Original Articles

# Development of a Deep Learning Model for Diagnosing Class III Malocclusion in Pediatric Patients Using Lateral Cephalometric Radiographs

**Chaypat Simsuchin[1] and Supattanawaree Thipcharoen[2]**

[1]Department of Orthodontic and Orthopedic Craniofacial Growth Modification, Faculty of Dentistry, Nation University, Chiang Mai, Thailand

[2]Department of Data Science and Digital Innovation, Faculty of Innovation Technology and Creativity, The Far Eastern University, Chiang Mai, Thailand



## Abstract

This research aims to develop and evaluate deep learning models for diagnosing Class III malocclusion in pediatric patients using lateral cephalometric radiographs. The study compared an artificial neural network (ANN) model and a convolutional neural network (CNN) model with image embedding and logistic regression. Radiographs from patients aged 3-12 years (2007-2023) were analyzed. Model performance was evaluated using classification accuracy, sensitivity, precision, F1 score, and area under the ROC curve (AUC). Contrary to expectations, the ANN model outperformed the CNN model. The ANN model achieved 90.3% classification accuracy, high sensitivity and precision, an F1 score of 0.902, and an AUC of 0.948, indicating excellent discrimination ability. In contrast, the CNN model showed a lower performance with 71.6% classification accuracy, an F1 score of 0.715, and an AUC of 0.750. Despite the underperformance of the CNN model, potential improvements include data augmentation, larger diverse datasets, and exploring advanced CNN architectures. The superior performance of the ANN model suggests its potential as a reliable diagnostic tool for general dentists in early screening of Class III malocclusion. This study demonstrates the promise of deep learning in orthodontic diagnosis, particularly the effectiveness of ANN models. Further research is needed to enhance CNN performance and validate findings with larger, diverse datasets. Developing such AI-based diagnostic tools could significantly impact early detection, timely referrals, and treatment planning for Class III malocclusion in pediatric patients.

**Keywords:** Artificial Neural Networks, Class III malocclusion, Convolutional Neural Networks, Deep learning, Pediatric orthodontics

**Received date:** Sep 26, 2024 **Revised date:** Jul 30, 2025 **Accepted date:** Aug 18, 2025

**Doi:**

**Correspondence to:**

Chaypat Simsuchin, Department of Orthodontic and Orthopedic Craniofacial Growth Modification, Faculty of Dentistry, Nation University, Chiang Mai 50210, Thailand Tel: 095-394-7419 Email: chaypat_sim@nation.ac.th

## Introduction

Class III Malocclusion is a common disorder in Asian populations, with a prevalence of up to 14%, primarily caused by maxillary deficiencies.[1] However, prevalence may vary across different population groups, with rates as high as 15% or more in some Southeast Asian countries.[2] A study by Ellis, McNamara, and Guyer found that 40% to 60% of patients had a correlation with small maxillary size, which is a significant factor contributing to this type of malocclusion.[3]

There are several approaches to treating Class III Malocclusion. In the deciduous dentition stage, early treatment using Prefabricated Functional Appliances (PFA) can induce mandibular growth retraction, correct occlusal discrepancies, and adjust perioral muscle function.[4,5] For patients in the mixed dentition phase with retruded maxilla, using a face mask to stimulate midface growth with orthopedic force is an effective method.[6] (Fig. 1)

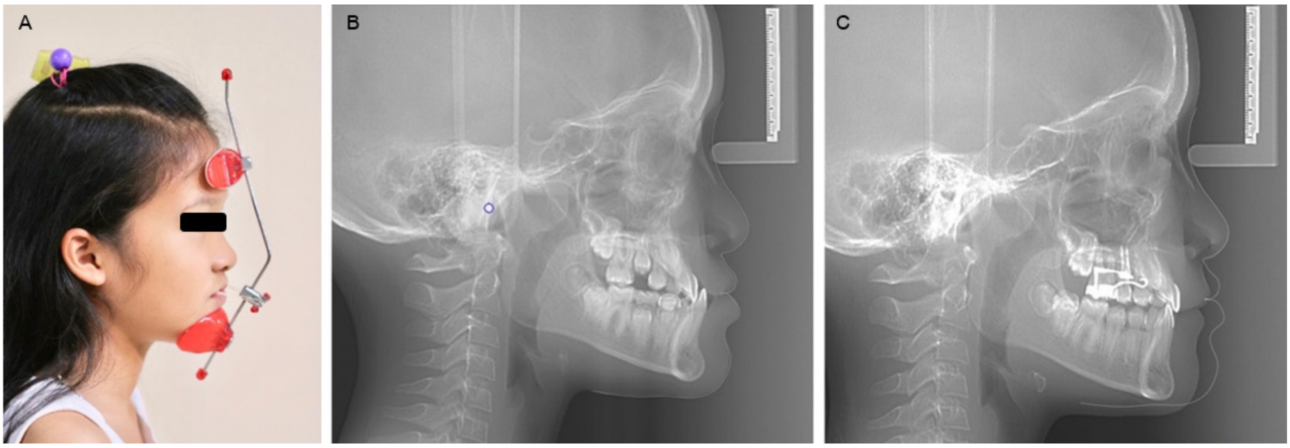


***Figure 1*** *Image A shows the use of a face mask to stimulate midface growth. Image B displays a radiograph before treatment, and Image C shows a radiograph after treatment using a face mask to stimulate midface growth.*

Early diagnosis and appropriate treatment are crucial. Dentists should be able to diagnose this condition from the late deciduous dentition stage or before the end of the early mixed dentition stage, which is the most suitable time for treatment. Delayed treatment may lead to difficulties in correction and potential surgical intervention in the future.[7]

Currently, diagnosing Class III malocclusion in pediatric orthodontics requires analyzing lateral cephalometric radiographs, a complex and time-consuming process. Although cephalometric analysis software is available, it still heavily relies on the skills, knowledge, and experience of orthodontists. Advancements in Artificial Intelligence (AI) and Deep Learning have opened up opportunities to develop models that can predict malocclusions directly from radiographs, without the need for traditional angle and distance measurements. In such cases, using Convolutional Neural Networks (CNN) in conjunction with Image Embedding techniques and Logistic Regression algorithms can efficiently classify and predict malocclusions.

Given these advancements, this study hypothesizes that deep learning models, specifically Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs), can accurately diagnose Class III malocclusion in pediatric patients using lateral cephalometric radiographs. Furthermore, these deep learning models are expected to effectively predict the necessity of using face masks to stimulate midface growth in pediatric patients with Class III malocclusion.

Artificial Intelligence is a technology that mimics human brain function, capable of analyzing data through algorithms, which are components of Machine Learning. Its capabilities are enhanced by Deep Learning through Neural Network algorithms.

Neural networks are mathematical models that simulate human learning through repetition and trial and error, similar to the brain's learning system.[8] They are machine learning algorithms with a basic structure consisting of an input layer, hidden layers, and an output layer. Each layer comprises processing units or neurons connected by weights Learning in artificial neural networks occurs through the process of adjusting these weights (Fig. 2) using various learning algorithms such as Backpropagation.[9]

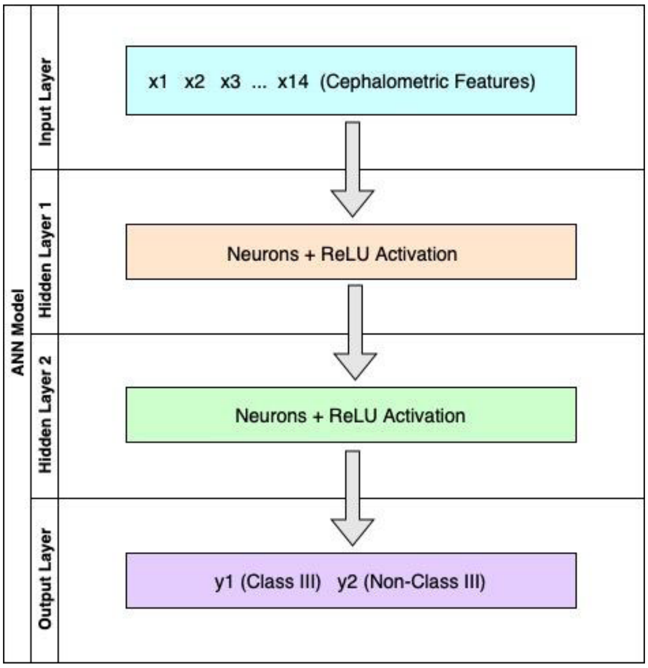


***Figure 2*** *Structure of the Feedforward Artificial Neural Network (ANN) used in this study.*
*The model consists of three layers:*
- *Input Layer (14 nodes): Receives cephalometric features (e.g., ANB angle, Wits appraisal, Overjet, etc.).*
- *Hidden Layer (optimized to 16 neurons via grid search): Each neuron applies weights ($w_i$), bias (b), and activation function (ReLU: $f(x) = max(0, x)$).*
- *Output Layer (2 nodes): Classifies images as Class III (1) or Non-Class III (0) using a Softmax activation function.*
- *Bias Node (+1): Included in each layer to improve learning flexibility.*
- *The connections between nodes represent weight calculations and learning processes in ANN.*



**Deep learning based on neural network characteristics can be categorized into three primary types:**

**1. Artificial Neural Networks (ANNs)** ANNs are basic feedforward networks designed to process one-dimensional data. These networks operate in a unidirectional flow, where inputs are passed forward through weighted connections without feedback loops. The general mathematical representation of an ANN is:

$$y=f(i=1\sum nwi\cdot xi+b)\} \qquad (1)$$

Where:
- y is the output of the neuron
- $x_i$ represents the input features
- $w_i$ are the corresponding weights
- b is the bias term
- f is the activation function (e.g., ReLU, sigmoid)

This equation illustrates how each input is multiplied by its weight, summed with the bias, and passed through an activation function to produce the output.

**2. Recurrent Neural Networks (RNNs)** RNNs are designed to handle sequential or time-series data. Unlike ANNs, RNNs maintain a hidden state that captures information from previous time steps, allowing the network to model dependencies across sequences. The basic RNN equation is:

$$h_t = f(W_h \cdot h_{t-1} + W_x \cdot x_t + b) \qquad (2)$$

Where:
- $h_t$: hidden state at time t
- $W_h$: weight for hidden
- $W_x$: weight for input $x_t$
- b: bias

This recurrent formulation enables RNNs to effectively retain and utilize prior contextual information throughout a sequence, thereby enhancing their ability to model temporal dependencies. As a result, RNNs have demonstrated strong performance in a variety of tasks that rely on sequential data processing, including language modeling, speech recognition, time-series forecasting, and other forms of sequential pattern prediction.

**3. Convolutional Neural Networks (CNNs)** accept 2-dimensional image data, with networks connected by convolutional layers, enhancing human-like vision capabilities, making them suitable for image and video processing. The convolution operation in CNNs can be represented as:

$$(f * g)(x, y) = \sum_i \sum_j f(i, j) \cdot g(x - i, y - j) \quad (3)$$

Where:

- $f(i, j)$ is the input image pixel at location $(i, j)$
- $g$ is the kernel (filter)
- $*$ denotes the convolution operation
- $(x, y)$ is the location of the output feature map

This operation allows CNNs to learn spatially local patterns such as edges, textures, and shapes, which are essential for medical image analysis, including cephalometric radiographs.

Neural Networks, as machine learning algorithms, can be applied to deep learning, increasing the capabilities of AI technology to work more like the human brain. With this potential, neural networks have gained significant attention, with Computer Vision being used to analyze medical images.[10] Studies have explored the potential for analyzing dental radiographs. However, a literature review reveals that while there have been studies using machine learning techniques to analyze orthodontic radiographs, such as Taraji *et al.*'s study[11] of identifying important cephalometric characteristics for predicting treatment decisions in Class III malocclusion in adults, and Gabriele *et al.*'s study[12] of attempting to predict treatment outcomes in prepubertal Class III malocclusion patients, there is a lack of studies focusing on using deep learning techniques to diagnose and predict the need for face masks to stimulate midface growth in pediatric patients with Class III malocclusions. Moreover, no studies have compared the efficiency of artificial neural network models and convolutional neural network models in analyzing cephalometric radiographs to diagnose Class III malocclusion in pediatric patients.

This study aims to develop and compare deep learning models, specifically Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs), for diagnosing Class III malocclusion in pediatric patients from lateral cephalometric radiographs and predicting the necessity of using face masks to stimulate midface growth (Fig. 3). The study hypothesizes that Convolutional Neural Networks (CNNs) will demonstrate superior efficiency and accuracy compared to Artificial Neural Networks (ANNs) in analyzing cephalometric radiographs for Class III malocclusion diagnosis and treatment prediction. The results of this study will help understand the advantages and limitations of each technique, leading to the development of highly efficient diagnostic tools to support the decision-making of orthodontists in diagnosis and treatment planning,[13,14] increase speed and accuracy in cephalometric radiograph analysis, reduce the workload of orthodontists, and minimize errors that may occur from traditional analysis.[15] It will also provide a pathway for developing automatic diagnostic systems for other dental abnormalities.[13]

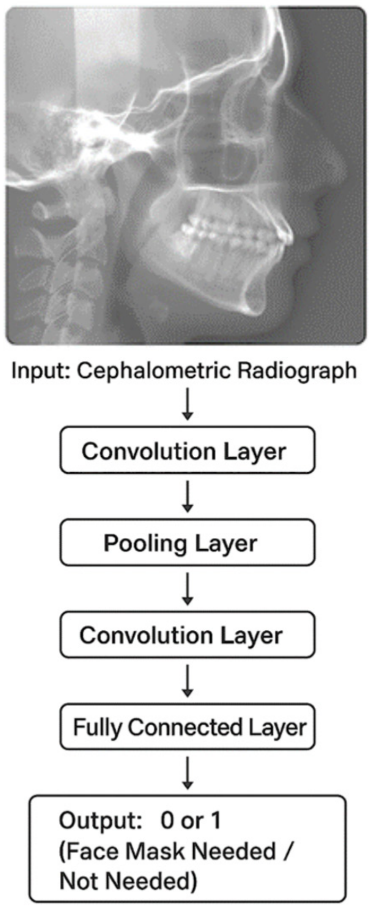


***Figure 3*** *A schematic flowchart illustrating the Convolutional Neural Network (CNN) model for binary classification of lateral cephalometric radiographs. The input cephalometric X-ray is processed through multiple convolutional and pooling layers, followed by a fully connected layer, to predict the necessity of using face masks to stimulate midface growth. The output is binary: 1 indicates that face masks are needed, while 0 indicates that they are not required.*

## Materials and Methods

This research is a comparative analytical study designed to evaluate and compare the efficiency of an Artificial Neural Network (ANN) and a Convolutional Neural Network (CNN) in classifying lateral cephalometric radiographs of pediatric patients with Class III malocclusion. As illustrated in the workflow diagram (Fig. 4), the research process consists of three main parts: developing the ANN model from structured data, developing the CNN model from image data, and comparing their performances.

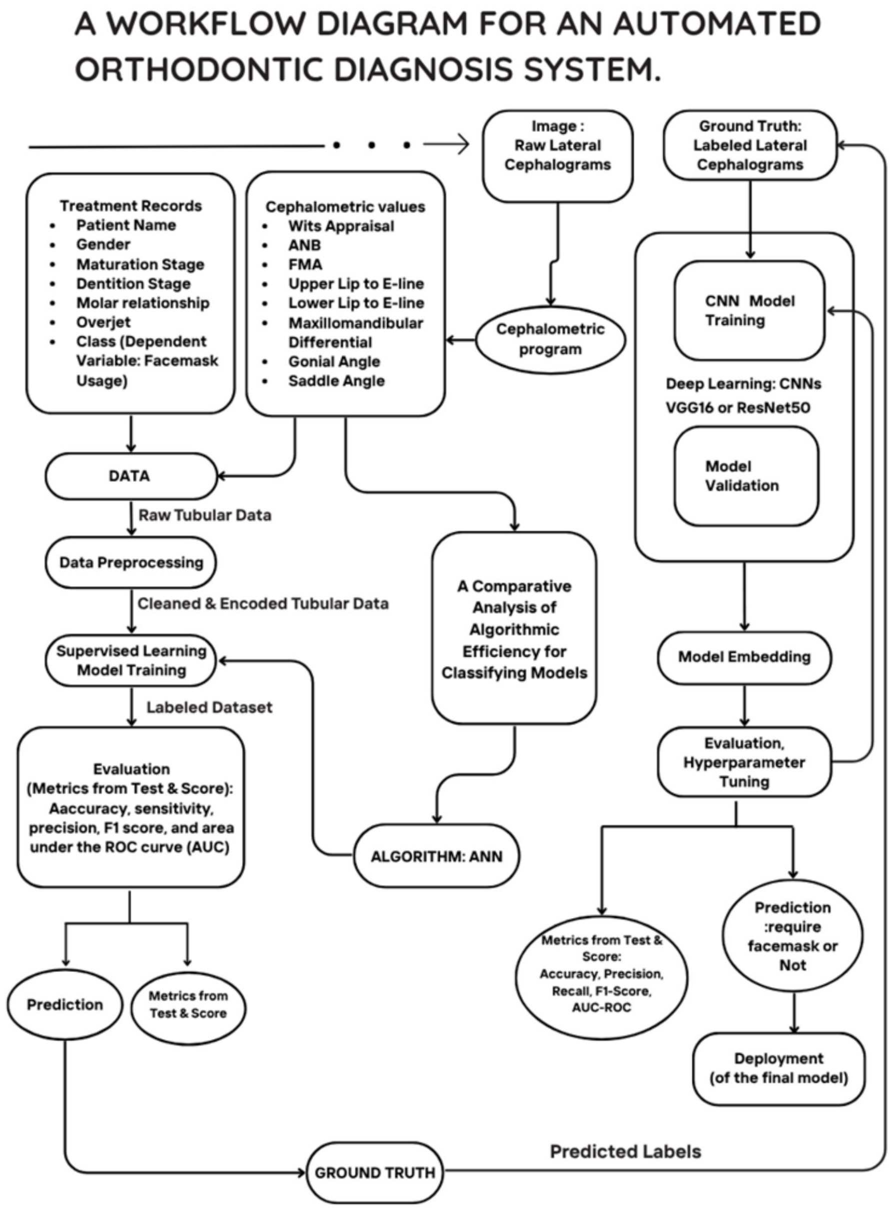


*Figure 4 A workflow diagram for an automated orthodontic diagnosis system.*

**Figure 4** presents the diagram that outlines a workflow for developing an automated orthodontic diagnosis system. It starts with extracting cephalometric values from treatment records, which are then analyzed using cephalometric imaging and deep learning. The results are used for model embedding and evaluation, feeding into an algorithm that performs comparative analysis to classify models. The predictions of the algorithm are assessed using additional metrics, leading to the final deployment of the system for Class III diagnosis and facemask treatment prediction.

**Part 1: Development of the Artificial Neural Network (ANN) Model**

**1.1 Data Collection and Preparation**

**1.1.1 Population, Sample, and Inclusion Criteria Rationale** The study population comprised pediatric patients who received orthodontic treatment at a private dental clinic in Chiang Mai Province, Thailand, between 2007 and 2023. Inclusion criteria were patients aged 3–12 years with good quality lateral cephalometric radiographs and complete clinical diagnostic data. Exclusion criteria included patients with a history of craniofacial surgery, congenital craniofacial abnormalities, or syndromes known to affect craniofacial growth.

The lower age limit of three years was deliberately chosen based on established clinical evidence and best practices. Skeletal discrepancies associated with Class III malocclusion, particularly maxillary deficiency, can be reliably detected at this early stage, as demonstrated by foundational studies.[3] This age aligns with the deciduous dentition stage, which is widely recognized as the optimal period for initiating interceptive orthopedic interventions like facemask therapy.[7] Early diagnosis allows for treatment before the puertal growth spurt limits the efficacy of non-surgical interventions. Furthermore, the use of modern digital lateral cephalometric radiography with low-dose radiation protocols ensures that the diagnostic benefits far outweigh the minimal risks for pediatric patients, a principle supported by guidelines from the American Academy of Pediatric Dentistry and established clinical literature.[21,22]

This research protocol was reviewed and approved by the Institutional Review Board of Ethics Committee of Research Involving Humans, Nation University, Chiang Mai, Thailand (Approval No. NTU. EC.1-041/2024). A final sample of 195 cases was yielded, consisting of 66 boys (33.85%) and 129 girls (66.15%).

**1.1.2 Ground Truth Verification and Data Labeling** To establish highly accurate ground truth labels essential for supervised learning, a rigorous two-step verification protocol was implemented. First, initial cephalometric parameters (e.g., ANB angle, Wits appraisal) were automatically generated for all 195 radiographs using specialized software. Following this, two experienced orthodontists meticulously reviewed each case. They integrated the software-generated skeletal values with a comprehensive clinical assessment of the dental relationship based on Angle's classification. Any discrepancies in classification between the two orthodontists were resolved through a consensus discussion until a definitive classification (Class III or Non-Class III) was reached. This meticulous, expert-validated process ensured the accuracy and reliability of the ground truth labels of the study, categorizing the dataset into 70 Class III malocclusion cases and 125 Non-Class III cases.

**1.1.3 Cephalometric Variable Extraction** For the ANN model, a set of 14 specific cephalometric variables was extracted from each radiograph using CephNinja® Application Software. These variables, forming the structured input for the ANN, were selected based on their established significance in standard orthodontic cephalometric analysis for evaluating skeletal relationships, facial prognathism, and growth patterns pertinent to Class III cases. The variables are detailed as follows:

**1.1.3.1 Skeletal Relationship:** SNA, SNB, and ANB Angles; Wits Appraisal; FMA and GoGn-SN Angles; Maxillary and Mandibular Lengths (Co-A, Co-Gn); Gonial Angle and Saddle Angle.

**1.1.3.2 Dental Relationship:** U1-SN and L1-NB Angles.

**1.1.3.3 Soft Tissue Profile:** Upper and Lower Lip to E-line.

**1.2 AI Model Development Platform and Frameworks**

The deep learning models in this study were developed and evaluated within a unified, high-level environment to ensure consistency and reproducibility. The primary platform used was Orange Data Mining Tools (version 3.39.0), an open-source visual programming software for machine learning and data analysis. While the modeling process was conducted visually, Orange utilizes industry-standard open-source libraries for its underlying computations. This allows for the power of complex frameworks to be leveraged through an intuitive and verifiable workflow:

**1.2.1 ANN Implementation:** The feedforward Artificial Neural Network (ANN) was constructed using the

"Neural Network" widget within Orange. This component functions as an interface to the Scikit-learn library, primarily utilizing its MLPClassifier implementation for network training and optimization.

**1.2.2 CNN Implementation:** The Convolutional Neural Network (CNN) was developed using the "Image Analytics" add-on in Orange. This module integrates with the TensorFlow/Keras framework, enabling the use of pre-trained architectures like ResNet50 for transfer learning and end-to-end image classification tasks directly within the visual workflow.

By using this integrated platform, we were able to rapidly prototype and compare both structured data models (ANN) and image-based models (CNN) while maintaining a high degree of methodological transparency.

**1.3 ANN Model Development and Training**

The structured dataset containing the 14 cephalometric variables was prepared for model training using the Orange Data Mining tool. A key step in data preparation involved normalizing all numerical data to a uniform [0, 1] range to enhance model performance and stability. The total dataset (N=195) was then divided into a training set (70%) and a testing set (30%) using stratified random sampling. This stratification was based on the primary outcome of the clinical need for facemask treatment to ensure that both categories were proportionally represented in the training and testing phases, thereby preventing biased training and ensuring a robust evaluation.

The ANN model was constructed as a feedforward network using the "Neural Network" widget in Orange. The architecture consisted of a 14-node input layer, a single hidden layer with the Rectified Linear Unit (ReLU) activation function, and a two-node output layer using a Softmax function for binary classification. The number of nodes in the hidden layer was optimized via a grid search technique. The model was trained using the Adam optimizer (learning rate = 0.001), categorical cross-entropy as the loss function, and L2 regularization ($\lambda$ = 0.01) to prevent overfitting. A 10-fold cross-validation strategy was employed over 1000 epochs with a batch size of 32 for a robust and generalizable performance evaluation.

**Part 2: Development of the Convolutional Neural Network (CNN) Model**

**2.1 Input Data and Pre-processing**

The CNN model was developed for end-to-end classification directly from the lateral cephalometric radiographic images, aiming to predict the necessity of facemask therapy from raw pixel data without prior extraction of numerical cephalometric values. The input for this model consisted of the same set of 195 radiographs used in Part 1.

The image pre-processing pipeline involved several crucial steps. All radiographs were first resized to a standard 224x224 pixels to ensure uniform input size for the network. Subsequently, pixel values were normalized to a [0, 1] range. Acknowledging the limited size of medical imaging datasets and its potential impact on CNN performance, data augmentation techniques were applied. This is a critical step to artificially increase the size and diversity of the dataset, allowing the CNN to learn more robust and generalizable features. The augmentation included Rotation (±20°), Horizontal Flipping, Shearing (±15%), and Zooming (0.8-1.2x). These techniques have been demonstrated in the literature to significantly improve CNN performance in cephalometric classification tasks.[7]

**2.2 CNN Model Architecture and Training**

A transfer learning approach was employed, utilizing a ResNet50 architecture pre-trained on the ImageNet dataset. This method is highly recommended for medical imaging tasks with limited data, as it leverages features learned from a massive dataset.[5] The model was fine-tuned for this specific task by replacing the final fully connected layers with a Logistic Regression classifier and implementing Global Average Pooling to reduce the risk of overfitting. The dataset was divided into an 80% training set and a 20% test set. The model was trained using the Adam optimizer with a fine-tuning learning rate of 0.0001 for 50 epochs, utilizing a 20% validation split and early stopping criteria to select the best-performing model iteration and prevent overfitting.

**Part 3: Comparative Performance Evaluation**

**3.1 Evaluation Strategy**

The performance of both the ANN and CNN models was comprehensively and separately evaluated to facilitate a direct and fair comparison of their efficiency. The evaluation was based on a standard suite of machine learning metrics, which were calculated from the predictions made by each model on their respective unseen test sets. These metrics were chosen to provide a multidimensional assessment of the diagnostic capability of each model:

**3.1.1 Classification Accuracy (CA):** Measures the overall proportion of correct predictions. (CA = (TP + TN) / (TP + TN + FP + FN))

**3.1.2 Recall (Sensitivity):** Assesses the ability of the model to correctly identify all actual positive cases, crucial for minimizing missed diagnoses. (Recall = TP / (TP + FN))

**3.1.3 Precision:** Evaluates the proportion of positive predictions that were actually correct, important for reducing false alarms. (Precision = TP / (TP + FP))

**3.1.4 F1-score:** The harmonic mean of Precision and Recall, providing a single, balanced measure of the performance of the model. (F1 = 2 × (Precision × Recall) / (Precision + Recall))

**3.1.5 Area Under the ROC Curve (AUC):** Quantifies the ability to distinguish between classes across all thresholds.

This rigorous evaluation strategy ensures a robust comparison, the results of which are presented in the following section to determine the most suitable model for implementation.

## Results

The research results analyze and compare the efficiency of deep learning models between Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) using Image Embedding and Logistic Regression in classifying cephalometric radiographs of pediatric patients with Class III malocclusion as follows:

**1. Efficiency of the Artificial Neural Network model**

The Artificial Neural Network model demonstrates high efficiency, achieving a classification accuracy of up to 90.3%. It exhibits strong performance in terms of both sensitivity and precision, which are crucial metrics for evaluating classification models. Additionally, the model boasts a high F1-score, indicating a well-balanced performance between precision and sensitivity. This combination of high accuracy, sensitivity, precision, and F1-score suggest that the Artificial Neural Network model is highly effective and reliable for its intended classification tasks.

**2. Efficiency of the Convolutional Neural Network model**

The Convolutional Neural Network model, which utilizes Image Embedding and Logistic Regression with Stratified Cross-validation using 20 folds, demonstrates notable efficiency. The Logistic Regression component of this model achieves an Area Under the ROC Curve (AUC) of 0.750, indicating good discriminative ability. The model also exhibits a Classification Accuracy of 0.716, suggesting that it correctly classifies over 71% of the instances. Furthermore, it attains an F1-score of 0.715, which represents a balanced harmony between precision and recall. These metrics collectively indicate that the model performs well in its classification task, showing a good balance between identifying positive cases and avoiding false positives.

**3. Efficiency Comparison**

The comparison reveals that the Artificial Neural Network (ANN) model significantly outperforms the Convolutional Neural Network (CNN) model using Image Embedding and Logistic Regression across all metrics. The ANN model achieves a notably higher Classification Accuracy of 90.3% compared to the CNN model of 71.6%. The performance improvement is substantial, ranging from 19% to 24% across all metrics. The most striking difference is observed in the Area Under the ROC Curve, where the ANN model shows a 24.13% improvement. The consistent superiority of the ANN model across all metrics underscores its significantly better performance, indicating that it is the more efficient and effective choice for the given task. Table 1

*Table 1 Comparison of performance parameters between Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs)*

| Performance parameters | ANNs | CNNs |
|---|---|---|
| Classification Accuracy (CA) | 0.903 | 0.716 |
| Recall (Sensitivity) | 0.903 | 0.716 |
| Precision | 0.903 | 0.714 |
| F1-score | 0.902 | 0.715 |
| Area Under the ROC Curve (AUC) | 0.948 | 0.750 |

**4. Additional Observations**

Additional observations reveal that the CNN model using Image Embedding and Logistic Regression may have room for improvement, such as adjusting hyperparameters or using more complex architectures. Other factors should also be considered, including processing speed and ability to work with different data sizes. Concluding the study, the Artificial Neural Network model demonstrates superior performance in classifying cephalometric radiographs of pediatric patients with Class III malocclusion. However, further studies should be conducted to improve the efficiency of the Convolutional Neural Network model using Image Embedding and Logistic Regression, and statistical tests should be performed to confirm the significance of the results.

## Discussion

While this study revealed that the Artificial Neural Network (ANN) model demonstrated superior performance with an accuracy of 90.3% compared to the Convolutional Neural Network (CNN) model's accuracy of 71.6%, the findings from the ANN model hold significant and immediate clinical applicability. The primary aim of this study was to predict the necessity of using face masks to stimulate midface growth in pediatric Class III malocclusion patients, and the ANN model has proven effective in this regard.

**Clinical Implications and Facemask Treatment Prediction**

The high accuracy of the ANN model (90.3%) in classifying Class III malocclusion and identifying cases that require facemask therapy positions it as a valuable tool for orthodontic practice. The integration of such a model into clinical workflows offers several direct benefits.

- **Identifying Patients at Risk:** The ANN model analyzes crucial cephalometric variables to help clinicians predict which pediatric patients exhibit skeletal characteristics indicative of Class III malocclusion that would benefit significantly from early facemask therapy. This predictive capability allows for proactive identification rather than reactive diagnosis.
- **Reducing Treatment Delays:** Early intervention is paramount for successful Class III correction, especially when skeletal growth modification is targeted. By integrating the ANN model into clinical workflows, clinicians can identify optimal treatment windows for initiating facemask therapy. This early identification is critical for improving prognosis and can significantly reduce the risk of more complex, late-stage interventions or the eventual necessity for orthognathic surgery in adulthood. This aligns with findings by Pattanaik & Mishra6 who emphasized the effectiveness of early intervention. Khan & Karra[7] also discussed the benefits of early treatment in Class III malocclusion, reinforcing the importance of timely intervention facilitated by predictive models.
- **Supporting Personalized Treatment Approaches:** Based on the comprehensive analysis of cephalometric parameters and the predictions of the model, orthodontists can leverage this information to customize highly individualized treatment plans for their patients. This data-driven approach enhances the precision of diagnosis and treatment strategy, moving towards more personalized orthodontics (Taraji *et al.,*[11]).

This expanded discussion, supported by relevant literature, clearly articulates the direct relevance and practical application of the ANN model in facilitating timely and effective facemask interventions for Class III patients. The ability of the ANN to provide accurate and consistent predictions can aid both general dentists in referral decisions and orthodontists in treatment planning.

**Why did ANN outperform CNN in this study?**

In this experiment, the ANN model achieved higher classification accuracy (90.3%) compared to the CNN model (71.6%), as well as superior performance across other evaluation metrics. Several key factors contributed to this outcome.

**Limited dataset size and CNN's data dependency**

A primary factor contributing to the performance discrepancy is the limited size of the dataset in the context of the data-intensive nature of a CNN. Deep CNN architectures are designed to learn robust feature representations by extracting complex spatial patterns from images, a process that requires a large and diverse dataset to be effective. This study, however, utilized a dataset of 195 cephalometric radiographs, which is insufficient to fully leverage the capabilities of a CNN, as these models often need thousands of labeled images to achieve strong generalization. This limitation is less impactful for ANN models, which can be trained effectively on smaller, structured tabular datasets. The strength of ANN lies in its ability to learn directly from numerical feature representations in this case, the 14 well-defined cephalometric measurements rather than from raw pixel data. This distinction is supported by existing literature; for instance, Gabriele et al. (2003) noted that CNNs require significantly more training data than traditional machine learning models for cephalometric classification. Similarly, a recent study by Zhang *et al.* (2024) found that while CNN performance tends to deteriorate on small medical imaging datasets, ANN models maintain their stability when trained on structured numerical inputs.

**Feature Representation and the Role of Data Augmentation**

Another critical distinction contributing to the performance disparity lies in the method of feature representation each model employs. The ANN model operates on structured numerical inputs—the 14 pre-defined cephalometric measurements. This approach makes it inherently less dependent on large datasets because the features provided are already high-level, curated, and known to be diagnostically relevant. In essence, the complex task of feature extraction was performed by human experts, allowing the ANN to focus solely on learning the patterns between these potent predictors.

In contrast, CNNs process raw pixel-based inputs, requiring the model to autonomously learn and extract a hierarchy of spatial features, from simple edges to complex anatomical shapes. This end-to-end feature extraction is a significant challenge for a CNN when working with a small dataset, as it may fail to learn robust and generalizable features. Consequently, this ANN model benefited directly from the expert-curated cephalometric features, leading to its superior accuracy. This finding aligns with existing literature, where studies by Schwendicke *et al.* (2020) emphasized that traditional ANNs remain highly effective for structured dental data,[10] and Aksoy *et al.* (2022) demonstrated that CNNs tend to underperform compared to ANNs when trained on limited orthodontic radiograph datasets.[19]

To overcome these challenges for CNNs, a viable and highly recommended solution is the application of data augmentation. By artificially increasing the size and diversity of the training set, data augmentation can help the CNN learn more robust feature representations, thereby enhancing its predictive performance even with an initially small dataset. For cephalometric radiographs, recommended techniques include Rotation (±20°) to account for positional variations, Horizontal Flipping to increase data diversity, Shearing (±15%) to simulate distorted projections, and Zooming (0.8-1.2x) to improve recognition of anatomical structures at different scales. The effectiveness of these methods is well-supported; for instance, Sabri *et al.* (2023) demonstrated a 12% improvement in CNN accuracy in dental image tasks, while Zhang *et al.* (2024) reported a 15% increase in performance after applying similar techniques on small medical imaging datasets.[17] This underscores the critical role of data augmentation in future efforts to successfully implement CNNs in this domain.

**AI Models vs. Traditional Human Diagnosis**

The traditional diagnosis of Class III malocclusion relies on the expertise of orthodontists in analyzing lateral cephalometric radiographs, a process that can be both time-consuming and subjective. In contrast, artificial intelligence (AI) models offer a promising alternative by providing automated, consistent, and objective assessments. A detailed comparison reveals the distinct advantages and limitations of each approach.

In terms of diagnostic accuracy, the ANN model developed in this study demonstrated a high performance of 90.3%, placing it on par with the typical accuracy range of experienced orthodontists, which is estimated to be between 85-95% depending on their experience.[11] This stands in contrast to the CNN model, which achieved a lower accuracy of 71.6%. A key advantage of the ANN is its high consistency, as it operates based on a standardized algorithm, thereby eliminating the inter-examiner variability commonly observed among human clinicians.[13] While AI models provide objective, data-driven results, the consistency of CNNs can be moderate and data-dependent, whereas human interpretation inherently varies between practitioners.

The most significant divergence is seen in efficiency and processing time. The ANN model can analyze an image in seconds, a stark contrast to the five to ten minutes required for a manual assessment by an orthodontist. This rapid processing capability presents a substantial opportunity to enhance clinical workflow efficiency, particularly in high-volume settings.[10] Furthermore, both the ANN and CNN models eliminate the element of subjectivity, which is a well-documented challenge in traditional cephalometric analysis.

However, human expertise remains indispensable, especially when considering the learning curve and the ability to manage complex cases. While a pre-trained ANN requires minimal setup and the CNN requires fine-tuning, an orthodontist needs years of clinical training and experience. This extensive training is crucial for detecting and managing borderline cases, where nuanced clinical judgment is paramount. The ANN model exhibits a strong ability for pattern recognition from structured data, but an experienced orthodontist excels by integrating a holistic view that includes patient history, growth prediction, and other contextual factors not available in a single radiograph a level of comprehensive decision-making that AI models currently cannot replicate.[11] Therefore, while AI, particularly the ANN model, shows great potential as a powerful diagnostic support tool, the irreplaceable clinical judgment of orthodontists remains central to patient care.

**Improving Deep Learning Approaches and Future Research Directions**

Despite the promising results of the ANN model, there is significant potential for further improving deep learning approaches, particularly concerning the CNN model, for diagnosing Class III malocclusion in pediatric patients. To enhance the performance of the CNN model, future research should focus on data augmentation techniques. As demonstrated by Zhang *et al.*[17] and Sabri *et al.*[18], implementing data augmentation methods such as rotation, flipping, zooming, shearing, and brightness adjustment can diversify the training dataset and address class imbalance issues. Zhang *et al.*[17] reported that data augmentation techniques improved the accuracy of their model by 15%, while Sabri *et al.*[18] found a 12% increase in model sensitivity using similar methods. It is proposed to utilize platforms like Google Colab, which offers free access to GPUs and cloud processing capabilities, to implement various augmentation techniques. These could include image rotation (±20°), horizontal flipping, zooming (0.8-1.2), shearing and translation (±0.2), and brightness and contrast adjustments (0.8-1.2). These techniques will not only increase data diversity but also mitigate the problem of imbalanced learning.

Moreover, the compilation of a larger and more diverse set of cephalometric radiographs, encompassing patients of various ages, genders, and ethnicities, as seen in studies by Aksoy *et al.*[19] and Kaya *et al.*[20] would be crucial. Aksoy *et al.*[19] demonstrated that increasing dataset diversity led to a 10% improvement in diagnostic accuracy across different patient demographics. Similarly, Kaya *et al.*[20] reported enhanced model generalizability when trained on a multi-center dataset. This step, combined with the exploration of more complex CNN architectures and transfer learning from pre-trained models, has the potential to significantly improve the accuracy and reliability of the CNN model. By addressing these aspects in future studies, the aim is to develop a more robust and clinically applicable AI system for orthodontic diagnosis. Such a system could potentially outperform the current ANN model and ultimately improve early detection and treatment planning for Class III malocclusion in pediatric patients.

**Impact of Gender on Class III Malocclusion**

While this study primarily focused on model performance, the influence of gender is a relevant factor in Class III malocclusion. Previous studies, such as Guyer *et al.* (1986)[3], have indicated a higher prevalence of Class III malocclusion in males, potentially due to more pronounced mandibular growth patterns. Although gender was included as a variable, further statistical analysis, such as a Chi-square test, could be performed in future work to investigate its significance as a predictor in our dataset. This would provide deeper, evidence-based insights into gender-related differences for diagnosis and treatment planning.

**Future research should also consider the following areas**

- **Feature extraction optimization** Investigating advanced feature extraction techniques specific to cephalometric landmarks could enhance the model's ability to identify key diagnostic indicators (Polizzi & Leonardi[15]).
- **Multi-modal approach** Integrating other types of diagnostic data, such as intraoral scans or facial photographs, alongside cephalometric radiographs could provide a more comprehensive diagnostic tool.
- **Longitudinal studies** Conducting studies that track patients over time could help in understanding the predictive capabilities of the model for malocclusion progression and treatment outcomes.
- **Explainable AI (XAI)** Developing methods to interpret the decision-making process of the model could increase trust and adoption among clinicians, as well as provide insights into previously unrecognized diagnostic patterns (Schwendicke *et al.,*[10,14]).
- **Clinical validation** Extensive clinical trials comparing the performance of the AI model against experienced orthodontists across various clinical settings would be crucial for validating its real-world applicability, particularly focusing on the accuracy and efficiency of facemask therapy prediction (Gabriele *et al.,*[12]).
- **Ethical Considerations and Implementation Challenges** Address potential barriers to widespread clinical adoption, including regulatory approval, data privacy concerns, integration into existing clinical software, and the need for comprehensive training for dental professionals.
- **Proposed Statistical** Comparison for Future Work: Although comprehensive comparative data with manual diagnoses from orthodontist was not available for this study, future research should include a statistical evaluation of AI model performance. If such data becomes available, a presentation is suggested.
  1. **Sensitivity & Specificity Analysis** How well does the ANN model detect true positive and true negative cases compared to human diagnosis?
  2. **Inter-examiner Agreement (Cohen's Kappa)** Does the ANN model provide more consistent results than human raters?
  3. **Time Efficiency Comparison** Measuring AI processing time vs. the average assessment time of an orthodontist.

In conclusion, while the current ANN model shows promising results, particularly in its capacity to predict the necessity of facemask treatment for Class III malocclusion, there is substantial room for improvement in deep learning approaches. By addressing the limitations of the CNN model and exploring advanced techniques, we can work towards developing a more accurate, reliable, and clinically valuable diagnostic tool. This improved AI system has the potential to significantly enhance early detection and treatment planning for Class III malocclusion in pediatric patients, ultimately leading to better orthodontic outcomes.

## Conclusion

This research demonstrates the superior performance of the Artificial Neural Network (ANN) model compared to the Convolutional Neural Network (CNN) model using Image Embedding and Logistic Regression in analyzing and classifying cephalometric radiographs of pediatric patients with Class III malocclusion. The ANN model showed outstanding results with an accuracy of 90.3%, meaning it could correctly classify images more than nine out of ten times. Moreover, it exhibited high sensitivity and precision, indicating a strong ability to identify true positive cases accurately and reduce misdiagnoses in terms of both false positives and false negatives. The high F1-score further confirms the balance between the sensitivity and the precision of the model. In contrast, the CNN model using Image Embedding and Logistic Regression showed inferior results. With an AUC of 0.750, which is considered good but not excellent, the model achieved a classification accuracy of 71.6% and an F1-score of 0.715. While these results are acceptable, they are significantly behind the performance of the ANN model. This comparison is crucial in the field of orthodontics, as accurate diagnosis of Class III malocclusion is essential for treatment planning and follow-up. The use of the ANN model could potentially increase diagnostic accuracy, reduce treatment errors, and lead to the development of highly efficient clinical decision support systems in the future.

## References


1. Kanas RJ, Carapezza L, Kanas SJ. Treatment classification of Class III malocclusion. ***J Clin Pediatr Dent*** 2008;33(2):175–85.

2. Alhammadi MS, Halboub E, Fayed MS, Labib A, El-Saaidi C. Global distribution of malocclusion traits: A systematic review. ***Dental Press J Orthod*** 2018;23(6):40.e1-40.e10.

3. Guyer EC, Ellis EE 3rd, McNamara JA Jr, Behrents RG. Components of Class III malocclusion in juveniles and adolescents. ***Angle Orthod*** 1986;56(1):7–30.

4. Simsuchin C, Chen Y, Huang S, Mallineni SK, Zhao Z, Hagg U, *et al*. Unilateral Scissor Bite Managed with Prefabricated Functional Appliances in Primary Dentition—A New Interceptive Orthodontic Protocol. ***Children (Basel)*** 2021;8(11):957.

5. Simsuchin C, Chen Y, Mallineni SK. Clinical effectiveness of vestibular shields in orthodontic treatment: a scoping review. ***Children (Basel)*** 2023;10(1):16.

6. Pattanaik S, Mishra S. Treatment of Class III with facemask therapy. ***Case Rep Dent*** 2016;2016:6390637.

7. Khan MB, Karra A. Early treatment of class III malocclusion: a boon or a burden?. ***Int J Clin Pediatr Dent*** 2014;7(2):130–6.

8. Beniwal S, Arora J. Classification and feature selection techniques in data mining. ***Int J Eng Res Technol*** 2012;1(6):1–6.

9. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. ***Nature*** 1986;323(6088):533–6.

10. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. ***J Dent Res*** 2020;99(7):769–74.

11. Taraji S, Choi DS, Petrov Y, Putra AE, Li G, Kuang Y, *et al.* Novel machine learning algorithms for prediction of treatment decisions in adult patients with class III malocclusion. ***J Oral Maxillofac Surg*** 2023;81(11):1391–402.

12. Gabriele S, Christopher JL, Angelika SE. Children with class III malocclusion: Development of multivariate statistical models to predict future need for orthognathic surgery. ***Angle Orthod*** 2003;73(2):136–45.

13. Moon JH, Hwang HW, Lee SJ. Evaluation of an automated superimposition method for computer-aided cephalometrics. ***Angle Orthod*** 2020;90(3):390–6.

14. Schwendicke F, Singh T, Lee JH, Gaudin R, Chaurasia A, Wiegand T, *et al.* Artificial intelligence in dental research: Checklist for authors, reviewers, readers. ***J Dent*** 2021;107:103610.

15. Polizzi A, Leonardi R. Automatic cephalometric landmark identification with artificial intelligence: An umbrella review of systematic reviews. ***J Dent*** 2024;149:105056.

16. Ratra R, Gulia P. Experimental evaluation of open source data mining tools (WEKA and Orange). ***Int J Eng Trends Technol*** 2020;68(8):30–5.

17. Zhang H, Liu C, Yang P, Yang S, Yu Q, Liu R. The concept of AI-assisted self-monitoring for skeletal malocclusion. ***Health Informatics J*** 2024;30(3):14604582241274511.

18. Fatin A, Natasya MS, Azliza MA, Aida NAA, Mohd AM, Afifah AS, *et al.* Classification of malocclusion using convolutional neural network and knowledge-based systems. In: 2023 International Conference on Recent Advances in Electrical and Electronics Engineering (ICRAIEE); 2023 Nov 22-23; Putrajaya, Malaysia. IEEE; 2023. p. 1–6.

19. Aksoy S, Kiliç B, Suzek TO. Comparative analysis of deep learning and machine learning models for early prediction of skeleton class III malocclusion from profile photos. medRxiv [Preprint]. 2022 Jul 27 [cited 2024 Jul 6]. Available from: https://www.medrxiv.org/content/10.1101/2022.07.26.22277593v1.

20. Kaya E, Güneç HG, Ürkmez ES, Aydin KC, Ates HF. Deep learning for diagnostic charting on pediatric panoramic radiographs. ***Int J Comput Dent*** 2023;26(2):107–13.

21. Proffit WR, Fields HW, Sarver DM. Contemporary Orthodontics. 6th ed. St. Louis, MO: Elsevier; 2019.

22. American Academy of Pediatric Dentistry. Guideline on oral radiography in children. In: The Reference Manual of Pediatric Dentistry. Chicago, IL: American Academy of Pediatric Dentistry; 2023. p. 306–12. Available from: https://www.aapd.org/research/oral-health-policies--recommendations/oral-health-guidelines/radiographs.