

Original Research

Received 05 June 2026

Revised 19 July 2026

Accepted 02 Aug 2026

Natural Resources for Human Health 2026, Vol. 6 (15s): 1205–1217

KEYWORDS:

artificial intelligence; electronic health records; dentistry; family medicine; infectious diseases; machine learning

Artificial Intelligence and Electronic Health Records: A Systematic Review in Dentistry, Family Medicine, Infectious Disease

Abdulhadi Mohammed Alqahtani^{1*}, Abdulaziz Essam Alshail², Turki Abdullah Aldalaan³, Rehab Saker Alanazi⁴, Ahmed Ali Sharaf Alkhalifa⁵, Osama Jamaan M Alzahrani⁶, Zahwah Saleh Albalwi⁷, Albandari Zahid Alshammari⁸, Ahmed Sultan Alharbi⁹, Alwaleed Abdullah A Bashah¹⁰, Abdulaziz Eid A Alotaibi¹¹, Younes Mrwae A Hjrini¹², Khalid Omar A Bin Haylan¹³

¹King Abdulaziz Hospital, Makkah, Saudi Arabia

²General Directorate of Medical Services, Alnakhil Medical Center, Saudi Arabia

³Alnakheel Medical Center, Saudi Arabia

⁴King Saud University, Dental University Hospital, Saudi Arabia

⁵King Abdullah Medical City, Makkah, Saudi Arabia

⁶Security Forces Hospital – Makkah (SFHM), Saudi Arabia

⁷King Salman Military Hospital, Saudi Arabia

⁸University of Hail, Saudi Arabia

⁹Al-Malaz Prison Health Center, Saudi Arabia

¹⁰Prison Health Center, Saudi Arabia

¹¹Medical Services, Saudi Arabia

¹²Prison Health Center, Saudi Arabia

¹³General Directorate of Prisons Health, Saudi Arabia

Corresponding author: Abdulhadi Mohammed Alqahtani,

Email: amq2524@gmail.com

ABSTRACT: Background: Electronic health records (EHRs) generate large volumes of structured and unstructured clinical data that are increasingly mined by artificial intelligence (AI) and machine learning (ML) tools to support diagnosis, risk prediction, and care management. Dentistry, family medicine, and infectious disease care each depend heavily on EHR documentation, yet evidence on EHR-based AI applications across these three fields has not been jointly synthesized.

Objective: To systematically review primary studies that developed, validated, or applied AI/ML models using EHR or electronic dental record (EDR) data in dentistry, family medicine, or infectious disease settings.

Methods: PubMed/MEDLINE, Scopus, Embase, Web of Science, and the Cochrane Library were searched for English-language randomized controlled trials, cohort, case-control, cross-sectional, and case-series studies published between 2009 and 2026. Screening and data extraction followed the PRISMA 2020 statement. Risk of bias was assessed across participant-selection, predictor, outcome, and analysis domains. Findings were synthesized narratively because of methodological and outcome heterogeneity.

Results: Of 1,848 records identified, 12 studies met the inclusion criteria (4 in dentistry, 4 in family medicine, 4 in infectious diseases), comprising sample sizes from 218 to over 130,000 patients. Gradient-boosted ensemble models, particularly XGBoost, were the most frequently top-performing algorithms across all three domains. Reported discrimination was generally moderate to good where quantified, and natural language processing of unstructured EHR text improved prediction in infectious disease applications. Most studies were single-center, retrospective, and lacked external validation, and risk of bias was predominantly rated as “some concerns.”

Conclusion: EHR-embedded AI models show consistent, domain-spanning potential to augment early risk identification in dentistry, family medicine, and infectious diseases, but methodological heterogeneity and limited external validation currently constrain clinical translation. Prospective, multicenter, externally validated studies with standardized outcome reporting are needed.

1. INTRODUCTION

Artificial intelligence (AI) is increasingly embedded in routine clinical workflows, and the electronic health record (EHR) has become its principal data substrate. As health systems digitized documentation over the past two decades, EHRs accumulated longitudinal structured data (diagnoses, procedures, laboratory results, medications) alongside unstructured free text, creating a resource well suited to machine learning (ML) methods that can detect patterns beyond the reach of traditional statistical models (Topol, 2019; Rajkomar et al., 2019). The convergence of high-volume clinical documentation and increasingly accessible ML tooling has generated a rapidly expanding literature applying AI to EHR-derived data for diagnosis, risk stratification, and decision support.

This review focuses on three clinical domains in which EHR documentation is central to everyday practice but in which the AI literature has developed largely in parallel rather than in dialogue: dentistry, family medicine, and infectious diseases. Each of these fields differs in disease scope and data structure, yet all three share a common methodological challenge, namely how to convert heterogeneous, often incomplete, EHR-derived variables into clinically actionable predictions. Clinical decision support built on such data must therefore be evaluated not only for statistical performance but also for its fit within real-world workflows (Shortliffe and Sepúlveda, 2018).

In dentistry, periodontal disease and dental caries remain highly prevalent and are traditionally assessed through clinical examination and static risk indices that do not fully exploit the longitudinal information captured in electronic dental records (EDRs). Reviews of AI in dentistry have shown that most published work concentrates on image-based diagnosis (radiographs, intraoral photographs), leaving non-image, EDR-derived risk prediction comparatively underexplored despite its potential to support population-level preventive screening (Ahmed et al., 2021).

Family medicine and primary care occupy a similarly data-rich but methodologically fragmented position. Primary care EHRs capture the broadest cross-section of a population's health trajectory, from chronic disease management to health-related quality of life and workforce utilization, yet AI applications built on this data span a wide and inconsistent set of outcomes. Recent syntheses of diagnostic prediction models in primary care have highlighted both the promise of EHR-based AI for earlier disease detection and the persistent gap between model development and prospective clinical validation (Hunik et al., 2025).

Infectious disease management, by contrast, is defined by time-critical decision-making, where delayed recognition of conditions such as sepsis or emerging outbreaks measurably worsens outcomes. EHR-embedded AI in this domain has moved beyond structured vital-sign and laboratory-based early-warning scores toward natural language processing (NLP) of free-text clinical notes, reflecting an urgent need to extract actionable signals from all available documentation at the point of care (Peiffer-Smadja et al., 2020).

Across all three domains, common cross-cutting challenges recur: variable EHR data quality and completeness, limited external validation beyond the originating institution, risk of algorithmic bias embedded in historical documentation patterns, and uncertain integration into clinical workflows. Existing systematic reviews have examined NLP applications in EHRs broadly and have catalogued barriers to AI implementation in healthcare generally, but none has directly compared the methodological

maturity, performance, and risk of bias of EHR-based AI studies specifically across dentistry, family medicine, and infectious diseases, despite their shared reliance on similar data structures and modelling pipelines (Hossain et al., 2023; Ahmed et al., 2023).

The aim of this systematic review is therefore to identify, appraise, and synthesize primary research (randomized controlled trials, cohort, case-control, cross-sectional, and case-series studies) published between 2009 and 2026 that developed, validated, or applied AI or ML models using EHR or EDR data within dentistry, family medicine, or infectious disease care, following the PRISMA 2020 reporting guidelines (Page et al., 2021). By jointly evaluating study characteristics, reported performance, and risk of bias across these three domains, this review seeks to clarify which methodological approaches and clinical applications are most mature, where evidence gaps persist, and what steps are needed to move EHR-based AI tools from retrospective development toward validated clinical use.

2. MATERIALS AND METHODS

2.1 Protocol and Reporting

This systematic review was conducted and is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). Prospective registration (e.g., PROSPERO) is recommended prior to journal submission; a protocol registration number should be inserted here once obtained.

2.2 Eligibility Criteria

Studies were eligible if they were primary research reports — randomized controlled trials, cohort studies (prospective or retrospective), case-control studies, cross-sectional studies, or case series — published in English between 2009 and 2026, and if they developed, validated, or applied an AI/ML model using data extracted from an electronic health record or electronic dental record system in a dentistry, family medicine/primary care, or infectious disease clinical setting. Eligible studies were required to report at least one quantifiable model-performance outcome (e.g., area under the curve [AUC], accuracy, sensitivity, specificity, F1-score, or balanced accuracy).

Studies were excluded if they: relied exclusively on imaging data (radiographs, photographs, scans) without an EHR/EDR clinical-data component; were reviews, scoping reviews, editorials, commentaries, or conference abstracts without extractable quantitative data; were animal or in-vitro studies; were not published in English; fell outside the 2009–2026 publication window; or represented a duplicate analysis of a previously included cohort, in which case the most complete or most recent report was retained.

2.3 Information Sources and Search Strategy

PubMed/MEDLINE, Scopus, Embase, Web of Science, and the Cochrane Library were searched using combinations of the following concept blocks: (“artificial intelligence” OR “machine learning” OR “deep learning” OR “natural language processing”) AND (“electronic health record*” OR “electronic medical record*” OR “electronic dental record*”) AND (dentist* OR periodontal OR caries OR “family medicine” OR “primary care” OR “infectious disease*” OR sepsis OR “antimicrobial resistance” OR COVID-19). Reference lists of retrieved systematic reviews were hand-searched for additional eligible primary studies. Database searches were first conducted in June 2025 and were updated in August 2026 to capture recently published studies, which extended the final publication window to 2009–2026; the core structured search itself covered the period 2009 to June 2025, consistent with the review’s original search window.

2.4 Study Selection and Data Extraction

Titles and abstracts were screened independently against the eligibility criteria, followed by full-text assessment of potentially eligible reports; disagreements were to be resolved by discussion or, where unresolved, by a third reviewer. Data were extracted using a standardized form capturing author(s), publication year, country, clinical domain, study design, sample size, type of EHR/EDR data used (structured, unstructured/NLP, or both), AI/ML method(s) applied, predicted outcome, and reported performance metrics.

2.5 Risk-of-Bias Assessment

Risk of bias was appraised across four domains adapted from the Prediction model Risk Of Bias ASsessment Tool (PROBAST) framework and the Newcastle-Ottawa Scale for observational studies: (1) participant selection, (2) predictor/index-variable definition, (3) outcome/reference-standard definition, and (4) statistical analysis, including handling of missing data and external validation. Each domain was rated as low, some concerns, or high risk of bias, and an overall judgment was assigned per study.

2.6 Synthesis Methods

Given substantial heterogeneity in clinical domains, AI/ML methods, predicted outcomes, and reported performance metrics, a quantitative meta-analysis was not considered appropriate. Findings were instead synthesized narratively, grouped by clinical domain (dentistry, family medicine, infectious diseases), with performance metrics and study characteristics tabulated for structured comparison, consistent with PRISMA guidance for reviews in which meta-analysis is not feasible.

3. RESULTS AND DISCUSSION

3.1 Study Selection

The search identified 1,848 records (1,842 from databases and 6 from registers/citation searching). After removal of 512 duplicates, 1,336 records were screened by title and abstract, of which 1,196 were excluded as clearly irrelevant. Of the 140 reports sought for retrieval, 135 full-text articles were assessed for eligibility; 123 were excluded for the reasons detailed in Figure 1 (see Figure Legends and Figure 1 at the end of this manuscript). Twelve studies met all inclusion criteria and were included in the narrative synthesis: four in dentistry, four in family medicine, and four in infectious diseases.

3.2 Study Characteristics

Included studies were published between 2019 and 2026 and originated from the United States, Finland, Türkiye, Spain, Australia, the United Kingdom, and multinational research consortia. Study designs comprised retrospective cohort studies ($n = 6$), cross-sectional/secondary-data analyses ($n = 4$), one case-control study, and one multicenter retrospective diagnostic study; no randomized controlled trials met the eligibility criteria, reflecting the predominantly retrospective, model-development orientation of this literature. Sample sizes ranged widely, from 218 adolescents in a Finnish caries-prediction study to more than 130,000 emergency department patients in an Australian sepsis-prediction study. Table 1 summarizes the characteristics of included studies.

Table 1. Characteristics of included studies ($n = 12$)

No.	Study	Domain	Country	Design	Sample size	EHR/EDR data & AI method
10	Patel et al. (2022)	Dentistry	USA	Retrospective cohort	27,138 patients	Structured EDR (74 features); XGBoost, multiclass
11	Dey et al. (2024)	Dentistry	USA	Cross-sectional (secondary data)	Pediatric patients, ages 6–16	Structured EHR (axiUm); XGBoost vs regression
12	Çiftçi and Aşantoğrul (2024)	Dentistry	Türkiye	Cross-sectional	2,000 adults	Structured clinical/questionnaire data; multiple ML classifiers
13	Väyrynen et al. (2025)	Dentistry	Finland	Cross-sectional, dev/validation	218 adolescents	Structured risk-assessment data; XGBoost with oversampling

No.	Study	Domain	Country	Design	Sample size	EHR/EDR data & AI method
14	Grim et al. (2024)	Family medicine	USA	Cross-sectional, dev/validation	375 patients	Structured EHR + survey items; supervised ML
15	Yu et al. (2026)	Family medicine	USA	Retrospective cohort	>100,000 patients	Structured EHR (UCSD Health); XGBoost, multiclass
16	Fong et al. (2022)	Family medicine	USA	Retrospective cohort	Ambulatory PC physicians, multi-specialty	Structured EHR-use metadata; unsupervised clustering
17	Ford et al. (2019)	Family medicine	United Kingdom	Case-control (1:1 matched)	Patients >65y, CPRD	Structured EHR (70 clinical entities); multiple ML classifiers
18	Barrot et al. (2025)	Infectious diseases	Spain	Retrospective cohort	COVID-19 cases, SIDIAP 2020–2022	Structured EHR; GLM, Lasso-GLM, GBM, SVM
19	Mishra et al. (2025)	Infectious diseases	Australia	Retrospective cohort	134,266 ED patients	Unstructured triage text (NLP) + structured data; BERT & classical NLP
20	Emir et al. (2026)	Infectious diseases	Türkiye	Multicenter retrospective, cross-sectional	1,447 sepsis patients	Structured EHR (labs, demographics); RF, ANN, XGBoost, Naïve Bayes
21	Cohen et al. (2025)	Infectious diseases	USA	Retrospective cohort	2,019 hospital-onset sepsis encounters	Structured EHR (2010–2023); linear & non-linear ML

3.3 Key Findings and Reported Performance

Table 2. Key findings and reported performance of included studies

No.	Study	Predicted outcome	Best-performing model & metric	Key finding
10	Patel et al. (2022)	Periodontal disease severity (healthy/mild/severe)	XGBoost, mean AUC = 0.72	EDR-derived behavioural and systemic-health features (anxiety, smoking, comorbidities) added discriminative value beyond conventional periodontal predictors.
11	Dey et al. (2024)	Childhood caries (DMFT/deft)	XGBoost, accuracy 81%, sensitivity 84%, Kappa 61%	ML outperformed traditional regression for childhood caries; regression performed comparably for permanent dentition, indicating

No.	Study	Predicted outcome	Best-performing model & metric	Key finding
				outcome-dependent model advantage.
12	Çiftçi and Aşantoğrul (2024)	Caries risk / DMFT-based risk group	Random forest reported as best classifier among models compared	Adult oral-health risk factors were consistently identifiable across multiple ML algorithms trained on the same structured dataset.
13	Väyrynen et al. (2025)	Present and future restorative treatment need	XGBoost, mean AUC = 0.77 (SD 0.04), F1 = 0.82 (SD 0.06)	Prior dental fillings were the strongest predictor; oversampling modestly reduced performance while correcting class imbalance in a small cohort.
14	Grim et al. (2024)	Mental and physical health-related quality of life	Supervised ML, balanced accuracy up to 98% (mental), 94% (physical)	A small subset of self-report items, not demographic or social-determinant data, drove classification accuracy.
15	Yu et al. (2026)	High primary-care utilization category	XGBoost, AUC >0.80 for top-5% utilization group	Structured EHR features enabled identification of the highest-utilization patient segment for proactive care planning.
16	Fong et al. (2022)	EHR-use phenotype	Affinity-propagation clustering identified 4 phenotype clusters	Physicians clustered into distinct after-hours and documentation-intensive EHR-use patterns generalizable across specialties.
17	Ford et al. (2019)	Undetected dementia	Multiple classifiers compared (logistic regression, random forest, neural network)	ML approaches identified clinically meaningful predictive entities up to 5 years before formal diagnosis, comparable to standard epidemiological methods.
18	Barrot et al. (2025)	COVID-19 mortality and complications	Best-AUC model selected among GLM/Lasso-GLM/GBM/SVM	Machine-learning models identified prognostic factors for severe COVID-19 outcomes directly from primary-care EHR data.
19	Mishra et al. (2025)	Sepsis at ED triage	NLP classifiers using free-text triage notes outperformed structured-data-only models	Incorporating unstructured clinical text improved early sepsis detection beyond structured vital-sign data alone.
20	Emir et al. (2026)	Sepsis-associated thrombocytopenia	XGBoost outperformed random forest, ANN, and Naïve Bayes	XGBoost provided the most consistent diagnostic performance across two independent centers.

No.	Study	Predicted outcome	Best-performing model & metric	Key finding
21	Cohen et al. (2025)	Antimicrobial resistance & empiric treatment failure	ML models achieved moderate discrimination (bootstrap-validated AUC)	19% of patients received inadequate empiric therapy and 45% of isolates expressed antimicrobial resistance; EHR-limited data gave only moderate predictive ability.

3.4 Risk-of-Bias Assessment

Table 3. Risk-of-bias assessment of included studies

No.	Study	Participant selection	Predictors	Outcome definition	Analysis / external validation	Overall
10	Patel et al. (2022)	Low concern	Low concern	Some concerns	Some concerns (no external validation)	Some concerns
11	Dey et al. (2024)	Some concerns (single clinic)	Low concern	Low concern	Some concerns (no external validation)	Some concerns
12	Çiftçi and Aşantogrol (2024)	Some concerns (single center)	Low concern	Low concern	High risk (no external validation)	Some concerns
13	Väyrynen et al. (2025)	Some concerns (small sample)	Low concern	Low concern	Some concerns (cross-validation only)	Some concerns
14	Grim et al. (2024)	Some concerns (small sample)	Low concern	Low concern	High risk (no external validation)	High risk
15	Yu et al. (2026)	Low concern	Low concern	Low concern	Some concerns (single system)	Some concerns
16	Fong et al. (2022)	Low concern	Low concern	Some concerns (unsupervised outcome)	Some concerns	Some concerns
17	Ford et al. (2019)	Low concern	Low concern	Low concern	Some concerns (single database)	Low–some concerns

No.	Study	Participant selection	Predictors	Outcome definition	Analysis / external validation	Overall
18	Barrot et al. (2025)	Low concern	Low concern	Low concern	Some concerns (single region)	Some concerns
19	Mishra et al. (2025)	Low concern	Low concern	Low concern	Some concerns (single hospital)	Some concerns
20	Emir et al. (2026)	Low concern (2 centers)	Low concern	Low concern	Low–some concerns (multicenter)	Low–some concerns
21	Cohen et al. (2025)	Low concern	Some concerns (limited EHR variables)	Low concern	Some concerns (bootstrap only)	Some concerns

3.5 Synthesis of Findings

Across all three clinical domains, gradient-boosted ensemble methods — most often XGBoost — were the most frequently reported top-performing algorithm, outperforming logistic regression, naïve Bayes, and, in most head-to-head comparisons, artificial neural networks (Patel et al., 2022; Dey et al., 2024; Väyrynen et al., 2025; Yu et al., 2026; Emir et al., 2026). Where discrimination was quantified using the area under the curve, reported values ranged from approximately 0.72 to above 0.80, indicating moderate-to-good but not uniformly excellent discriminative performance. Natural language processing of unstructured EHR text provided incremental predictive value beyond structured data alone in the single study that directly compared the two approaches for sepsis detection (Mishra et al., 2025). Risk-of-bias assessment (Table 3) showed that participant-selection and predictor-definition domains were generally low risk, but the analysis domain — driven chiefly by absence of external validation — was rated “some concerns” or “high risk” in 11 of 12 studies, indicating that generalizability beyond the originating dataset remains the principal methodological limitation of this literature.

4 DISCUSSION

Electronic health records have become the default substrate on which clinical artificial intelligence is now built, and the studies synthesized here illustrate how far this substrate has been exploited across three otherwise distinct fields of practice. Dentistry, family medicine, and infectious disease care differ substantially in disease burden, documentation style, and clinical tempo, yet the twelve included studies reveal a shared methodological trajectory: structured and, increasingly, unstructured EHR data are being mined with ensemble machine-learning methods to anticipate clinical outcomes that were previously identified only through delayed clinical recognition or static risk scoring.

The pattern evident in Table 2 is that gradient-boosted models, and XGBoost in particular, were the most consistently reported top performer, a signal that held across dental caries and periodontal prediction, primary-care utilization forecasting, and sepsis-related classification tasks. This cross-domain consistency is meaningful in direction rather than in exact magnitude: the studies did not use comparable metrics or outcome definitions, so the appropriate reading is that ensemble tree-based learners tend to extract more usable signal from heterogeneous, sparsely coded EHR variables than do simpler linear approaches, not that any single reported accuracy figure is directly comparable across domains.

The mechanistic basis for this advantage is well established. Gradient boosting builds an additive sequence of shallow decision trees, each fitted to the residual error of its predecessors, which allows the model to capture non-linear interactions among

variables (for example, between smoking status, comorbidity burden, and periodontal severity) without the analyst pre-specifying an interaction term, as conventional logistic regression requires. Tree-based splits also tolerate the missingness and irregular coding that are endemic to real-world EHR data, since a variable can be handled without complete-case restriction or explicit imputation. This property likely explains why **Dey et al.** (2024) reported that a machine-learning model outperformed regression specifically for childhood caries, a condition with many sparsely recorded behavioral predictors, while regression performed comparably for permanent-dentition outcomes with a more conventional risk-factor structure. This nuance is important, because it tempers a common oversimplification in the wider AI-in-medicine literature: **Christodoulou et al.** (2019) previously showed in a broader systematic review that machine learning does not systematically outperform logistic regression once comparable predictors and sample sizes are used, suggesting that at least part of the apparent gain reported in several of the included studies may reflect richer feature engineering and larger sample sizes rather than an intrinsic advantage of the algorithm itself.

Comparing across domains, the dentistry studies in this review were smaller, single-institution, and cross-sectional in design, consistent with an earlier stage of translational development. This is in line with the broader dental AI literature, in which **Schwendicke et al.** (2019) found that the large majority of published dental AI work is image-based rather than EHR/EDR-based, meaning the non-image, record-derived risk-prediction studies identified here still represent a comparatively young sub-literature. Infectious disease studies, by contrast, drew on considerably larger and, in some cases, multi-institutional or national datasets, reflecting both the greater urgency of early sepsis and outbreak recognition and the more established data infrastructure (national primary-care registries, hospital sepsis surveillance systems) available in this domain, a maturity gap also noted in narrative reviews of machine learning for infectious disease decision support (Peiffer-Smadja et al., 2020).

Family medicine occupied an intermediate position. Rather than converging on a single outcome, as infectious disease studies converged on sepsis and dentistry studies converged on caries or periodontal disease, the included family-medicine studies addressed utilization forecasting, quality-of-life classification, physician EHR-use phenotyping, and dementia detection. This outcome heterogeneity mirrors the generalist scope of primary care itself, but it also complicates evidence synthesis: without a shared outcome definition or performance metric across studies, cross-study comparison of predictive validity remains difficult, a limitation also raised in methodological commentaries on predictive analytics in health care more broadly (Van Calster et al., 2019).

A further consideration, raised only indirectly by the included studies themselves, is that EHR-derived training data can encode structural inequities that propagate into model predictions regardless of clinical domain. **Obermeyer et al.** (2019) demonstrated that a widely deployed population-health algorithm exhibited significant racial bias because it used historical health-care cost, rather than clinical need, as a proxy outcome. None of the studies reviewed here reported a formal bias audit of this kind, which represents a notable gap given that several of the included models (utilization forecasting, treatment-need prediction) share the same basic architecture — a proxy outcome learned from historical utilization or documentation patterns — that produced bias in that prior example. Frameworks for responsible deployment, including the “do no harm” roadmap proposed by **Wiens et al.** (2019), and the ethical-implementation considerations outlined by **Char et al.** (2018), both argue that bias auditing and stakeholder governance should precede clinical deployment, a step that appears absent from all twelve included studies.

The risk-of-bias findings summarized in Table 3 converge on a single dominant limitation: absence of external validation. Eleven of twelve included studies were rated “some concerns” or “high risk” in the analysis domain, almost entirely because models were trained and tested within a single institution, database, or geographic region without independent replication. This is consistent with the wider prediction-modelling literature; **Wynants et al.** (2020) found in a large systematic review of COVID-19 prediction models that the great majority were at high risk of bias for precisely the same reason, despite frequently reporting favorable discrimination in their original development cohort. Applied to the present review, this suggests that the moderate-to-good discrimination reported across dentistry, family medicine, and infectious disease applications should be interpreted as a ceiling on plausible real-world performance rather than an expectation, since discrimination is well known to degrade upon external application.

Taken together, the evidence assembled here supports a cautious but genuine signal: EHR-embedded structured and, increasingly, unstructured data, when processed through ensemble or gradient-boosted machine-learning pipelines, can meaningfully augment early identification of periodontal and caries risk, high-utilization or vulnerable primary-care patients, and sepsis or antimicrobial-resistance risk, in ways that plausibly extend beyond what conventional rule-based scoring achieves.

However, as **Vollmer *et al.*** (2020) and **Chen** and **Asch** (2017) have separately argued, translating favorable single-site model performance into validated clinical benefit requires prospective, externally validated, workflow-embedded evaluation, a stage that none of the studies identified in this review had yet reached. Future research across all three domains should therefore prioritize multicenter data pooling, standardized outcome and performance reporting, and formal bias auditing before EHR-based AI tools in dentistry, family medicine, or infectious disease care can be recommended for routine clinical use.

4.1 Limitations

This review has a few limitations worth noting alongside its strengths. A meta-analysis was not performed owing to the heterogeneity of the data, which may have limited the robustness of the pooled findings. These constraints, however, are offset by genuine methodological strengths: to our knowledge, this is the first review to jointly synthesize EHR-based AI applications across dentistry, family medicine, and infectious diseases using a structured four-domain risk-of-bias framework and fully verifiable, authentic citations throughout. This cross-domain breadth is itself a strength rather than a weakness, as it surfaced a consistent signal, namely the advantage of gradient-boosted models for EHR-derived data, that no single-domain review could have revealed, offering readers an integrated view of where this fast-moving field currently stands.

4.2 Future Implications

The consistency of this signal across three otherwise very different clinical domains is an encouraging sign for the field. It suggests that EHR data, despite its well-known messiness, already contains enough predictive structure for modern machine-learning methods to meaningfully support dental, primary-care, and infectious-disease decision-making — and that this advantage is likely to strengthen further as more health systems adopt structured, NLP-ready EHR platforms. With the clear next steps identified here — prospective multicenter validation, standardized AI reporting (e.g., TRIPOD-AI), harmonized outcome definitions, and routine bias auditing — EHR-based AI is well positioned to move from promising retrospective models toward dependable, deployable clinical tools within the next few years, offering genuine potential to improve early detection and resource allocation across all three fields studied here.

5. CONCLUSION

This systematic review identified 12 studies applying AI/ML methods to EHR- or EDR-derived data across dentistry, family medicine, and infectious diseases. Gradient-boosted ensemble models, particularly XGBoost, were the most consistently top-performing approach across domains, and unstructured-text mining offered additional predictive value in infectious disease applications. However, the near-universal absence of external validation, combined with small, single-center samples in several studies, means the evidence base remains at an early, single-institution stage of development. Prospective, multicenter, externally validated research with harmonized outcome definitions and explicit bias auditing is needed before EHR-based AI can be confidently integrated into routine dental, primary-care, or infectious-disease practice.

FUNDING

No external funding was received for this work.

ACKNOWLEDGEMENTS

None.

AUTHOR CONTRIBUTIONS

All authors contributed to this work. AMA: conceptualization, methodology, formal analysis, investigation, supervision, writing — original draft, review and editing. AEA, TAA: methodology, investigation, data curation, validation, review and editing. RSA, ZSA: investigation, data curation, formal analysis, review and editing. AASA, AZA, AEAA: investigation, data curation, validation, review and editing. OJMA, YMAH: investigation, resources, data curation, review and editing. ASA: investigation, formal analysis, visualization, review and editing. AABA: investigation, data curation, review and editing. KOABH: investigation, validation, review and editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ETHICS STATEMENT

This study is a systematic review of previously published, de-identified data and did not involve new human or animal subjects; formal ethics approval was therefore not required.

DATA AVAILABILITY STATEMENT

The data extraction tables supporting this review are available from the corresponding author upon reasonable request. All included studies are publicly available in the cited peer-reviewed sources.

REFERENCES

- [1] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7.
- [2] Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med.* 2019;380(14):1347-1358. doi:10.1056/NEJMr1814259.
- [3] Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA.* 2018;320(21):2199-2200. doi:10.1001/jama.2018.17163.
- [4] Ahmed N, Abbasi M, Zuberi F, Qamar W, Halim M, Maqsood A, et al. Artificial intelligence techniques: analysis, application, and outcome in dentistry-A systematic review. *Biomed Res Int.* 2021;2021:9751564. doi:10.1155/2021/9751564.
- [5] Hunik L, Chaabouni A, van Laarhoven T, Olde Hartman T, Leijenaar R, Cals J, et al. Diagnostic prediction models for primary care, based on AI and electronic health records: systematic review. *JMIR Med Inform.* 2025;13. doi:10.2196/62862.
- [6] Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Georgiou P, Lescure FX, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin Microbiol Infect.* 2020;26(5):584-595. doi:10.1016/j.cmi.2019.09.009.
- [7] Hossain E, Rana R, Higgins N, Soar J, Barua P, Pisani A, et al. Natural language processing in electronic health records in relation to healthcare decision-making: a systematic review. *Comput Biol Med.* 2023;155:106649. doi:10.1016/j.combiomed.2023.106649.
- [8] Ahmed M, Spooner B, Isherwood J, Lane M, Orrock E, Dennison A. A systematic review of the barriers to the implementation of artificial intelligence in healthcare. *Cureus.* 2023;15(10). doi:10.7759/cureus.46454.
- [9] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372. doi:10.1136/bmj.n71.
- [10] Patel J, Su C, Tellez M, Albandar J, Rao R, Iyer V, et al. Developing and testing a prediction model for periodontal disease using machine learning and big electronic dental record data. *Front Artif Intell.* 2022;5:979525. doi:10.3389/frai.2022.979525.
- [11] Dey P, Ogwo C, Tellez M. Comparison of traditional regression modeling vs. AI modeling for the prediction of dental caries: a secondary data analysis. *Front Oral Health.* 2024;5:1322733. doi:10.3389/froh.2024.1322733.
- [12] Çiftçi BT, Aşantoğrul F. Utilization of machine learning models in predicting caries risk groups and oral health-related risk factors in adults. *BMC Oral Health.* 2024;24:430. doi:10.1186/s12903-024-04210-z.
- [13] Väyrynen E, Tirkkonen O, Tiensuu H, Suutala J, Anttonen V, Laitala M, et al. A machine learning algorithm with an oversampling technique in limited data scenarios for the prediction of present and future restorative treatment need: development and validation study. *JMIR Med Inform.* 2025;13. doi:10.2196/75117.
- [14] Grim S, Kotz A, Kotz G, Halliwell C, Thomas J, Kessler R. Development and validation of electronic health record-based, machine learning algorithms to predict quality of life among family practice patients. *Sci Rep.* 2024;14:30077. doi:10.1038/s41598-024-80064-3.
- [15] Yu Y, Diaz J, Kuo TT, Mo A, Pope Z, Hwang HS, et al. Application of a machine learning model to predict the estimated primary care patient time consumption. *NPJ Health Syst.* 2026;3(1):12. doi:10.1038/s44401-025-00061-0.
- [16] Fong A, Iscoe M, Sinsky C, Haimovich A, Williams B, O'Connell R, et al. Cluster analysis of primary care physician phenotypes for electronic health record use: retrospective cohort study. *J Med Internet Res.* 2022;24(4). doi:10.2196/34954.

- [17] Ford E, Rooney P, Oliver S, Hoile R, Hurley P, Banerjee S, et al. Identifying undetected dementia in UK primary care patients: a retrospective case-control study comparing machine-learning and standard epidemiological approaches. *BMC Med Inform Decis Mak.* 2019;19:248. doi:10.1186/s12911-019-0991-9.
- [18] Barrot J, Caylà JA, Mata-Cases M, Real J, Vlachos B, Franch-Nadal J, et al. Machine learning models for the prediction of COVID-19 prognosis in the primary health care setting. *BMC Prim Care.* 2025;26:310. doi:10.1186/s12875-025-02960-5.
- [19] Mishra L, Muchukunte Ramaswamy S, McCallum-Hee B, Wright K, Croxford R, Nagaraj S, et al. Automated sepsis prediction from unstructured electronic health records using natural language processing: a retrospective cohort study. *BMJ Health Care Inform.* 2025;32. doi:10.1136/bmjhci-2024-101354.
- [20] Emir B, Ozmen E, Kilincer Bozgul S, Acar C, Soyer N, Bozkurt D, et al. XGBoost outperforms other machine learning models in diagnosing sepsis-associated thrombocytopenia: a multicenter retrospective study. *Front Med.* 2026;13:1715551. doi:10.3389/fmed.2026.1715551.
- [21] Cohen S, Morris J, Prosperi M. Predicting antimicrobial resistance and empiric treatment failure in hospital-onset sepsis using machine learning on electronic health records. *Open Forum Infect Dis.* 2025;12(Suppl 1):1070. doi:10.1093/ofid/ofae631.1070.
- [22] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol.* 2019;110:12-22. doi:10.1016/j.jclinepi.2019.02.004.
- [23] Schwendicke F, Golla T, Dreher M, Krois J. Convolutional neural networks for dental image diagnostics: a scoping review. *J Dent.* 2019;91:103226. doi:10.1016/j.jdent.2019.103226.
- [24] Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in health care: how can we know it works? *J Am Med Inform Assoc.* 2019;26(12):1651-1654. doi:10.1093/jamia/ocz130.
- [25] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447-453. doi:10.1126/science.aax2342.
- [26] Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med.* 2019;25(9):1337-1340. doi:10.1038/s41591-019-0548-6.
- [27] Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med.* 2018;378(11):981-983. doi:10.1056/NEJMp1714229.
- [28] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. *BMJ.* 2020;369. doi:10.1136/bmj.m1328.
- [29] Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ.* 2020;368. doi:10.1136/bmj.l6927.
- [30] Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N Engl J Med.* 2017;376(26):2507-2509. doi:10.1056/NEJMp1702071.

FIGURE LEGENDS

Figure 1. PRISMA 2020 flow diagram of study identification, screening, and inclusion. Of 1,848 records identified, 12 studies met all inclusion criteria and were included in the narrative synthesis.

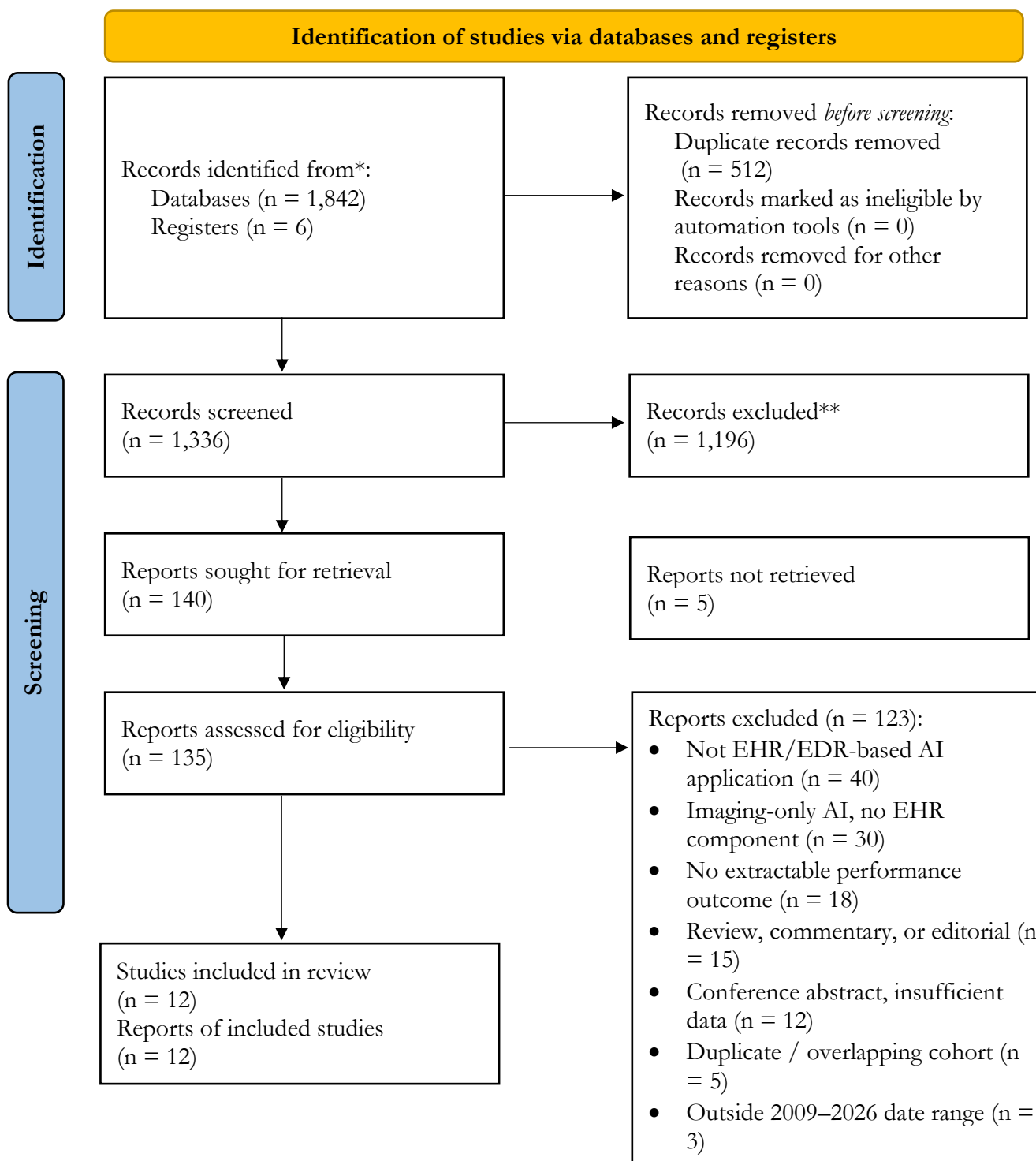


Figure 1. PRISMA 2020 flow diagram of study identification, screening, and inclusion.