Hindawi
Pain Research and Management
Volume 2021, Article ID 6659133, 9 pages
https://doi.org/10.1155/2021/6659133

Review Article

Machine Learning and Intelligent Diagnostics in Dental and Orofacial Pain Management: A Systematic Review

Taseef Hasan Farook,[1] **Nafij Bin Jamayet**,[2] **Johari Yap Abdullah**,[3] **and Mohammad Khursheed Alam**[4]

[1]*Maxillofacial Prosthetic Service, Prosthodontic Unit, School of Dental Sciences, Universiti Sains Malaysia, Health Campus, Kelantan 16150, Malaysia*
[2]*Division of Clinical Dentistry (Prosthodontics), School of Dentistry, International Medical University, Kuala Lumpur 57000, Malaysia*
[3]*Craniofacial Imaging and Additive Manufacturing Laboratory, School of Dental Sciences, Universiti Sains Malaysia, Health Campus, Kelantan 16150, Malaysia*
[4]*College of Dentistry, Jouf University, Sakaka, Saudi Arabia*

Correspondence should be addressed to Johari Yap Abdullah; johariyap@usm.my and Mohammad Khursheed Alam; dralam@gmail.com

Received 1 December 2020; Revised 11 March 2021; Accepted 17 April 2021; Published 26 April 2021

Academic Editor: Xue-Qiang Wang

Purpose. The study explored the clinical influence, effectiveness, limitations, and human comparison outcomes of machine learning in diagnosing (1) dental diseases, (2) periodontal diseases, (3) trauma and neuralgias, (4) cysts and tumors, (5) glandular disorders, and (6) bone and temporomandibular joint as possible causes of dental and orofacial pain. *Method*. Scopus, PubMed, and Web of Science (all databases) were searched by 2 reviewers until 29th October 2020. Articles were screened and narratively synthesized according to PRISMA-DTA guidelines based on predefined eligibility criteria. Articles that made direct reference test comparisons to human clinicians were evaluated using the MI-CLAIM checklist. The risk of bias was assessed by JBI-DTA critical appraisal, and certainty of the evidence was evaluated using the GRADE approach. Information regarding the quantification method of dental pain and disease, the conditional characteristics of both training and test data cohort in the machine learning, diagnostic outcomes, and diagnostic test comparisons with clinicians, where applicable, were extracted. *Results*. 34 eligible articles were found for data synthesis, of which 8 articles made direct reference comparisons to human clinicians. 7 papers scored over 13 (out of the evaluated 15 points) in the MI-CLAIM approach with all papers scoring 5+ (out of 7) in JBI-DTA appraisals. GRADE approach revealed serious risks of bias and inconsistencies with most studies containing more positive cases than their true prevalence in order to facilitate machine learning. Patient-perceived symptoms and clinical history were generally found to be less reliable than radiographs or histology for training accurate machine learning models. A low agreement level between clinicians training the models was suggested to have a negative impact on the prediction accuracy. Reference comparisons found non-specialized clinicians with less than 3 years of experience to be disadvantaged against trained models. *Conclusion*. Machine learning in dental and orofacial healthcare has shown respectable results in diagnosing diseases with symptomatic pain and with improved future iterations and can be used as a diagnostic aid in the clinics. The current review did not internally analyze the machine learning models and their respective algorithms, nor consider the confounding variables and factors responsible for shaping the orofacial disorders responsible for eliciting pain.

1. Introduction

Pain is a subjective sensation and has varying tolerance thresholds [1]. Orofacial pain has multiple origins and varying intensities. The pain may arise from exposed dentin (hypersensitivity pain) [2] or from carious infection of the dental pulp (pulpitis) [3]. Untreated dental pulp encourages the infection to spread through the root canals into the periodontal tissue (apical periodontitis) [4, 5] and may cause swelling, infection, and bone loss

(periapical abscess) [6]. Periodontal tissue can also be painfully infected without carious activity (gingivitis and periodontitis) [7]. Maxillofacial fractures [8], as well as iatrogenic trauma/infection during dental restorative/endodontic treatment [2], may elicit varying levels of pain. Bone diseases [9], temporomandibular joint disorders [10], space infections [11], salivary gland disorders [12, 13], and sinusitis [14] elicit pain. Furthermore, neuralgia and secondary sensory nerve compression due to growing cysts and tumors can elicit severe pain [15, 16]. These conditions are categorized as common diseases and disorders that elicit dental and orofacial pain in the dental clinic [17].

The clinician's ability to diagnose such events swiftly and accurately is pivotal in successful patient management. However, various studies have shown that incorrect diagnoses are fairly common among clinicians in such situations [5, 6, 18]. While pain itself might not be reliably quantified, machine learning/artificial intelligence (AI) has been recently deployed to detect and quantify various diseases which elicit pain within the orofacial region to aid in accurate diagnostics and management.

AI and computerized support, although not new to healthcare, have lately received a lot of attention within the sphere of dentistry. These reviews covered their potential dental applications [19], success in detecting precancerous lesions and metastases [20], effectiveness in improving the quality of maxillofacial radiology [21], success in orthodontic treatment [22], and orthopedic rehabilitation [23], as well as concurrent application with virtual reality to decrease anxiety in young patients [24]. However, the aforementioned reviews did not systematically explore the current diagnostic capabilities of AI in identifying common orofacial diseases and disorders and/or the subsequently elicited pain [17].

Therefore, the current review was conducted and narratively synthesized to explore the influence of machine learning in the following diagnostic roles: (1) pain associated with dental diseases, (2) pain associated with periodontal diseases, (3) pain associated with trauma and neuralgias, (4) pain associated with cysts and tumors, (5) pain associated with glandular disorders, and (6) pain arising from bone and temporomandibular joint. The clinical effectiveness of machine learning, potential variations and probable causes, and human versus machine comparisons were also explored. The effectiveness of AI's influence was quantified using accuracy (ability to correctly differentiate disease from control), sensitivity (correctly identifying diseased subjects), specificity (correctly identifying disease-free subjects), and precision (repeated correct diagnoses) as appropriate.

2. Materials and Methods

2.1. Research Design. The study adhered closely to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Diagnostic Test Accuracy (PRISMA-DTA) guidelines but followed a narration-based, qualitative approach to represent the included literature [25].

2.2. Eligibility Criteria. The following inclusion and exclusion criteria were developed for the current systematic review.

2.2.1. Inclusion Criteria

(1) Original articles describing the use of intelligent computer-guided decision-making to diagnose orofacial diseases that produce symptomatic pain in humans

(2) Studies that incorporated diagnostic management of pain and inflammation using deep learning and intelligent decision-making systems within all specialties of dentistry

(3) Studies of intelligent technologies for emotion and facial expression recognition applied in facial pain diagnostics and healthcare

2.2.2. Exclusion Criteria

(1) Literature demonstrating the application of expert systems, deep learning, and intelligent tools for anatomical and physiological morphology and radiomics quality analyses

(2) Studies on intelligent systems used to detect precancerous or metastatic cancerous lesions, monitor surgically intervened malignancies, or assess the quality of life changes following tumor metastasis and chemo/radiation therapy

(3) Editorials, reviews, book chapters, opinion letters, magazine issues, product advertisements, conference proceedings, social media and blog posts, and articles written in a foreign language without accompanying English translation

2.3. Specific Study Characteristics for Diagnostic Comparisons. Eligible and included studies that made human versus machine diagnostic comparisons were further screened according to the following criteria:

(i) Index test and evaluating parameters: the sensitivity and/or specificity of clinically trained machine learning models

(ii) Reference standards: diagnostic accuracy of clinicians in identifying target conditions

(iii) Target conditions: isolation of dental diseases that lead to symptomatic pain in the following conditions: dentinal, pulpal, periodontal, and alveolar inflammatory diseases; traumatic and cranial neuralgic disorders; odontogenic and nonodontogenic orofacial growths; orofacial glandular inflammation, obstruction, and impaired function; and facial bone and joint disorders

2.4. Information Source. All data were extracted from Scopus, PubMed, and Web of Science (all databases) by one clinician specializing in digital rehabilitation and one

computing and imaging specialist. The Web of Science databases included the WoS Core collection, Current Contents Connect, Derwent Innovations Index, KCI Korean Journal Database, Medline, Russian Science Citation Index, and SciELO Citation Index. The data was extracted from 2020 backward with no lower limits. The final search was made in early November 2020.

2.5. Electronic Search Strategy. The strategy was specifically formulated using Boolean Logic (AND) and wildcards (*) to allow for the same search terms to be applicable for all databases without requiring any modifications thereby maximizing data output [26]. The following combinations were used in the search:

[Big AND data AND dent* AND pain]; [Deep AND learning AND smart AND dent*]; [Expert AND system* AND dent]; [Expert AND system* AND maxill* AND pain]; [Machine AND learning AND dent* AND pain]; [Neural AND network AND dent* AND pain]; [Neural AND network AND maxill* AND pain]; [Generative AND adversarial AND dent*]; [Fuzzy AND network AND dent*]; [Artificial AND intelligen* AND dent* AND pain]; [Artificial AND intelligen* AND caries AND pain]; [Intelligen* AND ulcer AND pain]; [Smart AND dent* AND pain]; [Comput* AND Intelligen* AND pain AND diagnos* AND dent*]; [Smart AND diagnos* AND dent* AND pain]; [Smart AND diagnos* AND facial AND pain]; [Intelligen* AND pain AND face]; [Intelligen* AND pain AND dent*]; [Intelligen* AND device* AND dent* AND pain]; [Intelligen* AND Sensor* AND diagnos* AND dent* AND pain]; [Electr* AND Sensor* AND diagnos* AND maxill* AND pain]; [Intelligen* AND biosens* AND oral]; [Artificial AND Somatosensor* AND facial]; [Intelligen* AND Somatosensor* AND dent*]; [intelligen* AND inflam* AND facial]; [Tensor AND pain AND dent*]; [Comput* AND language AND inflam* AND face]; [Intelligen* AND oral AND carcinoma]; [Augment* AND reality AND dent* AND pain]; [Virtual AND dent* AND diagnos* AND pain]; [Artificial AND Intelligen* AND implant* AND pain]; [Deep AND learning AND maxil* AND surg*]; [Intelligen* AND ortho* AND pain AND dent*]; [Deep AND learning AND radio* AND oral]; [Deep AND learning AND radiol* AND pulp*]; [Deep AND learning AND radiol* AND periodon*].

2.6. Study Selection and Data Collection Process. Titles were screened for duplicates using Endnote v8.2, and the remaining manuscripts were then screened by abstract based on predefined eligibility criteria. The articles excluded during abstract screening were documented along with the theme of the study and the reasons for exclusion. The level of agreement between the two reviewers was measured using the kappa coefficient, and all disagreements were resolved by a face-to-face meeting. Finally, full papers were read, and ineligible articles were removed with the reason for removal being noted.

2.7. Data Extraction. The following data were extracted from the methodology and result sections of the selected papers: quantifications related to dental pain and the machine learning classification models used to develop the intelligent system; the number and conditional characteristics of the training dataset that was used to train the intelligent system; the number of test data used to evaluate the newly trained system with possible human comparisons along with their subsequent learning outcomes; and finally, the clinician's specific role in training or validating the machine learning model which was also documented.

2.8. Diagnostic Accuracy Measures. Specificity (Sp) and sensitivity (Sn) were measured along with accuracy (Ac) and precision (Pr) data which were collected. All obtained values were standardized to 0.00–1.00, and normalized data were given a 1-point standard deviation [27]. The number of learning data (n^L) and test data (n^T) was also collected. No eligible papers were excluded for not presenting one or more of the aforementioned summary measures.

2.9. Risk of Bias and Applicability. Studies that made a direct comparison to clinicians as reference standards were assessed for bias and applicability. The appropriateness of the machine learning model was evaluated using the Minimum Information about Clinical Artificial Intelligence Modeling (MI-CLAIM) checklist [28]. The risk of bias among studies and possible inconsistencies in the comparison were assessed using Joanna Brigg's Institute Critical Appraisal for Diagnostic Test Accuracy (JBI-DTA) checklist [29]. The findings from the MI-CLAIM and JBI-DTA were then used to evaluate the quality of the diagnostic evidence produced in the studies by using the Cochrane GradePro (GRADE approach) [30].

2.10. Additional Syntheses. A meta-analysis was deemed inappropriate due to the substantial functional differences and clinical heterogeneity present across the various disease classifications and machine learning models.

3. Results

3.1. Study Selection. During the screening process, the reviewers had a fair agreement ($k = 0.68$) in the screening process. 34 articles were eventually selected for full paper reading based on eligibility criteria (Figure 1).

3.2. Study Characteristics and Individual Results. The study characteristics and their individual findings have been tabulated and presented as supplementary documents with this manuscript. The papers and tables are categorized into the following subsections: (1) pain associated with dental diseases [1–3, 31–37] (Supplementary Table S1), (2) pain associated with periodontal diseases [4–7, 18, 38–41]

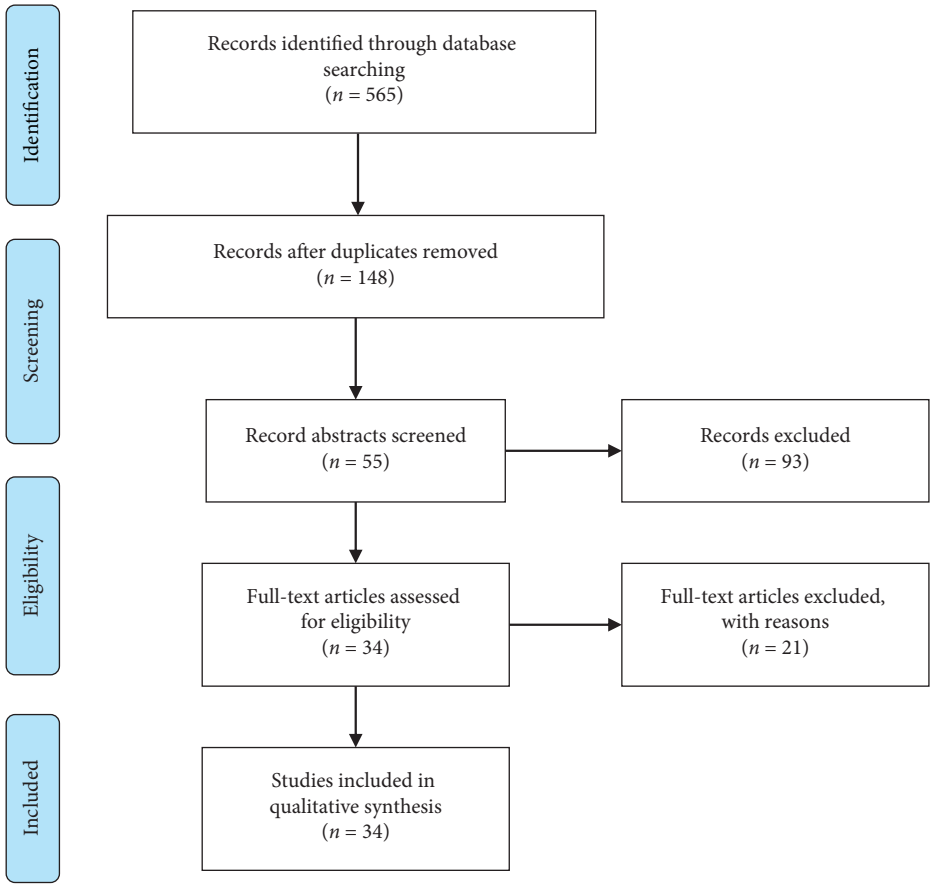

Figure 1: PRISMA flowchart of summary findings.

(Supplementary Table S2), (3) pain associated with trauma and neuralgias [8, 11, 16, 42] (Supplementary Table S3), (4) pain associated with cysts and tumors [15, 43, 44] (Supplementary Table S4), (5) pain associated with glandular disorders [12–14, 45] (Supplementary Table S5), and (6) pain arising from bone and temporomandibular joint [9, 10, 46–48] (Supplementary Table S6). The details of the articles excluded (and the entire study selection process) during systematic screening have been documented in Supplementary Material S7; Section 1.

3.3. Risk of Bias and Applicability. The current study of 34 published documents identified 8 articles [5, 6, 12–15, 31, 39] that made direct comparisons between the diagnostic accuracy of machine learning models and human clinicians. Of the 15 points evaluated from the MI-CLAIM checklist, all but one paper [39] scored over 13. JBI-DTA was assessed over 7 points where all papers scored 5 or more. Five of the 8 articles [5, 12–15, 39] could not avoid a case-control design as it was an integral part of the machine training process as found during MI-CLAIM. A "Range from studies" GRADE approach was undertaken to evaluate the collective diagnostic certainty of machine learning applicability. The GRADE approach suggested that a high certainty of diagnostic evidence for both positive and negative cases was present in machine learning. However, there were serious risks of collective bias and design inconsistencies among the cross-sectional cohorts that should be considered alongside the overall GRADE score. The conditions and explanations for all findings have been provided in Supplementary Material S7; Sections 2, 3, and 4.

3.4. Diagnostic Measure Comparisons. All 34 studies have been individually documented within Supplementary Tables S1 to S6. Only the articles that made direct comparisons to clinicians have been documented in Table 1. All the studies mentioned in Table 1 have also been discussed in detail within the supplementary tables.

4. Discussion

4.1. Summary of Findings. The current review explored the clinical influence, effectiveness, limitations, and human comparison outcomes of machine learning. The findings of all 34 papers included within the systematic review have been discussed in the following subsections: (1) pain associated with dental diseases, (2) pain associated with periodontal diseases, (3) pain associated with trauma and neuralgias, (4) pain associated with cysts and tumors, (5) pain associated with glandular disorders, and (6) pain arising from bone and temporomandibular joint.

Table 1: Summary outcomes of studies comparing diagnostic measures.

Author	Target condition definition	Testing sample size[a]	Index test outcomes[b]	Reference test outcomes[c]
Cantu et al. [31]	Extent and infiltration of proximal caries into dentinal tissue	141	Sn = 0.75, Sp = 0.83	Sn = 0.36, Sp = 0.91
Endres et al. [5]	Detect and classify periapical inflammation	102	Sn = 0.51	Sn = 0.51
Kise et al. [13]	Diagnose Sjogren syndrome in parotid and submandibular glands	40	*Parotid Gland* Sn = 0.90, Sp = 0.89 *Submandibular Gland* Sn = 0.81, Sp = 0.87	*Parotid Gland* Sn = 0.67, Sp = 0.86 *Submandibular Gland* Sn = 0.78, Sp = 0.66
Yang et al. [15]	Detect the presence of pathologic growth	181	Sn = 0.68	*Oral surgeons* Sn = 0.67 *General dentists* Sn = 0.64
Kim et al. [39]	Localize periodontal bone loss and classify apical lesions	800	Sn = 0.77, Sp = 0.95	Sn = 0.78, Sp = 0.92
Kise et al. [12]	Identify fatty degeneration within the salivary glands	100	Sn = 1.00, Sp = 0.92	*>3 years' experience* Sn = 0.99, Sp = 0.97 *<3 years' experience* Sn = 0.78, Sp = 0.89
Krois et al. [6]	To detect the extent of periodontal bone loss	353	Sn = 0.81, Sp = 0.81	Sn = 0.92, Sp = 0.63
Murata et al. [14]	Identify features of sinusitis	120	Sn = 0.86, Sp = 0.88	*>3 years' experience* Sn = 0.90, Sp = 0.89 *<3 years' experience* Sn = 0.78, Sp = 0.75

Sn: sensitivity; Sp: specificity; [a]Testing samples: medical imaging data (radiographs/ultrasound/computed tomography); [b]Index test: machine learning model; [c]Reference test: human clinicians.

4.1.1. Pain Associated with Dental Diseases. Real-time quantification of subjective dental pain demonstrated varying degrees of accuracy across multiple machine learning models when Hu et al. [1] attempted to detect (Ac = 0.80, Sn = 0.41, Sp = 0.89) and localize (Ac = 0.74, Sn = 0.54, Sp = 0.86) the source and intensity of dentin hypersensitivity pain arising from prefrontal and primary sensory cortices. The findings, in combination with Chattopadhyay's results [33], may contraindicate the implementation of an intelligent pain prediction system for perceived dental pain. Machine learning models based on clinically perceived pain produced less accurate outcomes for pulpal (Ac = 0.74–0.78, Sn = 0.48–0.71, Sp = 0.73–0.93) and periodontal diseases (Ac = 0.81, Sn = 0.78, Sp = 0.88), with the least accuracy (Ac = 0.64, Sn = 0.64, Sp = 0.96) for alveolar abscess [33]. Therefore, it can be argued that identifying the elusive source of dental pain is a more reliable estimate than quantifying pain as a symptom.

However, both proximal and periapical radiographs (Ac = 0.80, Sn = 0.75, Sp = 0.83) [31, 34] as well as histologically (Ac = 0.98, Pr = 0.98) trained models [2, 34] were able to reliably detect caries as a source for pain. While the aforementioned is considerably more efficient than clinicians (Ac = 0.71, Sn = 0.36, Sp = 0.91), dental specialists play an important role in training the machine from radiographs [3] or histological data [2]. Therefore, the prediction of the system may be directly dependent on the experience and agreement of the trainers.

Even periapical radiographs were capable of effectively (Ac = 0.82) detecting caries progression in posterior teeth [3]. Training dataset based on photographs ($n^L = 425$, Sn = 0.77–0.98, Sp = 0.84–0.96) [32] and photodetection ($n^L = 24$, Ac = 1.0) [37] produced varying outcomes when they were used to localize the progression of carious infiltration within the dentin layer [32]. This can be due to the funneling nature of caries progression as well as the small training datasets used. Many carious lesions, which visually appear negligible on the enamel surface, can funnel out within the dentin layer and cause sensitivity pain. Such factors were not considered in Rahman's study [37]. Researchers also attempted to provide camera-based intelligent solutions for end-users (patients). In such designs, video-learned systems ($n^L = 10{,}080$) produced reasonably reliable diagnoses of caries (Sn = 0.98, Sp = 0.93) and periodontitis (Sn = 0.97, Sp = 0.95) but were not very sensitive to painful microdefects like cracked teeth (Sn = 0.75, Sp = 0.99) [36].

4.1.2. Pain Associated with Periodontal Diseases. The majority of the periodontal pain was associated with periodontal bone loss and root attachment loss which were, therefore, the primary quantification parameters [5]. Clinicians' experience was assumed to play a critical role in dictating the overall accuracy of radiographic differential diagnosis in machine learning. This assumption was confirmed by Chang et al. [38], Kim et al. [39], and Krois et al. [6] who found clinicians to make poorer diagnoses (Ac = 0.76, Sn = 0.78–0.92, Sp = 0.63–0.92) than their intelligent prodigies (Ac = 0.81, Sn = 0.77–0.81, Sp = 0.81–0.95). This was eventually reflected on the deep learning model as less accurate results with more variations were obtained contradicting Endres et al. [5], who found no significant

correlation in their study. This could be due to the relatively low agreement ($k = 0.48$–0.52) between dental specialists [6, 18] in diagnosing a radiograph. Furthermore, Setzer's study [40] showed that the sensitivity of the machine in detecting periodontal diseases (Sn = 0.93, Sp = 0.88) was the same as the agreement between highly experienced specialists ($k = 0.93$). The clinicians themselves were inaccurate in diagnosing 31% of the time [5], and therefore, machine learning was deemed more specific. Periodontal conditions involved with larger bone defects [6] and indeciduous or crowded dentitions could affect predictive outcomes on panoramic radiographs (Sn = 0.84, Sp = 0.88, Pr = 0.81) [41]. Real-time/clinical machine learning, however, was less influenced by the operator's prowess [7] and heavily dependent on the accuracy of patient feedback (Ac = 0.82, Sn = 0.87, Sp = 0.76) during pain sensation [4].

4.1.3. Pain Associated with Trauma and Neuralgias. Pain associated with root fractures is difficult to diagnose without a clear radiograph. With machine learning applied to clear panoramic radiographs, the intelligent system was less sensitive to localizing fractures on anterior teeth (Sn = 0.53, Pr = 0.88) as opposed to the posterior teeth (Sn = 0.70, Pr = 0.95) [8]. This was probably due to the vertebral shadow superimposing on the dental root anatomy [8, 18]. Trauma is often accompanied by painful swelling. Zhang et al. [11] demonstrated that a trained machine with a detailed patient history was able to accurately predict (Ac = 0.94–0.98) which patients were likely to experience painful swelling after tooth extractions.

McCartney et al. [16] and Limonadi et al. [42] designed and compared questionnaire-based intelligent systems to diagnose the source of facial pain. While the systems were accurate in diagnosing typical trigeminal neuralgia (Sn = 0.84–0.92, Sp = 0.83–0.84), it was observed that deep learning was not very sensitive to atypical neuralgias (Sn = 0.50–0.63, Sp = 0.94–0.95) [16, 42]. This is partly due to the idiopathic nature of certain diseases, which cause varying clinical symptoms including pain. Such variations can cause further disagreement in differential diagnoses among specialists, whose opinions are in turn used to train and validate the intelligent systems [6, 18]. The questionnaire-based method of deep learning hinges on the patients' ability to accurately report their conditions and pain intensity and was therefore may not be preferable for evaluating dental pain [1, 33].

4.1.4. Pain Associated with Cysts and Tumors. Although most cysts, tumors, and other pathologic growths in the oral cavity are initially asymptomatic, growing lesions tend to elicit painful responses [15]. All the intelligent systems designed for tumor detection [15, 43, 44] were trained from panoramic radiographs by 2 expert radiologists. Watanabe et al. [44] carried out deep learning on larger (>10 mm) lesions, specifically radicular cyst lesions from panoramic radiographs ($n^L = 330$) where the authors found that the cortical thickness around the canine fossae and the maxillary sinus cavities drastically reduced prediction sensitivity (Sn = 0.46, Pr = 0.88 from Sn = 1.00, Pr = 0.92). Kwon's findings [43] agreed with Watanabe in that maxillary lesions were harder to predict. However, Kwon's results, which were based on a larger dataset ($n^L = 946$) and a pretrained neural network, saw comparatively better outcomes for radicular cysts (Ac = 0.96, Sn = 0.99, Sp = 0.83). This may indicate that the parameters used for machine learning in predicting oral tumors are more important than the experts who train the system. Deep learning produced better results for odontogenic keratocyst (Ac = 0.94, Sn = 0.70, Sp = 0.92, Pr = 0.63) when compared to diagnoses made by both surgeons (Sn = 0.67, Pr = 0.67) and general dentists (Sn = 0.64, Pr = 0.65) [15, 43]. This human-based discrepancy is probably due to the irregular shape and radiolucency of the tumor in respect to the rest of the mandibular anatomy. However, clinicians in Yang's study [15] were more sensitive (Sn = 0.36–0.45) to detecting well-defined ameloblastomas from radiographs than the trained machine (Sn = 0.33) [15].

4.1.5. Pain Associated with Glandular Disorders. Maxillary sinusitis is an important differential diagnosis when evaluating the source of maxillary anterior pain. This can be done clinically by observing mucus discharge or through radiographs exhibiting glandular thickening within the sinus lining [14]. Kim et al. [45] and Murata et al. [14] showed machine learning to accurately detect sinusitis from both Water's view paranasal sinus (PNS) (Ac = 0.94, Sn = 0.89, SP = 0.99) and panoramic radiographs (Ac = 0.88, Sn = 0.86, Sp = 0.88). Deep learning outcomes from panoramic radiographs were comparable to diagnoses made by radiologists who had >20 years of experience (Ac = 0.90, Sn = 0.90, Sp = 0.89) and better agreement ($k = 0.85$) in diagnoses. [14, 45] Kim also demonstrated that when multiple trained virtual machines unanimously ($k > 0.90$) diagnose an image (majority decision analysis system), they produce accurate results (Ac = 0.94) [45] comparable to radiologists with over 30 years of diagnostic experience (Ac = 0.98) [12].

When assessing glandular disorders, radiologists demonstrated better agreement ($k = 0.65$) for disorders of visibly larger glands (parotid) as opposed to smaller glands ($k = 0.51$) obstructed by bony anatomy (submandibular gland) [13]. This deemed machine learning more sensitive to glandular anomalies but was also equally prone to making mistakes. Kise developed deep-learned systems to diagnose Sjogren's syndrome from both ultrasound imaging (*parotid gland*: Ac = 0.89, Sn = 0.90, SP = 0.89; *submandibular gland*: Ac = 0.84, Sn = 0.81, Sp = 0.87) [13] and computed tomography (Ac = 0.96, Sn = 1.00, SP = 0.92) [12]. The authors found that only clinicians with >30 years of experience were able to compete (Ac = 0.98, Sn = 0.99, Sp = 0.97) with the deep learning algorithm (Ac = 0.96, Sn = 1.00, Sp = 0.92) in diagnosing salivary gland disorders from 3D CT images [12]. The outcomes for clinicians were, however, substantially poorer when made to diagnose 2D radiographs (*parotid gland*: Ac = 0.77, Sn = 0.67, Sp = 0.86; *submandibular gland*: Ac = 0.72, Sn = 0.78, Sp = 0.66) [13]. Regardless, deep learning was shown to be a valuable diagnostic support for inexperienced clinicians (Ac = 0.77–0.84, Sn = 0.78,

Sp = 0.75–0.89) [12, 14] to accurately diagnose gland-related orofacial pain.

4.1.6. Pain Arising from Bone and Temporomandibular Joint. Temporomandibular joint (TMJ) disorders can cause severe pain for the patients [10]. Some of the painful disorders addressed by machine learning include joint osteoarthritis (Ac = 0.82, Sn = 0.83, Pr = 0.81) [47, 48], osteoporosis (Ac = 0.93, Sn = 0.97, Sp = 0.86) [10], reducible disk displacements (*unilateral*: Sn = 0.80, SP = 0.95; *bilateral*: Sn = 1.00, Sp = 0.89), and nonreducible disk displacements (*unilateral*: Sn = 0.69, SP = 0.91; *bilateral*: Sn = 0.37, Sp = 1.00) [46]. However, machine learning is still in its infancy primarily due to the complex diagnostic criteria required to confirm diseases like osteoarthritis [47]. The disease requires diagnostic confirmations from clinical, radiological, and serological findings and thereby complicate the machine learning procedure. Furthermore, Nam et al. [9] found pericoronitis and alveolar abscess to commonly (44%) mimic TMJ disorders which could be accurately differentiated (Ac = 0.96, SP = 0.99, Sn = 0.69) from true cases based on clinical symptoms using machine learning [9].

4.2. Limitations of the Study. At the time of conceptualization and data collection, the review protocol and study design were not registered with any databases that indexed ongoing reviews. Past literature suggests that such registrations can guard against reporting biases and validate the integrity of the published protocol [49]. In addition to the aforementioned, the current study was limited by several other factors. Firstly, foreign articles without a formal translation were not manually translated in order to prevent misinterpretation of the technical content and, therefore, may indicate a certain degree of publication bias. Secondly, this review did not internally analyze the different machine learning models and their respective algorithms and primarily focused on the clinical parameters. Furthermore, the current study did not account for the confounding variables and factors responsible for shaping the orofacial disorders responsible for eliciting pain. The difficulty in quantifying pain encouraged focusing on specific target conditions commonly, but not solely, responsible for pain. Finally, while the diagnostic comparisons yielded high certainty and low bias, the risk of bias and quality of evidence were not evaluated across the remaining 26 studies due to missing standard reference (human clinicians) comparison.

4.3. Conclusions and Future Recommendations. Machine learning in orofacial healthcare is still emerging and has shown modest results in diagnosing oral diseases. However, such technology is far from replacing clinicians in rendering healthcare and can possibly serve as an "add-on" to the existing diagnostic tools. Various workflows and methods exist for diagnosing dental diseases that can benefit from future crossovers and randomized trials on larger pools of patients in the future.

Data Availability

All the supporting data have been provided as supplementary materials with the manuscript.

Conflicts of Interest

The authors declared no conflicts of interest.

Acknowledgments

The study was supported by the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia (grant no. 375213500) and Central Laboratory at Jouf University. The study was partially supported by Yayasan Haq through Universiti Sains Malaysia (grant no. 304/PPSG/6150182/Y115).

Supplementary Materials

Supplementary Table S1: summary findings of literature for dental diseases. Supplementary Table S2: summary findings of literature for periodontal diseases. Supplementary Table S3: summary findings of literature for dental trauma and neuralgias. Supplementary Table S4: summary findings of the literature on cystic and neoplastic lesions. Supplementary Table S5: summary findings of the literature on glandular disorders. Supplementary Table S6: summary findings of the literature on bone and joint disorders. Supplementary Material S7. (*Supplementary Materials*)

References

[1] X.-S. Hu, T. D. Nascimento, M. C. Bender et al., "Feasibility of a real-time clinical augmented reality and artificial intelligence framework for pain detection and localization from the brain," *Journal of Medical Internet Research*, vol. 21, no. 6, Article ID e13594, 2019.

[2] S. Javed, M. Zakirulla, R. U. Baig, S. M. Asif, and A. B. Meer, "Development of artificial neural network model for prediction of post-streptococcus mutans in dental caries," *Computer Methods and Programs in Biomedicine*, vol. 186, Article ID 105198, 2020.

[3] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," *Journal of Dentistry*, vol. 77, pp. 106–111, 2018.

[4] D. Teichmann, J. Klopp, A. Hallmann, K. Schuett, S. Wolfart, and M. Teichmann, "Detection of acute periodontal pain from physiological signals," *Physiological Measurement*, vol. 39, no. 9, Article ID 095007, 2018.

[5] M. G. Endres, F. Hillen, M. Salloumis et al., "Development of a deep learning algorithm for periapical disease detection in dental radiographs," *Diagnostics*, vol. 10, no. 6, p. 430, 2020.

[6] J. Krois, T. Ekert, L. Meinhold et al., "Deep learning for the radiographic detection of periodontal bone loss," *Scientific Reports*, vol. 9, no. 1, pp. 1–6, 2019.

[7] G. Meissner, B. Oehme, J. Strackeljan, and T. Kocher, "Influence of handling-relevant factors on the behaviour of a novel calculus-detection device," *Journal of Clinical Periodontology*, vol. 32, no. 3, pp. 323–328, 2005.

[8] M. Fukuda, K. Inamoto, N. Shibata et al., "Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography," *Oral Radiology*, vol. 36, no. 4, pp. 337–343, 2019.

[9] Y. Nam, H.-G. Kim, and H.-S. Kho, "Differential diagnosis of jaw pain using informatics technology," *Journal of Oral Rehabilitation*, vol. 45, no. 8, pp. 581–588, 2018.

[10] M. A. Alzubaidi and M. Otoom, "A comprehensive study on feature types for osteoporosis classification in dental panoramic radiographs," *Computer Methods and Programs in Biomedicine*, vol. 188, Article ID 105301, 2020.

[11] W. Zhang, J. Li, Z.-B. Li, and Z. Li, "Predicting postoperative facial swelling following impacted mandibular third molars extraction by using artificial neural networks evaluation," *Scientific Reports*, vol. 8, no. 1, pp. 1–9, 2018.

[12] Y. Kise, H. Ikeda, T. Fujii et al., "Preliminary study on the application of deep learning system to diagnosis of Sjögren's syndrome on CT images," *Dentomaxillofacial Radiology*, vol. 48, no. 6, Article ID 20190019, 2019.

[13] Y. Kise, M. Shimizu, H. Ikeda et al., "Usefulness of a deep learning system for diagnosing Sjögren's syndrome using ultrasonography images," *Dentomaxillofacial Radiology*, vol. 49, no. 3, Article ID 20190348, 2020.

[14] M. Murata, Y. Ariji, Y. Ohashi et al., "Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography," *Oral Radiology*, vol. 35, no. 3, pp. 301–307, 2019.

[15] H. Yang, E. Jo, H. J. Kim et al., "Deep learning for automated detection of cyst and tumors of the jaw in panoramic radiographs," *Journal of Clinical Medicine*, vol. 9, no. 6, p. 1839, 2020.

[16] S. McCartney, M. Weltin, and K. J. Burchiel, "Use of an artificial neural network for diagnosis of facial pain syndromes: an update," *Stereotactic and Functional Neurosurgery*, vol. 92, no. 1, pp. 44–52, 2014.

[17] T. Renton, "Dental (odontogenic) pain," *Reviews in Pain*, vol. 5, no. 1, pp. 2–7, 2011.

[18] T. Ekert, J. Krois, L. Meinhold et al., "Deep learning for the radiographic detection of apical lesions," *Journal of Endodontics*, vol. 45, no. 7, pp. 917–922, 2019.

[19] S. B. Khanagar, A. Al-ehaideb, P. C. Maganur et al., "Developments, application, and performance of artificial intelligence in dentistry—a systematic review," *Journal of Dental Sciences*, vol. 16, no. 1, pp. 508–522, 2020.

[20] H. Mahmood, M. Shaban, B. I. Indave, A. R. Santos-Silva, N. Rajpoot, and S. A. Khurram, "Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: a systematic review," *Oral Oncology*, vol. 110, Article ID 104885, 2020.

[21] R. Nagi, K. Aravinda, N. Rakesh, R. Gupta, A. Pal, and A. K. Mann, "Clinical applications and performance of intelligent systems in dental and maxillofacial radiology: a review," *Imaging Science in Dentistry*, vol. 50, no. 2, p. 81, 2020.

[22] E. Choi, B. Park, and H. Noh, "Efficacy of mobile health care in patients undergoing fixed orthodontic treatment: a systematic review," *International Journal of Dental Hygiene*, vol. 19, no. 1, pp. 29–38, 2020.

[23] N. Maffulli, H. C. Rodriguez, I. W. Stone et al., "Artificial intelligence and machine learning in orthopedic surgery: a systematic review protocol," *Journal of Orthopaedic Surgery and Research*, vol. 15, no. 1, pp. 1–5, 2020.

[24] P. Indovina, D. Barone, L. Gallo, A. Chirico, G. De Pietro, and A. Giordano, "Virtual reality as a distraction intervention to relieve pain and distress during medical procedures," *The Clinical Journal of Pain*, vol. 34, no. 9, pp. 858–877, 2018.

[25] J. Popay, H. Roberts, A. Sowden et al., "Guidance on the conduct of narrative synthesis in systematic reviews," *A Product from the ESRC Methods Programme*, vol. 1, p. b92, 2006.

[26] T. H. Farook, J. Radford, M. K. Alam, and N. B. Jamayet, "Plagiarism in dentistry—a systematic review," *British Dental Journal*, 2020.

[27] S. Patro and K. K. Sahu, "Normalization: a preprocessing stage," 2015, https://arxiv.org/abs/1503.06462.

[28] B. Norgeot, G. Quer, B. K. Beaulieu-Jones et al., "Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist," *Nature Medicine*, vol. 26, no. 9, pp. 1320–1324, 2020.

[29] Joanna Briggs Institute, *Critical Appraisal Tools*, Joanna Briggs Institute, Adelaide, Australia, 2020, https://joannabriggs.org/ebp/critical_appraisal_tools.

[30] H. Schünemann, J. Brożek, G. Guyatt, and A. Oxman, "Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach," 2013.

[31] A. G. Cantu, S. Gehrung, J. Krois et al., "Detecting caries lesions of different radiographic extension on bitewings using deep learning," *Journal of Dentistry*, vol. 100, Article ID 103425, 2020.

[32] E. D. Berdouses, G. D. Koutsouri, E. E. Tripoliti, G. K. Matsopoulos, C. J. Oulis, and D. I. Fotiadis, "A computer-aided automated methodology for the detection and classification of occlusal caries from photographic color images," *Computers in Biology and Medicine*, vol. 62, pp. 119–135, 2015.

[33] S. Chattopadhyay, R. M. Davis, D. D. Menezes, G. Singh, R. U. Acharya, and T. Tamura, "Application of Bayesian classifier for the diagnosis of dental pain," *Journal of Medical Systems*, vol. 36, no. 3, pp. 1425–1439, 2012.

[34] M. Hashem and A. E. Youssef, "Teeth infection and fatigue prediction using optimized neural networks and big data analytic tool," *Cluster Computing*, vol. 23, no. 3, pp. 1669–1682, 2020.

[35] H. Lee, Z. Idrees, and F. Kinsella, "Clostridium perfringens endophthalmitis following perforating eye injury," *Case Reports*, vol. 2009, no. 12, Article ID bcr0320091639, 2009.

[36] L. Liu, J. Xu, Y. Huan, Z. Zou, S.-C. Yeh, and L.-R. Zheng, "A smart dental health-IoT platform based on intelligent hardware, deep learning, and mobile terminal," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 898–906, 2019.

[37] H. A. Rahman, S. W. Harun, H. Arof et al., "Classification of reflected signals from cavitated tooth surfaces using an artificial intelligence technique incorporating a fiber optic displacement sensor," *Journal of Biomedical Optics*, vol. 19, no. 5, Article ID 057009, 2014.

[38] H.-J. Chang, S.-J. Lee, T.-H. Yong et al., "Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis," *Scientific Reports*, vol. 10, no. 1, pp. 1–8, 2020.

[39] J. Kim, H.-S. Lee, I.-S. Song, and K.-H. Jung, "DeNTNet: deep neural transfer network for the detection of periodontal bone loss using panoramic dental radiographs," *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019.

[40] F. C. Setzer, K. J. Shi, Z. Zhang et al., "Artificial intelligence for the computer-aided detection of periapical lesions in cone-

beam computed tomographic images," *Journal of Endodontics*, vol. 46, no. 7, pp. 987–993, 2020.

[41] B. Thanathornwong and S. Suebnukarn, "Automatic detection of periodontal compromised teeth in digital panoramic radiographs using faster regional convolutional neural networks," *Imaging Science in Dentistry*, vol. 50, no. 2, p. 169, 2020.

[42] F. M. Limonadi, S. McCartney, and K. J. Burchiel, "Design of an artificial neural network for diagnosis of facial pain syndromes," *Stereotactic and Functional Neurosurgery*, vol. 84, no. 5-6, pp. 212–220, 2006.

[43] O. Kwon, T.-H. Yong, S.-R. Kang et al., "Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network," *Dentomaxillofacial Radiology*, vol. 49, no. 8, Article ID 20200185, 2020.

[44] H. Watanabe, Y. Ariji, M. Fukuda et al., "Deep learning object detection of maxillary cyst-like lesions on panoramic radiographs: preliminary study," *Oral Radiology*, pp. 1–7, 2020.

[45] H.-G. Kim, K. M. Lee, E. J. Kim, and J. S. Lee, "Improvement diagnostic accuracy of sinusitis recognition in paranasal sinus X-ray using multiple deep learning models," *Quantitative Imaging in Medicine and Surgery*, vol. 9, no. 6, pp. 942–951, 2019.

[46] B. Bas, O. Ozgonenel, B. Ozden, B. Bekcioglu, E. Bulut, and M. Kurt, "Use of artificial neural network in differentiation of subgroups of temporomandibular internal derangements: a preliminary study," *Journal of Oral and Maxillofacial Surgery*, vol. 70, no. 1, pp. 51–59, 2012.

[47] J. Bianchi, A. C. de Oliveira Ruellas, J. R. Gonçalves et al., "Osteoarthritis of the temporomandibular Joint can be diagnosed earlier using biomarkers and machine learning," *Scientific Reports*, vol. 10, no. 1, pp. 1–14, 2020.

[48] P. de Dumast, C. Mirabel, L. Cevidanes et al., "A web-based system for neural network based classification in temporomandibular joint osteoarthritis," *Computerized Medical Imaging and Graphics*, vol. 67, pp. 45–54, 2018.

[49] L. Stewart, D. Moher, and P. Shekelle, "Why prospective registration of systematic reviews makes sense," *Systematic Reviews*, vol. 1, no. 1, pp. 1–4, 2012.