

PERSONALIZED MACHINE LEARNING-BASED RISK PREDICTION OF DENTAL DISEASES IN OLDER PATIENTS

Mavlyanova Nargiza Talatovna

Assistant, Department of Advanced Training in Therapeutic Dentistry,
with a course in Orthopedic Dentistry1.

dr.nargiza7909@gmail.com ORCID: <https://orcid.org/0000-0002-7145-2499>

Tashkent State Medical University1. Tashkent. Uzbekistan

<https://doi.org/10.5281/zenodo.23254706>

ABSTRACT

The review systematises machine learning models developed for the prediction of dental outcomes in older patients and appraises their methodological quality. Reported performance is shown to depend on study design more than on the choice of algorithm. Values range from an AUC of 0.71 to 0.99, and the highest of them belong to cross-sectional studies in which a prognostic task has been replaced by a diagnostic one. Moving to an independent sample lowers performance by 0.13–0.17 AUC units, whereas logistic regression that properly accounts for the structure of the data performs no worse than neural architectures. The specific features of the older patient are only partly represented in the models. Polypharmacy, hyposalivation, oral frailty and social isolation enter them fragmentarily, and the competing risk of death is almost never modelled. Conditions for clinical applicability are formulated, namely a longitudinal design with a predefined prediction horizon, external validation, assessment of calibration and reporting in line with TRIPOD+AI.

Keywords: machine learning, risk prediction, geriatric dentistry, root caries, periodontitis, tooth loss, personalized prevention.

Introduction. Untreated caries of permanent teeth has remained the most prevalent human disease for decades. In 2010, it affected 2.4 billion people, and global preventive efforts have yet to reverse this trend [27]. Seven years later, the situation was already being described as a crisis, prompting the World Health Organization to call for a reconsideration of the architecture of preventive programmes and to acknowledge their insufficient targeting [8].

For geriatric dentistry, this picture has a distinct interpretation. The advances in prevention during the second half of the twentieth century have meant that older patients now present with a retained, albeit functionally compromised, dentition rather than complete edentulism, as was common among their peers three decades ago. A 30-year analysis has shown a decline in age-standardized prevalence of complete edentulism alongside an increase in the absolute number of cases [33]. There are therefore more teeth to preserve, but less physiological reserve to support their long-term retention. The prevalence and severity of root caries differ markedly across generations, with root surfaces becoming a major target for demineralization in older adults [20].

The older population is highly heterogeneous, yet population-level prevention does not account for this heterogeneity. A 70-year-old patient with preserved salivary flow and regular oral hygiene and a peer taking seven medications with salivary flow below a critical threshold may receive exactly the same preventive intervention under a standard programme.

Personalized risk prediction is an attempt to allocate limited healthcare resources preferentially to those who are most likely to lose teeth in the coming years.

Regional studies confirm the urgency of the problem while also defining an important evidence gap. Older age groups have been reported to have a high prevalence of major dental diseases, with a DMFT index of 23.95 and a predominance of missing teeth [3]. Asadullaev and Rizaev (2017) described the clinical and immunological features of periodontal disease in older and very old adults, while the age distribution of pulpitis forms was investigated in a sample of 324 patients aged 45–65 years [1, 4]. These studies characterize oral health status but do not generate individualized prognostic models, and no prediction models trained on regional cohorts were identified in the available literature.

Machine learning methods appear to be a natural tool for this task. They do not require a priori assignment of predictor weights and can accommodate nonlinear interactions among dozens of predictors. One conceptual problem common in this literature, however, needs to be addressed from the outset. Detecting a carious lesion on a radiograph and predicting whether that lesion will develop three years later are fundamentally different tasks: diagnostic and prognostic, respectively. Of the 15 studies applying machine learning to caries research, only five addressed a genuinely prognostic question [36]. Conflating these tasks systematically overstates the maturity of the field.

The older patient as a prognostic target. The quality of a prediction model depends on how well its set of predictors matches the population in which the model is intended to be used. This condition is violated when models developed in general adult populations are transferred to older cohorts. It is not simply the strength of established risk factors that changes the set of relevant predictors itself changes.

The age-specific nature of this population is captured particularly well by the concept of oral frailty, defined as an age-related gradual decline in oral functions accompanied by deterioration in cognitive and physical functions. Measures of impaired dental status were most frequently associated with general frailty (52%), particularly a low number of remaining teeth (29%) [15]. Its prognostic relevance extends beyond the oral cavity. Oral frailty is an independent risk factor for physical frailty, sarcopenia, disability, and mortality, with risk increasing by 2.2–2.4-fold [41]. For model developers, this implies more than the inclusion of a convenient composite predictor. Death becomes a competing event, and ignoring it systematically inflates the estimated probability of tooth loss.

The second specific domain is medication burden. Medication-induced xerostomia and hyposalivation in older adults are driven primarily by the anticholinergic and sympathomimetic effects of commonly prescribed drugs [7]. The association between polypharmacy and dry mouth has been documented in a systematic review, while its range of oral consequences has been discussed in the geriatric literature [9, 39]. The modelling implication is straightforward. A binary predictor indicating whether medications are taken is of limited value because the effect is dose-dependent. The predictors should instead include the number of medications and their pharmacological classes.

The third domain is anatomical and topographical. Among independently living older adults with root caries, the strongest associations were observed for xerostomia (OR 18.49), inadequate plaque control (OR 9.59), 37 or more exposed root surfaces (OR 5.48), and the presence of two or more teeth with coronal caries (OR 4.50) [22].

The fourth domain is cognitive. In a meta-analysis of 21 observational studies, multiple tooth loss was associated with a higher frequency of dementia (OR 2.62; 95% CI 1.90–3.61) [17]. Cognitive impairment undermines independent oral hygiene and regular dental attendance and therefore modifies the effect of behavioral predictors.

The fifth domain is the social context. In the Japanese JAGES cohort, income, employment, and participation in community life contributed to the prediction of tooth loss to an extent comparable with clinical features a similar result was obtained from NHANES data [14, 16]. The picture is complemented by predictors included in a validated nomogram use of a removable partial denture, secondary caries, and the number of exposed roots as well as clustering results in which the high-risk group consisted of older adults with periodontal pockets, attachment loss, xerostomia, and systemic disease [24, 26]. The importance of multimorbidity is also supported in implant dentistry, where diabetes mellitus was among the leading predictors of peri-implantitis [35].

Table 1. Geriatric risk modifiers and their relevance to prediction models

Modifier	Relevance to prediction model	Sources
Polypharmacy and medication-induced hyposalivation	The model should include the number of medications and their pharmacological classes, together with objective sialometry. Xerostomia is the strongest indicator of root-caries risk (OR 18.49).	[7, 9, 22, 39]
Gingival recession and exposed root surfaces	The number of exposed root surfaces is among the strongest predictors (37 or more surfaces – OR 5.48).	[22, 24]
Oral frailty	Risk of disability and mortality is 2.2–2.4 times higher, making death a competing event.	[15, 41]
Cognitive impairment	Modifies the effect of behavioural predictors (OR for dementia associated with tooth loss 2.62); requires consideration of caregiver status.	[17]
Socioeconomic determinants	Improve tooth-loss models more strongly than the addition of clinical features.	[14, 16]
Multimorbidity	Important for prediction of periodontitis and complications of dental implantation.	[26, 35]

From deterministic risk scores to machine learning. The idea of individualized risk assessment in dentistry is not new. Cariogram, CAMBRA, and the Periodontal Risk Assessment (PRA) system have been used for decades, and their measurement properties are better established than those of most contemporary neural-network models. The evidence base for the predictive validity of these systems remains limited, and direct comparisons are scarce [42]. Quantitative clarity was provided by the meta-analysis of Su et al., which combined 22 studies [40]. The full Cariogram demonstrated an AUC of 0.78 (95% CI 0.68–0.85), whereas the shortened version yielded 0.72 (0.67–0.77), with observed-to-expected event ratios of 0.91 and 0.84, respectively, indicating acceptable, although not perfect, calibration.

The performance of the score in an older cohort has been studied separately. When Cariogram was evaluated in 334 dentate older adults, the AUC was 0.77 (0.70–0.83), with clear discriminatory ability. In the highest-risk group, new lesions developed in 55.6% of

participants compared with 3.8% in the lowest-risk group, while the mean increase in root caries was 2.00 versus 0.04 lesions [23]. In periodontology, PRA categories predicted recurrence of periodontitis at rates of 18.2%, 42.2%, and 49.2% for low-, moderate-, and high-risk groups, respectively, and also predicted tooth loss during supportive periodontal therapy [32].

The limitation of deterministic scores lies elsewhere. Predictor weights are expert-defined and fixed, and the output is expressed as a category rather than a continuous probability. A high-risk category does not answer the clinically relevant question of what the risk is for this particular patient or over what time horizon. The predictor set itself was established before geriatric specificity became a major focus of research. At the same time, the superiority of machine learning cannot be taken for granted. In the systematic review by Christodoulou et al., which included 71 studies and 282 paired comparisons, the difference in logit-transformed AUC under low risk of bias was 0.00 (95% CI -0.18 to 0.18), whereas an apparent advantage emerged only in studies at high risk of bias (0.34; 0.20–0.47). Notably, 68% of studies showed signs of validation bias, and 79% did not discuss calibration [10].

Algorithmic foundations of prediction. Logistic regression remains an essential benchmark, and in a number of studies it performs at least as well as more complex algorithms. The reason lies in the structure of the data. Repeated measurements from the same individuals require within-person correlation to be accounted for mixed-effects models do this by construction, whereas conventional flat machine-learning algorithms do not [43]. Requirements for minimum sample size when developing multivariable prediction models have been formalized and are far from consistently met in geriatric studies [37].

Ensembles of decision trees, particularly random forests and gradient boosting, generally provide the strongest performance on tabular data. They are robust to collinearity and outliers and offer an intrinsic measure of feature importance, but they are prone to optimistic internal performance estimates [18, 46]. Artificial neural networks consistently fail to meet expectations on small tabular datasets and may shift the operating point toward overdiagnosis [38]. Support vector machines and decision trees are used mainly in cross-sectional tasks, where their reported metrics do not have genuine prognostic meaning [25].

Survival-analysis models are most appropriate to the nature of dental prognosis because clinicians are interested not merely in whether an event occurs, but in when it occurs. These models accommodate censoring and estimate time to event, which is particularly important in geriatric cohorts [5, 47]. Unsupervised clustering occupies a separate niche by describing risk phenotypes in heterogeneous populations, although it does not provide an individual probability [26]. The general architecture of systems applied in dentistry is described in a national review [2].

Table 2. Algorithmic families and their applicability in geriatric dentistry

Algorithm	Advantages	Limitations in geriatric cohorts	Sources
Logistic regression, including mixed-effects models	Transparency, robustness in small samples, and readily interpretable calibration	Does not capture complex interactions without explicit specification	[24, 43]

Algorithm	Advantages	Limitations in geriatric cohorts	Sources
Random forest	Robust to outliers and collinearity; provides feature-importance estimates	Optimistic internal estimates and a marked decline in performance on external validation	[18, 35]
Gradient boosting (XGBoost, CatBoost)	Strong performance on large tabular datasets	Requires careful tuning and is sensitive to class imbalance	[14, 16, 46]
Artificial neural networks	Flexible for complex relationships and multimodal data	Overfitting on small tabular datasets; may shift the operating point	[38, 43]
Survival models (DeepSurv, survival random forest, Cox nomograms)	Provide time-to-event estimates and accommodate censoring	Require long follow-up and rarely account for the competing risk of death	[5, 47]
Unsupervised clustering	Describes risk phenotypes in heterogeneous populations	Does not provide individual probabilities; results depend on the distance metric	[26, 47]

Risk prediction across major disease groups. Root caries is particularly characteristic of older adults, and predictive models have been developed for it more actively than for other dental outcomes. The reported results, however, vary widely, and the divergence is methodological rather than biological. The highest performance, an AUC of 0.997 with 99.6% sensitivity, was reported by Hung et al., whose model was trained on cross-sectional NHANES data and essentially identified existing root caries, age was the strongest predictor [25]. A nomogram developed from 310 patients older than 60 years yielded an AUC of 0.840 in the training cohort and 0.834 in the validation cohort, with calibration assessment and decision-curve analysis. Predictors included age, history of coronal caries, secondary caries, use of a removable partial denture, and the number of exposed root surfaces [24]. Similar performance was reported in 423 patients aged 65–74 years, with accuracy of 0.81, sensitivity of 0.79, specificity of 0.83, and an AUC of 0.84 [26]. A multilayer perceptron trained on 2,000 adults achieved 95.8% accuracy and an F1 score of 0.96, but the sample was not restricted to older adults, limiting the transportability of the result [12]. Compared with Cariogram, which yielded an AUC of 0.77 in older adults, the realistic gain attributable to machine learning is therefore on the order of 0.06–0.07 AUC units rather than several tenths [23].

For periodontitis and treatment response, the comparison reported by Teza et al. is particularly instructive: classical mixed-effects regression achieved an AUC of 0.983 (95% CI 0.977–0.989), substantially outperforming neural-network architectures on the same data [43]. In the population-based NHANES setting, more moderate but credible performance was obtained, with an AUC of 84.5%, sensitivity of 78.8%, and specificity of 72.5%. The leading predictors were regularity of dental visits and age [46]. The study by Feher et al. is especially valuable because it included external testing. A random forest trained on 414 patients from a

South American cohort achieved an AUROC of 0.93 internally but only 0.76 on an independent sample of 78 patients from North America and Europe. The 0.17-point decline illustrates the cost of optimism in internal validation [18].

Prediction of tooth loss represents the most mature area. In NHANES 2011–2014 data, gradient boosting achieved an AUC of 88.7% for complete edentulism, 88.3% for absence of a functional dentition, and 83.2% for loss of any tooth [16]. A prospective study of 19,407 Japanese adults aged 65 years and older followed for six years reported a mean XGBoost accuracy of $90.5 \pm 0.9\%$ [14]. International validation was provided by a multicentre study of 3,157 molars in 515 patients, in which an ensemble model achieved an AUC of 0.726 and a neural network 0.724 [44].

The importance of the operating point was clearly demonstrated by Santamaria et al. [38], who compared a clinical prognostic system, logistic regression, and a neural network in a London cohort comprising 1,626 teeth from 69 patients, with external validation on 2,792 teeth from 116 patients in Pittsburgh. The clinical score showed 99.96% specificity with 0.85% sensitivity, whereas the neural network produced the opposite pattern, with 98.3% sensitivity and 52.3% specificity. Logistic regression occupied an intermediate position, with 91.5% sensitivity and 67.6% specificity. A RuleFit model applied to 7,840 patients with periodontitis produced more moderate AUROC and AUPRC values of 0.71 and 0.66, respectively, and identified age, diagnosis, baseline number of missing teeth, furcation involvement, and mobility as the leading predictors [29]. A systematic review with formal risk-of-bias assessment showed that models combining patient-level and socioeconomic predictors outperform models based solely on clinical features [21].

In dental implantology, a random forest applied to 942 implants in 398 patients achieved an AUC of 0.872 for implant failure and 0.840 for peri-implantitis [35]. The leading predictors included implant length and diameter, preoperative antibiotic prophylaxis, frequency of professional oral hygiene, and diabetes mellitus. In a cohort of 8,513 patients, two-step clustering followed by nomogram development yielded c-indices of 0.642, 0.781, and 0.715 for early, intermediate, and late periods of implant loss, respectively [47].

Cancer-related surveillance in older adults requires a separate prognostic framework. In 1,098 patients from two institutions, the DeepSurv architecture outperformed a survival random forest and a Cox proportional-hazards model, with c-indices of 0.95 internally and 0.82 on external validation [5]. A self-attention neural network trained on 1,094 patients from three institutions achieved an AUC of 0.988 compared with 0.880–0.902 for the Cox nomogram [30]. A systematic review of 15 studies reported overall accuracy ranging from 74% to 95.4%, with substantial methodological heterogeneity [45].



Table 3. Summary characteristics of models predicting dental outcomes

Predicted event	Population and sample size	Model performance	Validation	Source
Root caries	NHANES, adults	Support vector machine – AUC 0.997 in a cross-sectional design	Internal	[25]
Root caries	310 adults older than 60 years	Nomogram – AUC 0.840 and 0.834 with calibration assessment	Internal	[24]
Severe periodontitis	2,086 participants, 3,883 observations	Mixed-effects regression AUC 0.983 versus 0.712 for neural network	Internal	[43]
Response to periodontal treatment	414 and 78 participants	Random forest – AUROC 0.93 and 0.76	External	[18]
Tooth loss over 6 years	19,407 adults aged 65 years and older	XGBoost – mean accuracy $90.5 \pm 0.9\%$	Internal, prospective cohort	[14]
Molar loss over 10 years	515 participants, 3,157 molars, 4 centres	Ensemble AUC 0.726; neural network 0.724	External	[44]
Malignant transformation of potentially malignant lesions	1,098 participants, 2 institutions	DeepSurv – c-index 0.95 and 0.82	External	[5]

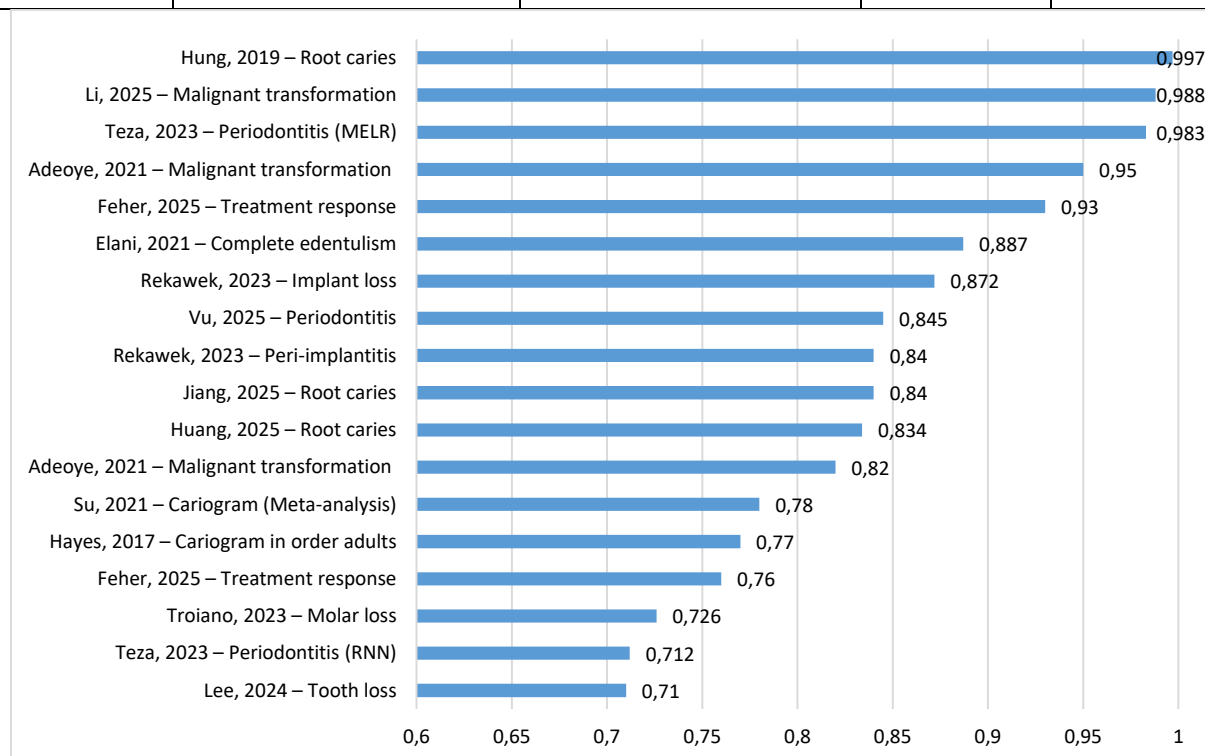


Figure. 1. Discriminatory performance of prognostic models compared with deterministic risk scores. Compiled by the author from the included studies.

Data sources and predictors. Model performance is determined first and foremost by the quality of the input data. Electronic health-record data and population-based surveys are the most readily available sources and underpin most of the studies reviewed here [14, 16, 46].

The laboratory domain is represented primarily by salivary biomarkers. Testing for activated matrix metalloproteinase-8 (aMMP-8) is suitable for chairside use, and mouthrinse samples provide better discrimination of periodontal status than saliva [34]. For prediction, such markers are valuable because they reflect ongoing tissue-destructive activity rather than its accumulated outcome and therefore may carry information about a future event that traditional indices cannot provide. The microbiological domain has been studied predominantly outside geriatric populations. Oral microbiota composition can predict the onset of early childhood caries [19], whereas translation of this approach to older adults remains an open research area.

Automated recognition of carious lesions on radiographs has achieved high diagnostic accuracy but this is a diagnostic task. In a prognostic model, radiographic information should not enter as an algorithm-generated diagnosis but as quantitative features, such as bone density and characteristics of the trabecular pattern, that may predict a future event. Conflating these roles is one of the major sources of methodological confusion [6].

Methodological quality, implementation barriers, and ethical considerations. When the findings from the preceding sections are considered together, a consistent pattern emerges that should guide interpretation of published performance figures. The more rigorous the study design, the more modest the reported model performance tends to be. In dental prognostic research, AUC values above 0.95 should almost always prompt scrutiny of study design rather than immediate acceptance as evidence of an exceptional algorithm.

The most common flaw is the substitution of a diagnostic task for a prognostic one. Cross-sectional designs are presented as prognostic studies, and the resulting metrics are unlikely to be attainable in a prospective setting [25, 36]. The second flaw is the absence of external validation. The two studies that included both internal and external assessments provide a quantitative measure of this problem: performance fell from 0.93 to 0.76 in Feher et al. and from 0.95 to 0.82 in Adeoye et al., corresponding to a decrease of 0.13–0.17 AUC units [5, 18]. Because the overwhelming majority of publications rely solely on internal validation, the real-world performance of these models is likely to be systematically lower than reported by approximately this magnitude.

The third flaw concerns calibration, which is simply not assessed in many studies. Discrimination answers the question of whether a model correctly ranks patients according to risk, but it does not establish whether a predicted probability of 30% corresponds to an actual event frequency of 30%. Calibration is therefore crucial for clinical decision-making, yet 79% of the studies in the cited review did not discuss it [10]. Exceptions in dentistry include the meta-analysis reporting observed-to-expected event ratios and the nomogram study that presented calibration curves [24, 40]. A fourth group of problems involves the operating point and class imbalance. The results shown above (Figure 2) demonstrate that the same task can be addressed in diametrically opposite ways depending on the classification threshold [38]. When the outcome is rare, averaged performance metrics can be misleading, reporting the area under the precision-recall curve is preferable [29].

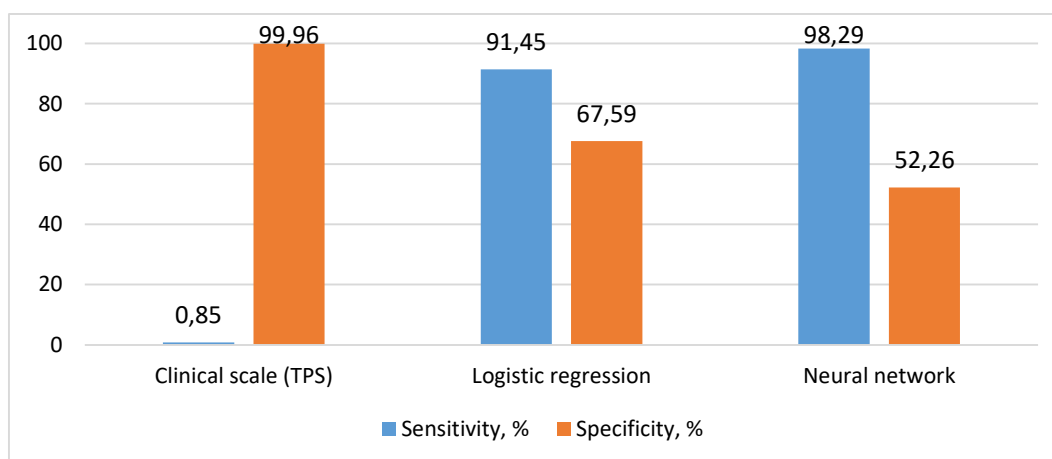


Figure 2. Operating-point shift in tooth-loss prediction [38]

The fifth group of issues is ethical and organizational. Older adults are systematically underrepresented in training datasets for medical models, creating age-related algorithmic bias, strategies for identifying and mitigating this bias have been summarized in a dedicated review [11]. A model trained predominantly on middle-aged patients is likely to systematically underestimate risk in the most vulnerable individuals. This has contributed to growing interest in explainable artificial intelligence [31]. Restrictions on transferring personal health data between institutions also make federated learning attractive, allowing models to be trained without moving patient-level data outside the clinical institution, its applicability to multicentre medical research has been addressed separately [28].

Table 4. Common methodological weaknesses in prediction models and approaches to their control

Weakness	Consequence	Control measure	Sources
Substitution of diagnosis for prognosis	Cross-sectional design is presented as prognostic; performance metrics are inflated	Longitudinal design with a prespecified prediction horizon	[25, 36]
Lack of external validation	Performance is overestimated by 0.13–0.17 AUC units	Independent cohort from another centre or country	[5, 18]
Failure to assess calibration	Predicted probabilities are biased while only AUC is reported	Calibration curves and observed-to-expected event ratios	[10, 40]
Class imbalance and arbitrary operating point	Overdiagnosis of risk or missed cases	Justification of the threshold based on error costs; area under the precision-recall curve	[29, 38]
Incomplete reporting	Results are not reproducible and external assessment is impossible	Adherence to TRIPOD+AI and assessment with PROBAST-AI	[13, 21]

Weakness	Consequence	Control measure	Sources
Age-related data bias	Risk is underestimated in the most vulnerable group	Age stratification and model fairness auditing	[11]
Ignoring the competing risk of death	Cumulative event probability is overestimated	Competing-risk models rather than standard survival analysis	[15, 41]

Conclusion. Risk prediction for dental diseases in older adults currently consists of a set of fragmented studies that are difficult to compare directly because they differ in disease outcome, data source, algorithm, and, most importantly, the rigor with which the models were evaluated. Nevertheless, a cross-study comparison, including comparison with deterministic risk scores, makes it possible to assess how close the field has come to clinical applicability.

First, machine learning in geriatric dentistry remains at the stage of accumulating disparate pilot findings rather than representing a mature technology. The reported performance range of AUC 0.71–0.99 reflects differences in study-design rigor more than differences between algorithms. The highest values come from cross-sectional studies with internal validation, whereas the more modest and credible estimates come from multicentre studies with external testing.

Second, the superiority of machine learning over conventional statistical methods is not automatic and is not supported at all in some studies. Logistic regression that appropriately accounts for the structure of the data is not an obsolete approach it remains an essential benchmark.

Third, the specific characteristics of older patients are only partially incorporated into existing models. Polypharmacy, objectively measured salivary flow, the number of exposed root surfaces, oral frailty, and social isolation are represented only fragmentarily, while the competing risk of death is almost never considered, despite being particularly relevant in this age group.

Fourth, clinical implementation depends less on model accuracy alone than on transparency, reproducibility, and fairness with respect to the target age group. A model embedded in an electronic health record should provide not only an estimated event probability but also an interpretable account of how that estimate was generated and should be evaluated for the absence of age-related bias.

Overall, personalized risk prediction using machine learning is a promising but not yet routine-ready tool in geriatric dentistry. Priority should be given to prospective multicentre studies with a prespecified prediction horizon, mandatory external validation, assessment of calibration, explicit handling of the competing risk of death, and reporting according to TRIPOD+AI. Another important task is the development and validation of such models in regional cohorts, because existing studies characterize the oral health status of older adults but do not provide individualized prognoses [1, 3].

References:

1. Асадуллаев Н, Ризаев Ж. Некоторые аспекты течения заболеваний пародонта у лиц пожилого и старческого возраста. *Stomatologiya*. 2017;1(66):82–5.
2. Долгалев АА, Мураев АА, Ляхов ПА, Ляхова УА, Чониашвили ДЗ, Золотаев КЕ и др. Архитектоника системы искусственного интеллекта и перспективы применения

технологий машинного обучения в стоматологии. Обзор литературы. Главный врач Юга России. 2022;(5):2–6.

3. Рыбак ОГ. Реализация стоматологических профилактических технологий для лиц старших возрастных групп. Дальневосточный медицинский журнал. 2023;(4):29–34. doi:10.35177/1994-5191-2023-4-5

4. Саматов УА. Комбинированные методы лечения пульпитов у лиц старческого возраста. Stomatologiya. 2021;1(82):22–4. doi:10.34920/2091-5845-2021-06

5. Adeoye J, Koohi-Moghadam M, Lo AWI, Tsang RK, Chow VLY, Zheng LW, et al. Deep learning predicts the malignant-transformation-free survival of oral potentially malignant disorders. Cancers (Basel). 2021;13(23):6054. doi:10.3390/cancers13236054

6. Albano D, Galiano V, Basile M, Di Luca F, Gitto S, Messina C, et al. Artificial intelligence for radiographic imaging detection of caries lesions: a systematic review. BMC Oral Health. 2024;24(1):274. doi:10.1186/s12903-024-04046-7

7. Barbe AG. Medication-induced xerostomia and hyposalivation in the elderly: culprits, complications, and management. Drugs Aging. 2018;35(10):877–85. doi:10.1007/s40266-018-0588-5

8. Benzia H, Watt R, Makino Y, Stauf N, Varenne B. WHO calls to end the global crisis of oral health. Lancet. 2022;400(10367):1909–10. doi:10.1016/S0140-6736(22)02322-4

9. Cannon I, Robinson-Barella A, McLellan G, Ramsay SE. From drugs to dry mouth: a systematic review exploring oral and psychological health conditions associated with dry mouth in older adults with polypharmacy. Drugs Aging. 2023;40(4):307–16. doi:10.1007/s40266-023-01017-5

10. 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

11. Chu C, Donato-Woodger S, Khan SS, Shi T, Leslie K, Abbasgholizadeh-Rahimi S, et al. Strategies to mitigate age-related bias in machine learning: scoping review. JMIR Aging. 2024;7:e53564. doi:10.2196/53564

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(1):430. doi:10.1186/s12903-024-04210-z

13. Cohen JF, Bossuyt PMM. TRIPOD+AI: an updated reporting guideline for clinical prediction models. BMJ. 2024;385:q824. doi:10.1136/bmj.q824

14. Cooray U, Watt RG, Tsakos G, Heilmann A, Hariyama M, Yamamoto T, et al. Importance of socioeconomic factors in predicting tooth loss among older adults in Japan: evidence from a machine learning analysis. Soc Sci Med. 2021;291:114486. doi:10.1016/j.socscimed.2021.114486

15. Dibello V, Zupo R, Sardone R, Lozupone M, Castellana F, Dibello A, et al. Oral frailty and its determinants in older age: a systematic review. Lancet Healthy Longev. 2021;2(8):e507–e520. doi:10.1016/S2666-7568(21)00143-4

16. Elani HW, Batista AFM, Thomson WM, Kawachi I, Chiavegatto Filho ADP. Predictors of tooth loss: a machine learning approach. PLoS One. 2021;16(6):e0252873. doi:10.1371/journal.pone.0252873

17. Fang WL, Jiang MJ, Gu BB, Wei YM, Fan SN, Liao W, et al. Tooth loss as a risk factor for dementia: systematic review and meta-analysis of 21 observational studies. *BMC Psychiatry*. 2018;18(1):345. doi:10.1186/s12888-018-1927-0
18. Feher B, de Souza Oliveira EH, Mendes Duarte P, Werdich AA, Giannobile WV, Feres M. Machine learning-assisted prediction of clinical responses to periodontal treatment. *J Periodontol*. 2025;96(11):1199–212. doi:10.1002/JPER.24-0737
19. Grier A, Myers JA, O'Connor TG, Quivey RG, Gill SR, Kopycka-Kedzierawski DT. Oral microbiota composition predicts early childhood caries onset. *J Dent Res*. 2021;100(6):599–607. doi:10.1177/0022034520979926
20. Hariyani N, Spencer AJ, Luzzi L, Harford J, Tan H, Mejia G, et al. The prevalence and severity of root surface caries across Australian generations. *Community Dent Oral Epidemiol*. 2019;47(5):398–406. doi:10.1111/cdoe.12459
21. Hasuike A, Watanabe T, Wakuda S, Kogure K, Yanagiya R, Byrd KM, et al. Machine learning in predicting tooth loss: a systematic review and risk of bias assessment. *J Pers Med*. 2022;12(10):1682. doi:10.3390/jpm12101682
22. Hayes M, Da Mata C, Cole M, McKenna G, Burke F, Allen PF. Risk indicators associated with root caries in independently living older adults. *J Dent*. 2016;51:8–14. doi:10.1016/j.jdent.2016.05.006
23. Hayes M, Da Mata C, McKenna G, Burke FM, Allen PF. Evaluation of the Cariogram for root caries prediction. *J Dent*. 2017;62:25–30. doi:10.1016/j.jdent.2017.04.010
24. Huang LR, Zhang CZ, Gong ML, Cheng XQ, Wu HK. Development of a nomogram for root caries risk assessment in a Chinese elderly population. *J Dent*. 2025;156:105624. doi:10.1016/j.jdent.2025.105624
25. Hung M, Voss MW, Rosales MN, Li W, Su W, Xu J, et al. Application of machine learning for diagnostic prediction of root caries. *Gerodontology*. 2019;36(4):395–404. doi:10.1111/ger.12432
26. Jiang L, Huang S, Reissmann DR, Schmalz G, Li J. Identification of risk group for root caries and analysis of associated factors in older adults using unsupervised machine learning clustering. *Clin Interv Aging*. 2025;20:483–93. doi:10.2147/CIA.S520229
27. Kassebaum NJ, Bernabé E, Dahiya M, Bhandari B, Murray CJ, Marcenes W. Global burden of untreated caries: a systematic review and metaregression. *J Dent Res*. 2015;94(5):650–8. doi:10.1177/0022034515573272
28. Kirienko M, Sollini M, Ninatti G, Loiacono D, Giacomello E, Gozzi N, et al. Distributed learning: a reliable privacy-preserving strategy to change multicenter collaborations using AI. *Eur J Nucl Med Mol Imaging*. 2021;48(12):3791–804. doi:10.1007/s00259-021-05339-7
29. Lee CT, Zhang K, Li W, Tang K, Ling Y, Walji MF, et al. Identifying predictors of the tooth loss phenotype in a large periodontitis patient cohort using a machine learning approach. *J Dent*. 2024;144:104921. doi:10.1016/j.jdent.2024.104921
30. Li JW, Zhang MJ, Zhou YF, Adeoye J, Pu JYJ, Thomson P, et al. Enhancing malignant transformation predictions in oral potentially malignant disorders: a novel machine learning framework using real-world data. *iScience*. 2025;28(3):112062. doi:10.1016/j.isci.2025.112062

31. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011–2022). *Comput Methods Programs Biomed.* 2022;226:107161. doi:10.1016/j.cmpb.2022.107161
32. Matulienė G, Studer R, Lang NP, Schmidlin K, Pjetursson BE, Salvi GE, et al. Significance of Periodontal Risk Assessment in the recurrence of periodontitis and tooth loss. *J Clin Periodontol.* 2010;37(2):191–9. doi:10.1111/j.1600-051X.2009.01508.x
33. Qin X, He J, He H, Yuan X, Su X, Zeng X. Long-term trends in the burden of edentulism in China over three decades: a Joinpoint regression and age-period-cohort analysis based on the Global Burden of Disease Study 2019. *Front Public Health.* 2023;11:1099194. doi:10.3389/fpubh.2023.1099194
34. Räisänen IT, Heikkinen AM, Nwhator SO, Umeizudike KA, Tervahartiala T, Sorsa T. On the diagnostic discrimination ability of mouthrinse and salivary aMMP-8 point-of-care testing regarding periodontal health and disease. *Diagn Microbiol Infect Dis.* 2019;95(4):114871. doi:10.1016/j.diagmicrobio.2019.114871
35. Rekawek P, Herbst EA, Suri A, Ford BP, Rajapakse CS, Panchal N. Machine learning and artificial intelligence: a web-based implant failure and peri-implantitis prediction model for clinicians. *Int J Oral Maxillofac Implants.* 2023;38(3):576–82. doi:10.11607/jomi.9852
36. Reyes LT, Knorst JK, Ortiz FR, Ardenghi TM. Machine learning in the diagnosis and prognostic prediction of dental caries: a systematic review. *Caries Res.* 2022;56(3):161–70. doi:10.1159/000524167
37. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE Jr, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: Part II – binary and time-to-event outcomes. *Stat Med.* 2019;38(7):1276–96. doi:10.1002/sim.7992
38. Santamaria P, Troiano G, Serroni M, Araùjo TG, Ravidà A, Nibali L. Exploring the accuracy of tooth loss prediction between a clinical periodontal prognostic system and a machine learning prognostic model. *J Clin Periodontol.* 2024;51(10):1333–41. doi:10.1111/jcpe.14023
39. Soto AP, Meyer SL. Oral implications of polypharmacy in older adults. *Clin Geriatr Med.* 2023;39(2):273–93. doi:10.1016/j.cger.2023.01.008
40. Su N, Lagerweij MD, van der Heijden GJMG. Assessment of predictive performance of caries risk assessment models based on a systematic review and meta-analysis. *J Dent.* 2021;110:103664. doi:10.1016/j.jdent.2021.103664
41. Tanaka T, Takahashi K, Hirano H, Kikutani T, Watanabe Y, Ohara Y, et al. Oral frailty as a risk factor for physical frailty and mortality in community-dwelling elderly. *J Gerontol A Biol Sci Med Sci.* 2018;73(12):1661–7. doi:10.1093/gerona/glx225
42. Tellez M, Gomez J, Pretty I, Ellwood R, Ismail AI. Evidence on existing caries risk assessment systems: are they predictive of future caries? *Community Dent Oral Epidemiol.* 2013;41(1):67–78. doi:10.1111/cdoe.12003
43. Teza H, Pattanateepapon A, Lertpimonchai A, Vathesatogkit P, J McKay G, Attia J, et al. Development of risk prediction models for severe periodontitis in a Thai population: statistical and machine learning approaches. *JMIR Form Res.* 2023;7:e48351. doi:10.2196/48351
44. Troiano G, Nibali L, Petsos H, Eickholz P, Saleh MHA, Santamaria P, et al. Development and international validation of logistic regression and machine-learning models for the prediction of 10-year molar loss. *J Clin Periodontol.* 2023;50(3):348–57. doi:10.1111/jcpe.13739

45. Uppal S, Kumar Shrivastava P, Khan A, Sharma A, Kumar Shrivastav A. Machine learning methods in predicting the risk of malignant transformation of oral potentially malignant disorders: a systematic review. *Int J Med Inform.* 2024;186:105421. doi:10.1016/j.ijmedinf.2024.105421
46. Vu GT, Mayya V, Mandhidi B, King C, Little BB, Gurupur V, et al. Application of machine learning to predict periodontal disease in US adults: a cross-sectional analysis of NHANES 2009–2014. *Health Informatics J.* 2025;31(4):14604582251394617. doi:10.1177/14604582251394617
47. Xie C, Li Y, Liu K, Liu J, Zeng J, Huang N, et al. A hybrid unsupervised clustering method for predicting the risk of dental implant loss. *J Dent.* 2024;149:105260. doi:10.1016/j.jdent.2024.105260

