

Systematic Review

Clinical performance and patient outcomes of artificial intelligence–assisted shade matching in aesthetic restorations: a systematic review

Halah M. Alturkstani^{1*}, Lama M. Allehyani², Rawan S. Alqahtani², Nuha A. Alkurdi²,
Hind J. Alrefai¹, Nouf H. Alrayiqi², Bashayer N. Alharbi¹

¹Department of Restorative Dentistry, King Abdulaziz University Dental Hospital, Jeddah, Saudi Arabia

²Department of Prosthodontics, King Abdulaziz University Dental Hospital, Jeddah, Saudi Arabia

Received: 21 July 2026

Accepted: 22 September 2026

*Correspondence:

Dr. Halah M. Alturkstani,

E-mail: h.m.tur@hotmail.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Shade selection for tooth restoration is a crucial step in achieving patient satisfaction in aesthetic dentistry. The practitioner should know the aspects of color, as well as the illumination of the object, the teeth and their surroundings and the observer himself. Given the highly subjective nature of the process, conventional shade guides were developed, along with modernized shade-taking devices, to achieve excellent color-matched restorations. Nevertheless, determining the correct shade of the restoration remains a significant challenge in routine dental practice. Artificial intelligence (AI) has emerged as a promising tool to improve the efficiency of shade matching. However, the clinical impact of its implementation and the limitations it poses remain inadequately explored. We conducted a systematic review, searching PubMed, ScienceDirect and Cochrane Library for studies. Eligible studies reported outcomes on AI performance compared with traditional triage methods. Risk of bias was assessed. The included studies (n=11) demonstrated AI's potential to reduce triage time, improve accuracy and enhance decision support. However, limitations included under-triage risks, variable accuracy and predominance of single-center studies. In addition to implementation challenges, including workflow integration barriers and clinician acceptance. The use of AI models in esthetic dentistry improved accuracy in shade selection and restoration color prediction compared with conventional visual methods. Nevertheless, given the variability in results and the predominance of laboratory-based and single-center validation studies, large-scale clinical validation studies and standardized outcome reporting are required to synthesize evidence on long-term impact, patient-related esthetic outcomes and ethical considerations regarding the routine use of AI models.

Keywords: Artificial intelligence, Aesthetic dentistry, Dental restoration, Shade matching

INTRODUCTION

Aesthetic dentistry is a branch of dentistry that focuses on the appearance of the teeth, gums and the overall smile, while preserving or improving the function and occlusion. It encompasses several Aesthetic procedures, such as external bleaching, veneers, crowns, periodontal contouring and orthodontics.¹ In Aesthetic dentistry, shade selection for the tooth restoration is a crucial step in achieving patient satisfaction. Therefore, determining the correct shade of the restoration has become a challenge in

routine dental practice. To achieve accurate shade matching, the dentist should have thorough knowledge of color and its different aspects.² Color is defined as a psychophysical sensation provoked in the eye by visible light and interpreted by the brain.

It has three dimensions, which are hue, chroma and value. Hue is the principal tone of the color and is determined by the wavelength of the reflected and transmitted light. It is often used to describe the primary color of the tooth.³ The intensity of the hue of the tooth that distinguishes

between the weak and strong color is defined as the chroma. Whereas the value is the greyscale of the color, which determines the lightness or darkness of the color.³ In addition to acquiring proper knowledge about the aspects of color, the practitioner should also take into consideration other factors, such as the illumination of the object, the teeth and their surroundings and the observer himself.

These factors, combined with the aspects of color, rendered the shade selection process very subjective, so in order to standardize the shades of the restoration and facilitate the shade selection process, conventional shade guides were produced, in addition to modernized shade-taking devices to achieve excellent color-matched restorations.⁴ Conventional shade guides consist of a set of fabricated color standards that mimic the color of dental structures and are used to match natural teeth visually. These shade guides are available and cost-effective. However, it completely depends on human eye observation.⁵ Additionally, the colors in shade guides differ in each manufacturing company and the porcelain used for restoration may differ from that used in a guide.

The shades in the guide do not cover the volume of color space that exists in natural dentition. Additionally, the shade guide tab reflects and transmits light, creating a vital appearance, which is the case with restorations that appear opaque and dense.⁶ Therefore, digital shade analysis devices were developed to get rid of subjectivity during the shade selection process, obtain reproducible, reliable and precise results, allow the practitioner and technician to replicate the desired shade if they use the same software and eliminate environmental influences.⁷

In this context, Kim et al, conducted a study in which they developed a digital shade-matching device for dental color determination using the support vector machine (SVM) algorithm.⁸ They determined shade matching using shade tabs, a modified intraoral camera with cross-polarization to reduce external light interference and a precise robot arm with ± 0.1 mm position to ensure accurate positioning of the shade guide. The results showed over 90% matching accuracy and less than 1% failure rate. Similarly, Wei et al conducted another study in which they used a novel computer color-matching (CCM) system based on the improved back-propagation neural network (BPNN) model.⁹

They found that the average color differences (ΔE) value of the CCM system was 1.89 ± 0.75 , which was lower than that of the visual approach (3.54 ± 1.11 , $p < 0.01$, resulting in greater accuracy in color reproduction within the given color space compared to the traditional visual approach. Moreover, Wahjuningrum et al conducted a comparative analysis of a color matching system, using K-nearest neighbors (KNN), neural network (NN) and decision tree (DT) algorithms to classify and recognize 16 types of dental images of persons that used several extracted features, from the shade guide of teeth, with a digital

camera, ranging from 250–300 lux lighting value.¹⁰ They found that the KNN algorithm in RGB characteristics achieved 97.5% within only 0.02 seconds of computation time, providing better performance when compared to the other approaches. Finally, Kose et al conducted a study to evaluate the accuracy of machine learning (ML) regression models in predicting the final color of leucite-reinforced glass computer-aided design and computer-aided manufacturing (CAD/CAM) ceramic veneer restorations based on substrate shade, ceramic shade, thickness and translucency.¹¹ They found that the decision tree regression model demonstrated the lowest mean absolute error and highest accuracy in predicting the shade based on the substrate shade, ceramic shade and thickness.

Despite these advancements in the shade selection process, several limitations persist, such as the cost, availability, translating the selected shade during the fabrication of the restoration and selecting the right combination of materials to accurately reproduce translucency.¹² Therefore, several AI assisted software were developed to overcome these limitations and facilitate the fabrication of restorations that replicate the shade of the natural teeth.¹³ This systematic review aims to systematically evaluate the accuracy, clinical performance and esthetic outcomes of AI-assisted shade matching compared to conventional methods in esthetic dental restorations.

METHODS

Reporting guidelines

The results were reported according to the PRISMA (preferred reporting items for systematic reviews and meta-analyses) checklist.

Information sources

A comprehensive search using a combination of electronic databases will be utilized, including PubMed, ScienceDirect and Cochrane Library. The search strategy employed a combination of keywords and Medical Subject Headings (MeSH) related to the PICO framework. Keywords or search terms were combined using Boolean operators (AND/OR) to capture relevant studies.

Search strategy

To conduct a comprehensive search for a systematic review on AI-assisted shade matching compared to conventional methods in esthetic dental restorations, a combination of electronic databases was utilized, including PubMed, Science Direct and Cochrane. The search strategy will employ a combination of keywords and Medical Subject Headings (MeSH) related to the PICO framework. Keywords or search terms were

combined using Boolean operators (AND/OR) to capture relevant studies.

Keyword search strategy for pubmed

((("Dental Restoration, Permanent"(MeSH) OR "Esthetic Dentistry"(MeSH) OR esthetic restoration OR aesthetic restoration OR cosmetic dentistry OR crown OR veneer OR prosthodontic) AND ("Artificial Intelligence"(MeSH) OR artificial intelligence OR AI OR machine learning OR deep learning OR neural network OR digital technolog) AND (shade matching OR tooth shade selection OR color matching OR dental color) AND (conventional OR manual OR visual OR spectrophotometer OR colorimeter) AND (clinical performance OR treatment outcome OR esthetic outcome OR aesthetic outcome OR accuracy OR patient satisfaction OR reliability OR validity))).

Keyword search strategy for cochrane

((("esthetic dentistry" OR "aesthetic dentistry" OR "dental restoration" OR "permanent dental restoration" OR crown OR veneer OR prosthodontic OR "esthetic restoration*" OR "aesthetic restoration") AND ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network" OR AI OR "digital technology") AND ("shade matching" OR "tooth shade selection" OR "color matching" OR "dental color" OR "tooth color selection") AND (conventional OR manual OR visual OR spectrophotometer OR colorimeter) AND ("clinical performance" OR "treatment outcome*" OR "esthetic outcome*" OR "aesthetic outcome*" OR accuracy OR reliability OR validity OR "patient satisfaction"))).

Keyword search strategy for science direct

("artificial intelligence" AND "shade matching" AND dentistry AND "aesthetic restoration" AND crown AND veneer AND accuracy AND "clinical outcome"). Filters were applied to include only English-language studies in the last 10 years and focus on specific study types, such as clinical trials, randomized controlled trials and retrospective studies, ensuring relevance to the research question.

Inclusion and exclusion strategy for study selection

Studies involving participants of any age or sex who are undergoing aesthetic dental restorations, such as ceramics, crowns, veneers and CAD/CAM, in which AI-assisted shade matching is utilized. AI-assisted shade matching utilizing machine learning models, including decision trees, regression models, neural networks and SMVs. Studies reporting any outcome, such as clinical performance, reliability, patient satisfaction and shade matching accuracy. Observational studies, clinical studies, surveillance studies and case series report AI-assisted shade matching. Review studies may be screened for references but will not be included in quantitative

synthesis. Only studies published in English were included (unless translation is feasible).

Studies published before 2016 were also excluded. Additionally, single case reports, editorials, commentaries, expert opinions without primary data, conference abstracts without full text and narrative reviews were excluded.

Data extraction

All the database search results were exported to EndNote (a reference management tool) and the duplicates were removed. The data extraction was conducted by two independent investigators. The extracted information encompasses study characteristics, participant details, shade matching accuracy and classification performance and color reproduction and predictive accuracy.

Quality assessment

The quality of the included studies was evaluated using established assessment tools. The Cochrane Risk of Bias Tool was employed for randomized controlled trials, while the Newcastle-Ottawa Scale was used for observational studies. The assessment focused on methodological aspects such as study design, sample size, data collection methods and risk of bias.

RESULTS

Search results

We executed the search methodologies outlined previously, resulting in the identification of a total of 31 citations, subsequently reduced to 29 following the removal of duplicates. Upon screening titles and abstracts, only 22 citations met the eligibility criteria for further consideration. Two studies were identified from reference searching. Through full-text screening, 11 articles, aligning with our inclusion and exclusion criteria, were included. Figure 1 provides an in-depth depiction of the search strategy and screening process.

The baseline characteristics of included studies

A total of 11 studies, published between 2018 and 2026, were included in this systematic review. The included studies demonstrated heterogeneity in design, setting, dataset characteristics, AI models and outcome measures. The majority of the studies were algorithm development and validation studies, experimental laboratory evaluations, retrospective machine-learning modelling studies and cross-sectional clinical agreement analyses were also included. The combined sample sizes and dataset characteristics ranged from standardized VITA 3D-Master shade guide image datasets of 26 samples per category to 634 CAD/CAM ceramic veneer restoration records. Geographically, the studies were conducted across East Asia, Europe, South America and the Middle

East. The AI and ML methods included traditional classification algorithms such as support vector machines and k-nearest neighbours, as well as predictive regression models, fuzzy decision frameworks and mobile-based random forest classifiers. Such methods were compared with conventional visual shade selection, clinical spectrophotometric devices and dental specialist visual assessment (Table 1).

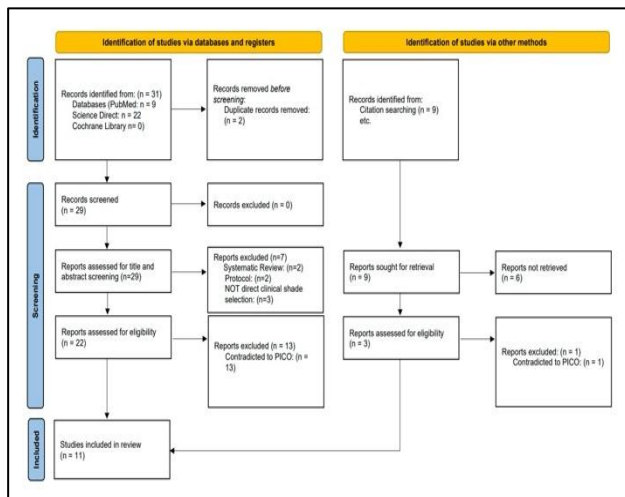


Figure 1: PRISMA flow.

Shade matching accuracy and classification performance

Having demonstrated superior classification performance across key shade-matching outcomes, AI models have been reported to be a reliable and effective method for dental color determination compared with conventional image processing methods, such as standard S-CIELAB and Euclidean distance approaches. In detail, the use of AI improved the classification accuracy, with several studies reporting strong to excellent gains in discrimination for several color models, including RGB, HSV and YCbCr formats. For instance, fuzzy decision models combining chrominance improved accuracy from 32.69% to 99.26%.

However, AI models may be well-suited for standardized shade matching, rather than that under uncalibrated lighting conditions, particularly when combined with reference-shade feature extraction. This can be attributed to the variable performance under uncontrolled illuminance. In this context, recent studies reported enhanced overall matching performance by improvements in model architecture, with SVM and Random Forest models achieving up to 96.9% and 97.1% accuracy, respectively.

Color reproduction and predictive accuracy

The use of ML models, including backpropagation neural networks and decision-tree regression models, was shown

to be feasible, effective and applicable for predicting color reproduction and final restoration shades. Some studies reported significant improvements in predictive accuracy, with decision-tree regressors achieving the mean absolute error (0.226 ± 0.012) and R^2 (0.997 ± 0.001). Moreover, evidence highlighted the importance of substrate characteristics, with substrate lightness (L^*) identified as the most influential predictor of final restoration color outcomes.

Other studies demonstrated clinically acceptable color differences, with AI-based computer color-matching systems producing significantly lower ΔE values (1.89 ± 0.75) compared to traditional visual assessment (3.54 ± 1.11 , $p < 0.01$). Overall, these findings suggest that ML can successfully complement conventional color reproduction and predictive methods in esthetic restorations (Table 3).

Reliability, clinical agreement and color difference

The use of AI models, including ChatGPT-4.0, Claude 3.5 and Microsoft Copilot, in dental shade selection was evaluated for agreement, repeatability and color difference in clinical assessment settings. In this context, cross-sectional agreement analyses demonstrated variable clinical performance in comparison to standard spectrophotometric and traditional visual methods. Notably, traditional spectrophotometers achieved significantly better color matching, with a lower ΔE_{00} (0.7), compared to AI applications (1.94-2.84). Moreover, a low level of agreement ($\kappa = 0.064$, $p = 0.287$) was observed between traditional methods and general AI models. However, ChatGPT-4.0 showed the best overall performance among AI models with near-expert color differences (3.96 ± 1.97) and fair repeatability ($\kappa = 0.33$), whereas models like Copilot exhibited poor and inconsistent clinical agreement ($\kappa = -0.12$) (Table 4).

The quality assessment demonstrated substantial methodological heterogeneity across the included studies. Most studies showed a low risk of bias for the reference standard because objective and validated color measurement methods, such as spectrophotometers, CIELab coordinates and ΔE calculations, were commonly used.

However, patient/specimen selection and index test domains frequently presented moderate to high risks of bias due to the use of small datasets, in vitro experimental designs, controlled imaging conditions, limited shade ranges and the absence of external validation.

Applicability concerns were generally low for the reference standard but moderate to high for patient selection and index test domains, as many investigations relied on shade tabs, ceramic specimens or standardized laboratory settings that may not fully reflect the complexity and variability of real clinical environments (Table 5).

Table 1: Baseline characteristics of included studies.

Study (Author, Year)	Study design	Sample/Data source	Setting/material	Primary focus
Kim et al⁸, 2018	Experimental validation study	Intraoral digital images of dental shades	Clinical dental imaging system	Accuracy and reliability of shade classification
Lin et al¹⁴, 2018	Algorithm development / validation study	VITA 3D-Master shade tab images (26 samples per dataset)	Standardized dental photography datasets	Improvement of dental shade matching accuracy
Wei et al⁹, 2018⁹	Experimental laboratory study	Dental ceramic color samples	Controlled color measurement environment	Color reproduction accuracy (ΔE reduction)
Justiawan et al¹⁰, 2019	Experimental ML classification study	Dental images with RGB/HSV/Lab color features	Image processing dataset	Effect of feature space on classification accuracy
Chen et al¹⁵, 2020	Experimental ML + fuzzy decision study	VITA 3D-Master shade images (datasets A–D)	Standardized clinical + lab imaging setup	Dental shade matching optimization using fuzzy AI
Kose et al¹¹, 2023	Retrospective ML modeling study	CAD/CAM ceramic veneer dataset	Digital prosthodontic color prediction environment	Prediction of final restoration color
Li et al¹⁶, 2025	Algorithm development study	Dental photography dataset	Standardized VITA 3D-Master system images	Improvement of dental image color standardization
Şahin et al¹⁷, 2025	Comparative experimental study	Clinical dental shade samples	Clinical dentistry setting	Comparison of AI vs clinical spectrophotometric accuracy
Polatoğlu et al¹⁸, 2025	Cross-sectional agreement study	Dental shade selection cases	Clinical + digital + AI-assisted environments	Agreement between AI and clinical methods
Ozdemir et al¹⁹, 2026	Comparative diagnostic accuracy study	Clinical shade selection cases	Dental clinical assessment setting	AI reliability and agreement in shade matching
Wanna et al²⁰, 2022	Experimental ML model development study	Mobile phone dental image dataset	Real-world uncontrolled lighting conditions	Mobile AI-based dental shade classification

Table 2: Shade matching accuracy and classification performance.

Study (Author, Year)	Outcome domain	AI/digital method	Comparator	Effect size / result	Unit	Direction of effect	Key finding
Kim et al⁸, 2018	Dental color determination matching accuracy and reliability	digital shade-matching device for dental color determination by applying the support vector machine (SVM) algorithm	Euclidean distance method; logistic regression, random forest and k-nearest neighbours (KNN), using the leave-pair-out cross-validation method to verify the classification performance of the SVM algorithm.	SVM accuracy: 96.9% (LPOCV); KNN: 96.1%; RF: 89.3%; LR: 77.2%; SVM device reliability: 95.0 (0.36), 91.9 (0.71) and 93.5 (0.76) for devices	% accuracy	Positive improvement - SVM significantly improved classification stability and accuracy vs Euclidean distance and traditional ML methods	SVM-based color classification with calibrated intraoral imaging achieved the highest and most stable shade-matching performance compared with Euclidean distance and other ML models
Lin et al, 2018	Dental shade matching accuracy & colorimetric analysis	Digital Image processing (conversion of the images in RGB format to YCbCr format, PSNR on Cb/Cr, CPSNR, S-CIELAB, fuzzy	S-CIELAB-based method and traditional PSNR/CPSNR approaches	Accuracy improved from 32.69% (S-CIELAB) to 92.31% (fuzzy decision); score improved from 79.36 to 98.74	% / score	Positive improvement - proposed methodology performs better than the previous	Fuzzy decision combining chrominance (Cb, Cr) and multiple image quality

Continued.

Study (Author, Year)	Outcome domain	AI/digital method	Comparator	Effect size / result	Unit	Direction of effect	Key finding
		decision model)				work and other methods	metrics significantly improved dental shade matching accuracy compared to S-CIELAB and single-metric methods
Justiawan et al ¹⁰ , 2019	Dental color matching accuracy	Color feature extraction (RGB, HSV and Lab color moment characteristic calculation);	K-nearest neighbors (KNN), neural network (NN) and decision tree (DT) algorithms	Best testing accuracy: 92.5% (RGB + KNN); HSV: 78.75%; Lab: 77.5%	% accuracy	Positive improvement (improved classification performance with RGB+KNN under stable lighting)	RGB color model with KNN achieved highest accuracy and fastest computation, while HSV/Lab were more robust under variable lighting conditions
Chen et al, 2020	Shade matching accuracy & image similarity performance	HSV color model, peak signal to-noise ratio (PSNR), structural similarity index (SSIM), composite peak signal-to-noise ratio (CPSNR) and S-CIELAB (Special International Commission on Illumination, L* for lightness, a* from green to red and b* from blue to yellow)	dataset A vs B; unweighted vs weighted models; YCbCr vs HSV	S-CIELAB (HSV): 89.57%; CPSNR (HSV): 90.53%; SSIM: 88.24%; PSNR(V): 83.58%; PSNR(Hw): 97.64%; PSNR(Sw): 99.93%; Fuzzy decision: 99.26%; baseline (YCbCr S-CIELAB): 32.69%, i.e., performance improvement: From 32.69% -> 99.26% (fuzzy decision method)	% accuracy	Positive increase (improvement in accuracy of dental shade matching)	The proposed methods (especially PSNR(Hw), PSNR(Sw) and fuzzy decision model) significantly increase classification accuracy compared to baseline methods
Li et al, 2025	Accuracy prediction, 2-tier color correction strategy	Machine learning (ML) models for color correction, including polynomial regression (PR), backpropagation neural network (BPNN) and extreme learning machine (ELM)	Alternative ML models	RMSE=2.2; R ² =0.996; accuracy=73.1%	RMSE, R ² , %	Positive effect of AI model	2-tier color correction strategy based on the ML models and the color space specified by the VITA 3D-Master system effectively standardized the color of dental

Continued.

Study (Author, Year)	Outcome domain	AI/digital method	Comparator	Effect size / result	Unit	Direction of effect	Key finding
							photographs and provided a more accurate and stable method
Wanna et al, 2022	Dental shade matching accuracy and classification performance	Mobile phone camera-based AI framework using ML models (Random Forest Classifier, Multi-class SVM Classification and XGBoost) with reference shade-guided feature extraction	Internal comparison across ML models; RGB-only vs RGB target+reference shade; subset vs full dataset training	Random Forest model outperformed the other models; the best model achieved 97.1% accuracy and 97.2% F1-score; XGBoost accuracy=92.7% ; SVM accuracy=59.6% (best subset setting with reference shade + outlier removal)	% accuracy, F1-score	AI model achieved high classification performance, strongly improved with reference-shade feature and outlier removal	DentShade AI framework using mobile phone images achieved high shade classification accuracy, with Random Forest outperforming other ML models; reference shade inclusion significantly improved performance under uncontrolled lighting conditions

Table 3: Color reproduction and predictive accuracy.

Study (Author, Year)	Outcome domain	AI/digital method	Comparator	Effect size/result	Unit	Direction of effect	Key finding
Wei et al⁹, 2018	Color reproduction accuracy	A CCM system based on the improved BPNN model constructed using MATLAB software; second-order algorithms: Levenberg-Marquardt)	Conventional visual shade selection method	CCM: mean color differences (ΔE)=1.89±0.75; Visual: mean ΔE =3.54±1.11; BPNN simulation ΔE =0.95-1.26; $p < 0.01$	ΔE (CIE color difference)	Lower ΔE indicates better accuracy	AI-based CCM system significantly improved color matching accuracy compared with visual assessment, producing clinically acceptable and lower color differences.
Kose et al¹¹, 2023	Prediction of final color of leucite-reinforced glass CAD/CAM ceramic veneer restorations	Supervised ML (28 regression models; decision tree regression model with grid search cross-validation (GridSearchCV, Scikit-learn))	Conventional statistical models (multiple regression-based ML models compared: linear, SVR,	decision tree regressor was the best model of all regression models, with the lowest mean absolute error (0.226±0.012) and the highest R^2 score (0.997±0.001);	R^2 , MAE, ΔE_{00}	Strong positive effect (high predictive accuracy with low error)	Decision tree model accurately predicted final restoration color; substrate L^* was the most influential

Continued.

Study (Author, Year)	Outcome domain	AI/digital method	Comparator	Effect size/result	Unit	Direction of effect	Key finding
	based on substrate shade, ceramic shade, thickness and translucency		random forest, etc.)	MAE=0.226±0.012; mean ΔE_{00} prediction error=0.382±0.316; only 2/634 cases exceeded the acceptability threshold ($\Delta E_{00} > 1.77$)			predictor, followed by ceramic thickness, demonstrating strong substrate dominance in color outcome

Table 4: Reliability, clinical agreement and color difference.

Study (Author, Year)	Outcome domain	AI / digital method	Comparator	Effect size/result	Unit	Direction of effect	Key finding
Şahin et al, 2025	color matching accuracy	ChatGPT-4, Gemini 1.5 Pro	Easy shade spectrophotometer	ΔE_{00} : AI=1.94-2.84 vs ES=0.7	ΔE_{00}	AI inferior effect	The Easyshade spectrophotometer was found to be more successful than AI applications in color matching. AI color matching has demonstrated success above the clinically acceptable threshold
Ünal & Polatoğlu, 2025	agreement among traditional, digital and AI-assisted shade selection methods	ChatGPT-4	TM, spectrophotometer, IOS	traditional method vs intraoral scanner (IOS): statistically moderate agreement ($\kappa=0.421$; $P<.001$), Traditional method vs AI: statistically low agreement ($\kappa=0.064$; $P=.287$). A slight yet statistically significant agreement was found among the 4 methods (Fleiss $\kappa=0.071$, $P<.05$).	Kappa	AI inferior effect	Very low agreement between AI and clinical methods
Ozdemir et al, 2026	Dental shade selection accuracy, repeatability, agreement with expert and color difference	ChatGPT-4.0 (OpenAI, San Francisco, CA, USA), Microsoft Copilot Pro (Microsoft)	Dental specialist visual assessment	Agreement: ChatGPT-4.0 64%, Claude 50%, Copilot 24%; Fleiss' κ : 0.33-0.45 (fair-moderate for ChatGPT/Claude), -0.12 (poor for	%, κ , ΔE_{00}	Mixed (ChatGPT-4.0 demonstrating fair repeatability ($\kappa=0.33$), Claude 3.5 showing	ChatGPT-4.0 showed the best overall performance among AI models with near-expert ΔE_{00} and higher agreement, while Copilot

Continued.

Study (Author, Year)	Outcome domain	AI / digital method	Comparator	Effect size/result	Unit	Direction of effect	Key finding
		Corporation, Redmond, WA, USA) and Claude 3.5 Sonnet (Anthropic, San Francisco, CA, USA)		Copilot); ΔE_{00} : Specialist 3.46 ± 1.61 vs ChatGPT 3.96 ± 1.97 , Claude 4.62 ± 2.76 , Copilot 5.31 ± 2.92		moderate repeatability ($\kappa=0.45$) and Microsoft Copilot exhibiting poor repeatability ($\kappa=-0.12$).	performed poorly and inconsistently in shade matching

Table 5: QUADAS-2 risk of bias assessment.

Study	Risk of bias				Applicability		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Chen et al ¹⁵ , 2020	High	High	High	High	High	High	Moderate
Justiawan et al ¹⁰ , 2019	High	High	High	High	Moderate	Moderate	Moderate
Kim et al ⁸ , 2018	Moderate	Low	Moderate	Low	Moderate	Low	Moderate
Kose et al ¹¹ , 2023	Moderate	Low	Low	Low	Moderate	Low	Low
Li et al ¹⁶ , 2025	Moderate	Low	Low	Low	Moderate	Low	Low
Lin et al ¹⁴ , 2018	Low	Low	Low	Low	Low	Low	Low
Ozdemir et al ¹⁹ , 2026	High	Moderate	Low	Low	High	Moderate	Low
Şahin et al ¹⁷ , 2025	Low	Moderate	Low	Low	Low	Moderate	Low
Polatoğlu et al ¹⁸ , 2025	Low	High	Moderate	Low	Low	High	Moderate
Wei et al ¹⁹ , 2018	Moderate	High	Low	Low	Moderate	High	Low
Wanna et al ²⁰ , 2022 ²⁰	Moderate	Low	Moderate	Low	Moderate	Moderate	Moderate

DISCUSSION

This systematic review synthesizes current evidence on the efficacy and reliability of AI in dental shade matching. Findings revealed a marked difference in performance between laboratory-based computational accuracy and real-world clinical application. AI demonstrates significant potential to enhance standardized shade determination and predictive modeling. However, given its current inconsistent performance, particularly in clinical agreement with visual assessment and standard spectrophotometry, it remains in a transitional phase rather than fully replacing conventional clinical methods.

Regarding shade classification, advanced algorithms and fuzzy decision models achieved excellent accuracy, outperforming conventional image processing methods. In this context, Chen et al highlighted that this substantial improvement in AI models over conventional methods can be attributed to their ability to extract high-dimensional features.¹⁵ Justiawan et al further noted that these improvements are particularly effective when

coupled with reference-shade markers that neutralize the deleterious effects of uncontrolled illuminance.¹⁰ Notably, the superior performance of systems utilizing SVMs results from their ability to process color data more effectively than standard baseline approaches. Overall, these findings suggest that AI is best utilized when imaging protocols are strictly standardized, rather than with uncalibrated datasets.

In terms of predictive accuracy and color reproduction, ML models, particularly backpropagation neural networks and decision-tree regressors demonstrated high feasibility for estimating restoration outcomes, achieving strong predictive correlations that surpass those of traditional statistical models, as reported by Wei et al and Kose et al, respectively.^{9,11} The improved performance observed in these AI-driven systems results from the models' ability to weigh complex substrate characteristics, with substrate lightness being the paramount determinant for final color outcomes, as demonstrated by Kose et al.¹¹ Unlike manual visual assessment, which is prone to metamerism and cognitive fatigue, these AI-driven systems provide consistent color differences that remain

within clinically acceptable limits. These findings suggest that applying algorithmic processing of optical properties to material selection provides a more reliable foundation for aesthetic restorations than the subjective visual assessment traditionally employed in clinical practice. On the contrary, analyses of clinical agreement and reliability highlight a significant difference between AI applications and established gold standards, such as the spectrophotometer. In this context, Şahin et al, found that the spectrophotometer consistently outperforms AI-based approaches.¹⁷ While general Language Models showed the best performance among AI cohorts, they still produced greater color differences than professional spectrophotometry did. In this context, Ünal and Polatoğlu reported a statistically weak agreement between these models and traditional clinical methods.¹⁸

Ozdemir et al, further reported inconsistent results with poor repeatability in some models.¹⁹ Such disparities can be attributed to general-purpose AI models lacking dedicated multispectral sensors and calibrated color-science engines inherent to the specialized clinical devices. These findings highlight the accessibility of generalist models but note that they lack the specialized processing power to achieve the reliability required in aesthetic restorative dentistry compared with professional spectrophotometric benchmarks.

Limitations and future recommendations

This systematic review is limited by methodological heterogeneity regarding training datasets, sample sizes, AI models, validation strategies, comparators and outcome measures in the included studies, which introduce inherent bias and limits direct comparisons between studies, thereby hindering the generalizability of findings. A gap exists due to the lack of long-term follow-up data on restoration survival, color stability and patient-centered esthetic outcomes. Of greater concern, the rapid evolution of generative AI models and the variety of validation approaches employed across studies require cautious interpretation, as current findings represent a may not account for future algorithmic improvements.

To facilitate the safe clinical integration of AI models into routine esthetic clinical settings, large-scale prospective studies with standardized image acquisition protocols and external validation frameworks are required. Therefore, future studies should focus on developing universal standards for image acquisition and lighting to ensure consistent performance across different clinical settings and on investigating the association between specialized ML algorithms and general-purpose models to bridge the gap in clinical agreement.

CONCLUSION

Having demonstrated improved accuracy in shade selection and restoration color prediction compared with conventional visual methods, the integration of AI into

esthetic dentistry offers a promising tool for improving the efficiency of dental restorative workflows. Given the variability in results and the predominance of laboratory-based and single-center validation studies, large-scale clinical validation studies and standardized outcome reporting are required to synthesize evidence on long-term impact, patient-related esthetic outcomes and ethical considerations regarding the routine use of AI models.

Funding: No funding sources

Conflict of interest: None declared

Ethical approval: Not required

REFERENCES

1. Goodluck E. The Art and Science of Esthetic Dentistry: Balancing Functionality and Aesthetics in Modern Practice. 2020.
2. Helmy N, Naguib AH, El-Naggar G. Patient satisfaction and dentist shade matching of Bio HPP compared to Lithium disilicate crowns in Anterior Aesthetic zone. Randomized controlled clinical trial). *IOSR J Dent Med Sci (IOSR-JDMS)*. 2020;19(1):1-10.
3. Norman RD, Scott WA. Color and affect: a review and semantic evaluation. *J General Psychol*. 1952;46(2):185-223.
4. Alayed MA, Alnasyan AS, Aljutayli AA, Alzaben MM, Alrusayni WM, Al Hujaylan AA. Considerations and implications in shade selection for dental restorations: a review. *J Pharm BioAllied Sci*. 2021;13(2):898-902.
5. Paravina RD. Performance assessment of dental shade guides. *J Dentist*. 2009;37:15-20.
6. Gokce HS, Piskin B, Ceyhan D, Gokce SM, Arisan V. Shade matching performance of normal and color vision-deficient dental professionals with standard daylight and tungsten illuminants. *J Prosth Dent*. 2010;103(3):139-47.
7. Sirintawat N, Leelaratrunguang T, Poovarodom P, Kiattavorncharoen S, Amornsettachai P. The accuracy and reliability of tooth shade selection using different instrumental techniques: An in vitro study. *Sensors*. 2021;21(22):7490.
8. Kim M, Kim B, Park B, Lee M, Won Y, Kim C-Y, et al. A digital shade-matching device for dental color determination using the support vector machine algorithm. *Sensors*. 2018;18(9):3051.
9. Wei J, Peng M, Li Q, Wang Y. Evaluation of a novel computer color matching system based on the improved back-propagation neural network model. *Journal of Prosthodontics*. 2018;27(8):775-83.
10. Justiawan, Wahjuningrum DA, Hadi RP, Nurhayati AP, Prayogo K, Sigit R, et al. Comparative analysis of color matching system for teeth recognition using color moment. *Medical Dev Evid Res*. 2019;6:497-504.
11. Kose Jr C, Oliveira D, Pereira PN, Rocha MG. Using artificial intelligence to predict the final color

- of leucite-reinforced ceramic restorations. *J Esth Rest Dent*. 2023;35(1):105-15.
12. Hein S, Modrić D, Westland S, Tomeček M. Objective shade matching, communication and reproduction by combining dental photography and numeric shade quantification. *J Esth Rest Dent*. 2021;33(1):107-17.
13. Awdaljan MW, Roque JC, Choi J, Rondón LF. Introducing a novel approach to dental color reproduction using AI technology. *J Esth Rest Dent*. 2024;36(12):1623-37.
14. Lin TL, Chuang CH, Chen SL, Lin NH, Miaou SG, Lin SY, et al. An efficient image processing methodology based on fuzzy decision for dental shade matching. *J Intel Fuzzy Sys*. 2019;36(2):1133-42.
15. Chen SL, Zhou HS, Chen TY, Lee TH, Chen CA, Lin TL, et al. Dental shade matching method based on hue, saturation, value color model with machine learning and fuzzy decision. *Sens Mater*. 2020;32(10):3185-207.
16. Li Y, Zhou Y, Lyu J, Cao X, Tan J, Liu X. Feasibility and accuracy of semi-active robot-assisted porcelain veneer tooth preparation: An in vitro study. *J Dentistry*. 2025;6:106012.
17. Şahin N, Kaleli N, Ural Ç. Evaluation of color matching accuracy using artificial intelligence applications and a spectrophotometer: A photometric analysis. *J Prost Dentist*. 2025;134(5):1955.
18. Ünal M, Polatoğlu S. Comparative agreement among traditional, digital and experimental AI-based shade selection methods in dentistry: A clinical study. *J Prost Dent*. 2025;4:67.
19. Ozdemir SB, Ozdemir B, Ural C. Repeatability of Artificial Intelligence Chatbots in Composite Shade Selection: Agreement with a Dental Specialist. *Applied Sciences*. 2026;16(5):2306.
20. Wanna Y, Wiratchawa K, Leenaracharongruang R, Sittiwong W, Panpisut P, Intharah T, editors. *DentShadeAI: a Framework for Automatic Dental Shade Matching through Mobile Phone Camera*. 2022 37th International Technical Conference on Circuits/Systems. Computers and Communications. 2022;6:43.

Cite this article as: Alturkstani HM, Allehyani LM, Alqahtani RS, Alkurdi NA, Alrefai HJ, Alrayiqi NH, et al. Clinical performance and patient outcomes of artificial intelligence–assisted shade matching in aesthetic restorations: a systematic review. *Int J Community Med Public Health* 2026;13:6055-65.