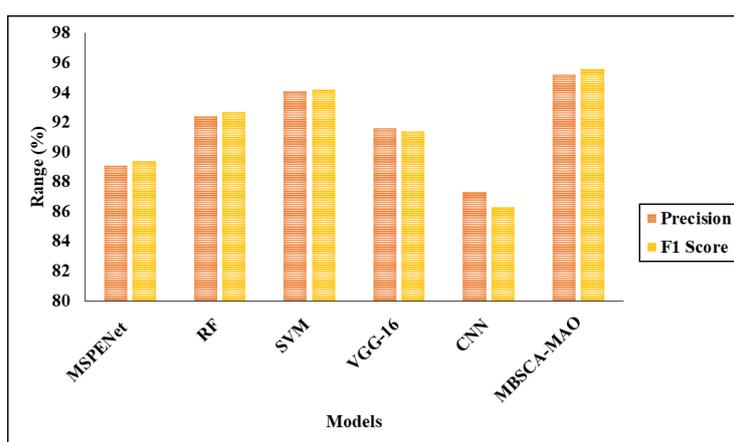


Figure 4. A graphical description of different models for DIS

6 Conclusions and Future Work

It was found that feature extraction architectures like VGG16, VGG19, and the MBSCA-MAO achieved satisfactory accuracy in identifying and categorizing fractured DIs. Notably, the automated DCNN architecture, utilizing input images, demonstrated the highest performance. The finely tuned CNNs, VGG16 and VGG19, excelled particularly in

classification tasks. Moreover, Grad-CAM analysis revealed an understanding of each network's convolutional layers regarding implant fixtures, which holds significance in identifying DI brands from input images. However, further clinical and prospective evidence is necessary to validate the effectiveness of DCNN construction practices.

The multi-branch methodology of MBSCA-MAO enables effective integration of a wide range of frequency information, encompassing both low- and high-frequency components. While the model performs well on images with clear borders and strong contrast, recognizing low-resolution photos with fuzzy borders remains challenging.

Several caveats and potential avenues for further research were proposed in this study. A key challenge is the scarcity of datasets containing fractured DI imaging, owing to the infrequent occurrence of DI fractures. Despite analyzing nearly 20,000 radiographs from two dental clinics, the dataset only included 194 images of fractured DIs. To enhance clinical applicability in implant dentistry, collecting a larger, higher-quality dataset from diverse dental institutions is imperative.

Another limitation of this study is that it uses poor-resolution picture datasets for training and validation of the proposed architecture. Due to resource constraints, such as storage space and processing power, the study had to rely on cropped and downscaled low-resolution panoramic and periapical images. Further investigation is warranted to assess whether a high-resolution image dataset could enhance classification accuracy. Further research could encompass several avenues. For example, advanced imaging techniques, such as cone beam computed tomography (CBCT) or three-dimensional (3D) imaging, could be incorporated to enhance the classification accuracy of DIS. These techniques can provide more detailed information about the implant structure and surrounding anatomy, which may improve the performance of the classification model.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] A. Kohlakala, J. Coetzer, J. Bertels, and D. Vandermeulen, "Deep learning-based dental implant recognition using synthetic X-ray images," *Med. Biol. Eng. Comput.*, vol. 60, pp. 2951–2968, 2022. <https://doi.org/10.1007/s11517-022-02642-9>
- [2] B. Çelik and M. E. Çelik, "Automated detection of dental restorations using deep learning on panoramic radiographs," *Dentomaxillofac. Radiol.*, vol. 51, no. 8, p. 20220244, 2022. <https://doi.org/10.1259/dmfr.20220244>
- [3] A. Alzaid, A. Wignall, S. Dogramadzi, H. Pandit, and S. Q. Xie, "Automatic detection and classification of peri-prosthetic femur fracture," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 17, pp. 649–660, 2022. <https://doi.org/10.1007/s11548-021-02552-5>
- [4] A. Chaurasia, A. Namachivayam, R. B. Koca-Ünsal, and J. H. Lee, "Deep-learning performance in identifying and classifying dental implant systems from dental imaging: A systematic review and meta-analysis," *J. Periodont. Implant Sci.*, vol. 53, 2023. <https://doi.org/10.5051/jpis.2300160008>
- [5] J. H. Lee, Y. T. Kim, J. B. Lee, and S. N. Jeong, "Deep learning improves implant classification by dental professionals: A multi-center evaluation of accuracy and efficiency," *J. Periodont. Implant Sci.*, vol. 52, no. 3, p. 220, 2022. <https://doi.org/10.5051/jpis.2104080204>
- [6] M. Hadj Saïd, M. K. Le Roux, J. H. Catherine, and R. Lan, "Development of an artificial intelligence model to identify a dental implant from a radiograph," *Int. J. Oral Maxillofac. Implants*, vol. 36, no. 6, pp. 1077–1082, 2020. <https://doi.org/10.11607/jomi.8060>
- [7] Q. Q. Nie, C. Li, J. Z. Yang, Y. D. Yao, H. X. Sun, T. Jiang, M. Grzegorzec, A. Chen, H. Chen, W. M. Hu, R. Li, J. W. Zhang, and D. N. Wang, "OII-DS: A benchmark oral implant image dataset for object detection and image classification evaluation," *Comput. Biol. Med.*, vol. 167, p. 107620, 2023. <https://doi.org/10.1016/j.combiomed.2023.107620>
- [8] H. Inamanamelluri, V. Pulipati, N. Pradhan, P. Chintamaneni, M. Manur, and R. Vatambeti, "Classification of a new-born infant's jaundice symptoms using a binary spring search algorithm with machine learning," *Rev. Intell. Artif.*, vol. 37, no. 2, pp. 257–265, 2023.
- [9] C. Y. Hsiao, H. Bai, H. Ling, and J. Yang, "Artificial intelligence in identifying dental implant systems on radiographs," *Int. J. Periodont. Restor. Dent.*, vol. 43, no. 3, pp. 363–368, 2023. <https://doi.org/10.11607/prd.5781>
- [10] M. Al-Sarem, M. Al-Asali, A. Y. Alqutaibi, and F. Saeed, "Enhanced tooth region detection using pretrained deep learning models," *Int. J. Environ. Res. Public Health*, vol. 19, no. 22, p. 15414, 2022. <https://doi.org/10.3900/ijerph192215414>

- [11] R. H. Putra, C. Doi, N. Yoda, E. R. Astuti, and K. Sasaki, "Current applications and development of artificial intelligence for digital dental radiography," *Dentomaxillofac. Radiol.*, vol. 51, no. 1, p. 20210197, 2022. <https://doi.org/10.1259/DMFR.20210197>
- [12] M. Bonfanti-Gris, A. Garcia-Cañas, R. Alonso-Calvo, M. P. S. Rodriguez-Manzaneque, and G. P. Ramiro, "Evaluation of an artificial intelligence web-based software to detect and classify dental structures and treatments in panoramic radiographs," *J. Dent.*, vol. 126, p. 104301, 2022. <https://doi.org/10.1016/j.jdent.2022.104301>
- [13] H. Mohammad-Rahimi, S. R. Motamedian, Z. Pirayesh, A. Haiat, S. Zahedrozegar, E. Mahmoudinia, M. H. Rohban, J. Krois, J. H. Lee, and F. Schwendicke, "Deep learning in periodontology and oral implantology: A scoping review," *J. Periodontol. Res.*, vol. 57, no. 5, pp. 942–951, 2022. <https://doi.org/10.1111/jre.13037>
- [14] N. N. Huang, P. Liu, Y. L. Yan, L. Xu, Y. D. Huang, G. Fu, Y. Q. Lan, S. Yang, J. L. Song, and Y. Z. Li, "Predicting the risk of dental implant loss using deep learning," *J. Clin. Periodontol.*, vol. 49, no. 9, pp. 872–883, 2022. <https://doi.org/10.1111/jcpe.13689>
- [15] X. Q. Yang, X. G. Li, X. C. Li, W. T. Chen, L. L. Shen, X. Li, and Y. Q. Deng, "Two-stream regression network for dental implant position prediction," *Expert Syst. Appl.*, vol. 235, p. 121135, 2024. <https://doi.org/10.1016/j.eswa.2023.121135>
- [16] R. A. Ramachandran, V. A. Barão, D. Ozevin, C. Sukotjo, and M. Mathew, "Early predicting tribocorrosion rate of dental implant titanium materials using random forest machine learning models," *Tribol. Int.*, vol. 187, p. 108735, 2023. <https://doi.org/10.1016/j.triboint.2023.108735>
- [17] P. Rekawek, E. A. Herbst, A. Suri, B. P. Ford, C. S. Rajapakse, and N. Panchal, "Machine learning and artificial intelligence: A web-based implant failure and peri-implantitis prediction model for clinicians," *Int. J. Oral Maxillofac. Implants*, vol. 38, no. 3, pp. 576–582, 2023.
- [18] W. Park, J. K. Huh, and J. H. Lee, "Automated deep learning for classification of dental implant radiographs using a large multi-center dataset," *Sci. Rep.*, vol. 13, no. 1, p. 4862, 2023. <https://doi.org/10.1038/s41598-023-32118-1>
- [19] J. H. Park, H. S. Moon, H. I. Jung, J. Hwang, Y. H. Choi, and J. E. Kim, "Deep learning and clustering approaches for dental implant size classification based on periapical radiographs," *Sci. Rep.*, vol. 13, no. 1, p. 16856, 2023. <https://doi.org/10.1038/s41598-023-42385-7>
- [20] Y. C. Chen, M. Y. Chen, T. Y. Chen, M. L. Chan, Y. Y. Huang, Y. L. Liu, P. T. Lee, G. J. Lin, T. F. Li, C. A. Chen, S. L. Chen, K. C. Li, and P. A. R. Abu, "Improving dental implant outcomes: CNN-based system accurately measures degree of peri-implantitis damage on periapical film," *Bioengineering*, vol. 10, no. 6, p. 640, 2023. <https://doi.org/10.3390/bioengineering10060640>
- [21] W. Park, F. Schwendicke, J. Krois, J. K. Huh, and J. H. Lee, "Identification of dental implant systems using a large-scale multicenter data set," *J. Dent. Res.*, vol. 102, no. 7, pp. 727–733, 2023.
- [22] D. W. Lee, S. Y. Kim, S. N. Jeong, and J. H. Lee, "Artificial intelligence in fractured dental implant detection and classification: Evaluation using dataset from two dental hospitals," *Diagnostics*, vol. 11, no. 2, p. 233, 2021. <https://doi.org/10.3390/diagnostics11020233>
- [23] E. von Elm, D. Altman, M. Egger, S. Pocock, P. C. Gøtzsche, and J. Vandenbroucke, "The strengthening the reporting of observational studies in epidemiology (strobe) statement: Guidelines for reporting observational studies," *J. Clin. Epidemiol.*, vol. 335, no. 7624, pp. 806–808, 2007. <https://doi.org/10.1136/bmj.39335.541782.AD>
- [24] J. Lee, Y. Kim, S. Jeong, N. Kim, and D. Lee, "Incidence and pattern of implant fractures: A long-term follow-up multicenter study," *Clin. Implant. Dent. Relat. Res.*, vol. 20, no. 4, pp. 463–469, 2018. <https://doi.org/10.1111/cid.12621>
- [25] H. Macherla, G. Kotapati, M. Sunitha, K. Chittipireddy, B. Attuluri, and R. Vatambeti, "Deep learning framework-based chaotic hunger games search optimization algorithm for prediction of air quality index," *Ing. Syst. Inf.*, vol. 28, no. 2, pp. 433–441, 2023. <https://doi.org/10.18280/isi.280219>
- [26] S. Baswaraju, V. U. Maheswari, K. K. Chennam, A. Thirumalraj, M. P. Kantipudi, and R. Aluvalu, "Future food production prediction using AROA based hybrid deep learning model in agri-sector," *Hum-Cent. Intell. Syst.*, vol. 3, pp. 521–536, 2023. <https://doi.org/10.1007/s44230-023-00046-y>
- [27] R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the International Conference on Computer Vision*, Venice, Italy, 2017, pp. 618–626. <https://doi.org/10.1109/ICCV.2017.74>
- [28] K. M. Mumenin, P. Biswas, M. A. M. Khan, A. S. Alammary, and A. A. Nahid, "A modified aquila-based optimized XGBoost framework for detecting probable seizure status in neonates," *Sensors*, vol. 23, no. 16, p. 7037, 2023. <https://doi.org/10.3390/s23167037>
- [29] H. Zerouaoui and A. Idri, "Reviewing machine learning and image processing based decision-making systems for breast cancer imaging," *J. Med. Syst.*, vol. 45, no. 1, p. 8, 2021. <https://doi.org/10.1007/s10916-020-01689-1>
- [30] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga,

A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in *the 33rd International Conference on Neural Information Processing Systems*, Vancouver, Canada, 2019, pp. 3933–3944.