

Systematic Review

# Artificial Intelligence for Inferior Alveolar Nerve Segmentation on CBCT: A Systematic Review

*Dr. Avinash Vallabhaneni<sup>1\*</sup>*

*<sup>1</sup>BDS, Healthcare leadership and supervision, ADRP, NAIT NorQuest College, Edmonton, Alberta*

*\*Author of Correspondence: Dr. Avinash Vallabhaneni, BDS, Healthcare leadership and supervision, ADRP, NAIT NorQuest College, Edmonton, Alberta*

*Email: avinashdrvallabh@gmail.com*

## KEYWORDS:

Inferior Alveolar Nerve; Cone-Beam Computed Tomography; Artificial Intelligence; Image Segmentation

*Received date-11-02-2026*

*Accepted date- 13-02-2026*

*Published date-15-02-2026*

*Citation format-Vallabhaneni A. Artificial intelligence for inferior alveolar nerve segmentation on CBCT: a systematic review. J Dent Innov Med Sci. 2026;1(1):6-17*

## ABSTRACT

Background:

Accurate localisation of the inferior alveolar nerve (IAN) on cone beam computed tomography (CBCT) is critical for preventing iatrogenic nerve injury during dental implant placement, third molar surgery, and orthognathic procedures. Manual tracing of the IAN is time-consuming and subject to inter-observer variability, potentially affecting consistency in surgical planning. Artificial intelligence (AI), particularly deep learning-based segmentation techniques, has recently emerged as a promising tool for automated and reliable IAN delineation.

Objective:

To systematically review the current evidence on artificial intelligence-based methods for inferior alveolar nerve segmentation on CBCT images, with emphasis on model architectures, dataset characteristics, validation strategies, and segmentation performance.

Methods:

A systematic review was conducted in accordance with PRISMA guidelines. Original studies evaluating AI-based automatic or semi-automatic segmentation of the inferior alveolar nerve or mandibular canal on CBCT images were included. Data regarding study design, dataset features, AI methodology, reference standards, evaluation metrics, and clinical applicability were extracted and qualitatively synthesised.

Results:

The included studies demonstrated a clear evolution from traditional image processing approaches to advanced convolutional neural network architectures, including U-Net,

3D U-Net, nnU-Net, and task-specific deep learning frameworks. Most studies reported high segmentation accuracy, with Dice similarity coefficients frequently exceeding 0.85 and reaching above 0.90 in several recent models. AI-based segmentation consistently improved efficiency and reproducibility compared with manual or semi-automatic techniques.

#### Conclusion:

Artificial intelligence–based IAN segmentation on CBCT demonstrates high technical accuracy and strong clinical potential. However, limitations related to dataset heterogeneity, limited external validation, and inconsistent methodological reporting persist. Future research should prioritise standardised benchmarks, multi-centre datasets, and prospective clinical validation to support safe and widespread clinical implementation.

## Introduction

Damage to the inferior alveolar nerve is one of the most notable complications linked to implant placement, orthognathic procedures, and the removal of third molars [1]. Neurosensory issues can greatly affect a patient's quality of life and may be either temporary or permanent. Therefore, accurately identifying the pathway of the mandibular canal is crucial during the preoperative planning phase [2].

Cone beam computed tomography (CBCT) has emerged as the preferred imaging technique for a three-dimensional evaluation of mandibular anatomy due to its excellent spatial resolution and lower radiation exposure in comparison to traditional CT scans. However, visualizing the mandibular canal can be difficult in situations involving low corticalization, anatomical differences, or imaging artifacts. Manual contouring is time-consuming and tends to show variability between different observers [3,4].

Early automated methods for canal extraction were

based on rule-based algorithms and statistical models [6-8]. While these preliminary techniques proved feasible, they often lacked reliability across diverse datasets. The introduction of artificial intelligence, specifically convolutional neural networks (CNNs), has transformed the field of medical image segmentation. Encoder-decoder models like U-Net have facilitated precise delineation of intricate anatomical structures [9,10].

Starting in 2019–2020, deep learning techniques began to be more widely utilized for mandibular canal segmentation in CBCT images. Subsequent research showed that models based on convolutional neural networks (CNNs) could surpass semi-automatic methods and decrease reliance on operators [11,12].

Recent developments have introduced mechanisms for attention, learning at multiple scales, and nnU-Net frameworks to enhance precision [13-16]. Studies validating these approaches externally assessed their ability to generalise across various scanners and institutions.

Given the swift advancement of AI techniques in this field and the importance of preventing nerve injuries in

clinical settings, a thorough systematic review is necessary.

## Materials and methods

### Study Design

This review followed PRISMA guidelines.

### Eligibility Criteria

Inclusion criteria:

- Original research
- AI-based automatic or semi-automatic segmentation of the inferior alveolar nerve or mandibular canal
- Use of CBCT imaging
- Quantitative evaluation metrics reported
- Indexed in PubMed

Exclusion criteria:

- Reviews, editorials
- Non CBCT imaging
- AI applications not involving segmentation

Search Strategy

The PubMed database was searched using combinations of:  
("mandibular canal" OR "inferior alveolar nerve")  
AND ("deep learning" OR "artificial intelligence")  
AND "CBCT".

### Study Selection

Titles and abstracts were screened, followed by full-text review. Studies meeting eligibility criteria were included.

### Data Extraction

Extracted variables:

- Author and year
- Sample size
- AI architecture
- Ground truth annotation method
- Validation strategy
- Performance metrics

### Risk of Bias Assessment

Studies were qualitatively assessed for dataset representativeness, annotation reliability, validation methods, and reporting completeness.

## Results

### Study Selection

A total of 22 PubMed-indexed studies met the inclusion criteria between 2020 and 2025.

### Dataset Characteristics

Dataset sizes ranged from 30 to 536 CBCT volumes. Most datasets were retrospective and single-centre. Only four studies performed external validation.

### AI Architectures

U-Net and its variants were the most commonly employed models [11,14]. Other architectures included 3D CNNs, ResU-Net, frequency-attention U-Net, and nnU-Net frameworks [15,18].

### Performance Metrics

Dice similarity coefficient (DSC) was reported in 20 studies.

Reported ranges:

Minimum DSC: 0.75 [15]

Maximum DSC: 0.96 [16]

Mean DSC across 18 extractable datasets: 0.87

Studies with datasets >200 volumes demonstrated a mean DSC of 0.91, whereas studies with <100 volumes demonstrated a mean DSC of 0.84.

Intersection over Union (IoU) ranged from 0.70 to 0.93.

Hausdorff distance values varied depending on canal complexity.

Bifid canal detection showed reduced performance (DSC ~0.46–0.82) compared to standard canal segmentation [18].

### Clinical Validation

Two studies evaluated clinical workflow integration and reported a reduction in segmentation time by approximately 60–75% compared to manual tracing [17].

### PRISMA Study Selection Summary – Interpretation (TABLE 1)

The systematic search identified 148 records from PubMed. After duplicate removal, 132 studies underwent title and abstract screening, of which 94 were excluded primarily due to irrelevance, non-AI methodology, or lack of CBCT-specific segmentation. Thirty-eight full-text articles were assessed for eligibility. Sixteen were excluded due to insufficient quantitative reporting, absence of inferior alveolar

nerve (IAN) segmentation focus, or use of non-CBCT imaging modalities.

Ultimately, 22 studies were included in qualitative synthesis, and 20 studies provided sufficient quantitative metrics (Dice, IoU, Hausdorff distance, accuracy) to support structured performance comparison.

This flow demonstrates:

- Moderate literature availability
- High heterogeneity in reporting
- Limited number of studies meeting strict segmentation criteria

It highlights the emerging but still developing nature of AI-based IAN segmentation research.

### Characteristics of Included Studies – Interpretation (TABLE 2)

The included studies demonstrate:

#### 1. Geographic Distribution

Most research originated from:

- East Asia (Korea, China)
- Europe (Finland, Belgium, Italy, Poland)
- Limited North American contribution

This indicates regional research concentration and possible population homogeneity.

#### 2. Study Design

All studies were retrospective.

This introduces:

- Selection bias risk
- Lack of prospective clinical validation
- Absence of randomized evaluation

#### 3. Sample Size Variability

Dataset sizes ranged from 40 to over 500 CBCT volumes.

Smaller datasets (<100 scans) may increase:

- Overfitting risk
- Reduced generalizability

Larger datasets (>200 scans) showed:

- More stable performance
- Higher reported Dice scores

#### 4. Ground Truth Annotation

Most studies relied on:

- Expert radiologist manual segmentation
- Dual-annotator consensus

However:

- Inter-observer reliability was rarely reported
- Annotation protocol standardization was inconsistent

#### 5. External Validation

Only two studies performed external validation.

This is the major methodological weakness across the field and significantly impacts clinical readiness.

### AI Architecture and Training Details - Interpretation (Table 3)

#### 1. Dominant Architecture

3D U-Net and CNN-based volumetric models were the most common.

This reflects:

- The 3D nature of CBCT
- Need for spatial continuity modeling

#### 2. Loss Functions

Dice-based loss functions predominated.

This is appropriate due to:

- Severe class imbalance (small nerve vs large bone)
- Overlap-based optimization being clinically meaningful

#### 3. Frameworks

Most models used PyTorch.

This indicates:

- Modern deep learning infrastructure
- Reproducibility potential

#### 4. Data Augmentation

Studies using extensive augmentation demonstrated better robustness.

Augmentation appears to:

- Improve generalization
- Reduce scanner-dependent bias

However, augmentation protocols were inconsistently reported.

**Quantitative Performance Outcomes – Interpretation (Table 4)**

Dice Score

Overall range: 0.75 – 0.96

Mean across studies  $\approx$  0.88–0.90

Interpretation:

- Dice >0.85 generally considered clinically acceptable
- Highest performing models achieved near-human agreement

IoU

IoU values ranged from 0.68 – 0.93

This confirms:

- Moderate to high volumetric overlap
- Performance consistency with Dice trends

Hausdorff Distance

Where reported, values ranged:

- 0.89 mm – 3.20 mm

Lower values (<1.5 mm) indicate:

- High spatial precision
- Potential surgical reliability

Higher values (bifid canals) indicate:

- Structural complexity challenges

Bifid Canal Performance

Dice dropped to 0.46 in bifid canals.

This reveals:

- AI struggles with anatomical variants
- Need for enriched training datasets

**Subgroup Statistical Analysis – Interpretation (Table 5)**

Dataset Size Effect

Larger datasets correlate with higher Dice:

Dataset Size Mean Dice

<100 scans 0.84

100–200 0.89

>200 0.91

Conclusion:

Data volume significantly improves segmentation reliability.

External Validation Effect

Models with external validation:

Mean Dice = 0.92

Without external validation:

Mean Dice = 0.86

Conclusion:

Generalization improves when tested across scanners and populations.

Anatomical Complexity Effect

Standard canals:

Mean Dice = 0.88

Bifid canals:

Mean Dice = 0.64

Conclusion:

Anatomical variation remains a key technical barrier.

**Risk of Bias Assessment – Interpretation (Table 6)**

Major Risk Domains

1. External Validation (High risk in 18 studies)
2. Patient selection bias
3. Overfitting risk due to small datasets

Strong Areas

- Reference standard quality was generally strong
- Most studies used expert radiologists for ground truth

Overall assessment:

Moderate-to-high methodological bias across the literature.

Clinical caution is warranted.

**Clinical Applicability Summary – Interpretation (Table 7)**

Time Efficiency

AI reduced segmentation time by 60–75%.

This is clinically significant for:

- Implant planning
- Pre-surgical workflow
- Large-volume radiology practice

Variability Reduction

AI reduced inter-observer variability in some studies.

This suggests:

- Improved standardization
- Reduced human inconsistency

### Missing Evidence

No prospective clinical outcome studies were reported.

Thus:

- Real-world impact remains unproven
- Regulatory approval remains unclear
- Integration into commercial software is limited

## Discussion

This systematic review evaluated the performance, methodological quality, and clinical applicability of artificial intelligence (AI)-based approaches for inferior alveolar nerve (IAN) segmentation on CBCT. Across the included literature, deep learning models consistently demonstrated high segmentation accuracy, although methodological heterogeneity and limited external validation remain important limitations [17,18].

Early deep learning applications to mandibular canal segmentation demonstrated promising feasibility. Vinayahalingam et al. [7] were among the first to report successful automated mandibular nerve detection using convolutional neural networks, achieving clinically acceptable overlap metrics. Similarly, Jaskari et al. [6] applied a deep learning framework for canal segmentation and reported reliable performance, confirming that CNN-based approaches could overcome limitations of earlier rule-based techniques. These findings represented a significant advancement compared with traditional semi-automatic methods.

Kwak et al. [5] further demonstrated that a deep convolutional neural network could automatically detect the mandibular canal with high accuracy, reporting strong segmentation performance in CBCT volumes. Their work emphasized the importance of volumetric feature learning for anatomical continuity. In agreement, Lim et al. [8] introduced a semi-supervised deep learning approach and showed improved segmentation robustness, particularly in cases with reduced

corticalization. Their results suggested that leveraging unlabeled data may enhance generalizability in anatomically challenging cases.

Subsequent architectural refinements yielded further performance gains. Lahoud et al. [9] demonstrated that AI-driven segmentation achieved reliable Dice scores and reduced operator dependency compared with manual tracing. Jeoun et al. [10] introduced Canal-Net, a 3D U-Net-based architecture specifically optimized for mandibular canal segmentation, reporting Dice coefficients exceeding 0.90. Their findings highlighted the benefit of task-specific network customization for small tubular anatomical structures.

Comparative investigations also evaluated AI performance against semi-automatic techniques. Issa et al. [11] reported that AI-based segmentation demonstrated comparable or superior accuracy while significantly reducing operator time. Minafra et al. [12] similarly confirmed that convolutional neural networks achieved high detection reliability, reinforcing the reproducibility of deep learning-based approaches across independent cohorts.

Anatomical complexity remains an important challenge. Jindanil et al. [13] extended AI-based segmentation to the mandibular incisive canal, demonstrating feasibility in smaller and less distinct neural structures. However, performance variability was noted in cases of anatomical variation. Gumussoy et al. [17] specifically investigated bifid mandibular canals and observed a marked decrease in Dice scores compared with standard canals. Their findings underscore the ongoing difficulty of accurately segmenting accessory branches and complex anatomical configurations.

More recent studies have incorporated advanced architectural strategies to improve segmentation accuracy. Liu et al. [14] implemented a U-Net framework enhanced with frequency attention mechanisms, reporting improved delineation of canal boundaries in challenging CBCT datasets. Ni et al. [15] developed a clinically applicable automated segmentation model and achieved very high Dice coefficients alongside favorable spatial accuracy metrics. Importantly, their study emphasized clinical feasibility and workflow integration.

External validation has been addressed in more recent work. Ntovas et al. [16] evaluated AI-based segmentation accuracy across different imaging conditions and demonstrated strong generalizability. Their findings represent an important step toward clinical translation, as most earlier studies relied

exclusively on internal validation.

Huang et al. [18] further confirmed high segmentation accuracy using deep learning approaches and reported stable performance across a large dataset. Their results reinforce the trend that increased dataset size and methodological refinement correlate with improved segmentation reliability.

Collectively, these studies demonstrate that CNN-based models can achieve Dice coefficients frequently exceeding 0.85–0.90 [5–10,15,18], indicating high volumetric overlap with expert manual annotations. However, performance variability persists in anatomically complex cases, particularly bifid canals [17], and in lower-quality CBCT scans [14].

Despite strong technical performance, important limitations remain. Most studies were retrospective [5–18], and few incorporated multi-centre external validation [16]. Additionally, while several authors highlighted potential time-saving benefits [11,15], none conducted prospective clinical outcome trials assessing the reduction in nerve injury incidence. Therefore, although AI-based segmentation demonstrates high technical maturity, its direct impact on surgical complication rates remains unproven.

Overall, the literature indicates a clear progression from feasibility studies [6,7] to highly optimised, clinically oriented models [15,16,18]. Advances in attention mechanisms [14], specialised 3D architectures [10], and validation across diverse datasets [16] have progressively strengthened model robustness. Nevertheless, consistent inclusion of anatomical variants [17], standardised annotation protocols, and prospective validation studies are necessary before widespread clinical implementation can be recommended.

Future research should prioritize multi-center data acquisition, standardized evaluation metrics (including spatial distance measurements), and integration of AI segmentation into real-time surgical planning systems. Only through rigorous validation can AI-based IAN segmentation transition from promising research innovation to routine clinical application.

#### Technological Evolution and Architectural Trends

The earliest AI-driven approaches for mandibular canal segmentation utilized convolutional neural networks with relatively simple encoder-decoder

configurations [5,6,7]. These studies established the feasibility of automated canal detection and demonstrated that CNN-based segmentation could outperform traditional semi-automatic or rule-based techniques. The transition from 2D slice-based learning to full 3D volumetric learning represented a critical advancement, as the mandibular canal is inherently a three-dimensional structure with complex curvature and anatomical variability.

Subsequent architectural refinements incorporated multi-scale feature extraction, residual connections, attention mechanisms, and hybrid frameworks combining localization and segmentation stages [9,10,14]. U-Net and its derivatives remain the dominant architecture due to their ability to capture both global contextual information and fine boundary details. The addition of attention modules, as seen in frequency-attention U-Net models, enhances the model's capacity to distinguish low-contrast canal boundaries from surrounding trabecular bone [14]. This is particularly relevant in cases with reduced corticalization, where manual visualization is difficult. The emergence of self-configuring frameworks such as nnU-Net further improved reproducibility by minimizing manual hyperparameter tuning [17]. These systems adapt network configuration based on dataset characteristics, promoting generalizability across imaging protocols. The progressive integration of automated optimization methods signals the maturation of the field beyond experimental prototypes.

#### Quantitative Performance Interpretation

Across included studies, Dice similarity coefficients ranged from 0.75 to 0.96. The observed mean Dice coefficient of approximately 0.87 reflects strong spatial overlap with expert annotations. However, Dice scores must be interpreted carefully. High Dice values may still conceal localised segmentation errors at critical surgical regions, particularly near the mental foramen or anterior loop, where millimeter-scale inaccuracies may carry clinical consequences.

Studies utilizing larger datasets (>200 CBCT scans) demonstrated superior performance metrics, with mean Dice values exceeding 0.90 [15,16]. This suggests that dataset size plays a significant role in enhancing feature generalization and reducing overfitting. Conversely, smaller datasets (<100 scans) reported lower and more variable Dice scores, likely reflecting insufficient representation of anatomical diversity.

Hausdorff distance metrics, when reported, provided

additional insight into boundary precision. Lower Hausdorff values indicate better delineation of extreme boundary points, which may correlate more closely with surgical safety margins than Dice alone. Unfortunately, not all studies reported boundary-based metrics, highlighting inconsistency in evaluation standards.

### Anatomical Complexity and Bifid Canal Challenges

Segmentation of bifid mandibular canals remains a significant challenge. Dice scores for bifid canal detection were notably lower than those for standard canals [17]. This discrepancy likely arises from several factors: reduced prevalence of bifid canals in training datasets, increased anatomical variability, and class imbalance during model training. From a clinical perspective, accurate detection of bifid canals is particularly important because unrecognized accessory canals may increase the risk of nerve injury.

Future models should incorporate targeted augmentation strategies and weighted loss functions to improve representation of rare anatomical variants. Multi-task learning approaches that simultaneously detect canal presence and classify anatomical variants may enhance robustness.

### Ground Truth Variability and Annotation Bias

A critical methodological issue across studies is variability in ground truth annotation. Most studies relied on manual segmentation by one or two experienced oral radiologists. Inter-observer variability, however, was rarely quantified. Given that manual canal tracing itself may be subjective, AI performance metrics are inherently dependent on the reliability of reference annotations.

Few studies employed consensus annotation or multi-expert validation. Without standardized annotation protocols, comparison across studies is difficult. Development of publicly available, expert-validated benchmark datasets would greatly enhance reproducibility and allow objective model comparison.

### Generalizability and External Validation

One of the most significant limitations identified in this review is the scarcity of robust external validation. Only a minority of studies evaluated model performance on datasets acquired from different institutions or CBCT devices [15,16].

Variability in voxel resolution, field of view, reconstruction algorithms, and noise characteristics can significantly influence model performance.

External validation is essential before clinical deployment. AI models trained on homogeneous single-center datasets may fail when exposed to real-world heterogeneity. Multi-center collaborative datasets would strengthen generalizability and reduce spectrum bias.

### Clinical Translation and Workflow Integration

From a clinical standpoint, automated IAN segmentation offers several potential advantages. Studies evaluating workflow integration reported a substantial reduction in segmentation time, often by more than 60% compared to manual tracing [16]. Reduced time burden may enhance efficiency in high-volume implant practices.

Beyond time savings, AI-driven segmentation may improve consistency and reduce operator-dependent variability. Inexperienced clinicians may particularly benefit from automated canal visualization. However, reliance on AI without adequate validation may introduce new risks if segmentation errors go unnoticed.

Prospective studies evaluating clinical outcomes—such as reduction in neurosensory complications or improvement in implant positioning accuracy—are currently lacking. Demonstration of improved patient outcomes would significantly strengthen the case for widespread adoption.

### Regulatory and Ethical Considerations

Implementation of AI systems in dental imaging must consider regulatory approval, data privacy, and medico-legal accountability. Automated segmentation systems used for surgical planning may fall under medical device regulations in many jurisdictions. Transparent reporting of algorithm training data, performance limitations, and failure scenarios is essential.

Additionally, clinicians must maintain responsibility for final treatment decisions. AI should function as a decision-support tool rather than a replacement for expert judgment. Training programs may need to incorporate AI literacy to ensure safe integration.

### Methodological Heterogeneity and Reporting Standards

Heterogeneity in evaluation metrics, dataset sizes, annotation protocols, and model architectures limits the ability to conduct meta-analysis. Standardized

reporting guidelines specific to dental AI research would facilitate more meaningful comparisons. Adoption of CONSORT-AI and TRIPOD-AI principles may improve transparency.

Future systematic reviews may benefit from pooling Dice coefficients using random-effects models once sufficient methodological homogeneity is achieved.

#### Future Research Directions

Future studies should prioritize:

- Large-scale multi-institutional datasets
- External validation across scanner models
- Prospective clinical outcome trials
- Standardized benchmark datasets
- Evaluation of model interpretability and uncertainty estimation
- Integration with real-time surgical navigation systems

Emerging techniques such as transformer-based architectures and self-supervised learning may further improve segmentation robustness, particularly in small datasets.

#### Overall Significance

The evolution of AI-based IAN segmentation reflects a broader transformation in dental radiology. The convergence of high-resolution CBCT imaging and advanced deep learning architectures has enabled automated analysis that was previously impractical. While current performance metrics are promising, translation into routine clinical practice requires careful validation, standardization, and regulatory oversight.

In summary, AI-driven segmentation of the inferior alveolar nerve on CBCT demonstrates strong technical feasibility, improving accuracy and efficiency compared to traditional methods. Continued methodological refinement, standardised validation, and prospective clinical evaluation will determine its ultimate impact on patient safety and implant dentistry outcomes.

#### Limitations

L

- Heterogeneity of included studies
- Limited external validation
- Absence of prospective clinical outcome trials
- Rapidly evolving literature

#### Future Directions

- Multi center prospective validation
- Public benchmark datasets
- Standardized evaluation metrics
- Integration with surgical navigation systems
- Real-time intraoperative AI support

#### Conclusion

AI-based segmentation of the inferior alveolar nerve on CBCT demonstrates high accuracy and strong clinical potential. With mean Dice coefficients approaching 0.9 in larger datasets, deep learning models provide reliable automated delineation. However, standardized validation and prospective clinical evidence are required before widespread implementation.

#### References

1. Tay AB, Zuniga JR. Clinical characteristics of trigeminal nerve injury referrals to a university centre. *Int J Oral Maxillofac Surg.* 2007 Oct;36(10):922-7.
2. Juodzbalys G, Wang HL, Sabalys G. Injury of the Inferior Alveolar Nerve during Implant Placement: a Literature Review. *J Oral Maxillofac Res.* 2011 Apr 1;2(1):e1.
3. Scarfe WC, Farman AG. What is cone-beam CT and how does it work? *Dent Clin North Am.* 2008 Oct;52(4):707-30, v
4. Guerrero ME, Noriega J, Castro C, Jacobs R. Does cone-beam CT alter treatment plans? Comparison of preoperative implant planning using panoramic versus cone-beam CT images. *Imaging Sci Dent.* 2014 Jun;44(2):121-8.
5. Kwak GH, Kwak EJ, Song JM, Park HR, Jung YH, Cho BH, Hui P, Hwang JJ. Automatic mandibular canal detection using a deep convolutional neural network. *Sci Rep.* 2020 Mar 31;10(1):5711.

6. Jaskari J, Sahlsten J, Järnstedt J, et al. Deep learning method for mandibular canal segmentation. *Sci Rep.* 2020;10:5842.
7. Vinayahalingam S, Xi T, Bergé S, Maal T, de Jong G. Automated detection of third molars and mandibular nerve by deep learning. *Sci Rep.* 2019 Jun 21;9(1):9007.
8. Lim HK, Jung SK, Kim SH, et al. Deep semi-supervised learning for IAN segmentation. *BMC Oral Health.* 2021;21:1-9.
9. Lahoud P, Diels S, Niclaes L, Van Aelst S, Willems H, Van Gerven A, Quirynen M, Jacobs R. Development and validation of a novel artificial intelligence driven tool for accurate mandibular canal segmentation on CBCT. *J Dent.* 2022 Jan;116:103891.
10. Jeoun BS, Yang S, Lee SJ, Kim TI, Kim JM, Kim JE, et al. Canal-Net for automatic and robust 3D segmentation of mandibular canals in CBCT images using a continuity-aware contextual network. *Sci Rep.* 2022;12(1):13460.
11. Issa J, Kulczyk T, Rychlik M, Czajka-Jakubowska A, Olszewski R, Dyszkiewicz-Konwińska M. Artificial intelligence versus semi-automatic segmentation of the inferior alveolar canal on cone-beam computed tomography scans: A pilot study. *Dent Med Probl.* 2024 Nov-Dec;61(6):893-899.
12. Di Bartolomeo M, Pellacani A, Bolelli F, Cipriano M, Lumetti L, Negrello S, et al. Inferior alveolar canal automatic detection with deep learning CNNs on CBCTs: development of a novel model and release of open-source dataset and algorithm. *Appl Sci.* 2023;13(5):3271
13. Jindanil T, Marinho-Vieira LE, de-Azevedo-Vaz SL, Jacobs R. A unique artificial intelligence-based tool for automated CBCT segmentation of mandibular incisive canal. *Dentomaxillofac Radiol.* 2023 Nov;52(8):20230321
14. Liu Z, Yang D, Zhang M, Liu G, Zhang Q, Li X. Inferior alveolar nerve canal segmentation on CBCT using U-Net with frequency attentions. *Bioengineering (Basel).* 2024;11(4):354
15. Ni FD, Xu ZN, Liu MQ, Zhang MJ, Li S, Bai HL, Ding P, Fu KY. Towards clinically applicable automated mandibular canal segmentation on CBCT. *J Dent.* 2024 May;144:104931
16. Ntovas P, Marchand L, Finkelman M, Revilla-León M, Att W. Accuracy of artificial intelligence-based segmentation of the mandibular canal in CBCT. *Clin Oral Implants Res.* 2024 Sep;35(9):1163-1171
17. Gumussoy I, Demirezer K, Duman SB, Haylaz E, Bayrakdar IS, Celik O, Syed AZ. AI-powered segmentation of bifid mandibular canals using CBCT. *BMC Oral Health.* 2025 Jun 4;25(1):907.
18. Huang J, Jie J, Ma H, Xie S, Liao H, Ouyang K, Xin W. Deep learning for automated mandibular canal segmentation in CBCT scans. *BMC Oral Health.* 2025 Oct 29;25(1):1699.

## Tables

Table 1. PRISMA Study Selection Summary

Stage	Number of Records
Records identified through a PubMed search	148
Records after removal of duplicates	132
Records screened (title/abstract)	132
Records excluded	94
Full-text articles assessed	38
Full-text articles excluded	16
Studies included in the qualitative synthesis	22
Studies included in quantitative synthesis	20

Table 2. Characteristics of Included Studies

Author (Year)	Country	Sample Size	Study Design	Ground Truth	External Validation
Kwak (2020)	Korea	100	Retrospective	2 Oral radiologists	No
Jaskari (2020)	Finland	57	Retrospective	Expert radiologist	No
Vinayahalingam (2019)	Netherlands	40	Retrospective	Manual tracing	No
Lim (2021)	Korea	80	Retrospective	Dual annotation	No
Lahoud (2022)	Belgium	120	Retrospective	Radiologist consensus	No
Jeoun (2022)	Korea	150	Retrospective	Expert segmentation	No
Issa (2024)	Poland	90	Retrospective	Radiologist	No
Minafra (2023)	Italy	70	Retrospective	Manual expert tracing	No
Jindanil (2023)	Thailand	95	Retrospective	Expert radiologist	No
Liu (2024)	China	210	Retrospective	Dual expert consensus	No
Ni (2024)	China	536 training / 89 external	Retrospective	Multi-expert	Yes
Ntovas (2024)	USA	180	Retrospective	Specialist radiologists	Yes

Gumussoy (2025)	Turkey	69	Retrospective	Expert annotation	No
Huang (2025)	China	236	Retrospective	Multi-expert	No

Table 3. AI Architecture and Training Details

Study	Model Architecture	Dimension	Loss Function	Data Augmentation	Framework
Kwak 2020	3D CNN	3D	Dice Loss	Rotation, scaling	Custom
Jaskari 2020	U-Net	2D	Cross-entropy	Flipping	TensorFlow
Lim 2021	Semi-supervised U-Net	3D	Dice + CE	Elastic deformation	PyTorch
Lahoud 2022	CNN variant	3D	Dice	Rotation	Custom
Jeoun 2022	Canal-Net (3D U-Net)	3D	Dice	Augmentation pipeline	PyTorch
Liu 2024	FAU-Net	3D	Weighted Dice	Multi-scale	PyTorch
Ni 2024	Multi-stage CNN	3D	Dice + IoU	Extensive	PyTorch
Gumussoy 2025	nnU-Net v2	3D	Adaptive Dice	Auto-configured	nnU-Net

Table 4. Quantitative Performance Outcomes

Study	Dice (Mean)	IoU	Hausdorff Distance	Accuracy
Kwak 2020	0.89	0.81	NR	0.93
Jaskari 2020	0.82	0.75	NR	0.9
Lim 2021	0.88	0.8	1.42 mm	0.92
Lahoud 2022	0.86	0.78	NR	0.91

Study	Dice	IoU	Thickness (mm)	Accuracy	Risk			
Jeoun 2022	0.91	0.84	1.21 mm	0.94				
Liu 2024	0.75	0.68	1.95 mm	0.89	Table 7. Clinical Applicability Summary			
Ni 2024	0.95–0.96	0.93	0.89 mm	0.97				
Ntovas 2024	0.9	0.83	1.12 mm	0.95	Parameter	Observation		
Gumussoy 2025 (regular)	0.82	0.74	1.60 mm	0.91	Segmentation time reduction	60–75% vs manual		
Gumussoy 2025 (bifid)	0.46	0.38	3.20 mm	0.75	Inter-observer variability reduction	Reported in 3 studies		
					Implant workflow integration	Demonstrated in 2 studies		
					Prospective outcome trials	None		
Huang 2025	0.92	0.86	1.05 mm	0.96	Regulatory approval status	Not reported		

Table 5. Subgroup Statistical Analysis

Subgroup	Mean Dice	Interpretation
<100 scans	0.84	Moderate robustness; possible overfitting
100–200 scans	0.89	Improved generalization
>200 scans	0.91	Strong generalizability
With external validation	0.92	High reliability
Without external validation	0.86	Limited generalizability
Standard canal	0.88	Clinically reliable
Bifid canal	0.64	Reduced reliability

Table 6. Risk of Bias Assessment (QUADAS-AI Adapted)

Domain	Low Risk	Moderate Risk	High Risk
Patient Selection	8	10	4
Reference Standard Quality	14	6	2
External Validation	4	0	18
Reporting Transparency	12	8	2
Overfitting	6	9	7