

FROM PREDICTION TO PREVENTION: A SYSTEMATIC REVIEW OF AI-DRIVEN PREDICTIVE ANALYTICS IN HEALTHCARE

Soon Eu Hui¹, Marcus Teunissen², Yash Bafna^{*3}, Nishu Ayedee⁴

1. Faculty of Business and Communications, INTI International University, Negeri Sembilan, Malaysia

2. Rushford Business School, Switzerland

3. Narsinhbhai Patel Dental College and Hospital Sankalchand Patel University (SPU), Visnagar, Gujarat, India

4. Confab 360 Degree and INTI International University, India

*Correspondence: ybafna.fds@spu.ac.in

ABSTRACT

PURPOSE:

This study examined how Artificial Intelligence (AI)-enabled predictive analytics moves from technical prediction towards actionable preventive healthcare and identified the conditions that enable or constrain this translation.

DESIGN/METHODOLOGY/APPROACH:

A systematic literature review of 12 peer-reviewed studies published between 2021 and 2026 was conducted using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-guided screening, Critical Appraisal Skills Programme (CASP) quality assessment, and open, axial, and selective coding with frequency analysis.

FINDINGS:

Three themes emerged: predictive performance does not equal preventive value, timeliness and actionability shape preventive utility, and implementation conditions shape translation into practice. The findings show that preventive value depends on the alignment of predictive validity, timely action, workflow integration, decision ownership, and governance.

ORIGINALITY/VALUE:

The study introduces the prediction-to-prevention translation gap and develops a decision-oriented framework for evaluating AI beyond model performance.

KEYWORDS:

Artificial Intelligence, predictive analytics, preventive healthcare, machine learning, clinical decision-making, systematic literature review

INTRODUCTION

Healthcare systems increasingly use predictive analytics to identify health risks before they develop into severe and costly conditions. A major example of this challenge is noncommunicable diseases. World Health Organisation [1] reported at least 43 million deaths globally in 2021 caused by these diseases, including 18 million deaths among people younger than 70 years. Therefore, early detection and timely treatment are important for preventive healthcare. Artificial Intelligence

(AI) and Machine Learning (ML) offer new possibilities in this area because of their ability to analyse large and diverse health datasets to estimate future risk, identify emerging deterioration and support earlier healthcare decisions.

AI-enabled prediction is increasingly moving beyond experimental modelling into regulated healthcare settings. The United States Food and Drug Administration (FDA) maintains a public list of AI-enabled medical devices authorised for marketing in the United States [2]. These comprise technologies utilised in all areas of cardiovascular and neurological care, screening, as well as continuous monitoring, including devices for use outside the standard healthcare environment [2]. This trend simply indicates predictive and monitoring technologies that increasingly make their appearance in regulated healthcare settings. However, greater availability of tools does not necessarily mean AI-enabled tools provide preventive value. Such a differentiation matters because in healthcare, technical performance and usefulness measure different things. Performance indicators like accuracy, discrimination and others indicate whether a model can distinguish risk, but they do not establish whether its output arrives early enough to change care, corresponds to a feasible intervention or can be integrated into routine decision-making. However, complex ML approaches were not automatically superior to conventional modelling in predicting the risk of pancreatic cancer among adults with newly diagnosed diabetes, as found by Cliff et al. [3]. Hinson et al. went a step further, however, using the ML model in a prospective validation study and integrating its results into workflows in the emergency department [4]. The difference between these studies highlights a broader issue. A model may predict well without demonstrating that the prediction can be translated into useful healthcare action.

Current assessment approaches of AI-based predictive analytics may then tend to focus on the predictive performance rather than on what happens once the prediction is made. Preventive value depends on a connected process. The model must produce sufficiently reliable information, the information must arrive within a useful intervention window, an appropriate response must exist, and the healthcare system must be capable of acting on it. Workflow integration, decision responsibility, interpretability and governance can consequently influence value just as strongly as technical performance.

The paper conceptualises the distance between predictive capability and preventive action as the prediction-to-prevention translation gap. Its contribution lies in examining AI-enabled predictive analytics not only by how accurately systems predict risk, but by the conditions under which prediction becomes timely and actionable preventive healthcare. For this, the paper addresses two Research Questions (RQs):

RQ1. How does predictive performance translate into actionable preventive healthcare across AI-enabled predictive analytics studies?

RQ2. What technological, clinical, organisational, and governance conditions enable or constrain this translation?

For this review, preventive healthcare includes anticipatory applications that identify emerging risk, support screening or early detection, enable timely monitoring or intervention, or inform prospective healthcare decisions before adverse outcomes intensify.

THEORETICAL LENS

The paper is guided by Socio-Technical Systems (STS) Theory, which argues that organisational performance depends on how well technical and social systems work together, rather than on technology alone [5]. Technical system components of AI-powered predictive healthcare are data quality, model design, validation, and predictive performance. Clinical judgement, workflow, decision responsibility, organisational readiness and monitoring are components of the social system. This matters because the prediction must have a preventative effect that the healthcare system can use to act on it. Meanwhile, even the best clinical processes cannot overcome an unreliable prediction. STS helps explain the prediction-to-prevention translation gap as an alignment problem between predictive technology and the people, processes, and organisational conditions needed to turn predictions into action.

LITERATURE REVIEW

The literature review underscores significant advancements in the development of AI-driven predictive health models; however, it also reveals a gap between their predictive capabilities and the potential benefit for prevention. Lin et al. [6] used large electrocardiogram (ECG) datasets to evaluate cardiovascular event prediction, and Lee et al. [7] used retinal images with conventional risk factors and evaluated the model in an independent population. Kalmady et al. [8] also analysed cardiovascular prediction from hospitals and patient subgroups. These studies help build confidence in model performance beyond a single development dataset. However, stronger technical performance does not mean greater clinical value. Cliff et al. found that more complex machine-learning approaches did not provide better clinical utility than a conventional model for pancreatic cancer risk prediction [3]. This challenges the assumption that more sophisticated algorithms will yield superior preventive decisions.

Another concern is when a prediction turns into a supported action and a question arises about its usefulness. The shift toward screening and monitoring studies in the near future is toward preventive care, since these provide an opportunity to intervene while risk is not yet an issue. Khattab et al. [9] performed an external validation of the prediction process for personalised diabetes screening, whereas Hughes et al. [10] demonstrated that early prediction of cardiovascular risk using AI can change the classification of some patients, which in turn can inform treatment decisions. Decision curve analysis has also been applied to evaluate community-based screening systems [11] in addition to assessing model accuracy. But remote monitoring goes one step further in helping to determine predictive risk and then convert it into warning intensities that aid in early action [12]. Nevertheless, earlier predictions do not necessarily lead to better prevention. The value of the prediction is limited if it is taken outside the window of meaning, or outside of what is going to be done about it. Recent studies of wearable monitoring also show that prediction accuracy may vary depending on the predicted time [13].

Yet another issue pertaining to translation is its integration into the routine practice of health care. Imrie et al. [14] combined automated model development with interpretability and assessment of clinical net benefit. Hinson et al. [4] implemented a process for using ML to support their risk assessment in the emergency department and monitored this process in a prospective study. Coombs et al. [15] related risk classification to the perspectives of clinicians and future decisions in oncology. These studies illustrate the additional complexity of the process of implementation. Technically valid predictions cannot be directly translated into healthcare decisions without clinicians, a well-fitting workflow, interpretation, responsibility for taking action and monitoring after deployment. However, evidence on model implementation is less prevalent than performance-centric studies. The literature thus reflects the merits of predictive analytics but also confirms that the link between prediction and preventive action is not yet that clear.

RESEARCH GAP

Existing research provides strong evidence that AI can predict health risks but less is known about how those predictions become preventive healthcare actions. Predictive validity, timely actionability, and implementation are often investigated independently. The relationship between these stages is therefore missing. The paper fills this gap by examining the interaction between technical performance and timing, workflow, decision responsibilities, and organisational context. It then integrates these relationships into a Prediction-to-Prevention Translation Framework that explains how predictive capability can move towards preventive value.

METHODS

RESEARCH DESIGN

The study used a Systematic Literature Review (SLR) methodology to synthesise how AI-enabled predictive analytics contributes to preventive and anticipatory healthcare. An SLR was appropriate because the evidence spans different prediction targets, data environments, algorithms, and healthcare settings, making direct statistical pooling inappropriate. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [16] were followed to ensure methodological transparency and rigour, as shown in Figure 1.

SOURCE AND SEARCH SELECTION

Digital database research was undertaken for an extensive literature search, including PubMed and Web of Science. The research approach was supported by the extensive coverage of peer-reviewed articles based on the inclusion and exclusion criteria shown in Table 1. Boolean search string used was: ("artificial intelligence" OR "machine learn*" OR "deep learn*" OR "neural network*" OR "automated machine learn*" OR AutoML) AND (predict* OR prognos* OR "risk strat*" OR "risk scor*" OR "risk assess*" OR screen* OR "early detect*" OR "early warn*" OR monitor*) AND (health* OR clinic* OR patient* OR disease* OR hospital* OR "population health"). The asterisk was used for right-hand truncation as shown in Table 2 to capture lexical variants. Preventive terminology was not made a mandatory search field for databases.

TABLE 1. INCLUSION AND EXCLUSION CRITERIA

Inclusion criteria	Exclusion criteria
Peer-reviewed primary empirical studies published between 2021 and 2026	Reviews, editorials, commentaries, protocols, conference abstracts, or other non-primary studies
Studies published in English with full text available	Non-English or unavailable full-text studies
Studies applying AI or machine learning to human healthcare data	Studies without an AI or machine-learning component
AI used for prediction, risk stratification, screening, early detection, monitoring, or prospective clinical decision support.	AI used only for classification, segmentation, data extraction, or administrative purposes without predictive or anticipatory relevance.
Studies with preventive or anticipatory healthcare relevance, including early risk identification or support for timely clinical action	Studies with no preventive, anticipatory, or clinically actionable relevance
Studies reporting model performance, validation, clinical utility, or implementation evidence	Studies providing insufficient evidence to evaluate model performance, validation, utility, or implementation

Source: Compiled by Authors

TABLE 2. TRUNCATION

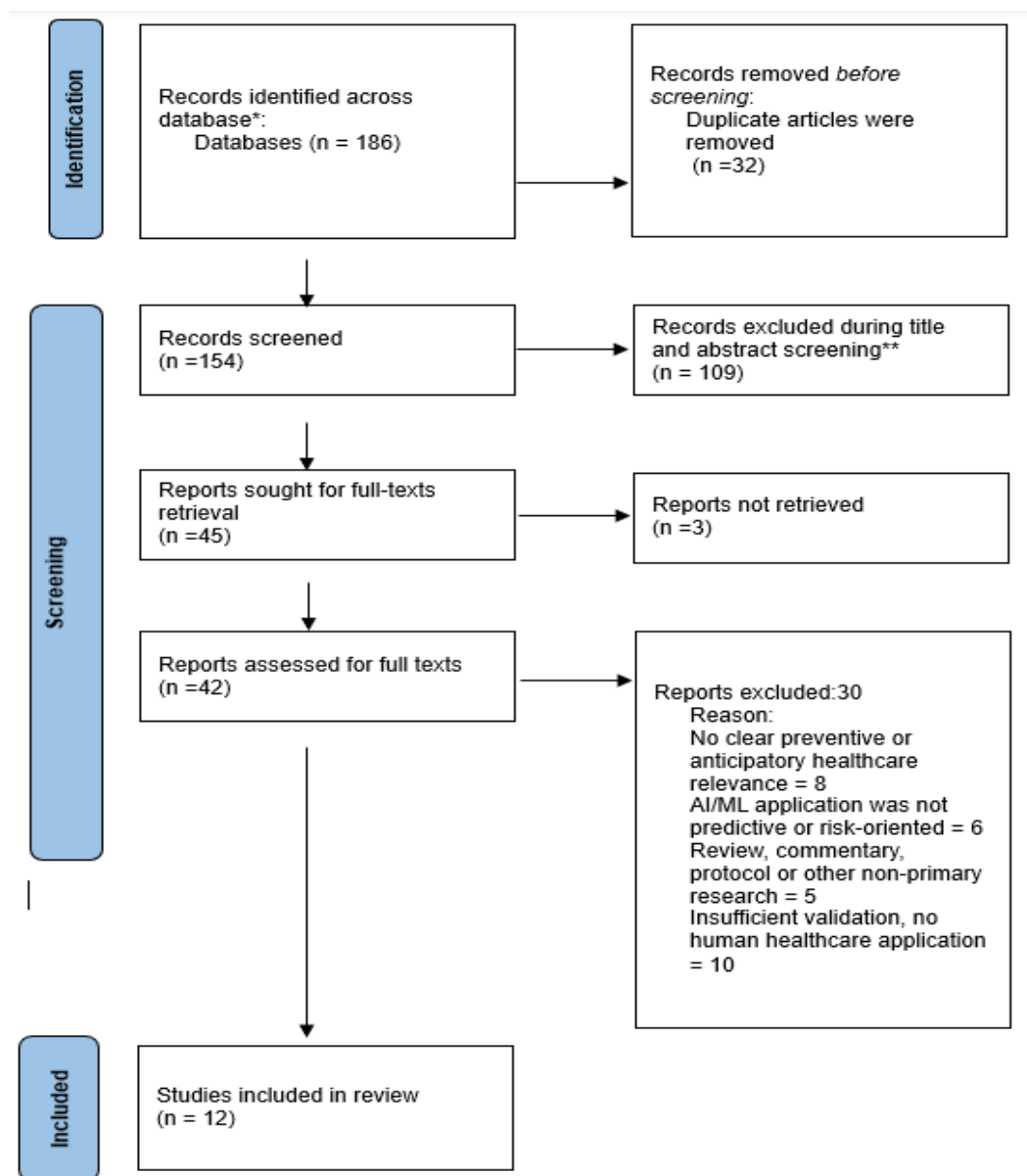
Search term	Captures examples such as	Purpose
machine learn*	machine learning, machine learner	ML terminology
deep learn*	deep learning	Deep learning terminology
neural network*	neural network, neural networks	Singular/plural
automated machine learn*	automated machine learning	AutoML
predict*	predict, predicted, predicting, prediction, predictive	Prediction variants
prognos*	prognosis, prognostic, prognostication	Prognostic variants
risk strat*	risk stratification, risk stratified	Risk stratification
risk scor*	risk score, risk scores, risk scoring	Risk scoring
risk assess*	risk assessment, risk assessments	Risk assessment
screen*	screen, screening, screened	Preventive screening
early detect*	early detection, early detecting	Early identification
early warn*	early warning, early warnings	Early-warning systems
monitor*	monitor, monitoring, monitored	Continuous/remote monitoring
health*	health, healthcare	Healthcare variations
clinic*	clinic, clinical, clinician	Clinical contexts
patient*	patient, patients	Patient populations
disease*	disease, diseases	Disease contexts
hospital*	hospital, hospitals	Healthcare settings

Source: Compiled by Authors

STUDY SELECTION AND PRISMA PROCESS

A total of 186 records were identified across the databases. After removing 32 duplicates, 154 records were screened based on their title and abstract. Of these, 109 were excluded because they did not meet the predefined eligibility criteria. Forty-five full-text articles were sought for full-text retrieval; three could not be retrieved. Forty-two full-text reports were assessed for eligibility, of which 30 were excluded for the reasons shown in Figure 1. The final qualitative synthesis included 12 studies that became part of the SLR table shown in Table 3.

FIGURE 1. PRISMA FLOWCHART



Source: Adapted from PRISMA 2020 [16]

Data were extracted from each included study using a structured matrix covering healthcare application, data or setting, AI/ML approach, key evidence, validation or performance focus, Critical Appraisal Skills Programme (CASP) score, and contribution to the thematic synthesis.

QUALITY ASSESSMENT

An adapted CASP framework was used to assess the methodological quality of the included studies [17]. Eight criteria were assessed: clarity of research aim, appropriateness of study design, adequacy of data and sample, transparency of AI/ML methods, model validation, consideration of bias, clarity of findings, and relevance to preventive or anticipatory healthcare. Each criterion was scored 1 if adequately addressed and 0 if not, for a maximum score of 8. Studies scoring

7–8 were classified as high quality, 5–6 as moderate quality, and 0–4 as low quality. Quality scores guided the strength of interpretation during evidence synthesis rather than excluding studies.

TABLE 3. SUMMARISATION OF LITERATURE INCLUDED FOR ANALYSIS

Study	Healthcare application	Data / setting	AI/ML approach	Key evidence / contribution	Validation / performance focus	CASP	Theme (T)
Khattab et al. (2026)	Type 2 diabetes screening	All of Us Research Program and UK Biobank	Two-stage ML / XGBoost	Supported personalised identification of individuals at risk of developing type 2 diabetes	External validation across independent cohorts; subgroup assessment	8, High	T2
Lin et al. (2025)	Major cardiovascular event prediction	Large-scale ECG data from multiple medical centres	Multitask deep learning	Demonstrated prediction of several future cardiovascular outcomes from ECG data	Independent external validation and comparison with conventional cardiovascular risk assessment	7, High	T1
Hughes et al. (2023)	Long-term cardiovascular risk	ECG datasets from multiple healthcare centres	Deep-learning ECG risk score	AI-generated risk information supported cardiovascular risk reclassification with potential implications for preventive treatment	Validation across independent healthcare centres	8, High	T2
Lee et al. (2023)	Cardiovascular risk prediction	Fundus images and conventional cardiovascular risk factors	Multimodal deep learning	Combining imaging and traditional risk factors improved cardiovascular risk estimation	External validation and comparison with conventional approaches	7, High	T1
Imrie et al. (2023)	Diagnostic and prognostic modelling	UK Biobank and clinical prediction applications	Automated machine learning	AutoML supported accessible model development, interpretation and clinical decision-	Comparison with established risk scores and assessment of clinical net benefit	7, High	T3

				support applications			
Zhang et al. (2024)	Community-based eye disease screening	Real-world community screening data	RETFound-enhanced deep learning	Demonstrated the value of AI for community screening across different service environments	Real-world evaluation and decision-curve analysis	7, High	T2
Capstick et al. (2024)	Early detection of urinary tract infection	Remote monitoring and home-sensor data	Explainable ML	Risk estimates were converted into actionable warning levels for early identification	Evaluation of sensitivity, specificity and alert burden	6, Moderate	T2
Gadaleta et al. (2021)	Early COVID-19 detection	Wearable sensor data	Explainable ML	Passive physiological monitoring supported detection before or around conventional clinical identification	Performance evaluated across different prediction windows	7, High	T2
Hinson et al. (2022)	Hospital admission and deterioration decisions	Multisite emergency department setting	Workflow-integrated ML	Model predictions were incorporated into electronic clinical workflows and used to support admission decisions	Prospective validation and implementation monitoring	8, High	T3
Coombs et al. (2022)	Oncology risk prediction and prospective care decisions	Oncology clinical setting	ML risk-prediction framework	Connected prediction with clinician review and prospective care decisions	Retrospective and prospective validation with clinician confirmation	8, High	T3
Cliff et al. (2024)	Pancreatic cancer risk in adults with new-onset diabetes	Population-based healthcare data	Cox regression, XGBoost and neural networks	More complex ML did not necessarily provide superior clinical value	Internal-external validation, calibration and decision-curve analysis	8, High	T1

				compared with conventional modelling			
Kalmady et al. (2024)	Population-level cardiovascular prediction	Large multisite ECG dataset	ML algorithms	Predictive performance differed across cardiovascular outcomes and settings, highlighting limits of performance metrics alone	Patient-level and cross-hospital validation with subgroup evaluation	8, High	T1

Source: Compiled by Authors

THEMATIC SYNTHESIS AND FREQUENCY ANALYSIS

For thematic synthesis, a three-step coding procedure was followed, including open coding, axial coding, and selective coding across 12 studies, as shown in Table 4. These studies were grouped together according to their conceptual relationships based on recurring concepts as identified by open coding. Selective coding was then used to integrate these categories around the central issue of how AI-enabled predictive analytics translated into preventive healthcare value. Cross-study comparison was also used to identify convergent and contradictory findings, while CASP scores informed the strength placed on each pattern during interpretation. Each study was assigned to the theme representing its dominant contribution to the synthesis. A study-level frequency analysis was subsequently conducted, with each study counted once, to show the distribution of evidence across the three themes.

TABLE 4. THEMATIC SYNTHESIS AND FREQUENCY ANALYSIS

Open codes identified across studies	Axial coding category	Selective coding interpretation	Final theme	Studies, n	Frequency
Predictive accuracy, discrimination, calibration, internal validation, external validation, generalisability, model comparison, subgroup performance	Predictive validity and robustness	Strong prediction establishes technical capability, but predictive performance alone does not demonstrate preventive or clinical value	T1. Predictive Performance Does Not Equal Preventive Value	4	33.3%
Early detection, screening, risk stratification, prediction horizon, early warning, timely alerts, intervention opportunity, clinical net benefit	Timeliness and actionability	Prediction becomes more useful for prevention when it is generated early enough and is linked to a feasible healthcare response	T2. Timeliness and Actionability Shape Preventive Utility	5	41.7%
Workflow integration, clinician involvement, decision support, explainability, organisational readiness, bias assessment,	Implementation and governance conditions	Translation from predictive capability into healthcare action depends on the clinical and organisational	T3. Implementation Conditions Shape Translation into Practice	3	25.0%

implementation monitoring, human review		conditions surrounding model use			
		Total		12	100%

Source: Compiled by Authors

FINDINGS AND ANALYSIS

THEME 1. PREDICTIVE PERFORMANCE DOES NOT EQUAL PREVENTIVE VALUE

Findings indicate that strong technical performance creates predictive capability, but not preventive value. External validation showed that models could be applied usefully to achieve scores beyond those from their original development datasets. The results also indicated that more refined and complicated models do not necessarily lead to more clinical value. However, Cliff et al. [3] found that more complex ML methods did not generate greater clinical utility than conventional modelling. This differentiation is directly related to RQ1. Accuracy and discrimination are not enough to infer Preventive value. Reliability of the identification of risk does not mean the model changes the outcome of the patient. In an STS perspective, these measures reflect the technical system's strength but not whether the social and clinical system will utilise the information. The finding thus translates the concept of predictive validity not to the end of the translation process, but rather at the start. Technical reliability is important, but only if and when the prediction can inform a decision or intervention that allows for prevention.

THEME 2. TIMELINESS AND ACTIONABILITY SHAPE PREVENTIVE UTILITY

The synthesis shows that the utility of the prediction changes according to what's being predicted, when the information arrives and whether taking some action could follow. Prediction has become a possibility to prevent action with personalised diabetes screening [9], cardiovascular risk reclassification [10], community screening [11], as well as remote warning systems [12] and wearables [13]. Previous prediction, however, should not be assumed to be superior to prevention. A risk signal has limited utility if it comes too late, too early to support a reliable decision, or if there is no feasible response. Actionability, in turn, is an important connection between the technical prediction and the use of healthcare. This addresses RQ1 as it demonstrates that translation is more about a meaningful intervention window than predictive performance. It also helps to answer RQ2, as time, communication, and how the participant can respond are turning conditions for success in translation. The STS perspective considers actionability to be the point of convergence between the technical and the social systems. The function of the model is to provide information; however, preventive value can only be gained when the information is appropriate and in time for the healthcare decision context.

THEME 3. IMPLEMENTATION CONDITIONS SHAPE TRANSLATION INTO PRACTICE

There is an important gap in maturity, as evidenced by a few studies that focus on implementation. While evidence of routine use of technical prediction in healthcare processes seems less developed, prediction seems to be more advanced than technical prediction. This directly answers RQ2. The extent to which a prediction becomes an operational decision depends on workflow fit, clinicians' involvement, responsibility for action, interpretability and post-deployment monitoring. Related governance evidence was found in the form of biases assessed, human review and monitoring, but not as advanced as technical validation. STS helps explain this pattern clearly. Predictive technology cannot generate preventive value without being embedded in a particular healthcare system. Therefore, the process of translation should ensure that the outputs of the models correspond to the people, responsibilities, and processes involved in their use. The discovery changes the deployment problem to one of effectiveness at the time of deployment.

PREDICTION-TO-PREVENTION TRANSLATION FRAMEWORK

Table 5 outlines the Prediction-to-Prevention Translation Framework to guide the translation of predictive capability and value in the context of preventive healthcare driven by AI. The themes identified in the review were then organised in accordance with the logic of STS Theory. Predictive validity underscores the need to predict before preventing, as the

framework begins with that criterion. Then it turns to the challenges of preventative actionability and timely delivery that enable an actionable response to the prediction. The second phase, decision and workflow alignment, focuses on the intended target audience for the prediction, who is carrying out the action, and how the target audience's healthcare process involves the prediction. Governing implementation concludes with human oversight, control, bias assessment, accountability, and continuous monitoring. This framework is not to be followed in a step-by-step sequence. Failure can occur at any stage of the translation process, and the model, workflow, or decision will be subject to change as a result of feedback from the implementation. The predictive efficacy, therefore, is not enough to guarantee the preventive value; it is necessary to understand the social conditions together with the technical conditions.

TABLE 5. PREDICTION-TO-PREVENTION TRANSLATION FRAMEWORK

Framework stage	Key question	Main condition	Decision rule
Predictive validity	Can the prediction be trusted in the intended population and setting?	Reliable performance, validation, subgroup robustness	Do not move towards implementation if predictive performance is weak, unstable, or not adequately validated
Preventive actionability	Does the prediction identify a risk that can realistically be acted upon?	Clear link between predicted risk and a feasible healthcare response	High predictive performance alone is insufficient when no meaningful preventive action follows
Timely delivery	Does the prediction arrive within a useful intervention window?	Appropriate prediction horizon and timely communication	A prediction has limited preventive value if it arrives too late or too early to support a reliable response
Decision and workflow alignment	Who receives the prediction and who is responsible for acting on it?	Clear decision ownership, workflow fit, clinician involvement	Predictive outputs should be linked to an identified decision maker and response pathway before routine use
Governed implementation	Can the system be used safely and responsibly over time?	Human oversight, bias assessment, accountability, monitoring	Deployment should include review of model performance, unintended effects, subgroup differences, and workflow consequences
Preventive value	Has prediction been translated into useful healthcare action?	Alignment across technical and social conditions	Preventive value is achieved only when valid prediction, timely actionability, workflow integration, and governance operate together.

Source: Compiled by Authors

IMPLICATIONS

Predictive AI needs to be considered as a part of the healthcare management system and not just a tool or technology. Organisations need to address four practical questions before using a model: 1) Is the model validated, 2) can any action that may result in a decreased outcome to the prediction be taken, 3) will any person take the action, and 4) will the action be carried out within a meaningful time to the model or prediction. Reaction, alert burden, workflow and side effects should, therefore, be part of the assessment, in addition to accuracy when pilot testing. For instance, if a hospital implements an algorithm to stratify patients who are at risk of diabetes, the algorithm's output should be associated with a standardised and formalised protocol, e.g. linked to a recommendation for confirmatory testing and then a call to a clinician or to a diabetes prevention counselling pathway. If prevention is used, it will have value only if assertive follow-up is directly involved in the process. The framework can thus be applied to procurement, pilot testing, and pilot implementation decisions.

LIMITATIONS

Different performance evaluations in the reviewed studies were combined into one synthesis, including various diseases, datasets, and healthcare contexts, which provides broader coverage but may limit direct comparability of model performance. Evidence of prediction and validation was also more well-developed than evidence of long-term implementation, governance and post-deployment outcomes. The differences reflect the fact that the field is still a maturing one. Future studies should therefore pay greater attention to plans for future implementation, cross-setting validation of the impact of the outputs of the predictions on the various facets of the preventive decision, workflow, equity and patient outcomes.

CONCLUSION

This paper shows that accurate prediction alone does not create preventive healthcare value. Predictive validity provides the technical foundation, but value depends on whether the prediction is timely, actionable, integrated into workflow, assigned to a responsible decision maker, and appropriately governed. The Prediction-to-Prevention Translation Framework brings these conditions together and explains where translation can succeed or break down. By separating predictive capability from preventive value, the study moves the discussion beyond model performance and towards the practical conditions required for AI-enabled prediction to support preventive healthcare.

References

1. World Health Organization. Noncommunicable diseases [Internet]. Geneva: World Health Organization; 2025. Available: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>
2. U.S. Food and Drug Administration. Artificial Intelligence-Enabled Medical Devices [Internet]. Silver Spring (MD): FDA; 2026. Available: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices>
3. Clift AK, Tan PS, Patone M, Liao W, Coupland C, Bashford-Rogers R, et al. Predicting the risk of pancreatic cancer in adults with new-onset diabetes: development and internal-external validation of a clinical risk prediction model. *Br J Cancer*. 2024;130(12):1969-1978. doi:10.1038/s41416-024-02693-9.
4. Hinson JS, Klein E, Smith A, Toerper M, Dungarani T, Hager D, et al. Multisite implementation of a workflow-integrated machine learning system to optimize COVID-19 hospital admission decisions. *NPJ Digit Med*. 2022;5(1):94. doi:10.1038/s41746-022-00646-1.
5. Trist EL, Bamforth KW. Some social and psychological consequences of the longwall method of coal-getting: an examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Hum Relat*. 1951;4(1):3-38. doi:10.1177/001872675100400101.
6. Lin CH, Liu ZY, Chu PH, Chen JS, Wu HH, Wen MS, et al. A multitask deep learning model utilizing electrocardiograms for major cardiovascular adverse events prediction. *NPJ Digit Med*. 2025;8(1):1. doi:10.1038/s41746-024-01410-3.
7. Lee YC, Cha J, Shim I, Park WY, Kang SW, Lim DH, et al. Multimodal deep learning of fundus abnormalities and traditional risk factors for cardiovascular risk prediction. *NPJ Digit Med*. 2023;6(1):14. doi:10.1038/s41746-023-00748-4.
8. Kalmady SV, Salimi A, Sun W, Sepehrvand N, Nademi Y, Baine K, et al. Development and validation of machine learning algorithms based on electrocardiograms for cardiovascular diagnoses at the population level. *NPJ Digit Med*. 2024;7(1):133. doi:10.1038/s41746-024-01130-8.
9. Khattab A, Chen SF, Javedani Sadaei H, Wineinger NE, Torkamani A. Development and validation of a two-stage machine learning model for personalised type 2 diabetes screening in the All of Us Research Program and UK Biobank. *BMJ Open*. 2026;16(1):e103171. doi:10.1136/bmjopen-2025-103171.
10. Hughes JW, Tooley J, Torres Soto J, Ostropelets A, Poterucha T, Christensen MK, et al. A deep learning-based electrocardiogram risk score for long term cardiovascular death and disease. *NPJ Digit Med*. 2023;6(1):169. doi:10.1038/s41746-023-00916-6.
11. Zhang J, Lin S, Cheng T, Xu Y, Lu L, He J, et al. RETFound-enhanced community-based fundus disease screening: real-world evidence and decision curve analysis. *NPJ Digit Med*. 2024;7(1):108. doi:10.1038/s41746-024-01109-5.

12. Capstick A, Palermo F, Zakka K, Fletcher-Lloyd N, Walsh C, Cui T, et al. Digital remote monitoring for screening and early detection of urinary tract infections. *NPJ Digit Med*. 2024;7(1):11. doi:10.1038/s41746-023-00995-5.
13. Gadaleta M, Radin JM, Baca-Motes K, Ramos E, Kheterpal V, Topol EJ, et al. Passive detection of COVID-19 with wearable sensors and explainable machine learning algorithms. *NPJ Digit Med*. 2021;4(1):166. doi:10.1038/s41746-021-00533-1.
14. Imrie F, Cebere B, McKinney EF, van der Schaar M. AutoPrognosis 2.0: Democratizing diagnostic and prognostic modeling in healthcare with automated machine learning. *PLOS Digit Health*. 2023;2(6):e0000276. doi:10.1371/journal.pdig.0000276.
15. Coombs L, Orlando A, Wang X, Shaw P, Rich AS, Lakhtakia S, et al. A machine learning framework supporting prospective clinical decisions applied to risk prediction in oncology. *NPJ Digit Med*. 2022;5(1):117. doi:10.1038/s41746-022-00660-3.
16. 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:n71. doi:10.1136/bmj.n71.
17. Critical Appraisal Skills Programme. CASP Clinical Prediction Rule Checklist [Internet]. Oxford: CASP; 2024. Available: <https://casp-uk.net/casp-tools-checklists/clinical-prediction-rule-checklist/>